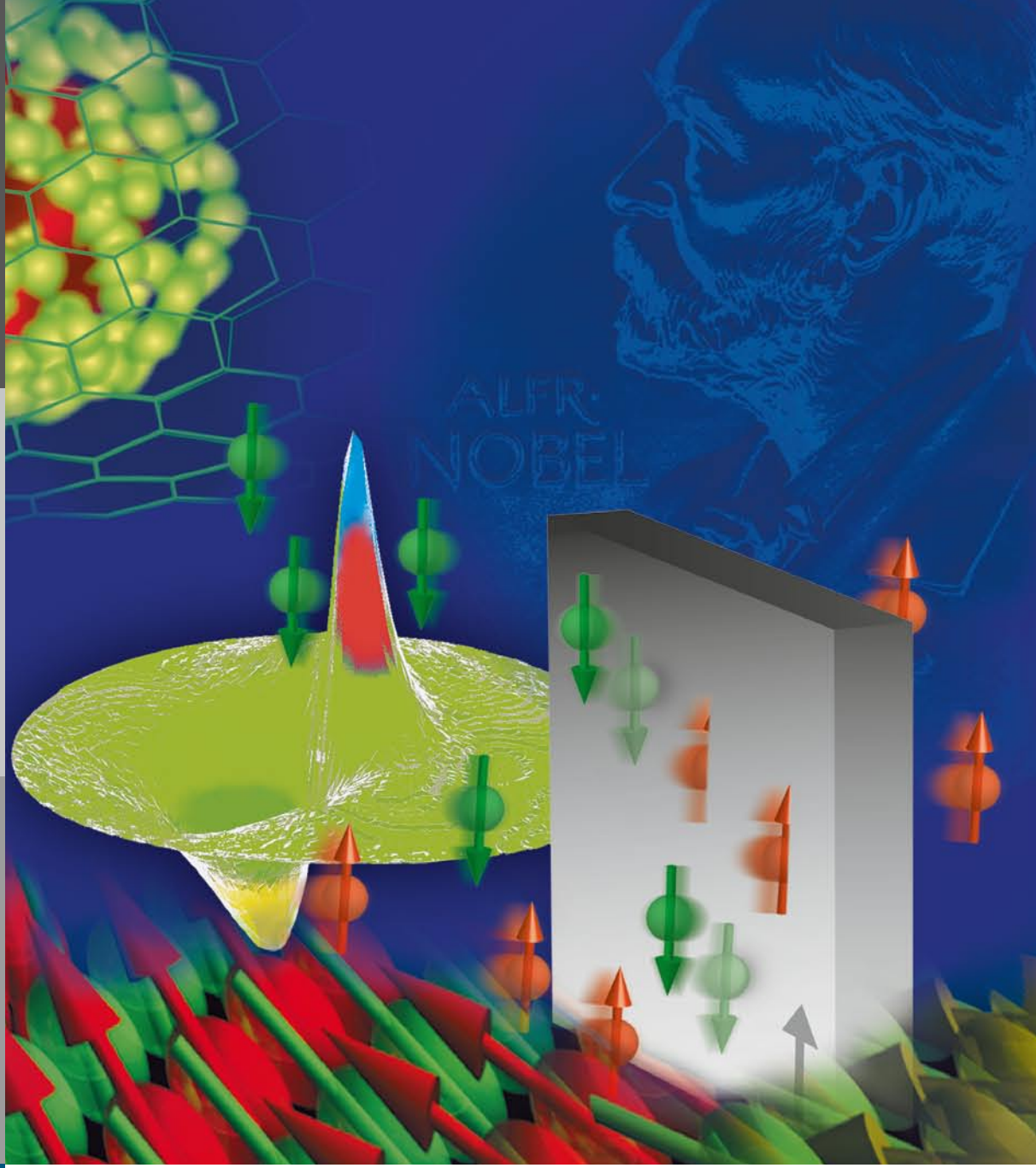


40th IFF Springschool 2009

Spintronics

From GMR to Quantum Information

Lecture Notes

JÜLICH
FORSCHUNGSZENTRUM

Mitglied der Helmholtz-Gemeinschaft

Contents

Preface

I. Introduction

I1 *Introduction (C.M. Schneider)*

A. Fundamentals of Electron Theory

A1 *Electronic States in Solids (G. Bihlmayer)*
A2 *Electronic Basis of Magnetism (R. Zeller)*
A3 *Magnetism in Reduced Dimensions (S. Blügel)*
A4 *Highly Correlated Electron Materials: Dynamical Mean Field Theory (A. Liebsch)*
A5 *Electronic Structures of Transition-Metal Oxides (F. Ishii)*
A6 *Introduction to Multiferroics (M. Ležaić)*
A7 *Berry Phase in Solid State Physics (M.-C. Chang)*

B. Spin-Dependent Transport Processes: Basics

B1 *Spin Polarization: From the EPR Paradox to the GMR Effect (P. Grünberg)*
B2 *Theory of Tunneling Magnetoresistance (D. Wortmann)*
B3 *Spin Injection into Semiconductors (B. Beschoten)*
B4 *Spin Filter (M. Müller)*
B5 *Spin-Orbit Coupling in Semiconductor Devices (A. Bringer)*
B6 *Spin Relaxation in Nonmagnetic Metals and Semiconductors (P. Mavropoulos)*
B7 *Galvanomagnetic Transport: From Hall Effect to AMR (P.S. Bechthtold)*

C. Spin-Dependent Transport: Advanced Concepts

C1 *The Standard Model of Spin Injection (J. Fabian)*
C2 *Anomalous Hall Effect (Y. Mokrousov)*
C4 *Quantum Spin Hall Effect (H. Buhmann)*
C5 *Electron Spins in Carbon Based Materials (M. Morgenstern)*
C6 *Spin and Transport through Quantum Dots (M.R. Wegewijs)*

D. Spin Dynamics

D1 *Micromagnetism (R. Hertel)*
D2 *Spin Waves in Confined Systems (S.O. Demokritov)*
D3 *Spin-Transfer Torque Dynamics (D.E. Bürgler)*
D4 *Femtosecond Opto-Magnetism: Ultrafast Laser Manipulation of Spins (T. Rasing)*

E. Spin-Based Quantum Information

E1 *Fundamental Concepts of Quantum Information Processing (T. Schäpers)*
E2 *Quantum Computing with Semiconductor Quantum Dots (C. Meyer)*
E3 *Donors for Quantum Information Processing (M.S. Brandt)*
E4 *Quantum algorithms (M. Richter)*

F. Applications

F1 *Technology of Hard Disk Drives (A. Berger)*
F3 *Magnetic Logic (G. Reiss)*

Index

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 10

Forschungszentrum Jülich GmbH
Institute of Solid State Research

Lecture Notes of the
40th Spring School 2009

Stefan Blügel, Daniel Bürgler, Markus Morgenstern,
Claus M. Schneider, Rainer Waser (Eds.)

Spintronics – From GMR to Quantum Information

This spring school was organized
by the Institute of Solid State Research
of the Forschungszentrum Jülich
on 9 – 20 March 2009.

In collaboration with JARA-FIT, universities,
research institutes and the industry.

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 10

ISSN 1866-1807 ISBN 978-3-89336-559-3

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche
Nationalbibliografie; detailed bibliographic data are available in the
Internet at http://dnb.d-nb.de

Publisher:	Forschungszentrum Jülich GmbH Institut für Festkörperforschung D-52425 Jülich Phone +49 (0)2461 61 6048 • Fax +49 (0)2461 61 2410
Cover Design:	Grafische Medien, Forschungszentrum Jülich GmbH
Printer:	Druckerei Gehler, Düren-Birkesdorf
Copyright:	Forschungszentrum Jülich 2009
Distributor:	Forschungszentrum Jülich Zentralbibliothek, Verlag D-52425 Jülich Phone +49 (0)2461 61 5368 • Fax +49 (0)2461 61 6103 e-mail: zb-publikation@fz-juelich.de Internet: http://www.fz-juelich.de

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 10

ISSN 1866-1807
ISBN 978-3-89336-559-3

Contents

Preface

I. Introduction

I1 *Introduction (C.M. Schneider)*

A. Fundamentals of Electron Theory

A1 *Electronic States in Solids (G. Bihlmayer)*
A2 *Electronic Basis of Magnetism (R. Zeller)*
A3 *Magnetism in Reduced Dimensions (S. Blügel)*
A4 *Highly Correlated Electron Materials: Dynamical Mean Field Theory (A. Liebsch)*
A5 *Electronic Structures of Transition-Metal Oxides (F. Ishii)*
A6 *Introduction to Multiferroics (M. Ležaić)*
A7 *Berry Phase in Solid State Physics (M.-C. Chang)*

B. Spin-Dependent Transport Processes: Basics

B1 *Spin Polarization: From the EPR Paradox to the GMR Effect (P. Grünberg)*
B2 *Theory of Tunneling Magnetoresistance (D. Wortmann)*
B3 *Spin Injection into Semiconductors (B. Beschoten)*
B4 *Spin Filter (M. Müller)*
B5 *Spin-Orbit Coupling in Semiconductor Devices (A. Bringer)*
B6 *Spin Relaxation in Nonmagnetic Metals and Semiconductors (P. Mavropoulos)*

C. Spin-Dependent Transport: Advanced Concepts

C1 *The Standard Model of Spin Injection (J. Fabian)*
C2 *Anomalous Hall Effect (Y. Mokrousov)*
C4 *Quantum Spin Hall Effect (H. Buhmann)*
C5 *Electron Spins in Carbon Based Materials (M. Morgenstern)*
C6 *Spin and Transport through Quantum Dots (M.R. Wegewijs)*

D. Spin Dynamics

D1 *Micromagnetism (R. Hertel)*
D2 *Spin Waves in Confined Systems (S.O. Demokritov)*
D3 *Spin-Transfer Torque Dynamics (D.E. Bürgler)*
D4 *Femtosecond Opto-Magnetism: Ultrafast Laser Manipulation of Spins (T. Rasing)*

E. Spin-Based Quantum Information

E1 *Fundamental Concepts of Quantum Information Processing (T. Schäpers)*
E2 *Quantum Computing with Semiconductor Quantum Dots (C. Meyer)*
E3 *Donors for Quantum Information Processing (M.S. Brandt)*
E4 *Quantum algorithms (M. Richter)*

F. Applications

F1 *Technology of Hard Disk Drives (A. Berger)*
F3 *Magnetic Logic (G. Reiss)*

Index

Preface

The discovery of Giant Magnetoresistance (GMR) in 1988 laid the foundation to a whole new and very active research field – Spinelectronics or Spintronics – which strives to exploit the electron spin and electron spin currents as the basic carriers for the device functionality and information transfer in electronic devices. The pioneering work of Peter Grünberg (IFF) and Albert Fert (Université Paris-Sud), changed our view of the role of the electron spin in electrical transport and has been honored by the 2007 Nobel Prize in Physics, partly also because of its enormous technological and economical impact. Only 10 years after the discovery of the effect in the laboratory, GMR-based hard disk read heads hit the market as first generation spintronic devices and revolutionized the magnetic mass storage industry. Since its advent 20 years ago, Spintronics continues to provide us with a wide variety of spin-dependent transport and transfer processes, novel materials, phenomena and concepts, and many open questions and challenges. The emphasis in current spintronics research is threefold:

- First, it aims to achieve a control of and the ability to manipulate spin transport on very small length scales down to the level of single spins, which will open a pathway to quantum information applications. This control also includes the active switching of the magnetization by means of spin-polarized currents.
- Second, in order to obtain the best of both worlds spinelectronics may be combined with advanced semiconductor nanoelectronics. A crucial step in this direction is the realization of an efficient electrical spin-injection into semiconductors.
- Third, the next generation of spintronic devices should combine passive and active functionalities, thereby enabling magnetologic circuits and even magnetoprocessors.

On the way to meet these challenges many fundamental questions have to be solved and many new materals and materials combinations will be developed and explored. Among others this concerns the microscopic interactions and mechanisms leading to spin dephasing, the manipulation of spins by spin-orbit interactions, the understanding of spin transfer torque mechanisms, and the utilization of the spin Hall effect. On the material side, dilute magnetic semiconductors, highly spin-polarized oxides and half-metals, but also graphene and multiferroics are currently in the focus of interest.

The present course continues a series of Spring Schools in thin film magnetism and nanomagnetism taking place in the years 1993, 1999, and 2005, and particularly addresses the new developments in the field of spintronics. The lectures will first build a basis for the understanding of the major fundamental phenomena and aspects in magnetism and spin-dependent transport. This includes new theoretical concepts as well as the theoretical framework for a quantitative description. The School will then advance to contemporary aspects of the anomalous, spin and quantum spin Hall effect, and of spin transfer processes at interfaces and in nanostructures down to the quantum level. Exchange interactions and spin effects in highly correlated materials are another important topic covered by the lectures. Finally, we will also discuss current and future technological applications of spintronics concepts.

The topics of the lectures cover:

- Fundamentals of Magnetism
- Spin-dependent Interactions
- Quantum Transport
- Spin Transport Phenomena
- Spin Injection and Coherence
- Spin Transfer Torque
- Electronic Correlations
- Magnetization and Spin Dynamics
- Multiferroics
- Spin Hall Effect
- Magnetic Storage, Memory and Logics
- Quantum Information Processing

For the first time, this year the IFF Spring School takes place under the umbrella of the Jülich-Aachen Research Alliance (JARA). JARA combines complementing expertise at the RWTH Aachen University and the Forschungszentrum Jülich to address research fields identified at an international level in a targeted manner. The research and education activities in the area of "Fundamentals of Future Information Technology" are bundled in the section JARA-FIT. Furthermore, the school has been organized in collaboration with the Universities of Cologne and Duisburg-Essen and is integrated in the curricula of these universities.

We are grateful to all contributors from the Institut für Festkörperforschung (IFF), the Institut für Bio- und Nanosysteme (IBN), and the Institute for Advanced Simulation (IAS) in the Forschungszentrum Jülich, as well as the Institut für Experimentalphysik of the RWTH Aachen University for the time and effort they spent to prepare the manuscripts and the lectures, and for their spontaneous help and support of this spring school:

Dr. B. Beschoten (RWTH)
Dr. G. Bihlmayer (IFF-1 & IAS)
Prof. P. A. Grünberg (IFF-9)
Dr. M. Ležaić (IFF-1 & IAS)
Dr. Ph. Mavropoulos (IFF-1& IAS)
Dr. Y. Mokrousov (IFF-1 & IAS)
Dr. M. Richter (IAS-JSC)
JunProf. M. Wegewijs (IFF-3)
Dr. R. Zeller (IFF-3)
Prof. P. Bechthold (IFF-9)
Dr. A. Bringer (IFF-1)
PD Dr. R. Hertel (IFF-9)
Dr. A. Liebsch (IFF-1)
Dr. C. Meyer (IFF-9)
Dr. M. Müller (IFF-9)
Dr. Th. Schäpers (IBN-1)
Dr. D. Wortmann (IFF-1 & IAS)

A large number of courses is offered during the school. We are grateful to our colleagues from the Research Centre Jülich and the RWTH Aachen University for their willingness to conduct these courses. The Jülich and Aachen divisions involved in the lectures are: IFF-1: Quantum Theory of Materials, IFF-3: Theory of Structure Formation, IFF-6: Electronic Materials, IFF-9: Electronic Properties, IBN-1: Semiconductor Nanoelectronics, IAS-JSC: Jülich Supercomputing Centre, and RWTH: II. Institute of Physics A & B.

We are very glad that several colleagues from external universities and (industry) research laboratories have agreed to contribute to the program of the school:

Dr. A. Berger	CIC Nanogune, Donostia, Spain
Prof. M. Brandt	Walter-Schottky-Institut, Technische Universität München
Prof. H. Buhmann	Physikalisches Institut EP-3, Universität Würzburg
Prof. M.-C. Chang	Department of Physics, National Taiwan Normal University, Taiwan
Prof. S. Demokritov	Institut für Angewandte Physik, Universität Münster
Prof. J. Fabian	Institut für Theoretische Physik, Universität Regensburg
Prof. A. Fert	Unité Mixte de Physique CNRS/THALES, Orsay, France
Prof. F. Ishii	Department of Computational Science, Kanazawa University, Japan
Dr. S. S. P. Parkin	IBM Research Laboratories, Almaden, USA
Prof. T. Rasing	Inst. f. Molecules and Materials, Radboud University Nijmegen, The Netherlands
Prof. G. Reiss	Physik Department, Universität Bielefeld

Without the participation of all these colleagues, the program would not be as interesting, versatile, and attractive. We would like to express our thanks to all of them for the effort and enthusiasm which they have put into the preparation and presentation of their lectures and manuscripts. We are very grateful to the board of directors of the Forschungszentrum Jülich for the continuous organizational and financial support, which we have received for the realization of the IFF Spring School and the production of this book of lecture notes. Finally, our special thanks go to Dipl.-Ing. R. Hölzle for the general management, to Mrs. A. Wenzik for taking care of the public relation issues, and to Mrs. L. Snyders for her help in compiling the Lecture Notes.

C. M. Schneider, D. E. Bürgler, S. Blügel, R. Waser, M. Morgenstern

February 2009

I Introduction

Claus M. Schneider
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 The Phenomenon Magnetism **2**

2 Currents: Electrons and Spins **3**

3 The Discovery of Giant Magnetoresistance (GMR) **5**

4 Beyond GMR – the Age of Spintronics **7**
4.1 Spin transport effects 7
4.1.1 Tunneling Magnetoresistance 7
4.1.2 Spin Injection 8
4.1.3 Spin Hall Effect 8
4.1.4 Spin transfer effects 9
4.2 Materials 11
4.2.1 Halfmetallic ferromagnets 11
4.2.2 Dilute magnetic semiconductors 12
4.2.3 Multiferroics 13
4.2.4 Carbon-Based Materials 14

5 Future developments **15**

Having its roots in magnetism – or more specifically – thin film magnetism, over the last two decades spinelectronics has matured into a research field in its own right. Nowadays it includes aspects not only from basically all areas of condensed matter physics, but also from electrical engineering, chemistry, and even biology. Spintronics features a wealth of spin-based effects and involves an extremely broad material basis. In this situation the Spring School has to limit itself to a selection of topics only. The purpose of this introduction is therefore to give some historical background and provide an overview of the main phenomena and material classes.

1 The Phenomenon Magnetism

Magnetism poses an intriguing phenomenon which is known to mankind already for a long time. It was noted more than 2000 years ago that iron is attracted to pieces of a certain mineral – the loadstone or lodestone. The loadstone contains a mixture of different iron oxides, one of them being the ferrimagnet Magnetite, Fe_3O_4 (Fig. 1). The name goes back to the Greek region Magnisia, where this mineral was found. The large remanent magnetization of naturally occuring loadstone is attributed to the strong magnetic fields, which surround lightning bolts striking the ground. These fields align the magnetic moments in the material causing a permanent magnetic behavior.

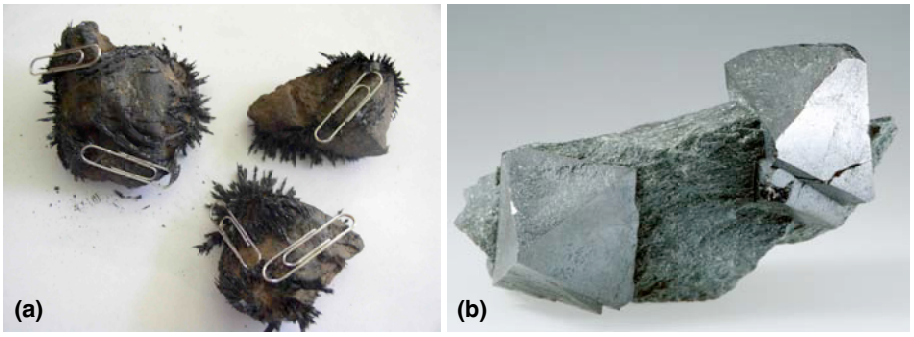


Fig. 1: *Lumps of Magnetite rocks attracting Iron filings and paper clips (a) and natural Magnetite (Fe_3O_4) single crystals (b).*

The first rudimentary application of magnetism dates back already to the 12^{th} century when a compass device was reported to be used by Chinese military (Fig. 2). It took up to the 15^{th} century for the first truly scientific study of magnetism to be conducted, described by *W. Gilbert* in his book *"De Magnete"* [1]. With this growing knowledge of quantitative interrelations in physics and the development of accompanying mathematical tools, the interpretation of magnetic phenomena shifted slowly from metaphysical to a more analytical reasoning.
Maxwell's equations represent an important milestone in this course and opened the pathway to treat magnetic phenomena in the framework of classical electrodynamics [2]. The concept of magnetic fields interacting with matter, thereby producing mechanical forces and electrical fields, forms the basis of our modern technology and laid the foundation for a phenomenological description of the various types of magnetic order. The microscopic mechanisms leading to magnetism as a solid state property, however, became only accessible after the advent of quantum mechanics, which introduced a very important property of the electron – the *spin*. Magnetism is a many-electron phenomenon involving interactions on very different length scales between spins, spins and lattice, or spins and external magnetic fields. Therefore, magnetic

Fig. 2: *Replica of a chinese compass. The handle of the spoon always facing to the south, this device was called "zhi-nan-zhen", which means "south pointer".*

systems develop a wealth of magnetization and spin structures covering length scales of more than eight orders of magnitude down to atomic dimensions.
Particularly in the last century, magnetism and magnetic materials have exerted an enormous impact on the technological development in very different areas. Electrical motors and generators form the fundament of our electrical power engineering and many other areas, such as automation, automotion, etc. Magnetic resonance imaging (MRI) has become an indispensible tool in medical diagnostics, and bioassays with magnetic labeling schemes are currently being developed for lab-on-chip analytics in chemistry, biology, and biochemistry.
Above all, the entire field of information technology would be unthinkable without mass data storage devices, which exploit various facets of magnetism. In fact, the first magnetic recording device dates back to 1898, when V. Poulsen developed the "telegraphone" [3] (Fig. 3). The telegraphone recorded analogue data (sound) onto a magnetic wire, which was moved in front of a read/write head. Due to the unreliability of the wire spools, however, the telegraphone found only limited use. At the beginning of the 1930's recording on magnetic tape was introduced. This principle was in widespread use for audio recordings for more than 70 years. In the 1950's magnetic tapes were first used by IBM to store digital data in order to replace punch cards. In 1956 IBM introduced the first random access system (RAMAC) in an effort to improve the data access time. This hard disk principle features a rotating disk coated with a magnetic film into which data are stored as small magnetic domains ("bits"). It is still successfully employed today, but features a more than eight orders of magnitude higher recording density than in 1950, presently approaching 1 Tbit/in^2. This breathtaking evolution has been and still is made possible by the fundamental research in nanomagnetism and spintronics.

2 Currents: Electrons and Spins

The applications discussed above are making extensive use of the static and dynamic magnetic properties, such as spontaneous magnetization, anisotropy, magnetic order, exchange coupling, etc., occurring in specific materials. The microscopic origin of these properties in a solid rests on quantum mechanical principles and is nowadays understood as a particular consequence of spin-dependent interactions between Fermions in a many-electron system. These interactions are provided by the exchange and spin-orbit coupling. In a closer view, we find that the most important contribution to the magnetism comes from the electrons in the vicinity of the Fermi level, i.e. the same electrons which are also responsible for the fundamental electrical and optical properties of matter. This immediately leads to the interesting question: Is there an "interference" between the magnetically ordered state of matter and its electrical or optical behavior?

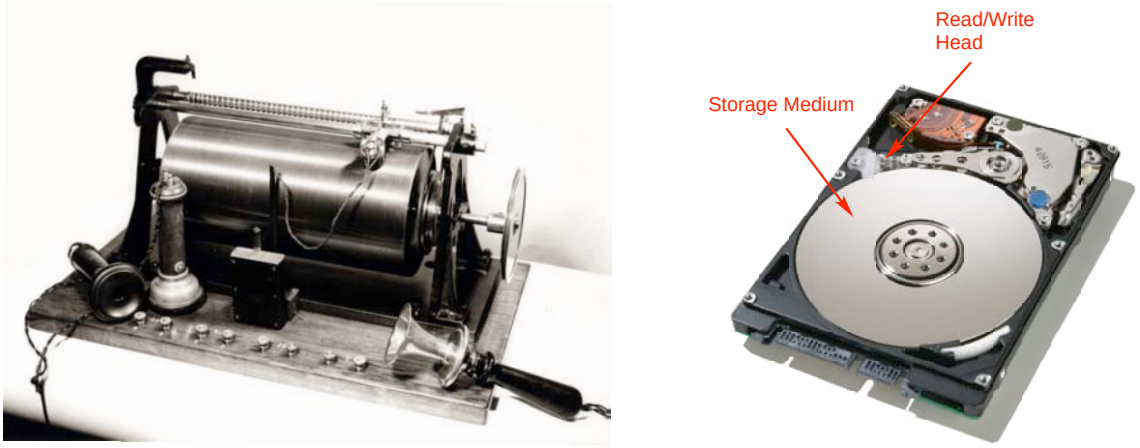


Fig. 3: *More than 100 years of magnetic recording.* ***Left:*** *Magnetic wire recorder developed by V. Poulsen in 1898. The wire is wound on the drum, with the read/write head being guided by the thread rod.* ***Right:*** *High-capacity hard disk drive from 2008. Vital components, such as the read/write head and the storage medium are based on nanomagnetism and spintronics.*

The question was answered – at least in a phenomenological manner – already more than 150 years ago. An interaction between light and magnetically ordered systems was first noted by Michael Faraday in 1846 [4]. The Faraday effect describes the rotation of the polarization plane of linearly polarized light, which passes through a magnetized crystal. Most importantly, the rotation depends on the remanent state of the crystal and occurs also in the absence of any external magnetic fields. About 10 years later (1856) William Thomson, later to become Lord Kelvin, observed a magnetic effect on the electrical conductivity [5]. The phenomenon *magnetoresistance* caused the electrical conductivity of a ferromagnetic material to depend on the orientation of the remanent magnetization with respect to the direction of the flowing current.
A first microscopic explanation of the ordinary magnetoresistance was given by Sir Nevill Mott in 1936 [6]. In his two-current model he considered the electrical current to consist of two contributions, which are scattered differently – a *spin-up* and a *spin-down* contribution. The main scattering channel involves $s-d$ scattering, which becomes different for spin-up and spin-down electrons due to the exchange splitting of the d-states. In this way Mott pointed out the importance of the quantum mechanical property spin in electrical transport effects and created a strong connection between electrical phenomena and magnetism. This concept was later confirmed by Fert and Campbell [7], who also pointed out the importance of spin-flip scattering processes. Of technological importance is the *anisotropic magnetoresistance* (AMR). The microscopic mechanism behind the AMR is spin-orbit scattering of the spin-polarized charge carriers, which results in an anisotropic angular dependence of the current density $\vec{j}$ with respect to the magnetization $\vec{M}$ [8]. The maximum AMR signal is defined as

$$\frac{\Delta R}{R} = \frac{R(\vec{j}||\vec{M}) - R(\vec{j} \perp \vec{M})}{R(\vec{j} \perp \vec{M})} \tag{1}$$

Compared to the exchange interaction, spin-orbit coupling is relatively weak, which renders AMR a small effect, reaching a resistance change $\Delta R/R_0$ of a few percent in certain Ni-based alloys at best. With respect to technical applications the relatively small effect magnitude was

certainly a drawback. Nevertheless, in order to maintain and push the areal density growth rate in magnetic data storage, the inductive read heads were replaced by magnetoresistive sensors exploiting the AMR effect at the beginning of the 1990s [9]. At that time, AMR already looked back to more than hundred years of research. It is also important to note that this progress in magnetic recording technology was made possible by the strong scientific interest in magnetism in reduced dimensions leading to a new and extremely fruitful research field – thin film and surface magnetism.

3 The Discovery of Giant Magnetoresistance (GMR)

The study of thin film magnetism went hand in hand with a breathtaking improvement of the preparation methods. By transferring first the technique of molecular beam epitaxy from semiconductors to metallic systems, magnetic layers and even their surfaces and interfaces could be controlled down to an atomic level. In this way it became possible to realize artificial magnetic systems composed of ferromagnetic films separated by ultrathin nonmagnetic layers. Also sputtering deposition techniques have been perfected to permit the fabrication of such nanomagnetic systems on an industrial scale.
The confinement of the individual layer thicknesses in the nanometer regime gives rise to novel quantum effects, affecting the magnetic properties, such as the magnetic moment or the ordering temperature. A first breakthrough was marked by the discovery of the interlayer exchange coupling (IEC) in 1986 [10]. This coupling mechanism forms between two ferromagnetic layers separated by a nonferromagnetic interlayer and is mediated through spin-polarized quantum-well states developing in the metallic interlayer. As a characteristic property, both the coupling strength and sign vary strongly in a damped oscillatory manner with the interlayer thickness t_{NM}. A very specific feature of this coupling is the antiparallel orientation (AP) of the layer magnetizations forming at selected values of t_{NM}. These antiparallel configurations represent magnetic ground states, which are very useful for magnetotransport experiments. In particular material combinations, for example, Co/Ru, the interlayer coupling may be extremely strong [11]. This property is widely employed nowadays to define a stable reference magnetization direction via a *synthetic antiferromagnet (SAF)*, which is a Co/Ru/Co trilayer in the first AP coupling maximum at around $t_{Ru} \approx 0.8$ nm.

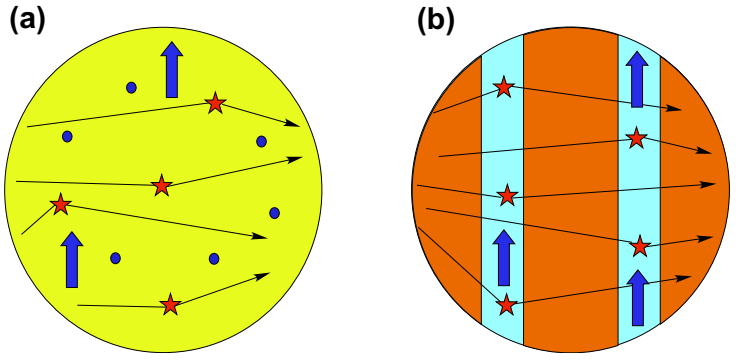


Fig. 4: *Schematic picture of the spatial distribution of scattering centers (marked by stars) in AMR (a) and GMR (b).*

The statistical distribution of the scattering centers and their strong coupling to the magnetic matrix in AMR materials excludes a separate optimization of the spin-polarized scattering mech-

anisms and the magnetic properties (Fig. 4). The layered geometry in heteromagnetic thin film stacks, however, offers a chance to modify the spin-dependent scattering effects independently of the magnetic properties of the layer stack. This hypothesis was impressively proven by the discovery of the giant magnetoresistance (GMR) in Fe/Cr multilayers by Albert Fert's group [12] and in Fe/Cr/Fe trilayers by Grünberg's group [13]. In both cases a strong change in resistivity R was observed when switching the magnetization of the layer stack from the antiparallel (AP) configuration at zero magnetic field to the parallel (P) configuration in the applied saturation field H_S. The magnetoresistance is then defined as

$$\frac{\Delta R}{R} = \frac{R(H=0) - R(H_S)}{R(H_S)} \tag{2}$$

The finding of GMR initiated a broad research activity on the magnetotransport in many different magnetic layer systems. From these studies the Cobalt-Copper material combination emerged as among the most interesting ones for transport applications. It combines a strong magnetoresistive response with moderate interlayer coupling strength and favorable growth behavior. In multilayers a GMR ratio exceeding 60% at room temperature and 110% at cryogenic temperatures can be achieved (Fig. 5).

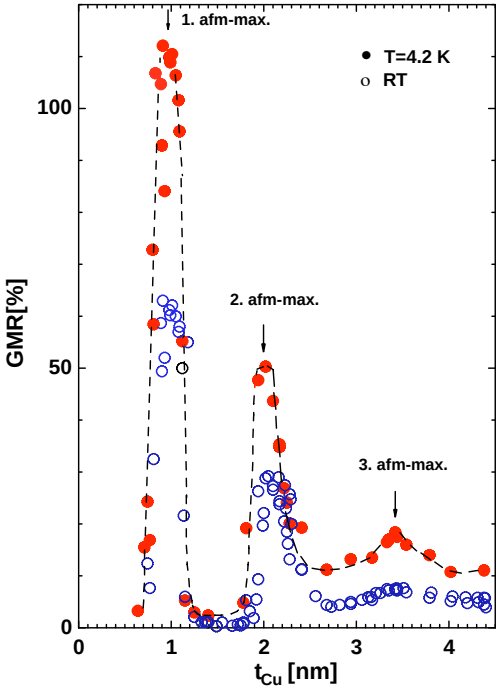


Fig. 5: *Giant magnetoresistance measured in the current-in-plane geometry in Co-Cu multilayer structures. The GMR signal takes a maximum for the antiparallel magnetization configuration obtained at Cu interlayer thicknesses t_{Cu} at around 1 nm, 2 nm, and 3.5 nm.*

The reason for this strong effect in Co/Cu structures is threefold: *(i)* the spin-dependent scattering of the charge carriers is caused by the exchange interaction, rather than the spin-orbit coupling, *(ii)* the major spin-scattering contribution stems from the interfaces of the layer structure, and *(iii)* the Cu Fermi surface is very similar to that of the Co spin-up electrons (electronic matching), but considerably different from the Co spin-down electrons. This unique combination of effective spin-dependent scattering and pronounced spin-dependent confinement is the key to a high magnetoresistance ratio.

The first generation GMR-based hard disk read heads which were introduced in 1999 employed the current-in-plane (CIP) geometry. At present they have already been replaced by sensors based on tunneling magnetoresistance. The further increase in storage density and the transition

to perpendicular magnetic recording may eventually lead to a revival of the GMR read head, however, involving a current-perpendicular-to-plane (CPP) configuration.

4 Beyond GMR – the Age of Spintronics

In order to perceive the impact that the discovery of GMR had on science and technology, we have to recall the situation in 1988: AMR was already known for more than 100 years and was subject to intense research since the 1960's. Despite some improvement particularly in thin films, however, the AMR effect magnitude remained in the region of a few percent. The hard disk industry had already decided to introduce thin film AMR read heads to meet the demands for higher storage densities – they finally entered the market around 1990. The finding of the giant magnetoresistance came therefore very unexpected and started a revolution in the hard disk business. Largely based on the thin film technology originally developed for AMR read heads, GMR read heads were developed to marketability in a period of only ten years after the first observation of this effect.
Even more important than its economic impact were the ramifications of GMR on the research field magnetism. GMR considerably changed our understanding of spin-dependent transport processes – not only in magnetic systems. In this way it opened an avenue from magnetism to microelectronics and initiated a new research area – *spinelectronics* or *spintronics*.

4.1 Spin transport effects

One of the central concepts in spintronics is the spin-scattering length λ_s, i.e. the distance a spin-polarized electron can travel, before its spin direction is changed by a scattering process. It can vary from a few nanometers in transition metals to micrometers in semiconductors. The dimension of the functional component, i.e. the thickness of the nonmagnetic interlayer in GMR systems, must be adapted to λ_s in order to observe sizable spin transport effects.

4.1.1 Tunneling Magnetoresistance

The *tunneling magnetoresistance* (TMR) is another important effect, which illustrates this concept. TMR describes the magnetization dependence of the current through an insulating tunneling barrier and was first observed by Jullière in 1975 at low temperatures [14]. The first TMR experiments at room temperature were successfully performed in 1995 [15, 16] and employed Al_2O_3 barriers. The thickness of the tunneling barrier is a compromise between structural perfection and sufficient conductivity and lies typically in the $1-3$ nm regime. The studies of TMR effects have not only improved our understanding of the electronic properties of ferromagnets, but also stimulated the search for new magnetic materials. The reason is the strong dependence of the TMR signal on the spin polarization of the charge carriers, which for the simplest case is expressed by the Jullière formula

$$\frac{\Delta R}{R} = \frac{R_{AP} - R_P}{R_P} = \frac{2 \cdot P_1 P_2}{1 - P_1 P_2} \tag{3}$$

with P_1 and P_2 denoting the spin polarization of the electrons at the Fermi level for the two magnetic electrodes. Implementing materials with a high spin polarization into magnetic tunneling junctions may thus offer TMR values exceeding 100 %.

The other element that may strongly affect the TMR signal is the insulating barrier itself. This was pointed out by theoretical treatments which predicted extremely high TMR values, if selected electronic states in the barrier match one of the spin states in the electrodes [17]. This concept was impressively confirmed by experiment for the Fe/MgO/Fe system [18, 19]. Further development of materials and layer structures has produced TMR values exceeding 600% at room temperature up to now [20].

4.1.2 Spin Injection

The present microelectronics technology rests on charge-based device concepts, which are realized by means of semiconductors, mostly silicon. In addition, semiconductors also strongly interact with light and thus provide optical functionalities (optoelectronics). Combining independent spin and charge control in a single device promises a wide range of new phenomena and applications [21]. To successfully incorporate spins into existing semiconductor technology, however, one has to resolve several issues such as the efficient injection, transport, control and manipulation, and detection of spin-polarized currents. Despite recent successes, the all-electrical spin injection and detection still remains the holy grail of spin electronics [22].

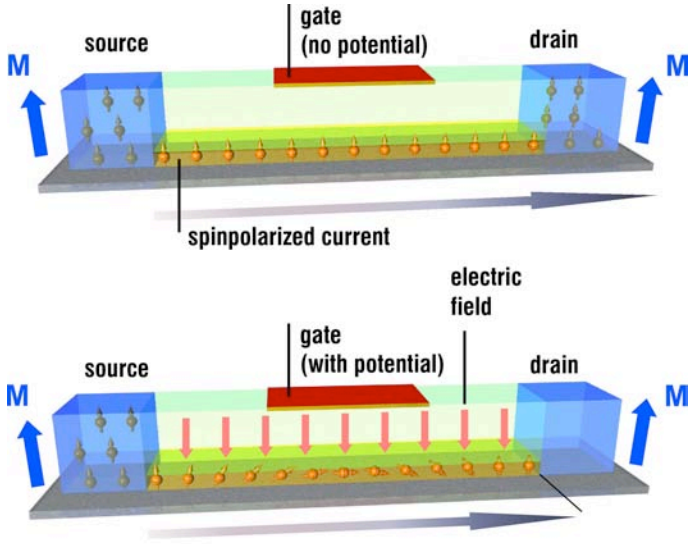


Fig. 6: *Rotation of the spin polarization in a field-effect type device due to the Rashba effect (taken from [23]).*

Realizing new functionalities in spintronic devices requires the spin control and manipulation of the charge carriers passing through the semiconductor. This manipulation can be achieved, for example, by the Rashba effect [24]. This effect is due to a spin-orbit like interaction, which is generated by an electric field in a system with broken inversion symmetry (Fig. 6). It rotates the spin quantization axis of the charge carriers traversing the electric field. A simple device following the field-effect transistor (FET) scheme was proposed by Datta and Das [25]. This proposal has stimulated a broad research activity, leading to a variety of competing device concepts, for example, based on bipolar circuit schemes [26].

4.1.3 Spin Hall Effect

In 1971 Dyakonov and Perel predicted a spin transport phenomenon in semiconductors, which was later termed *Spin Hall Effect (SHE)* [27]. If we assume a slab-like geometry as sketched in Fig. 7, the SHE generates a spin accumulation on the long sides of a current-carrying sample.

The sign of the spin direction is opposite for the opposing boundaries. A reversal of the current direction reverses also the spin polarization at the boundaries. This behavior bears some similarity to the classical Hall effect, where an applied magnetic field exerts a Lorentz force on the moving charge carriers and separates carriers of opposite sign to opposite sides of the slab. In contrast to the classical Hall effect, however, a magnetic field is detrimental in the case of SHE. The magnetic field causes the spins at the boundaries to precess, thereby destroying the spin polarization. The microscopic origin of the SHE is the spin-orbit interaction. It couples spin and charge currents in a peculiar manner: an electrical current induces a transverse spin current (a flow of spins) and vice versa. The SHE has been confirmed experimentally only recently by magnetooptical techniques [28]. The opposite effect of a spin current causing an electrical current is also known (inverse spin Hall effect). The spin Hall effect may be seen as another means to manipulate spins electrically.

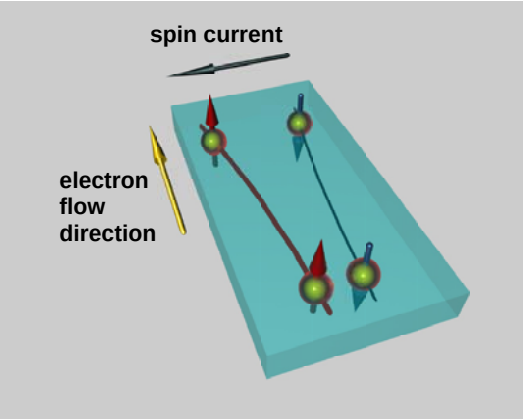


Fig. 7: *Principle of the spin Hall effect: electrons with opposite spin are separated into opposite directions.*

The *Quantum Hall Effect (QHE)* describes a Hall conductance quantization in a two-dimensional electron gas (2DEG). It originates from two main ingredients, namely the Landau quantization of the electronic levels and some disorder in the system. Of particular importance are the electronic states at the boundary of the 2DEG, the socalled edge states. These edge states provide one-dimensional conduction paths with peculiar properties: (i) all electrons on one sample edge move in the same direction, with electrons on the opposite sample edge moving in the opposite direction, and (ii) there is no backscattering, i.e. the elastic mean free path of the electrons approaches infinity. In particular, there is no scattering of charge carriers from one edge to the other, which leads to the quantization of conductance levels. This picture has been extended recently extended to spins [29], predicting a *Quantum Spin Hall Effect (QSHE)* to occur in certain very thin insulating layers. It involves spin-up electrons conducting along one edge of the insulator, with spin-down electrons conducting along the other side. Despite being insulators in the bulk, conduction is allowed at the edges because the interaction between the spin and orbital angular momentum of the electrons reduces the energy gap between the valence and conduction bands to zero for spin-polarized electrons. The QSHE was experimentally realized in mercury (II) telluride (HgTe) semiconductors [30]. In a simple picture it may be considered as edge currents of opposite spins flowing in opposite directions (Fig. 8).

4.1.4 Spin transfer effects

The effects discussed in Sects. 4.1.1 to 4.1.3 have in common that the (magnetic) state and configuration of the sample determines the magnitude and direction of the spin current. In

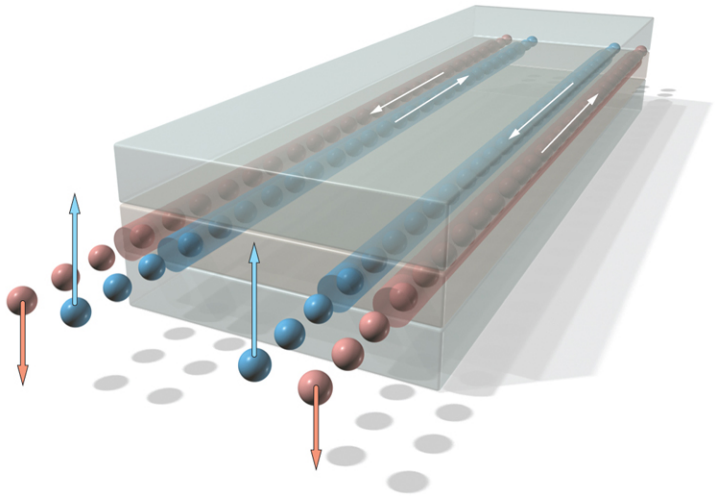


Fig. 8: *The Quantum Spin Hall Effect: In a two-dimensional insulator edge currents of opposite spin run in opposite directions.*

particular, the spin transport is manipulated through the sample and the magnetization $\vec{M}$ can be considered a stable quantity. However, in 1996 Berger and Slonczewski independently raised the question, whether there may be a retroaction of the spin current on the magnetization [31, 32]. The microscopic mechanism for such a retroaction is basically the same spin-dependent scattering process in the magnetic layers that leads to GMR and TMR. During the scattering event, there is a transfer of angular momentum (spin) $\vec{S}$ to the magnetic layer, which must be compensated by the magnetization $\vec{M}$. Berger and Slonczewski predicted that the angular momentum or *spin transfer* should lead to an excitation of the magnetization, which is then dissipated via spin waves.
The magnitude of the spin transfer depends mainly on the current density $\vec{j}$ and the spin scattering asymmetry β. Provided that the current density is sufficiently high ($j = 10^8 ... 10^9$ A/cm^2), the spin transfer may be strong enough to even reverse the magnetization direction. This has been demonstrated for the first time in Co/Cu/Co nanocontacts in 1999 [34]. Subsequent theoretical and experimental investigations established close links between the spin transfer process and the magnetization dynamics described by the Landau-Lifshitz-Gilbert equation. In the microsopic picture, the electrical current through the magnetic layer system causes a continuous spin transfer to the magnetic layer, which in turn acts like a torque onto $\vec{M}$. This torque drives the magnetization vector into a large-angle precessional motion, which may result into a magnetization reversal. Even a multi-level switching of the magnetization vector becomes possible in this way [33].

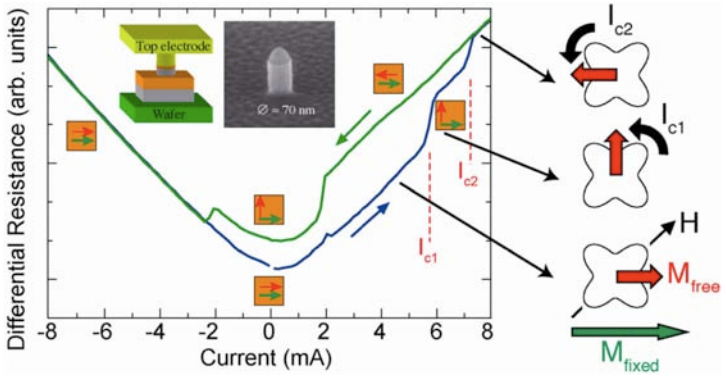


Fig. 9: *Multi-level current-induced magnetic switching in an epitaxial Fe/Ag/Fe nanopillar. Inset: Device structure and SEM micrograph of the free-standing nanopillar prior to application of the insulation and top electrode layer [33].*

The spin transfer effects offer a very exciting alternative to a magnetization reversal driven by external magnetic fields. Particularly in nanoscale magnetic elements, the current-induced

magnetization reversal becomes much more effective than field-induced magnetization reversal. This is a crucial issue in the realization of magnetic random access memories (MRAM). Another manifestation of the spin transfer processes is found in the current-driven domain wall motion [35]. This phenomenon forms the functional basis for novel magnetic logic and memory devices [36, 37, 38].

4.2 Materials

Spin-dependent transport phenomena are often linked to specific material systems and classes. Consequently, the wealth of spin effects which have emerged during the last twenty years has been closely accompanied by a strong efforts in materials synthesis and an enormous broadening of the materials basis. Nowadays it covers the full range from metals to insulators, from ferromagnets to antiferromagnets, and from thin films to molecules.

4.2.1 Halfmetallic ferromagnets

The spin polarization of the charge carriers in the vicinity of the Fermi level $P(E_F)$ is a central quantity in spintronics, as it often decides about the efficiency of spin transport and transfer processes and the effect magnitude, which can be obtained. This can be directly seen, for example, in the simple Jullière model of the tunneling magnetoresistance, the quantitative form of which is given in Eq. 3. If we increase the spin polarization of the electrode materials from, say $P_i = 0.4$ (a typical value for Co) to $P_i = 0.8$, the TMR ratio will rise by almost a factor of 10 from $\Delta R/R = 38\%$ to $\Delta R/R = 356\%$. Another illustrative example is found in spin injection from metal electrodes into semiconductors. Due to the large conductivity mismatch at the metal-semiconductor interface, electrons injected from a metallic ferromagnet into the semiconductor will be depolarized. In diffusive transport, the ferromagnet must provide a spin polarization of close to 100% in order to provide a sizable spin injection effect [39] or the resistivity mismatch must be bypassed by means of a tunneling barrier inserted at the interface [40].

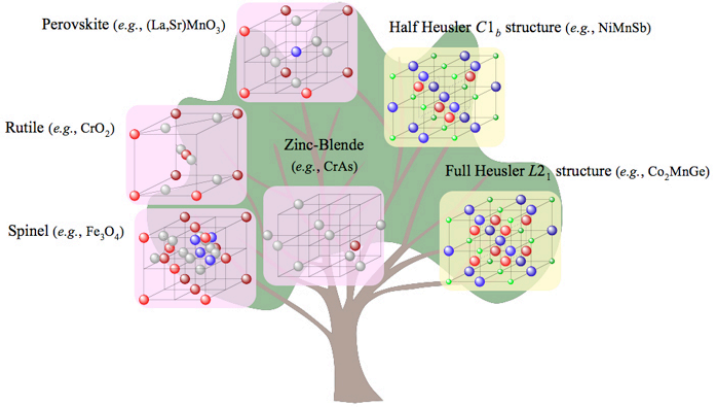


Fig. 10: *Diversity of materials with half-metallic properties discussed in the context of spintronics applications (taken from [41]).*

In order to obtain a pure spin state at the Fermi level, the electronic structure of the respective material has to have a gap around E_F in either the spin-up or spin-down subbands. This property of half-metallicity has been found by now in a number of different material classes [42] (Fig. 10). The formation of the gap relates to specific hybridization phenomena in the electronic structure. Well-known half-metallic ferromagnets are the half Heusler (XYZ) and full Heusler alloys (X_2YZ), the latter class providing some members with high Curie temperatures. Typical

representatives are NiMnSb and Co_2MnGe, with the gap being ascribed to Ni-Mn interactions in the first, and to symmetry-split Co minority spin states in the second material [43]. The gap in full-Heusler alloys is smaller than in half-Heuslers, i.e. the half-metallic state is more sensitive to chemical disorder and the formation of surface or interface electronic states.
Several classes of oxides also are exhibiting half-metallic properties. A strongly magnetic half-metal is CrO_2, its magnetic properties resulting from a combination of ferromagnetic superexchange and double-exchange mechanisms. The origin of the half metallicity rests in a gap in the minority spin bands forming between the occupied oxygen $2p$ states and empty chromium $3d$ states. The half-metallic character of CrO_2 depends mainly on the chromium valence and the crystal field splitting, and is thus considered to be relatively robust. The double-exchange mechanisms also determine the magnetic properties in perovskites, for example, $La_{1-x}Sr_xMnO_3$ (LSMO). These materials have been important in the context of colossal magnetoresistance [44]. Being an antiferromagnetic insulator for $x = 1$, a reduction to $x = 0.3$ leads to a ferromagnetic metallic state with half-metallic character, which manganese assuming a high-spin state. The crystal field splits the electronic levels into t_{2g} and e_g manifolds. The half-metallic character is caused by an interplay of exchange- and crystal-field splitting. The majority-spin t_{2g} state is rather localized and filled. The majority spin e_g state is more spread out and partially occupied, whereas the more localized minority spin t_{2g} states are positioned at higher energy, thus being unoccupied.
An example for a material with a majority spin gap is provided by Magnetite (see Sect. 1). Fe_3O_4 is a weakly ferromagnetic (ferrimagnetic) half metal with narrow bands and strong correlation effects. The half-metallic state is attributed to a single $3d$ minority spin electron hopping among the $3d$ majority spin cores on octahedral sites. A majority spin gap appears also in double perovskites, with Sr_2FeMoO_6 being one of the first examples investigated in spintronics context. As in the perovskites, the interplay of exchange- and crystal-field splitting is responsible for the half-metallic gap. In contrast to the perovskites, however, in the double perovskites the majority t_{2g} and e_g manifolds a fully occupied, while the minority t_{2g} states are partially filled. As a further difference, the Curie temperature may be significantly higher than in conventional perovskites.

4.2.2 Dilute magnetic semiconductors

A very interesting problem in spintronics is the simultaneous control of charge and spin in the same system, i.e. the combination of semiconducting and ferromagnetic properties in the same material. This combination provides two fascinating perspectives. First, assuming that the ferromagnetism in the material is directly related to the density of the charge carriers (*carrier-induced ferromagnetism*), the magnetic state may be easily manipulated by an electric field $\vec{E}$ acting on the charge density. Second, the semiconducting property is related to small band gaps and provides a direct pathway to interactions with photons, which may lead to spin-optoelectronics or optically induced magnetism. These opportunities to create multifunctional materials drive the research on dilute magnetic semiconductors (DMS). DMS are nonmagnetic semiconductors, which have been doped with transition metal atoms carrying a high atomic magnetic moment, typically chromium or manganese. The probably most intensely studied DMS and one of the very few proven examples for carrier-induced ferromagnetism is {Ga,Mn}As [45, 46]. The drawback of this material, however, is its low Curie temperature.
The origin of ferromagnetism in magnetic semiconductors is attributed to a Zener exchange mechanism. Mean-field theoretical treatments within the Zener picture have predicted partic-

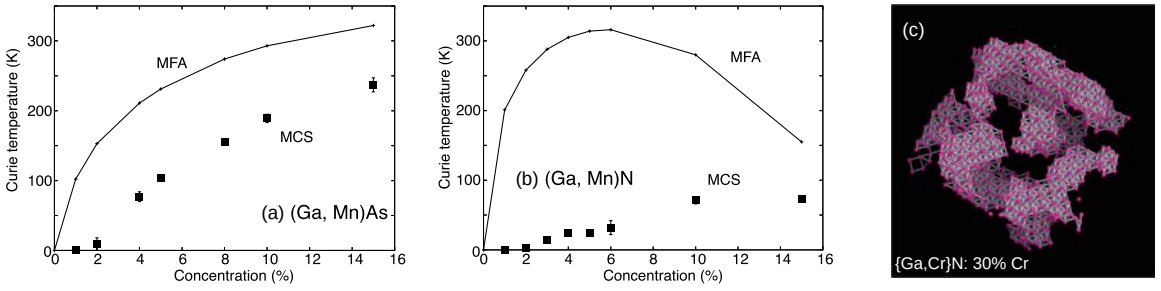


Fig. 11: *Curie temperatures of {Ga,Mn}As (a) and {Ga,Mn}N (b) calculated within the mean-field approximation (MFA) and by Monte-Carlo simulations (MCS). (c) Calculated spinodal nano-decomposition in {Ga,Mn(30%)}N, leading to $T_C^{calc} = 700$ K (purple dots mark Cr atoms). Taken from [48].*

ular III-V and II-VI compound semiconductors, namely GaN and ZnO to exhibit Curie temperatures well above 300 K [47]. These predictions have initiated a large materials research activity on DMS. Although a high Curie temperature has been observed for very high doping (in the range of several percent of the magnetic element), the origin of ferromagnetism in these systems is far from being understood, and frequently may be attributed to the formation of inhomogeneities and magnetic clusters. More elaborate theoretical models show the exchange interactions to be very short-ranged, which inhibits a magnetic percolation and results in very low Curie temperatures for a homogeneous spatial distribution of the magnetic dopant atoms [48] (Fig. 11a, b) – in contrast to the predictions from mean-field approximations. In order to explain the higher T_C-values observed in some experiments, for example, spinodal decomposition mechanisms have been proposed leading to the enrichment of nanosized regions with the magnetic dopant and thereby enabling a magnetic percolation in one (Konbu phase) or three dimensions (Dairiseki phase) [48].

4.2.3 Multiferroics

Multifunctionality is found also in other material classes. Of particular interest for spintronics is the phenomenon of *multiferroicity*, which describes the simultaneous presence of ferroic properties, such as ferromagnetism, ferroelectricity and ferroelasticity [49]. This situation involves a complex interplay of a spontaneous magnetization $\vec{M}$ with a spontaneous electric polarization $\vec{P}$ and ferroelastic lattice distortions $\vec{\epsilon}$ (Fig. 12). This coupling of physical parameters gives rise to fascinating effects, such as the *magnetoelectric effect*, which creates a magnetization $\vec{M}$ through an electric field $\vec{E}$. Vice versa an electric polarization $\vec{P}$ may be created by subjecting the material to a magnetic field $\vec{H}$. We may thus think about entirely new device paradigms, for example, electric field controlled magnetic data storage based on multifunctional tunneling contacts [50]. However, attempts to design multiferroics that combine ferromagnetism and ferroelectricity in the same phase have proved unexpectedly difficult, as either the ferromagnetic or ferroelectric Curie temperature T_C^M or T_C^E is usually well below room temperature. Exemplary materials, which have been experimentally addressed show a coexistence of antiferromagnetism and ferroelectricity, such as $HoMnO_3$ ($T_N^M = 72$ K, $T_C^E = 875$ K) or $BiFeO_3$ ($T_N^M = 640$ K, $T_C^E = 1100$ K). In $BiFeO_3$ a clear influence of electric fields on the antiferromagnetism has

been shown already [51].

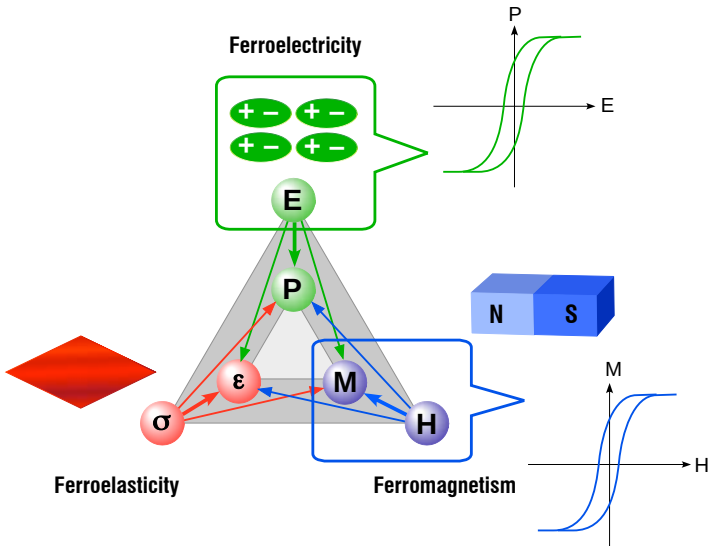


Fig. 12: *Interplay of elastic, electric and magnetic interactions in a multiferroic material, leading to interdependencies of the spontaneous magnetization $\vec{M}$, the spontaneous electric polarization $\vec{P}$ and the strain $\vec{\epsilon}$.*

The recent advances in the preparation and characterization of well-defined thin oxide films have enabled to a new research strategy in the field of multiferroicity, building upon layer stacks of ferromagnetic and ferroelectric thin films [52]. This approach promises an independent optimization of the ferroelectric and ferromagnetic properties and thus a chance to taylor the functionality of the thin film stack within a broader range of the physical parameters.

4.2.4 Carbon-Based Materials

Carbon as an element is a particularly interesting substance for spintronics, mainly because of two reasons. First, it provides a wide variety of allotropes (fullerenes, nanotubes, graphite, diamond), all of which have already been studied to some extent with respect to magnetism and spin-dependent transport. Second, carbon has a low nuclear number, rendering it a material with small spin-orbit coupling and therefore potentially good spin-transport properties. In fact, an efficient spin-injection into *carbon nanotubes* and spin transport through tubes of several 100 nm length has been successfully demonstrated already [53, 54]. This also includes the electric field control of the spin transport by means of electrical gates [55]. Carbon nanotubes can also be employed to define quantum dots and are therefore predestined to study the interplay of quantum and spin transport.
The allotrope graphite has a peculiar structure, as it is composed of *graphene* layers, which are only weakly bound to each other by v. d. Waals forces. Graphene consists of a monatomic planar sheet of carbon atoms, which are arranged in a hexagonal lattice. Representing a two-dimensional electronic system, graphene has a very unique electronic structure with the Fermi surface consisting of Dirac points [56]. Electrons which are excited into the conduction band close to the Fermi level thus behave like massless Dirac fermions. First studies on the spin transport through single graphene layers have recently revealed extremely large spin relaxation lengths in the μm range even at room temperature [57]. This finding makes graphene an extremely compelling system for spintronics applications.
Carbon is also the basis of all organic chemistry, which is employed to synthesize new molecular entities. Organic thin films have already found major applications in microelectronics, for example, in organic light-emitting diodes (OLED) used in modern displays [58]. Another area

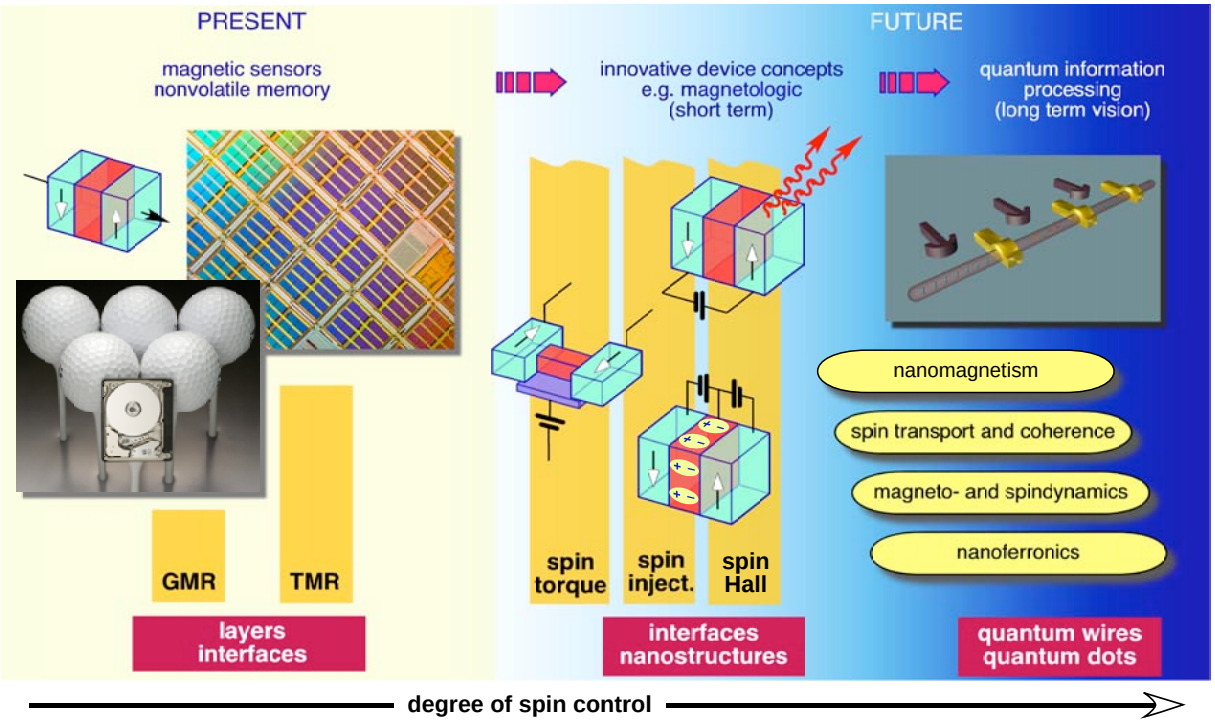


Fig. 13: *Present status and potential future developments in the field of spintronics.*

of current research concerns molecular electronics, in which the charge transport through single molecules is investigated [58]. On the other hand, the field of molecular magnetism would be unthinkable without metallorganic chemistry [59]. However, spin-dependent transport through organic layers and molecules is just moving into the focus now. There are considerable improvements in the theoretical understanding of spin transport phenomena through molecules [60]. Despite some successful observations of magnetoresistance effects in certain systems [61], however, the experimental realizations are still lagging behind. This is mainly due to the structural and chemical complexity of the interfaces in such organic layer systems and contacts, which often impairs the sample quality. Careful in-depth characterization of these interfaces will be needed to arrive at well-defined and reproducible synthesis procedures [62].

5 Future developments

Today, both giant and tunneling magnetoresistance have reached the market. The second generation of spintronics-based hard disk read heads is already based on magnetic tunnel junctions. Strong efforts are directed towards the development of MRAMs (Fig. 13). The future evolution of spintronics aims at improving the degree of spin control with a long-term perspective towards quantum information processing. On this pathway, there are four major topics to be dealt with. First, the reduction of lateral and vertical dimensions of the magnetic elements requires the investigation of fundamental questions of magnetism on the nanoscale. Second, the microscopic mechanisms of spin transport and transfer processes and particularly the issue of spin coherence needs to be explored. Third, the dynamical behavior of magnetic and spin systems must be understood in order to overcome the limitations of magnetic switching processes. Fourth, the area of nanoferronics addresses the microscopic mechanisms of multiferroicity and

the implementation of multifunctional components into spin transport devices. Of course, these four topics are strongly interrelated and will be cross-fertilizing each other.

References

[1] W. Gilbert, *De Magnete*, english translation by P. Fleury Mottelay (Dover Publications, New York, 1958).

[2] J. C. Maxwell, *Treatise of Electricity and Magnetism*, reprint of the 3rd edition (Dover Publications, New York, 1954).

[3] V. Poulsen, US PAT No. 661619 *"Method of Recording and Reproducing Sounds or Signals"*.

[4] M. Faraday, *On the magnetization of light and the illumination of magnetic lines of force*, Phil. Trans. R. Soc. London **136**, 1 (1846).

[5] W. Thomson, *On the electro-dynamic qualities of metals: effects of magnetization on the electric conductivity of nickel and of iron*, Proc. Roy. Soc. (London) **8**, 546, (1856/1857).

[6] N. F. Mott, *The Electrical Conductivity of Transition Metals*, Proc. Roy. Soc. (London), Ser. A **153**, 699 (1936).

[7] A. Fert and I. A. Campbell, *Two-current conduction in nickel*, Phys. Rev. Lett. **21**, 1190 (1968).

[8] I. A. Campbell and A. Fert, *Transport properties of ferromagnets*; in: *Ferromagnetic Materials Vol. 3*, ed. by E. P. Wohlfarth (North-Holland, Amsterdam, 1982).

[9] D. A. Thompson and J. S. Best, *The future of magnetic data storage technology*, IBM J. Res. Develop. **44**, 311 (2000).

[10] P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky, and H. Sowers, *Layered magnetic structures: Evidence for antiferromagnetic coupling of Fe layers across Cr interlayers*, Phys. Rev. Lett. **57**, 2442 (1986).

[11] S. S. P. Parkin, *Systematic variation of strength and oscillations period of indirect magnetic exchange coupling through 3d, 4d and 5d transition metals*, Phys. Rev. Lett. **67**, 3598 (1991).

[12] M. N. Baibich, J. M. Broto, A. Fert, F. N. V. Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, *Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices*, Phys. Rev. Lett. **61**, 2472 (1988).

[13] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, *Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange*, Phys. Rev. B **39**, 4828 (1989).

[14] M. Jullière, *Tunneling between ferromagnetic films*, Phys. Lett. **54A**, 225 (1975).

[15] T. Miyazaki and N. Tezuka, *Spin polarized tunneling in ferromagnet/insulator/ferromagnet junctions*, J. Magn. Magn. Mater. **151**, 403 (1995).

[16] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, *Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions*, Phys. Rev. Lett. **74**, 3273 (1995).

[17] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, *Spin-dependent tunneling conductance of Fe/MgO/Fe sandwiches*, Phys. Rev. B **63**, 054416 (2001).

[18] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, *Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers*, Nature Materials **3**, 862 (2004).

[19] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, *Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions*, Nature Materials **3**, 868 (2004).

[20] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, *Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature*, Appl. Phys. Lett. **93**, 082508 (2008).

[21] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. v. Molnár, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, *Spintronics: A spin-based electronics vision for the future*, Science **294**, 1488-1495 (2001).

[22] I. Appelbaum, B. Huang, and D. J. Monsma, *Electronic measurement and control of spin transport in silicon*, Nature **447**, 295 (2007).

[23] D. D. Awschalom, M. E. Flatté, and N. Samarth, *Spintronics*, Scientific American, June issue, 67 (2002).

[24] E. I. Rashba, *Properties of semiconductors with an extremum loop*, Sov. Phys. Solid State **2**, 1109 (1960).

[25] S. Datta and B. Das, *Electronic Analog of the electro-optic modulator*, Appl. Phys. Lett. **50**, 665 (1990).

[26] I. Zŭtic, J. Fabian, and S. C. Erwin, *Bipolar spintronics: Fundamentals and applications*, IBM J. Res. Develop. **50**, 121 (2006).

[27] M. I. Dyakonov and V. I. Perel, *Current-induced spin orientation of electrons in semiconductors*, Phys. Lett. A **35**, 459 (1971).

[28] Y. K. Kato, R. C. Myers, A. C. Gossard, and D. D. Awschalom, *Observation of the spin Hall effect in semiconductors*, Science **306**, 1910 (2004).

[29] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, *Quantum spin Hall effect and topological phase transition in HgTe quantum wells*, Science **314**, 1757 (2006).

[30] M. König, S. Wiedmann, C. Brune, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, *Quantum spin Hall insulator state in HgTe quantum wells*, Science **318**, 766 (2007).

[31] L. Berger, *Emission of spin waves by a magnetic multilayer traversed by a current*, Phys. Rev. B **54**, 9353 (1996).

[32] J. C. Slonczewski, *Current-driven excitation of magnetic multilayers*, J. Magn. Magn. Mater. **150**, L1 (1996).

[33] R. Lehndorff, M. Buchmeier, D. E. Bürgler, A. Kakáy, R. Hertel, and C. M. Schneider, *Asymmetric spin-transfer torque in single-crystalline Fe/Ag/Fe nanopillars*, Phys. Rev. B **76**, 214420 (2007).

[34] E. B. Myers, D. C. Ralph, J. A. Katine, R. N. Louie, and R. A. Buhrman, *Current-induced switching of domains in magnetic multilayer devices*, Science **285**, 867 (1999).

[35] J.-E. Wegrowe, D. Kelly, Y. Jaccard, P. Guittienne, and J.-P. Ansermet, *Current-induced magnetization reversal in magnetic nanowires*, Europhys. Lett. **45**, 626 (1999).

[36] D. A. Allwood, G. Xiong, M. D. Cooke, C. C. Faulkner, D. Atkinson, N. Vernier, and R. P. Cowburn, *Submicrometer ferromagnetic NOT gate and shift register*, Science **296**, 2003 (2002).

[37] S. S. P. Parkin, US PAT No. 10/458554 *"Shiftable magnetic shift register"*.

[38] S. S. P. Parkin, M. Hayashi, and L. Thomas, *Magnetic domain-wall racetrack memory*, Science **320**, 190 (2008).

[39] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip, and B. J. v. Wees, *Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor*, Phys. Rev. B **62**, R4790 (2000).

[40] E. I. Rashba, *Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem*, Phys. Rev. B **62**, R16267 (2000).

[41] see: `http://www.riken.jp/lab-www/nanomag/`.

[42] M. I. Katsnelson, V. Y. Irkhin, L. Chioncel, A. I. Lichtenstein, and R. A. de Groot, *Half-metallic ferromagnets: From band structure to many-body effects*, Rev. Mod. Phys. **80**, 315 (2008).

[43] I. Galanakis, P. H. Dederichs, and N. Papanikolaou, *Slater-Pauling behavior and origin of the half-metallicity of the full-Heusler alloys*, Phys. Rev. B **66**, 174429 (2002).

[44] R. von Helmolt, J. Wecker, B. Holzapfel, L. Schultz, and K. Samwer, *Giant negative magnetoresistance in perovskitelike $La_{2/3}Ba_{1/3}MnO_x$ ferromagnetic films*, Phys. Rev. Lett. **71**, 2331 (1993).

[45] H. Ohno, *Making nonmagnetic semiconductors ferromagnetic*, Science **281**, 951 (1998).

[46] F. Matsakura, H. Ohno, and T. Dietl, *III-V ferromagnetic semiconductors*; in *Handbook of Magnetic Materials Vol. 14*, ed. by K. H. J. Buschow (Elsevier Science Publishers, Amsterdam, 2002).

[47] T. Dietl, H. Ohno, F. Matsukura, J. Cibert, and D. Ferrand, *Zener model description of ferromagnetism in zinc-blende magnetic semiconductors*, Science **287**, 1019 (2000).

[48] H. Katayama-Yoshida, K. Sato, T. Fukushima, M. Toyoda, H. Kizaki, V. A. Dinh, and P. H. Dederichs, *Theory of ferromagnetic semiconductors*, phys. stat. sol. (a) **204**, 3 (2007).

[49] N. A. Spaldin and M. Fiebig, *The renaissance of magnetoelectric multiferroics*, Science **309**, 391 (2005).

[50] E. Y. Tsymbal and H. Kohlstedt, *Tunneling across a ferroelectric*, Science **313**, 181 (2006).

[51] T. Zhao, A. Scholl, F. Zavaliche, K. Lee, M. Barry, A. Doran, M. P. Cruz, Y. H. Chu, C. Ederer, N. A. Spaldin, R. R. Das, D. M. Kim, S. H. Baek, C. B. Eom, and R. Ramesh, *Electrical control of antiferromagnetic domains in multiferroic $BiFeO_3$ films at room temperature*, Nat. Mater. **5**, 823 (2006).

[52] R. Ramesh and N. A. Spaldin, *Multiferroics: progress and prospects in thin films*, Nat. Mater. **6**, 21 (2007).

[53] K. Tsukagoshi, B. W. Alphenaar, and H. Ago, *Coherent transport of electron spin in a ferromagnetically contacted carbon nanotube*, Nature (London) **401**, 572 (1999).

[54] B. Zhao, I. Mönch, H. Vinzelberg, T. Mühl, and C. M. Schneider, *Spin-coherent transport in ferromagnetically contacted carbon nanotubes*, Appl. Phys. Lett. **80**, 3144 (2002).

[55] S. Sahoo, T. Kontos, J. Furer, C. Hoffmann, M. Gräber, A. Cottet, and C. Schönenberger, *Electric field control of spin transport*, Nature Physics **1**, 99 (2005).

[56] A. H. C. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, *The electronic properties of graphene*, Rev. Mod. Phys. **81**, 109 (2009).

[57] N. Tombros, C. Jozsa, M. Popinciuc, H. T. Jonkman, and B. J. van Wees, *Electronic spin transport and spin precession in single graphene layers at room temperature*, Nature **448**, 571 (2007).

[58] See, for example, *Nanoelectronics and Information Technology*, ed. R. Waser (Wiley-VCH, Weinheim, 2005).

[59] D. Gatteschi, R. Sessoli, and J. Villain, *Molecular Nanomagnets* (Oxford University Press, Oxford, 2006).

[60] S. Sanvito and A. R. Rocha, *Molecular spintronics: The art of driving spin through molecules*, J. Comp. Theor. Nanoscience **3**, 642 (2006).

[61] Z. H. Xiong, D. Wu, Z. V. Vardeny, and J. Shi, *Giant magnetoresistance in organic spin-valves*, Nature **427**, 821 (2004).

[62] Z. Valy Vardeny, *Organics strikes back*, Nature Materials **8**, 91 (2009).

A 1 Electronic States in Solids

Gustav Bihlmayer
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Electrons in a lattice **2**
2.1 Translation symmetry 3
2.2 Nearly free electrons 6
2.3 Bandstructures of selected systems 8

3 Interacting electrons **10**
3.1 The Hartree and the Hartree-Fock approximation 11
3.2 Density functional theory 14
3.3 Extensions to DFT: the LDA+U method 17
3.4 Quasiparticles and the GW approximation 21

4 Relativistic effects **22**
4.1 Spin-orbit coupling 23
4.2 The Rashba- and the Dresselhaus effect 25
4.3 Magnetic anisotropy 28

1 Introduction

Since the word "spintronics" comprises both the spin and its carrier, the electron, in a word that promises application – like electronics – it is natural to start a series of lectures on spintronics with a general introduction on the electronic states in solids. In a periodic crystal these states are characterized by their momentum, their spin, and one additional quantum number (the band index) which set the playground for important effects in spintronics, like magnetoresistance or spin-flip scattering, and constitute the materials to build spin-filters, spin-valves, spin-FETs etc. In the last forty years "band-theory" became an indispensable tool for the description and characterization for many states of condensed matter, be it (band) insulators, semiconductors or metals. Also in so-called strongly correlated electron systems, where a subset of states shows correlation effects that are typically not covered by band-theory, a proper description of the more delocalized bands is indispensable since the unique material properties are determined by the coupling of the localized, correlated states to these bands. While these theories started as more conceptual tools for the understanding of the solid state, nowadays advances both in theory and (computer) simulation techniques made it possible to predict electronic (and magnetic) properties on a quantum mechanical basis. Most notably, density functional theory developed into a reliable tool for material scientists which aim at designing materials with selected properties. In particular for spintronic applications we have to meet very specific demands which can only be accomplished through an fundamental understanding of electronic states in solids.
Of course there are many materials – often with promising functional properties – that need a theoretical description beyond the one given in this lecture. Some of the subsequent lectures will cover part of these, like strongly correlated oxides. Mesoscopic quantum systems and their properties, effects on more mesoscopic length scales (e.g. Hall effects) etc. will be discussed at a later stage. Here, we deal mainly with the electronic ground state of matter from an atomistic point of view. The discussion of magnetic phenomena will be left out except in connection to spin-orbit coupling, where it natural to include also magnetic materials in the discussion. Other aspects will be covered by the contribution on the "electronic basis of magnetism".
In a first section we will look at the basic properties of electrons in an infinite periodic lattice. We will ignore their mutual interactions but incorporate the proper symmetry that defines the quantum numbers (constants of motion) of the system in a nonrelativistic context. The interaction between electrons is then the topic of the second section, where methods will be discussed to treat Coulomb- and exchange interactions e.g. Hartree-Fock and density functional theory or the GW approximation. While up to that point the spin of the electron was just considered as an additional quantum number, in the last section we will introduce the spin as a real vectorial quantity that can be manipulated in utilized as required for spintronic applications. Since spin is an inherently relativistic phenomenon, a theory on the level of the Dirac equation will be required, which is outlined in the appendix.

2 Electrons in a lattice

The electronic properties of a periodic solid are to a large part determined by the symmetry of the lattice that is formed by the atomic nuclei. They create a periodic potential in which the electrons (in particular the most loosely bound valence and conduction electrons) are moving. The constants of motion of such a system are determined by the symmetry, here in particular the translation symmetry in the crystal. Therefore, we start with a discussion of the symmetry

properties of a crystal that leads us to Bloch's theorem and illustrate these considerations for the case of a nearly free electron gas. To be specific, we present a couple of prototypical band structures to see how – even in the presence of electron-electron interactions – many properties can be inferred from the crystal symmetry. For a more extended treatment of these topics, the reader is referred to standard textbooks on solid state physics, e.g. Ref. [1].

2.1 Translation symmetry

The structure of an infinite periodic crystal can be considered as a space filling repetition of non-overlapping units cells in three dimensions. The origin, $\mathbf{R}$, of an unit cell can be written as

$$\mathbf{R_n} = \underline{A}\,\mathbf{n} \quad ; \quad \text{where} \quad \underline{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \text{and} \quad \mathbf{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} \tag{1}$$

where the n_i are integer numbers labeling the unit cell and $\underline{A}$ is the Bravais matrix of the crystal. In each unit cell $\mathbf{n}$, a finite number of atoms (denoted by α) are located at positions

$$\mathbf{r_{n},\alpha} = \mathbf{R_n} + \boldsymbol{\tau}_\alpha \tag{2}$$

and the vectors $\boldsymbol{\tau}_\alpha$ are called the basis of the lattice. A crystal, that can be described by equations (1) and (2), is invariant under an infinite set of symmetry operations which can be classified as translations, $\mathcal{T}$, and (proper and improper) rotations, $\mathcal{R}$, and combinations of these two. Here, we will focus on the translations, which act on some function in real space, $f(\mathbf{r})$:

$$\mathcal{T}_{\mathbf{R_n}} f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R_n})\,. \tag{3}$$

The Hamiltonian of the electrons in a periodic solid consists of three parts: the kinetic energy of the electrons, T, their mutual Coulomb repulsion, $V_{\mathrm{e-e}}$ and the potential created by the nuclei, V_{ext}. The first two parts of the Hamiltonian are invariant with respect to any translation, but the latter term will only be unchanged if the translation vector is a lattice vector, $\mathbf{R_n}$:

$$\mathcal{T}_{\mathbf{R_n}} V_{\mathrm{ext}}(\mathbf{r}) = V_{\mathrm{ext}}(\mathbf{r} + \mathbf{R_n}) = V_{\mathrm{ext}}(\mathbf{r})\,. \tag{4}$$

Therefore, the total Hamiltonian, $\mathcal{H}$, commutes with the translation operator $\mathcal{T}_{\mathbf{R_n}}$ and both operators will have common eigenfunctions.
To find the eigenvalues γ of the translation operator, we consider the successive action of two translation operators on a function:

$$\begin{aligned} \mathcal{T}_{\mathbf{R_{n'}}} \mathcal{T}_{\mathbf{R_n}} f(\mathbf{r}) &= \mathcal{T}_{\mathbf{R_{n'}}} \gamma(\mathbf{R_n}) f(\mathbf{r}) = \gamma(\mathbf{R_{n'}})\gamma(\mathbf{R_n}) f(\mathbf{r}) \\ \mathcal{T}_{\mathbf{R_{n'}}} \mathcal{T}_{\mathbf{R_n}} f(\mathbf{r}) &= \mathcal{T}_{\mathbf{R_{n'}}+\mathbf{R_n}} f(\mathbf{r}) = \gamma(\mathbf{R_{n'}} + \mathbf{R_n}) f(\mathbf{r})\,. \end{aligned} \tag{5}$$

The fact that $\gamma(\mathbf{R_{n'}})\gamma(\mathbf{R_n}) = \gamma(\mathbf{R_{n'}}+\mathbf{R_n})$ suggests, that the vectors $\mathbf{R}$ appear in an exponential form in γ, i.e.

$$\gamma(\mathbf{R_n}) = \mathrm{e}^{\mathbf{R_n}\cdot\mathbf{P}}\,. \tag{6}$$

If we consider, that f is a normalized function, $\phi(\mathbf{r})$, that should not grow or vanish exponentially in an infinitely extended solid by applying a translation, we can assume that $\mathbf{P}$ is an

imaginary quantity and write it as $i\mathbf{k}$. We can use this vector $\mathbf{k}$ to label the functions ϕ according to

$$\mathcal{T}_{\mathbf{R_n}}\phi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{i\mathbf{R_n}\cdot\mathbf{k}}\phi_{\mathbf{k}}(\mathbf{r})\,. \tag{7}$$

The matrix elements of the Hamilton operator with two such functions should be invariant to a lattice translation, i.e.

$$\begin{aligned}\langle\phi_{\mathbf{k}'}(\mathbf{r})|\mathcal{H}|\phi_{\mathbf{k}}(\mathbf{r})\rangle &= \langle\mathcal{T}_{\mathbf{R_n}}\phi_{\mathbf{k}'}(\mathbf{r})|\mathcal{H}|\mathcal{T}_{\mathbf{R_n}}\phi_{\mathbf{k}}(\mathbf{r})\rangle = \left\langle \mathrm{e}^{i\mathbf{R_n}\cdot\mathbf{k}'}\phi_{\mathbf{k}'}(\mathbf{r})|\mathcal{H}|\mathrm{e}^{i\mathbf{R_n}\cdot\mathbf{k}}\phi_{\mathbf{k}}(\mathbf{r})\right\rangle \\ &= \mathrm{e}^{i\mathbf{R_n}\cdot(\mathbf{k}-\mathbf{k}')}\,\langle\phi_{\mathbf{k}'}(\mathbf{r})|\mathcal{H}|\phi_{\mathbf{k}}(\mathbf{r})\rangle \end{aligned}\tag{8}$$

which means, that either the exponential factor is unity or the matrix element must vanish. From the translation symmetry properties of the lattice we can thus conclude, that we only get non-vanishing matrix elements, if $\mathbf{R_n}\cdot(\mathbf{k}-\mathbf{k}') = 2\pi N$, if N is some integer number. The latter condition can be brought into a slightly different form, if we write

$$\mathbf{R_n} = \underline{A}\,\mathbf{n} \quad ; \quad \mathbf{k}-\mathbf{k}' = \underline{B}\,\mathbf{m} \quad \text{and} \quad \underline{A}\,\underline{B} = 2\pi\underline{1} \tag{9}$$

where $\underline{1}$ is the 3×3 unit matrix and $\underline{B}$ defines a lattice, where the lattice vectors are given by $\mathbf{K_m} = \underline{B}\,\mathbf{m}$ for integer vectors $\mathbf{m}$. This lattice is called the reciprocal lattice and $\mathbf{K_m}$ is called a reciprocal lattice vector. To get non-vanishing matrix elements, $\mathbf{k}-\mathbf{k}'$ must be a reciprocal lattice vector.
This result will help us in two ways: firstly, this results holds for all kinds of operators, that commute with the translation operator, i.e. $\langle\phi_{\mathbf{k}'}|\mathcal{O}|\phi_{\mathbf{k}}\rangle$ is only non-vanishing, if $\mathbf{k}-\mathbf{k}' = \mathbf{K_m}$. E.g. if $\mathcal{O}$ describes some excitation of the crystal and the ϕ's are wavefunctions of the ground- and excited state, we can derive selection rules from this symmetry. Secondly, if we consider that $\phi_{\mathbf{k}}$ is a trial function for the solution of the Dirac or the Schrödinger equation, we can immediately block-diagonalize the Hamiltonian in blocks of wavefunctions, where the difference of two wavevectors $\mathbf{k}$ and $\mathbf{k}'$ is a reciprocal lattice vector. This allows us to restrict the values of $\mathbf{k}$ to the smallest ones in each block and to use this vectors as quantum numbers that label the wavefunctions in the solid. The volume filled by these $\mathbf{k}$-vectors is called Brillouin zone and it is the equivalent of the Wigner-Seitz cell in real space, but now in reciprocal space.
The reciprocal lattice is particularly useful to describe lattice periodic functions:

$$u(\mathbf{r}) = \sum_{\mathbf{K_m}} \mathrm{e}^{i\mathbf{K_m}\cdot\mathbf{r}}u(\mathbf{K_m}) \quad \text{since} \quad u(\mathbf{r}+\mathbf{R_n}) = \sum_{\mathbf{K_m}} \mathrm{e}^{i\mathbf{K_m}\cdot(\mathbf{r}+\mathbf{R_n})}u(\mathbf{K_m}) = u(\mathbf{r})\,. \tag{10}$$

We can now write the eigenfunctions of the translation operator according to equation (7) as

$$\mathcal{T}_{\mathbf{R_n}}\phi_{\mathbf{k}}(\mathbf{r}) = \mathcal{T}_{\mathbf{R_n}}(\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r})) = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{R_n}}\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{R_n}}\phi_{\mathbf{k}}(\mathbf{r})\,. \tag{11}$$

Functions of the form $\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r})$ are called Bloch functions. They are the eigenfunctions of the translation operator $\mathcal{T}_{\mathbf{R_n}}$ and play an important role in the electron theory of periodic solids. Since we know that $\mathcal{T}_{\mathbf{R_n}}$ and the Hamiltonian commute, also the eigenfunctions of $\mathcal{H}$ can be written in this form:

$$\mathcal{H}\phi_{\mathbf{k},\nu}(\mathbf{r}) = \varepsilon_{\mathbf{k},\nu}\phi_{\mathbf{k},\nu}(\mathbf{r}) \quad ; \quad \phi_{\mathbf{k},\nu}(\mathbf{r}) = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k},\nu}(\mathbf{r})\,. \tag{12}$$

This is called Bloch's theorem. We introduced an additional quantum number ν to distinguish the different solutions that belong to the same vector $\mathbf{k}$. They will correspond to different values of $u_{\mathbf{k},\nu}(\mathbf{K_m})$ in equation (10).

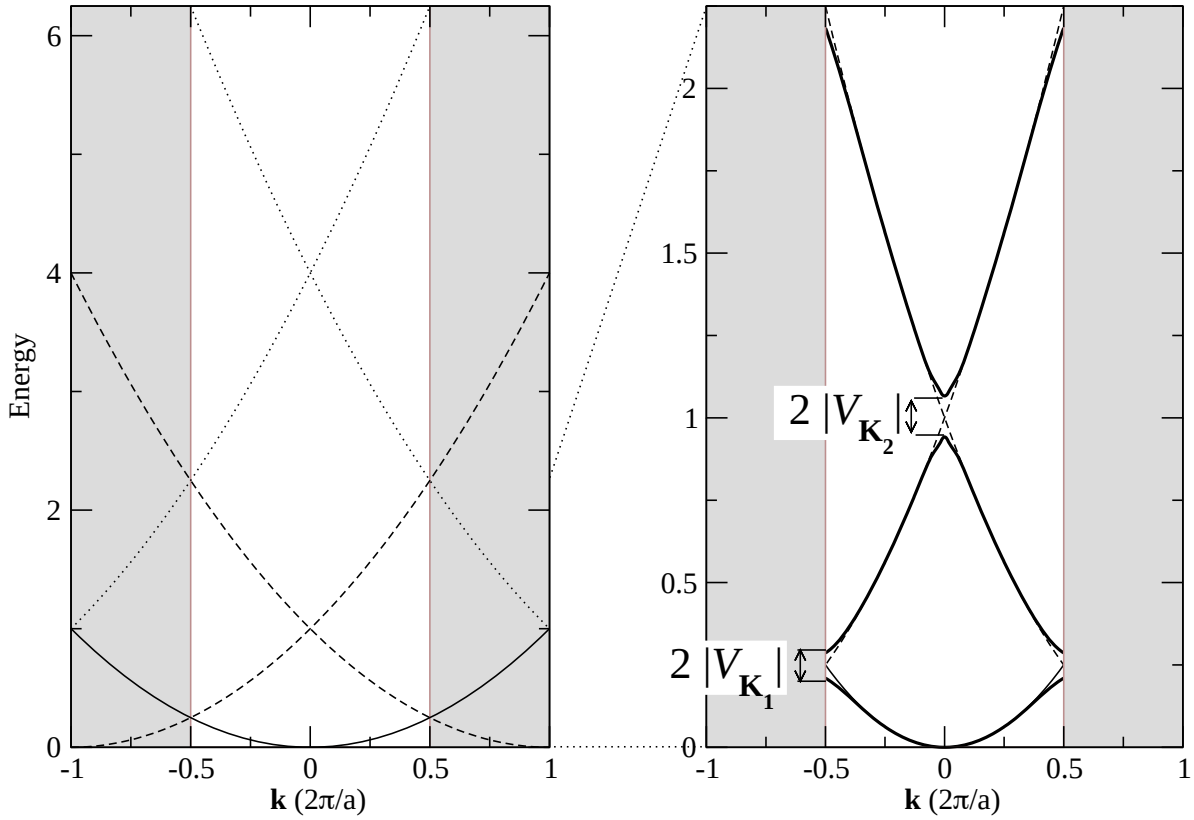


Fig. 1: *Band structure of a free electron gas (left) with three parabolic bands, originating from reciprocal lattice points outside the Brillouin zone (BZ, white). In the case of a nearly-free electron gas, i.e. in the presence of a periodic potential, the degeneracy of the bands at the BZ boundaries and at the origin will be lifted (thick lines, right panel).*

Let us illustrate these points with the simplest possible example, a non-interacting electron gas in an uniform potential, V_0. Since the wavefunction is separable, it is sufficient to study the Hamiltonian for a single particle, that is of the form

$$\mathcal{H} = -\frac{\hbar^2}{2m_\mathrm{e}}\boldsymbol{\nabla}^2 + V_0\,. \tag{13}$$

Omitting the constant potential and using atomic units ($\hbar = 1$, $m_\mathrm{e} = 1$) we can write the Schrödinger equation with Eq. (10) and Eq. (12) as

$$\begin{aligned} -\frac{1}{2}\boldsymbol{\nabla}^2\left(\sum_{\mathbf{K_m}} \mathrm{e}^{i(\mathbf{k}+\mathbf{K_m})\cdot\mathbf{r}} u_{\mathbf{k},\nu}(\mathbf{K_m})\right) &= \\ \frac{1}{2}\sum_{\mathbf{K_m}}(\mathbf{k}+\mathbf{K_m})^2 \mathrm{e}^{i(\mathbf{k}+\mathbf{K_m})\cdot\mathbf{r}} u_{\mathbf{k},\nu}(\mathbf{K_m}) &= \varepsilon_{\mathbf{k},\nu}\sum_{\mathbf{K_m}} \mathrm{e}^{i(\mathbf{k}+\mathbf{K_m})\cdot\mathbf{r}} u_{\mathbf{k},\nu}(\mathbf{K_m})\,. \end{aligned} \tag{14}$$

To fulfill Eq. (14) for each $\mathbf{K_m}$ we get a solution

$$\varepsilon_{\mathbf{k},\nu} = \frac{1}{2}(\mathbf{k}+\mathbf{K_m})^2\,, \tag{15}$$

i.e. the eigenvalues can be described as parabolas in $\mathbf{k}$-space originating at reciprocal lattice points. This is illustrated in figure 1 in one dimension: if we restrict our description to the first

Brillouin zone (BZ), we observe that for each **k**-vector we obtain an infinite, but discrete set of eigenvalues. At each **k**-point, we can label these eigenvalues with the band index ν increasing with energy. A set of eigenvalues with the same index ν is called a band. At the boundaries of the BZ we observe band crossings, which will – in general – disappear for more realistic potentials (see next subsection).
Before we study the effect of a non-constant potential, it is instructive to study the problem of the free electron gas with a different choice for the wavefunctions. Although planewaves are a perfect solution to the free electron problem, in many real situations the eigenfunctions can be regarded to be derived from atomic wavefunctions, $s, p, d,$ or f-like. Of course in a crystal they form linear combinations that have to fulfill Bloch's theorem, i.e. if we start from atomic functions $\chi(\mathbf{r})$ centered on lattice sites $\mathbf{R_n}$, we choose a form

$$\phi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{i\mathbf{kr}} \sum_{\mathbf{n}} \chi(\mathbf{r} - \mathbf{R_n}) \,. \tag{16}$$

E.g. if χ is a spherical s-like function, it is modulated by the **k**-dependent "Bloch factor" with a period of $2\pi/k$ throughout the crystal. In the spirit of the tight-binding approximation, it is convenient to write

$$\phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{n}} \mathrm{e}^{i\mathbf{kR_n}} \chi(\mathbf{r} - \mathbf{R_n}) \,, \tag{17}$$

where N is the number of lattice sites in the sum. This allows us to estimate the energy as the expectation value of the Hamiltonian:

$$\varepsilon(\mathbf{k}) = \frac{1}{N} \sum_{\mathbf{n},\mathbf{n'}} \left\langle \mathrm{e}^{i\mathbf{kR_n}} \chi(\mathbf{r} - \mathbf{R_n}) | \mathcal{H} | \mathrm{e}^{i\mathbf{kR_{n'}}} \chi(\mathbf{r} - \mathbf{R_{n'}}) \right\rangle = \sum_{\mathbf{n}} \mathrm{e}^{i\mathbf{kR_n}} \left\langle \chi(\mathbf{r}) | \mathcal{H} | \chi(\mathbf{r} - \mathbf{R_n}) \right\rangle \,. \tag{18}$$

Assuming a linear chain of atoms with a lattice constant a, where only the nearest neighbor atoms have significant overlap, this reduces to

$$\varepsilon(\mathbf{k}) = \langle \chi(\mathbf{r}) | \mathcal{H} | \chi(\mathbf{r}) \rangle + \mathrm{e}^{i\mathbf{ka}} \langle \chi(\mathbf{r}) | \mathcal{H} | \chi(\mathbf{r} - \mathbf{a}) \rangle + \mathrm{e}^{-i\mathbf{ka}} \langle \chi(\mathbf{r}) | \mathcal{H} | \chi(\mathbf{r} + \mathbf{a}) \rangle \,. \tag{19}$$

Since the nearest-neighbor integrals are identical, we can write

$$\varepsilon(\mathbf{k}) = \alpha + 2\beta \cos(ka) \quad \text{where} \quad \alpha = \langle \chi(\mathbf{r}) | \mathcal{H} | \chi(\mathbf{r}) \rangle \quad \text{and} \quad \beta = \langle \chi(\mathbf{r}) | \mathcal{H} | \chi(\mathbf{r} \pm \mathbf{a}) \rangle \,. \tag{20}$$

For a s-type wavefunction, β is negative and ε is lowest at $k = 0$. With increasing k the energy increases and reaches its maximum at $k = \pi/a$. The situation is similar to planewave description, although the energy is too high now at the zone boundary, where the Bloch factor forces the s-type wavefunction to disappear at every second atom, since it cannot describe the node at this position properly. At this k-point, a p-type wavefunction would be required. Such "odd" functions have a positive β and their energy decreases from $k = 0$ towards the zone boundary, like indicated in the second band in figure 1. As we will see later, realistic bandstructures of simple metals indeed start at low energies with a parabolic, s-type band, followed by three inverted parabolas that correspond to p-type states. Many properties of the valence electrons of these metals can be found in the simple free-electron picture of this section.

2.2 Nearly free electrons

Of course, the electrons in a crystal feel a periodic potential that differs considerably from out constant model potential. The attractive potential of the nuclei is partially screened by energetically low-lying core electrons, i.e. the valence electrons "feel" a potential that is smoothed by

electrons that are bound closely by the nuclei. This is the external potential, $V_{\rm ext}$, encountered in Eq. (4). Since it has lattice periodicity, it can also be expanded in reciprocal lattice vectors:

$$V_{\rm ext} = \sum_{\mathbf{K_m}} \mathrm{e}^{i\mathbf{K_m}\cdot\mathbf{r}} V(\mathbf{K_m})\,. \tag{21}$$

Adding this potential term to the Hamiltonian modifies Eq. (14) and we get

$$\sum_{\mathbf{K_m}} \left(\varepsilon_{\mathbf{k},\nu} - \frac{1}{2}(\mathbf{k}+\mathbf{K_m})^2 \right) u_{\mathbf{k},\nu}(\mathbf{K_m}) \mathrm{e}^{i(\mathbf{k}+\mathbf{K_m})\mathbf{r}} = \sum_{\mathbf{K_m}} \sum_{\mathbf{K'_m}} V(\mathbf{K'_m}) u_{\mathbf{k},\nu}(\mathbf{K_m}) \mathrm{e}^{i(\mathbf{k}+\mathbf{K_m}+\mathbf{K'_m})\mathbf{r}}. \tag{22}$$

Introducing $\mathbf{K''_m} = \mathbf{K'_m} + \mathbf{K_m}$, we write the right part of Eq. (22) as

$$\sum_{\mathbf{K''_m}} \sum_{\mathbf{K'_m}} V(\mathbf{K'_m}) u_{\mathbf{k},\nu}(\mathbf{K''_m} - \mathbf{K'_m}) \mathrm{e}^{i(\mathbf{k}+\mathbf{K''_m})\mathbf{r}}. \tag{23}$$

Substituting back $\mathbf{K_m} \leftarrow \mathbf{K''_m}$ and comparing the coefficients with the left side of Eq. (22) we obtain

$$\left(\varepsilon_{\mathbf{k},\nu} - \frac{1}{2}(\mathbf{k}+\mathbf{K_m})^2 \right) u_{\mathbf{k},\nu}(\mathbf{K_m}) = \sum_{\mathbf{K'_m}} V(\mathbf{K'_m}) u_{\mathbf{k},\nu}(\mathbf{K_m} - \mathbf{K'_m})\,. \tag{24}$$

For the case of a constant potential, we set $V(\mathbf{0}) = V_0$ and all other Fourier coefficients to zero. In this case, Eq. (24) reduces to

$$\left(\varepsilon_{\mathbf{k},\nu} - \frac{1}{2}(\mathbf{k}+\mathbf{K_m})^2 \right) u_{\mathbf{k},\nu}(\mathbf{K_m}) = V_0 u_{\mathbf{k},\nu}(\mathbf{K_m}) \tag{25}$$

which corresponds, apart from an additional constant V_0, to Eq. (15). Eigenfunctions are again planewaves with wave vector $\mathbf{K_m}$, i.e. the Fourier coefficients for the expansion of the wavefunction, $u_{\mathbf{k},\nu}(\mathbf{K_m})$ for a certain state ν are unity for a specific $\mathbf{K_m}$ and zero otherwise. If we consider $u_{\mathbf{k}}$ as a matrix with dimensions ν and $\mathbf{K_m}$, we find that for the case of a constant potential $\underline{u_{\mathbf{k}}}$ is the unit matrix $\underline{1}$ (we denote matrix quantities here and in the following with an underline).

If the potential is of general shape, the expansion coefficients $u_{\mathbf{k},\nu}(\mathbf{K_m})$ can be obtained from Eqs. (24). We can rewrite these equations using $\frac{1}{2}(\mathbf{k}+\mathbf{K_m})^2 = \varepsilon^0_{\mathbf{k},\mathbf{K_m}}$ in the form

$$\left(\varepsilon_{\mathbf{k},\nu} - \varepsilon^0_{\mathbf{k},\mathbf{K_m}} \right) u_{\mathbf{k},\nu}(\mathbf{K_m}) = \sum_{\mathbf{K'_m}} V(\mathbf{K'_m} - \mathbf{K_m}) u_{\mathbf{k},\nu}(\mathbf{K'_m})\,. \tag{26}$$

If we write V in matrix form and consider u and ε^0 as vectors, this equation can be rewritten in the form of a standard eigenvalue problem:

$$\left(\underline{V} + \boldsymbol{\varepsilon}^0_{\mathbf{k}} \underline{1} \right) \mathbf{u}_{\mathbf{k},\nu} = \varepsilon_{\mathbf{k},\nu} \mathbf{u}_{\mathbf{k},\nu}\,. \tag{27}$$

Let us finally analyze, how a weakly varying potential affects the bandcrossings at $k = \pi/a$ in figure 1. The lowest bandcrossing is formed by a parabola originating at $\mathbf{K_m} = \mathbf{0}$ and a parabola that has its minimum at $\mathbf{K_m} = \mathbf{K}_1$. In the vicinity of the crossing, we denote the (unperturbed) eigenvalues of these two states as $\varepsilon^0_+ = V_0 + \varepsilon^0_{\mathbf{0}}$ and $\varepsilon^0_- = V_0 + \varepsilon^0_{\mathbf{K}_1}$. Ignoring all

other states, the potential matrix $\underline{V}$ will be a 2×2 matrix which has diagonal elements V_0 and off-diagonal elements $V_{-\mathbf{K}_1}$ and $V_{\mathbf{K}_1}$. Eq. (27) has then the form

$$\begin{pmatrix} V_0 + \varepsilon^0_{\mathbf{0}} & V_{\mathbf{K}_1} \\ V_{-\mathbf{K}_1} & V_0 + \varepsilon^0_{\mathbf{K}_1} \end{pmatrix} \begin{pmatrix} u_{\mathbf{0}} \\ u_{\mathbf{K}_1} \end{pmatrix} = \varepsilon \begin{pmatrix} u_{\mathbf{0}} \\ u_{\mathbf{K}_1} \end{pmatrix} . \tag{28}$$

This problem can be solved by setting the determinant to zero:

$$\begin{vmatrix} \varepsilon^0_+ - \varepsilon & V_{\mathbf{K}_1} \\ V_{-\mathbf{K}_1} & \varepsilon^0_- - \varepsilon \end{vmatrix} = 0 \tag{29}$$

resulting in

$$\varepsilon_{1,2} = \frac{\varepsilon^0_+ + \varepsilon^0_-}{2} \pm \sqrt{\left(\frac{\varepsilon^0_+ - \varepsilon^0_-}{2}\right)^2 + |V_{\mathbf{K}_1}|^2} . \tag{30}$$

In case the two eigenvalues ε^0_+ and ε^0_- coincide, the non-constant potential will lift this degeneracy and lead to a splitting of $\pm|V_{\mathbf{K}_1}|$. If $|V_{\mathbf{K}_1}|$ is small compared to $\varepsilon^0_+ - \varepsilon^0_-$, e.g. $\mathbf{k}$ is far from the zone boundary, the effect will be small. Likewise, the interaction with other bands, that are energetically far away, will be small and our assumption to consider just two bands near a crossing will be justified.
Of course, in realistic bandstructures more than two bands can be energetically close and not all potential Fourier coefficients will be small, so that more complicated interactions will occur. In this case, it is necessary to solve Eq. (27) in full.

2.3 Bandstructures of selected systems

To visualize the effects of the crystal lattice, we shortly discuss here two simple examples of bandstructures. First, we consider the sodium crystal, which is body-centered cubic and has a very low valence electron concentration. In contrast, the face-centered cubic copper crystal contains eleven valence electrons, some of them rather closely bound, being rather far from the free electron gas limit.
First we discuss the electronic structure of sodium, which is a simple metal from the first column of the periodic table. All the metals in this row crystallize in a bcc lattice and with a single valence electron per atom. Thus, the Na bandstructure is prototypical for these elements, like K or Rb [2]. Due to its low electron density (the s-electrons are typically very delocalized in metals), it is already very close to an almost free electron gas in a periodic lattice.
The bandstructure in figure 2 has been obtained by density functional theory (DFT), that will be outlined in the next section (3.2). As can be seen from this figure, the bottom of the occupied band is almost parabolic, as expected for a free-electron like dispersion. We see, however, that gaps are opening at the boundaries of the Brillouin zone, e.g. at the N-point. Above the s-band three downward dispersing p-bands can be observed with very different dispersions. E.g. in ΓN direction, only one band reaches down to the s-band, while the other two bands remain above 8 eV.
We should notice, that even in this very simple case one has to be careful when comparing singe-particle eigenvalues (figure 2) with experimental photoemission results: While the Fermi surface, i.e. the surface that is created by all states $\mathbf{k}, \nu$ which fulfill the condition $\varepsilon_{\mathbf{k},\nu} = E_{\mathrm{F}}$, is in very good agreement, the bottom of the s-band is too low as compared to the experiment. Photoemission results show that 2.5 eV are required to excite an electron at the Γ-point to the

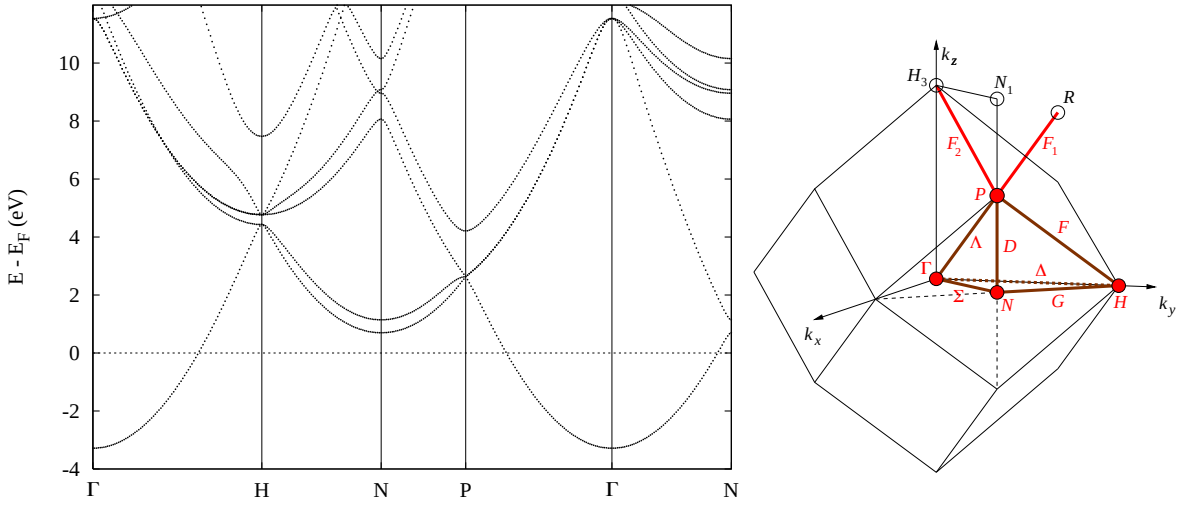


Fig. 2: *Band structure of sodium (left) and the reciprocal unit cell of the bcc lattice with high symmetry points and lines (right). The right image was taken from the Bilbao Crystallographic Server [3].*

Fermi level, while the DFT eigenvalue is at about -3.2 eV [4]. Methods to calculate excitation spectra will be shortly discussed in subsection 3.4.

As a second example, we show in figure 3 the band structure and Fermi surface of fcc Cu. Again, we can see that the bottom of the valence band is formed by a parabolic s-type band, but between -1.5 and -5.0 eV flat d-bands are crossing. Due to symmetry some of these five bands are degenerate at the displayed high-symmetry lines and points, e.g. at the Γ-point a 3-fold and a 2-fold degenerate state (corresponding to t_{2g} and e_g symmetry) can be seen. It might be interesting to notice, that the bandstructures of the three coinage metals Cu, Ag, and Au are rather similar, only in the case of Ag the d-band is shifted about 2 eV further away from the Fermi level. This accounts for the different colors of these metals, i.e. the smaller the energy

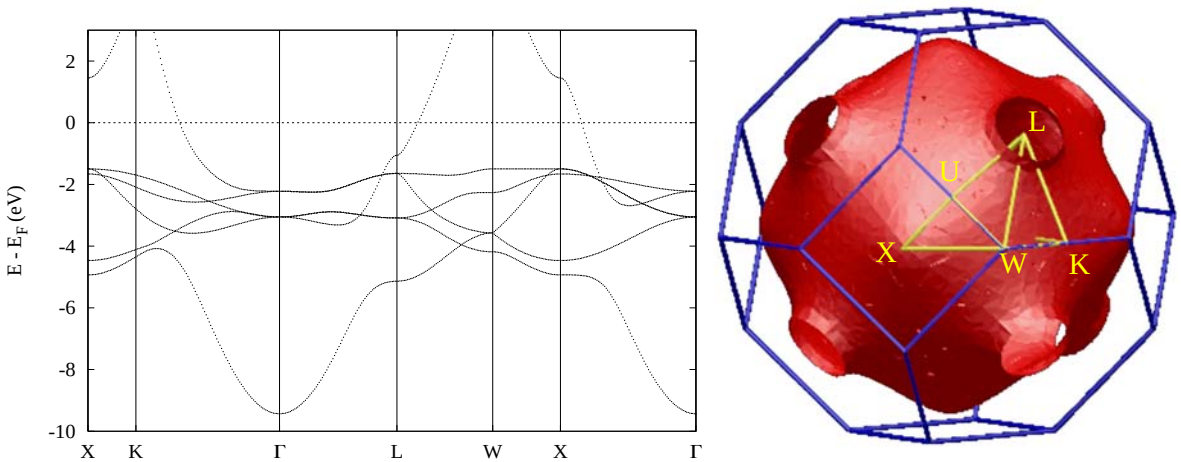


Fig. 3: *Band structure of copper (left) the Fermi surface (right, red) with the reciprocal unit cell of the fcc lattice in blue and the high symmetry points and lines (yellow).*

difference between Fermi level and the d-band, the more reddish the color. Density functional theory reproduces this trend reliably, provided that in case of Au scalar-relativistic corrections (cf. Appendix) are included in the Hamiltonian [5]. The Fermi surface of Cu is not spherical, as in the case of Na, but shows characteristic necks which can also be observed in De Haas-Van Alphen oscillations experimentally. Again, the electronic structure at the Fermi level is quite reliably described by the calculation.

3 Interacting electrons

In any realistic calculation, we cannot simply ignore the mutual Coulomb repulsion of the electrons and we have to include the term

$$V_{\mathrm{e-e}} = \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{31}$$

in the Hamiltonian (since we work in atomic units, $e^2 = 1$). This destroys the separability of the wavefunction and the straightforward quantum mechanical treatment of the electronic degrees of freedom is limited to a very small number of particles. This is mainly due to the appearance of the manybody wavefunction $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, which contains a tremendous amount of information and is difficult to handle for N larger than a few dozen or so.
One can try to construct Ψ from single particle wavefunctions and combine them to manybody wavefunctions of different complexity: a simple product Ansatz, $\Psi = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\ldots\phi_N(\mathbf{r}_N)$, leads to the so called Hartree approximation. Unfortunately, this form of the wavefunction is not compatible with the Pauli principle, i.e. interchanging two arguments of Ψ does not lead to $-\Psi$. In the Hartree-Fock (HF) method, the wavefunction has the form of a determinant of a $N \times N$ matrix of single particle wavefunctions $\phi_\mu(\mathbf{r}_\nu)$ with $1 \leq \mu, \nu \leq N$, which ensures that the Pauli principle is fulfilled. Therefore, the HF method leads to better results (e.g. binding energies) than the Hartree method. The energy contribution missing in the latter method as compared to the former one is called *exchange energy*. Although the HF method is numerically quite complicated, the obtained energies are still often quite far from the true ground state energies. What is missing is called *correlation energy* and the results can be improved by e.g. constructing the manybody wavefunction as a linear combination of many determinant functions. These so called configuration-interaction (CI) methods can be systematically improved, but the numerical effort is huge. While the HF method scales nominally like N^4, calculational schemes that include correlation scale with N^5 (second order Møller-Plesset perturbation theory) or N^7 (Coupled Cluster theory). A good account of these quantum-chemical methods can be found in the article of V. Staemmler in Ref. [6].
A completely different approach is taken by the density functional theory (DFT): although in most cases the true wavefunction is impossible to access, this poses no fundamental limitation since normally we are not interested in Ψ, but in a limited number of physical observables. Density functional theory therefore bypasses the troublesome manybody wavefunction and starts directly from the density of the particles in question (in our case electrons) allowing thereby the treatment of a large number of particles.
As we will see, DFT is quite suitable to describe the many aspects of the electronic structure. Also structural properties, like lattice parameters, that can be obtained from total energies are very well accessible in DFT, since the total energy is a quantity that has a definite meaning in this theory. In metals also other electronic properties are reproduced well, mainly due to the

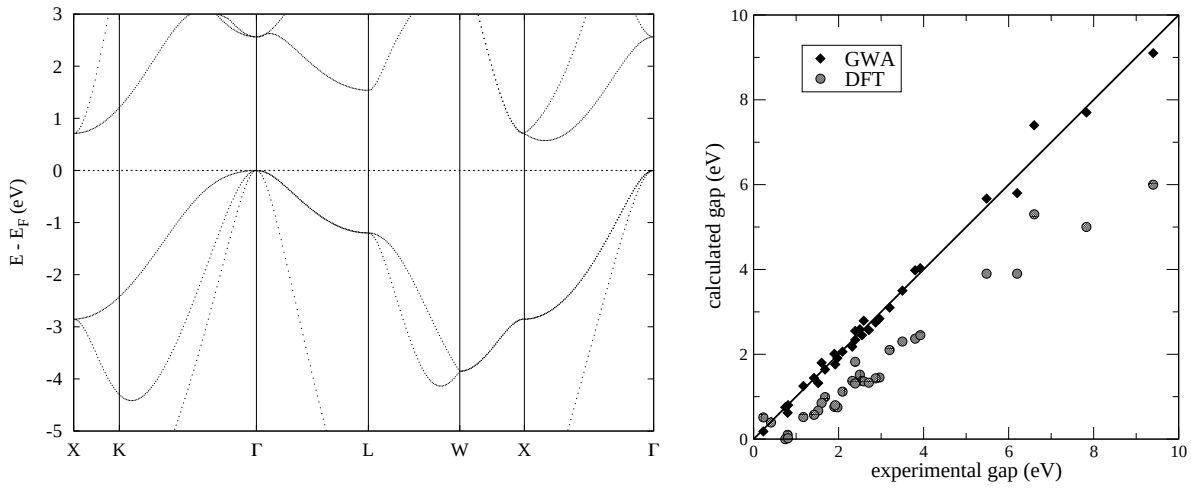


Fig. 4: *Left: LDA bandstructure of Si around the Fermi level. Note, that the fundamental gap is indirect, as found experimentally, but the experimental gap is a factor two larger than the calculated one. Right: Calculated bandgaps in DFT (LDA) and the GW-approximation for several insulators and semiconductors as a function of the experimental gap (data taken from Ref. [7]).*

fact that they are dominated by states near the Fermi level. Evidence, that these electrons can be qualitatively described in an independent particle description (similar to the "particles" in DFT) comes from Landau's Fermi liquid theory [1].

In insulators or semiconductors, we are typically interested in states far away from the Fermi level which cannot be expected to bear very close resemblance with the single particle states described by DFT (and also other methods that will be discussed below). As an example, we show in figure 4 the bandstructure of silicon. Although DFT can describe many properties of Si reliably (e.g. the lattice constant or the Γ-point phonon turn out to be in good agreement with the experiment), the experimental bandgap is a factor two larger than the energy difference between the highest occupied and the lowest unoccupied state. This does not mean, that the DFT bandstructure would be useless, it "predicts" correctly the nature of the bandgap as indirect between the Γ-point and the ΓX-line, but the valence and the conduction band are too close by the same amount throughout the BZ. This close resemblance between calculated bandstructure and experimental data lead to the fact, that the too small bandgaps in DFT are often called a "DFT problem" but of course, these bandstructures simply do not describe the electron-removal or electron-addition process that defines the bandgap. Other methods, like the GW approximation to manybody perturbation theory are available for this purpose and will be described at the end of this section.

3.1 The Hartree and the Hartree-Fock approximation

Let us start with the many-body Schrödinger equation for the electrons

$$\sum_i \left(h_i + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right) \Psi = \varepsilon \Psi \qquad \text{with} \qquad h_i = -\frac{1}{2} \boldsymbol{\nabla}_i^2 + V_{\text{ext}}(\mathbf{r}_i) \qquad (32)$$

where $V_{\rm ext}(\mathbf{r})$ includes the potential arising from the interaction with the nuclei and other possible external potentials. Assume that we found solutions to the single-particle Hamiltonian h_i and denote them $\phi_i(\mathbf{r}_i)$. Then, we can try to construct Ψ from these single particle wavefunctions and combine them to a manybody wavefunctions by a simple product Ansatz, $\Psi = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\ldots\phi_N(\mathbf{r}_N)$.
Ignoring the interaction part in Eq.(32) for the moment, we study a system of independent electrons. Then, if we multiply from the left with all single particle wavefunctions except one (e.g. ϕ_i) and integrate over all $\mathbf{r}$'s except $\mathbf{r}_i$, we arrive at a set of equations

$$h_i\phi_i(\mathbf{r}_i) = (\varepsilon - \sum_j \epsilon_j)\phi_i(\mathbf{r}_i) \qquad \text{where} \qquad h_j\phi_j(\mathbf{r}_j) = \epsilon_j\phi_j(\mathbf{r}_j)\,. \tag{33}$$

In this case, the eigenvalue of the manybody wavefunction, ε, is obviously the sum of all single particle eigenvalues, ϵ_i. Given the fact that electrons are fermions and cannot occupy a state more than once, this means that the ground state of our system will be the one that has the lowest N single particle states occupied.
Now, if we reintroduce the electron-electron interaction and go through the same steps, we get a coupled set of equations

$$\left(h_i + V_i(\mathbf{r}_i)\right)\phi_i(\mathbf{r}_i) = (\varepsilon - \sum_j \epsilon_j)\phi_i(\mathbf{r}_i) \qquad \text{with} \qquad V_i(\mathbf{r}_i) = \sum_{j\neq i}\left\langle \phi_j(\mathbf{r}_j)|\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}|\phi_j(\mathbf{r}_j)\right\rangle \tag{34}$$

where $V_i(\mathbf{r}_i)$ is the potential created by all electrons except the one described by ϕ_i. Solving these equations is already a complicated task, but a considerable simplification can be achieved if we assume that in an infinite solid there are so many electrons in the system, that we can assume that every electrons "sees" the same potential arising from all the states

$$V_{\rm H}(\mathbf{r}) = \sum_j \left\langle \phi_j(\mathbf{r}_j)|\frac{1}{|\mathbf{r} - \mathbf{r}_j|}|\phi_j(\mathbf{r}_j)\right\rangle\,. \tag{35}$$

This leaves a singe equation for all states

$$\left(h + V_{\rm H}(\mathbf{r})\right)\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \tag{36}$$

which has to be solved self-consistently. This means, since $V_{\rm H}$ – the Hartree potential – depends on the states ϕ_i, first a guess for this potential has to be made (e.g. for states calculated in the independent electron approximation), and then Eq.(36) can be solved initially. With the solutions in the next iteration a new, better guess for $V_{\rm H}$ can be obtained and this process can be repeated until the potential does not change any more from one iteration to another.
What we can learn from this so-called Hartree method are two things: in some approximation we can retain the notion of single-particle states and occupy them by an Aufbau-principle to construct a manybody wavefunction. In a self-consistent scheme, equations for these single-particle states can be solved to obtain a solution iteratively. The other lesson to learn is, that already a simple product Ansatz is very difficult to handle unless we make approximations, like substituting the state-dependent potential V_i by the Hartree potential, $V_{\rm H}$, thereby introducing some self-interaction of the single-particle states.
However, this is not the most severe shortcoming of the Hartree method: as mentioned above, the biggest approximation we introduced in the beginning by choosing a simple product Ansatz.

This construction of the many-body wavefunction of N non-interacting electrons still suffers from a serious problem in the treatment of the fermionic nature of the electrons. The simple product of single-particle states does not fulfill a basic requirement for fermions which states that the many-body wavefunction has to be anti-symmetric under the exchange of two particles

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N) = -\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N) \tag{37}$$

where we introduced $\mathbf{x} = (\mathbf{r}, \sigma)$ to denote the combination of the spatial and spin degrees of freedom. For the moment it is sufficient to consider the spin, σ, simply as a label that can assume two values.
However, it was realized early by Slater [8], that an anti-symmetric linear combination of product wavefunctions can be constructed, which has the desired property. This construction is known as a 'Slater determinant' as it can be expressed in terms of a determinant of a matrix containing the single-particle states

$$\begin{aligned} \Psi_{\text{Slater}}(\mathbf{x}_1 \ldots \mathbf{x}_N) &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \ldots & \phi_1(\mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{x}_1) & \ldots & \phi_N(\mathbf{x}_N) \end{vmatrix} \\ &= \frac{1}{\sqrt{N!}} \sum_P (-1)^P P\left(\phi_1(\mathbf{x}_1) \ldots \phi_N(\mathbf{x}_N)\right). \end{aligned} \tag{38}$$

In this notation the sum is performed over all permutations P acting on the indices i of the ϕ_i. The factor $(-1)^P$ ensures the required anti-symmetry.
The Slater-determinants as given in Eq. (38) form an anti-symmetric solution of the non-interacting Schrödinger equation. It can be shown that using all possible combinations of single-particle wavefunctions these determinants form a basis of the space of the N-body wavefunctions so that the interacting many-body wavefunction can be expressed as a linear combination of Slater determinants. Such an expansion forms the basis of complicated and expensive computational methods like so called configuration interaction calculations which yield high accuracy but can be performed only for a small number of electrons N.
Using these determinant functions to find a solution for the Hamiltonian as shown in Eq.(32) is the essence of the Hartree-Fock method. The derivation of the equations is somewhat lengthy but rather straightforward. The strategy is to vary the functions ϕ to make the expectation value of the many body Hamiltonian $\langle\Psi|\mathcal{H}|\Psi\rangle$ an extremum under the constraint that the functions ϕ are normalized to unity. This constraint is introduced by an Lagrange multiplier $\varepsilon_{i,\sigma}$ for each $\phi_{i,\sigma}$. This leads then to the so called Hartree-Fock equation

$$\left(-\frac{1}{2}\boldsymbol{\nabla}^2 + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r})\right)\phi_{i,\sigma}(\mathbf{r}) + \sum_{j,\sigma'} \int \frac{\phi^*_{j,\sigma'}(\mathbf{r}')\phi_{i,\sigma'}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \phi_{j,\sigma}(\mathbf{r}) = \varepsilon_{i,\sigma}\phi_{i,\sigma}(\mathbf{r}). \tag{39}$$

It is the last term on the left side of Eq.(39) that introduces the physics missing in the Hartree method. It can be rewritten to give the equation a more familiar form:

$$\left(-\frac{1}{2}\boldsymbol{\nabla}^2 + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{ex}}(\mathbf{r}; i\sigma)\right)\phi_{i,\sigma}(\mathbf{r}) = \varepsilon_{i,\sigma}\phi_{i,\sigma}(\mathbf{r}). \tag{40}$$

In addition to the non-interacting single-particle Hamiltonian two additional single-particle potential terms appear which describe the Coulomb interaction. The first one we know already as

the Hartree potential. The second term, the so called exchange potential, is a combined effect of Coulomb interaction and the antisymmetry condition and can, therefore, not be interpreted classically. This term can be written

$$V_{\text{ex}}(\mathbf{r};i\sigma) = -\frac{1}{\phi^*_{i,\sigma}(\mathbf{r})\phi_{i,\sigma}(\mathbf{r})}\sum_{j,\sigma'}\left\langle \phi_{i,\sigma}(\mathbf{r})\phi_{j,\sigma'}(\mathbf{r}')|\frac{1}{|\mathbf{r}-\mathbf{r}'|}|\phi_{i,\sigma}(\mathbf{r}')\phi_{j,\sigma'}(\mathbf{r})\right\rangle' \tag{41}$$

where the integration ($\langle\rangle'$) is assumed over $\mathbf{r}'$. If the summation in Eq.(41) would be restricted to the term $j = i$ only, we would recover Eq.(34) indicating that the exchange potential in the Hartree-Fock method contains a self-energy correction. Again we arrive at a state dependent potential, that can be thought to originate from a (nonlocal) charge density

$$n^{i\sigma}_{\text{ex}}(\mathbf{r},\mathbf{r}') = \sum_{j,\sigma'}\frac{\phi^*_{i,\sigma}(\mathbf{r})\phi^*_{j,\sigma'}(\mathbf{r}')\phi_{i,\sigma}(\mathbf{r}')\phi_{j,\sigma'}(\mathbf{r})}{\phi^*_{i,\sigma}(\mathbf{r})\phi_{i,\sigma}(\mathbf{r})}\,. \tag{42}$$

n_{ex} has the property that it integrates to unity, so it corresponds to the charge of a single electron. Furthermore, for a spin σ and the limit $\mathbf{r} = \mathbf{r}'$ it reduces to the state-independent value $\sum_j \phi^*_{j,\sigma}(\mathbf{r})\phi_{j,\sigma}(\mathbf{r})$. This charge density is also called the exchange hole, describing the influence of a state i,σ when moving through the ensemble of all states in the system. We will encounter the exchange hole once more in the context of density functional theory where it appears in a state-independent form, actually very similar to Slaters idea [9] of a state-averaged version of Eq.(42) that inspired also the conception of the first exchange-correlation potentials for DFT.
Interestingly, even though this simple construction of the many-body wavefunction as a Slater determinant only describes non-interacting systems, the description of the complicated many-body interaction in terms of single-particle states ϕ_i is a rather useful and powerful concept frequently also used for the case of interacting particles. This success of the single-particle description of the interacting system is even more surprising if one considers the both very strong and very long-range character of the Coulomb interaction. It can be understood and explained by the theory of Fermi-liquids introduced by Landau [10].

3.2 Density functional theory

While many researchers were working on more tractable versions of the Hartree-Fock method, in the middle of the sixties Hohenberg and Kohn [11] worked out two central theorems that form the basis of a conceptually different approach, the density functional theory: Consider a system of N particles (e.g. electrons) moving in an external potential $V(\mathbf{r})$ (caused by e.g. nuclei). In a non-degenerate ground state (i) the many-body wavefunction Ψ and $V(\mathbf{r})$ are uniquely determined by the particle density distribution $n(\mathbf{r})$ and (ii) there exists an energy functional of this density, $E[n(\mathbf{r})]$, which is stationary with respect to variations of the ground-state density. These two theorems allow – at least in principle – the determination of the ground-state density and energy of a N-particle system by searching for the density that minimizes the energy functional. Extracting the classical Coulomb interaction energy, this Hohenberg-Kohn energy functional takes the form

$$E[n(\mathbf{r})] = \int V_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int\int\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + G[n(\mathbf{r})] \tag{43}$$

where the functional $G[n(\mathbf{r})]$ contains all other contributions.

If we succeed to find the functional $G[n(\mathbf{r})]$ or a good approximation to it, the immediate advantage of DFT is that, instead of dealing with the full manybody wavefunction, $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, we can work with the much more tractable density, $n(\mathbf{r})$. Although more information is directly accessible from the wavefunction than from the density,

$$n(\mathbf{r}) = \int d\mathbf{r}_2 \ldots \int d\mathbf{r}_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{44}$$

in DFT many physical quantities, like the structural properties or bond strength can be obtained for large systems, where a manybody wavefunction would be impossible to access. E.g. calculations of the ground-state energies for different external potentials, as they result from a variation of the lattice parameters in a periodic solid, allow the determination of the equilibrium lattice constant, which is nowadays possible to within a few percents. Early attempts to use the density as a key parameter for calculations of periodic solids were made by Lenz [12] based on the statistical method of Thomas [13] and Fermi [14]. In this approach, $G[n(\mathbf{r})]$ was considered to contain the kinetic energy density (taken to be proportional to $[n(\mathbf{r})]^{\frac{5}{3}}$). In the Thomas-Fermi-Dirac method $G[n(\mathbf{r})]$ even contains an exchange energy density term proposed by Dirac [15] (proportional to $[n(\mathbf{r})]^{\frac{1}{3}}$). Although the Thomas-Fermi theory has still its applications today, it never became useful as a theoretical method for the prediction of materials properties [16].
The key idea, that made DFT a success, was to extract from $G[n(\mathbf{r})]$ the kinetic energy T_0 of a non-interacting electron gas in its ground state which has the same density distribution, $n(\mathbf{r})$, as the interacting system. In this Kohn-Sham theory [17] a new functional

$$E_{\mathrm{xc}}[n(\mathbf{r})] = G[n(\mathbf{r})] - T_0[n(\mathbf{r})] \tag{45}$$

appears, that remains to be determined. E_{xc} is a much smaller term than G and is called exchange-correlation energy functional, since – as we will see below – without E_{xc} our energy functional E would yield just the energy in the Hartree approximation. If we take into account that particle conservation, i.e. $N = \int n(\mathbf{r})d\mathbf{r}$, has to be ensured, we can formulate the stationarity of E in equation (43) with respect to variations of the ground-state density, n, as

$$\frac{\delta T_0}{\delta n(\mathbf{r})} + V(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{\mathrm{xc}}}{\delta n(\mathbf{r})} - \lambda = 0 \tag{46}$$

where the Lagrange parameter λ ensures the particle conservation. Expressing the kinetic energy of the non-interacting particles via their wavefunctions, ϕ_i, we can recast equation (46) in the form of an effective single particle Schrödinger equation, the Kohn-Sham equation:

$$\left[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{\mathrm{xc}}}{\delta n(\mathbf{r})}\right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \tag{47}$$

which has to be solved self-consistently since $n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2$. From this point of view, the structure of the Kohn-Sham equations is very similar to the Hartree approach outlined in the last subsection. The index i combines now the k-point, $\mathbf{k}$, and the band index, ν. Note, that without E_{xc} equation (47) reduces to the Hartree equation. Therefore, this last term of the Hamiltonian is called the exchange-correlation potential, since exchange and correlation are exactly what is missing in the Hartree approximation.
Although λ was introduced as a Lagrange multiplier and also the ε_i's should be strictly be interpreted in this way, it is usual to derive from the ε_i's the bandstructure of a crystal and use the

wavefunctions $\phi_i(\mathbf{r})$ as approximations to true quasiparticle wavefunctions. Some justification will be given below and comparison with experimental data often confirms this point of view, but there are also well known examples, where this interpretation leads to significant "errors", like in the comparison of the bandgaps of semiconductors and insulators with bandstructures derived from these ε_i's.

One of the first interpretations of the term, which emerged as the exchange-correlation potential in DFT, was given by Slater [16] in the context of the Thomas-Fermi method and later in connection with the Hartree-Fock method [9]. Essentially, it describes the afore mentioned interaction of a particle with the "hole" that is created by its own presence in the gas of the other particles. This means, that the probability of finding an electron at a position $\mathbf{r}$ reduces the probability of finding another electron at a position $\mathbf{r}'$ nearby, depending of course also on the spin of the two particles (therefore, in the Hartree-Fock method this hole, Eq. (42), has been given the name "exchange hole").

To derive some properties of the exchange correlation energy functional, it is useful to write this "hole" (exchange-correlation hole in DFT), n_{xc}, in terms of a two-particle correlation function, $g(\mathbf{r},\mathbf{r}')$ [18]:

$$n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = n(\mathbf{r}')\int_0^1 d\xi[g_n(\mathbf{r},\mathbf{r}',\xi)-1] \equiv n(\mathbf{r}')h(\mathbf{r},\mathbf{r}'). \tag{48}$$

Here, $g_n(\mathbf{r},\mathbf{r}',\xi)$ is the correlation function of a system of charged particles where the Coulomb interaction is scaled by a factor ξ and a ξ-dependent potential has been added, so that the density, $n(\mathbf{r})$, is independent of ξ. Additionally, the so called hole function, $h(\mathbf{r},\mathbf{r}')$, was introduced. The exchange correlation energy can then be written as

$$E_{\mathrm{xc}}[n(\mathbf{r})] = \frac{1}{2}\int d\mathbf{r} n(\mathbf{r}) \int d\mathbf{r}' \frac{1}{|\mathbf{r}-\mathbf{r}'|} n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}'). \tag{49}$$

Although the exchange-correlation hole can be very complicated in shape, it was soon realized, that only its radial dependence enters in the exchange correlation energy [19]. This means that in practice E_{xc} is rather insensitive to details of shape of n_{xc}. Some properties of the exchange-correlation hole can be derived from the definition via the correlation function g. E.g. there is a sum rule, which states that n_{xc} corresponds exactly to one electron, i.e. that

$$\int d\mathbf{r}' n_{\mathrm{xc}}(\mathbf{r},\mathbf{r}') = -1 \tag{50}$$

has to be fulfilled. Such relations can guide the construction of exchange-correlation functionals or help to judge the validity of existing approximations to E_{xc}.

One of the big surprises in the early days of density functional theory was certainly the fact, that even a simple exchange-correlation functional like the local density approximation (LDA) leads to relatively convincing results. The LDA starts from the limit of the homogeneous electron gas, assuming E_{xc} rather as a function than as a functional of $n(\mathbf{r})$. Its success can now be explained by the fact, that the exchange-correlation hole in the local density approximation is of the form

$$n_{\mathrm{xc}}^{\mathrm{LDA}}(\mathbf{r},\mathbf{r}') = n(\mathbf{r}')h_0(|\mathbf{r}-\mathbf{r}'|;n(\mathbf{r}')) \tag{51}$$

where $h_0(|\mathbf{r}-\mathbf{r}'|;n)$ is the hole function of an uniform interacting electron gas of density n. For an uniform density, this exchange-correlation hole satisfies equation (50). For a non-uniform density the sum rule should be at least approximately fulfilled and [20] showed, that in LDA

this is on average the case. This, together with the fact that $E_{\rm xc}$ depends only on the spherical average of $n_{\rm xc}$, is mainly responsible for the success of the LDA.
Also modern, exchange-correlation functionals including gradient corrections are constructed in such a form, that they fulfill certain conditions that are known exactly in different limits (like high or low density, constant or slowly varying density etc.). In this way, exchange-correlation potentials are improved on a parameter-free basis. Alternatively, the functionals (or parts of the functionals, e.g. the correlation energy) can be fitted to numerical results from manybody calculations. Another strategy – often used in the chemical literature – is to adjust the functional to yield best results (like bond-length, dissociation energies etc.) for a given set of systems.

3.3 Extensions to DFT: the LDA+U method

Dealing with f and some d transition metals and their compounds it was realized that, while the s, p and some d electrons can successfully be described in standard DFT methods, for the strongly localized electrons a more atomic-like description (e.g. Hartree-Fock) is appropriate [21]. Taking into account the different atomic potentials and the stronger screening in the metal, an atomic theory [22] for these localized states can describe the situation quite satisfactorily. Following this approach, Anisimov et al. [23] merged this atomic picture with band theory (i.e. standard DFT), to get a "band approach" to Hubbard-type models: For the localized d and f states, the Coulomb interaction of the electrons is formulated in the spirit of the Anderson model:

$$E_{\rm ee} = \frac{1}{2} U \sum_{i \neq j} n_i n_j \tag{52}$$

where the n's are here the d-orbital occupation numbers and U is the famous Hubbard parameter, describing the on-site Coulomb interaction. In the local density approximation to this model the energy of the $d-d$ interaction is [24]

$$E_{\rm ee}^{\rm LDA} = \frac{1}{2} UN(N-1) \quad \text{where} \quad N = \sum_i n_i \ . \tag{53}$$

If we add E_{ee} from equation (52) to the LDA energy functional, $E_{\rm ee}^{\rm LDA}$ should be subtracted, so that

$$E^{\rm LDA+U} = E^{\rm LDA} + \frac{1}{2} U \sum_{i \neq j} n_i n_j - \frac{1}{2} UN(N-1) \ . \tag{54}$$

This is a simple version of the LDA+U method. Such a modification of the LDA results in a shift of the LDA eigenvalues:

$$\epsilon_i = \frac{dE}{dn_i} = \epsilon_i^{\rm LDA} + U \left(\frac{1}{2} - n_i \right) \tag{55}$$

i.e. more than half-filled bands are shifted down in energy, while less then half-filled bands are shifted up. Despite the formal similarity with the Stoner model, it should be noted that the physical background of this model is quite different [23]. A simple example is given in figure 5, where the LDA+U method was used to correct the positions of the $4f$ states in ferromagnetic bcc Eu. It is easy to see, that the correction has almost no effect on the s and p states, but shifts down the occupied $4f$ states (an enhanced localization of these states can be seen by the narrowing of the band) and pushes the unoccupied $4f$ levels to higher energies. How large this

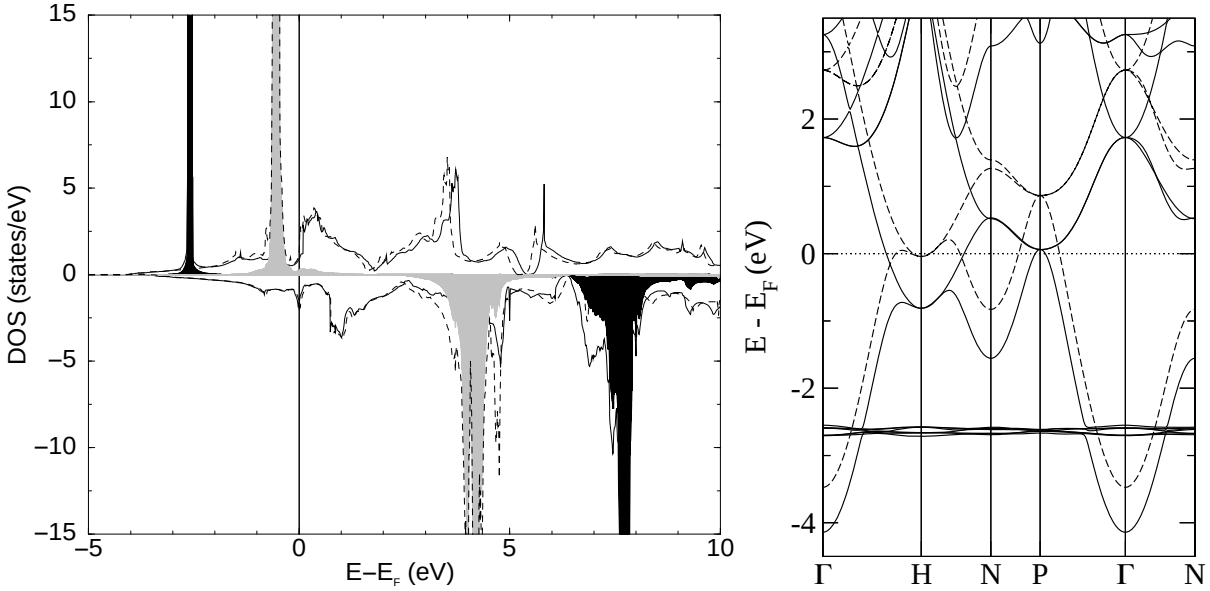


Fig. 5: *Left: Density of states of bcc Eu calculated with standard LDA (dashed line) and the LDA+U (full line) method. A Hubbard U of about 6 eV was used to correct the positions of the $4f$ states. The local partial $4f$ DOS as obtained in LDA is shown as grey shaded area, the LDA+U result in black. Right: Eu bandstructure obtained in the LDA+U method: majority spin states (full lines) and minority spin states (broken lines).*

shift is, depends of course on the chosen U. Before we turn to the question how to obtain a reasonable estimate for U, we have to refine the model to see, how we can apply the LDA+U method on a certain set of states (e.g. $4f$) at a given atom.
To separate the localized orbitals from the itinerant states, for which the LDA provides already a good description, one chooses a site-centered, $\{l, m\}$ dependent orbital basis, $|\nu, l, m\rangle$, where ν is the site-index of the selected atom and l and m are the angular and azimuthal quantum numbers, respectively. If the density is given by Kohn-Sham orbitals like

$$n^{(+)}(\mathbf{r}) = \sum_i w_i^{(+)} |\phi_i^{(+)}(\mathbf{r})|^2 \quad \text{and} \quad n^{(-)}(\mathbf{r}) = \sum_i w_i^{(-)} |\phi_i^{(-)}(\mathbf{r})|^2 \tag{56}$$

where the weights, $w_i^{(\pm)}$, determine the occupation of the states, we can define a density matrix for spin α in m, m'-space:

$$n_{mm'}^{\alpha\nu} = \sum_i w_i^\alpha \langle \nu, l, m | \phi_i^\alpha \rangle \langle \phi_i^\alpha | \nu, l, m' \rangle. \tag{57}$$

E.g. if we want to apply the LDA+U method on $4f$ states, we need for each spin a 7×7 density matrix, where the diagonal elements give the occupancy of the $l = 3, m = -3, -2, \ldots, 3$ orbitals of the selected atom. Using this density matrix, the electron-electron interaction energy can be formulated as [25]

$$E_{\text{ee}} = \frac{1}{2} \sum_\nu \sum_{mm'pq}^{\alpha,\beta} n_{mm'}^{\alpha\nu} \left[\langle m, p | V_{\text{ee}} | m', q \rangle - \langle m, p | V_{\text{ee}} | q, m' \rangle \delta_{\alpha\beta} \right] n_{pq}^{\beta\nu} \tag{58}$$

and used instead of the simpler version, equation (52). Here, the electron-electron interaction can be expressed in terms of an angular part, contained in a_k, and the radial part that is given by the effective Slater integrals [22], F_k:

$$\langle m,p|V_{\rm ee}|m',q\rangle = \sum_k a_k(m,p,m',q)F_k \quad ; \quad 0 \le k \le 2l \tag{59}$$

In terms of the screened Coulomb- and exchange parameters, U and J, the Slater integrals can be approximated, e.g. for $l = 2$, as

$$U = F_0 \quad ; \quad J = \frac{F_2 + F_4}{14} \quad \text{and} \quad \frac{F_4}{F_2} = \frac{5}{8}, \tag{60}$$

and the a_k are sums of integrals of the angular part of the wavefunction with spherical harmonics. Then, we can define an orbital selective potential,

$$V^{\alpha\nu}_{mm'} = \sum_{pq\beta} \left[\langle m,p|V_{\rm ee}|m',q\rangle - \langle m,p|V_{\rm ee}|q,m'\rangle\delta_{\alpha\beta}\right] n^{\beta\nu}_{pq} - \left[U(n^{\nu} - \frac{1}{2}) - J(n^{\alpha\nu} - \frac{1}{2})\right]\delta_{mm'} \tag{61}$$

where $n^{\alpha\nu} = \sum_m n^{\alpha\nu}_{mm}$ and $n^{\nu} = \sum_\alpha n^{\alpha\nu}$. This spin-, site- and l,m-dependent potential enters now the Kohn-Sham equation via

$$\left[-\nabla^2 + V^{\alpha}_{LDA}(\vec{r})\right]\phi^{\alpha}_i + \sum_{\nu}\sum_{mm'} V^{\alpha,\nu}_{mm'}\frac{\delta n^{\alpha,\nu}_{mm'}}{\delta\phi^{\alpha}_i} = \epsilon^{\alpha}_i\phi^{\alpha}_i. \tag{62}$$

Thus, we have introduced a Hartree-Fock like potential term that acts on a certain subset of the orbitals, leaving the others (in a first approximation) unchanged. Equation (62) has to be solved self-consistently, until both the density and the density matrix are converged. If the Kohn-Sham equations are solved by expanding the wavefunction into some basis set, for different types of basis sets also a different orbital basis, $|\nu,l,m\rangle$, will be convenient. It is clear, that also the result of the LDA+U calculation will depend to some extent on the choice of the orbital basis, but in practice for the same parameters U and J also qualitatively the same answers are reached. As can be seen from figure 5, when applied to $4f$ metals like Eu or Gd, the LDA+U method can be used to shift the position of the $4f$ as a function of U. A comparison to the position where they are spectroscopically measured can be used to determine a value for U, although other methods will be described below. One of the problems of LSDA, the prediction of an antiferromagnetic groundstate for hcp Gd, is resolved when the LDA+U method is applied in this way [26]. Another improvement due to the LDA+U approach can be seen in the case of Eu, where LDA would predict a much too small lattice constant, while LDA+U removes this overbinding caused by the $4f$ states [27].

More complicated physical phenomena, like orbital ordering, can be introduced in a DFT calculation by the LDA+U method. In perovskite materials a transition metal atom sits in an octahedral cage of oxygen atoms and its electron system is rather isolated from the rest of the electronic structure. If a single d electron is left on this atom, e.g. on the V in $SrVO_3$, this electron is in an almost atomic-like state forming a very narrow band and feels just a small octahedral crystal field from the neighboring anions. Thus, the d states are split in e_g and t_{2g} states, populated by a single electron. Due to the remaining degeneracy, we would expect a Jahn-Teller (JT) distortion to form but conventional exchange-correlation potentials fail to capture this phenomenon. One of the first examples, where this was studied, was $KCuF_3$ where a

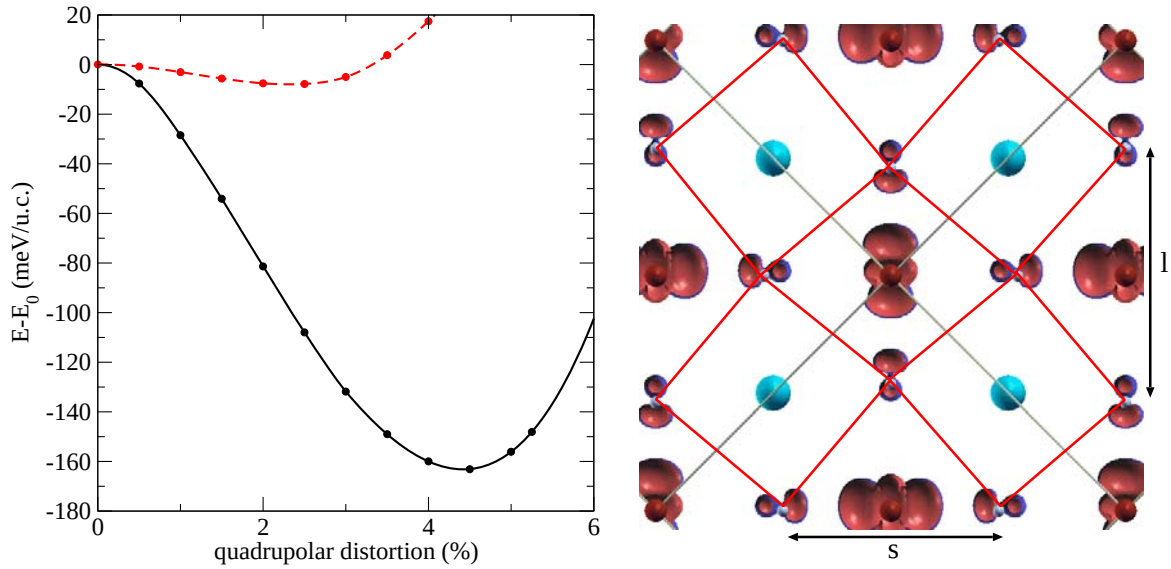


Fig. 6: *Left: Total energy as a function of the Jahn-Teller distortion in (paramagnetic) $KCuF_3$ as calculated with the GGA without (dashed red line) and with a Hubbard-U (full black line). The experimental (cubic) lattice constant and an a-type orbital ordering was assumed. Within the LDA+U method a quadrupolar distortion is stabilized. Right: Orbital ordering in a d-type antiferromagnetic structure of $KCuF_3$. The distorted octahedra, formed by the fluorine atoms, are indicated by red lines, the unit cell in gray. Blue spheres are K atoms, the magnetization density (mostly on the Cu atoms) is shown in red.*

single d-hole sits in a fluorine octahedron, inducing a sizable distortion [25]. While LDA gives no JT distortion at all, GGA leads to a very shallow energy minimum at finite distortions as can be seen from figure 6. Here the distortion is measured by the fraction $(l-s)/(l+s)$ where l and s are the long and short axis of the deformed octahedron. In contrast to these results, with the LDA+U method (U was chosen to be 7.5 eV) a paramagnetic orbitally ordered state with a sizeable JT distortion can be stabilized (left of figure 6). Several different orbital orderings can be obtained, one of them shown in the right of figure 6, and their energies can be compared. We see that in the LDA+U method it is also possible to stabilize solutions which are not the electronic ground state - in contrast to non-spin polarized DFT. Since it is a ground-state method, LDA+U predicts an antiferromagnetically and orbitally ordered ground-state in good agreement with experiment [28]. The structural parameters are very similar to the results obtained from dynamical mean-field theory (DMFT) calculations [29], although it needs this more elaborate theory to capture the electronic structure of this material in its (high temperature) paramagnetic state. Despite the enormous computational expenses of DMFT calculations for even a few bands or sites (cluster-DMFT), detailed insight into the physics of these strongly correlated materials is nowadays possible [30]. More details on the "electronic structure of transition metal oxides" and on "highly correlated electron systems" will be given in subsequent lectures.

Although the LDA+U method is rather simple and quite successful, it faces the problem that it introduces an external parameter and thus destroys the "*ab initio*" character of the conventional LDA approach. Therefore, concepts to calculate U within constrained DFT [31] and with the GW method [32] (next subsection) have been developed. Fortunately, in many cases the results do not depend too sensitively on the exact values of U and J. But there are also systems,

like $YMnO_3$, where depending on the value of U different magnetic ground-states can be stabilized [33]. A collection of applications of the LDA+U method can be found in reference [24].

3.4 Quasiparticles and the GW approximation

Up to now, we relied on the concept of single-particle states which we inherited from the independent electron approximation. Of course in a many-body system it is not at all clear in how far this concept still is meaningful. As we have seen at the beginning of this section, in a system of independent particles the energy to remove a single electron can be determined as the eigenvalue of a single particle equation like Eq.(33). So we can ask whether it is possible to create an equation similar in structure that yields as an eigenvalue the energy to remove or add a single electron to a many-body system. These energies would be what is typically obtained in experiments like photoemission or inverse photoemission.
It would lead to far to introduce all necessary theoretical concepts to develop an appropriate theory to study this problem. To outline the basic difficulties we will follow here an early paper of many-body perturbation theory [34]. It starts from a single-particle Hamiltonian, e.g. Eq.(36), and assumes that the difference between this Hamiltonian and the true, many-body Hamiltonian can be treated as a perturbation. The energy needed to add a single particle to the N-electron state will differ from the $(N+1)$th eigenvalue of a single particle Hamiltonian, h_0, by an amount which is called the self-energy of this particle. It can be shown that it is possible to add a term to h_0 so that this self-energy vanishes. The resulting Hamiltonian has the form

$$h_0\phi_i(\mathbf{r}) + \int \Sigma(\mathbf{r},\mathbf{r}',\varepsilon_i)\phi_i(\mathbf{r}')d\mathbf{r}' = \varepsilon_i\phi_i(\mathbf{r}) \tag{63}$$

with the non-local, energy dependent self-energy operator. The eigenvalues are now excitation energies, i.e. the energy differences between a N and a $N+1$ particle system (or a N and $N-1$ particle system).
Formally, we can notice a similarity between Eq.(63) and the Hartree-Fock Eq.(39) by writing an energy independent self-energy

$$\Sigma^{\mathrm{HF}}(\mathbf{r},\mathbf{r}') = \sum_{j,\sigma'} \phi^*_{j,\sigma'}(\mathbf{r}')\phi_{j,\sigma'}(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|} = iG(\mathbf{r},\mathbf{r}',-\eta)v(\mathbf{r},\mathbf{r}')\,. \tag{64}$$

In the last step we wrote the sum over the single-particle states as a Green function with η being an infinitesimally small (positive) time, so that G reduces to the density matrix.
In many body perturbation theory it turns out that – in a certain approximation – the self energy operator, when Fourier transformed from the energy to the time domain, can be written in a rather similar form:

$$\Sigma(\mathbf{r},\mathbf{r}';\tau) = iG(\mathbf{r},\mathbf{r}',\tau)W(\mathbf{r},\mathbf{r}',\tau+\eta) \tag{65}$$

where G is now the full Green function and W is a screened Coulomb interaction. Generally, $iG(\mathbf{r},\mathbf{r}',\tau)$ describes the probability to find a particle inserted into a many-body system at a position $\mathbf{r}$ after some time τ at a position $\mathbf{r}'$. The screened Coulomb interaction W is related to the bare Coulomb interaction $v(\mathbf{r},\mathbf{r}')$ via the dielectric function ϵ,

$$W(\mathbf{r},\mathbf{r}',\varepsilon) = \int \epsilon^{-1}(\mathbf{r},\mathbf{r}'',\varepsilon)v(\mathbf{r},\mathbf{r}'')d\mathbf{r}'' = v(\mathbf{r},\mathbf{r}') + \int n_{\mathrm{ind}}(\mathbf{r},\mathbf{r}'',\varepsilon)v(\mathbf{r}',\mathbf{r}'')d\mathbf{r}'' \tag{66}$$

Again, we see the effect that the Coulomb potential of the electron repels neighboring charges to give rise to a positive induced charge, n_{ind}, that modifies (screens) the bare Coulomb interaction. This behavior reminds to the exchange hole, Eq.(42), of the Hartree-Fock theory or the exchange-correlation hole, Eq.(48), of DFT that gives rise to the exchange-correlation energy, Eq.(49).

It should be noticed that Eq.(65) is a kind of Hartree-Fock (HF) approximation for quasiparticles, while Eq.(64) is the HF approximation for electrons. So, despite the formal similarity to the HF equations we have to keep in mind a couple of important differences: Eq.(63) contains an energy-dependent non-Hermitian self-energy operator. The eigenvalues, ε_i, are complex numbers and the imaginary part leads to a damping term in the time-dependent Schrödinger equation, meaning that the quasiparticles, described by Eq.(63), have a finite lifetime that is proportional to the inverse of the value of the imaginary part. An electron or hole that is added to a many body system keeps its particle-character for some time, until it "dissipates" into the many-body ensemble.

Hedin [35] provided a set of equations that link all these quantities like the self-energy (containing so-called vertex corrections), the Green function, screened Coulomb interaction and dielectric function. These equations can – in principle – be solved self-consistently. In practice, however, the solution of these equations is far too complicated and commonly an approximation to this equations is solved, which takes Eq.(65) for the self-energy and substitutes the G in this equation by a Green function constructed by Kohn-Sham wavefunctions. Also the dielectric function, ϵ, is calculated from these wavefunctions in the random phase approximation. This scheme is commonly termed GW approximation [7] and leads to quite reliable excitations energies, e.g. for bandgaps of semiconductors (cf. figure 4).

It should be mentioned that this is a method to calculate excitations in a many-body system where the particle number is changed by one. There are also excitations which leave the particle number unchanged and are accessible by generalizations of density functional theory, like time-dependent DFT (TDDFT), which provide a way to calculate these types of spectra and are active research fields today [36].

4 Relativistic effects

Having established various methods to deal with the electronic structure of solids, in this section we will focus on the spin of the electron. Up to now, we encountered the spin just as a label attached to electron wavefunctions that has to be taken into account in context of the Pauli principle. This additional quantum number can lead to important consequences, e.g. favoring magnetic ground states in transition metals, but this is the topic of a separate contribution on the "electronic basis of magnetism". We have to note here, that the spin introduced in this way is effectively just a number. To make full use of the electrons spin, e.g. to store or transmit information, it is crucial to focus on spin as a vectorial quantity that is carried along with the electron, having a definite orientation in space that can be manipulated and used in a spintronic device.

Therefore, in this section we will try to give spin a physical interpretation and study its consequences. Semi-classically, the electrons "spinning" around its own axis can be thought to be the source of the spin magnetic moment. This should not be confused with the orbital moment, arising from the precessional (orbital) motion of the electron. If we will denote the wavefunction

and the spin-label (referred to as spin-up or spin-down) as

$$\psi(\mathbf{r}) = \phi(\mathbf{r})\chi \qquad \text{with} \qquad \chi = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{67}$$

we can express the spin, $\mathbf{S}$ as the expectation value of the spin-operator, $\boldsymbol{\sigma}$,

$$\mathbf{S} = \langle\psi|\boldsymbol{\sigma}|\psi\rangle \qquad ; \qquad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{68}$$

Of course the Schrödinger equation will provide the wavefunctions ψ, but tells us nothing about the orientation of $\mathbf{S}$. In a collinear case, i.e. when all the spins are oriented along the same direction, for convenience the spins are assumed to be aligned in z-direction.
To give this spin-vector an absolute orientation in space, we first have to introduce a new term in the Hamiltonian that connects the spin-orientation with the axes of the crystal. This spin-orbit coupling term, which will be discussed on a general basis in the first subsection, leads then to a preferential spin orientation in the crystal, which is essential for every magnetism-based device, be it a simple compass or a modern hard-disk drive.
Spin-orbit coupling is a crucial effect in magnetic systems. In most solid state systems, however, due to chemical bonding the number of spin-up and spin-down wavefunctions are equal, so that the total spin is zero. Interestingly, even in these spin-compensated systems, that are in total non-magnetic, spin-dependent processes can be observed, e.g. the Rashba effect. These effects are fundamental for combining semiconductor systems with "classical" magnetic structures and will be introduced in the second subsection. Finally, we will shortly discuss spin-orbit coupling in magnetic systems, that leads to the magnetic anisotropy.

4.1 Spin-orbit coupling

As a consequence of the Lorentz transformation, an electron that is traveling with a velocity $\mathbf{v}$ on a classical trajectory around the nucleus, experiences an electric field $\mathbf{E}$ (from the potential gradient that arises due to the screened nucleus) as a magnetic field, $\mathbf{B} = \frac{1}{c}(\mathbf{v} \times \mathbf{E})$. This field will couple to the spin, $\boldsymbol{\sigma}$, of the electron as $-\boldsymbol{\sigma} \cdot \mathbf{B}$.[1] To include this effect on a quantum-mechanical basis, it is necessary to start from relativistic one-electron theory, the Dirac equation. In the Schrödinger equation – even for a magnetic system – there is no term that explicitly includes the spin-operator. But if we include a certain term from the Pauli equation (a two-component approximation to the Dirac equation [38], see Appendix) we get

$$\left[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \frac{\mu_\mathrm{B}}{2c}\boldsymbol{\sigma} \cdot (\mathbf{p} \times \mathbf{E}(\mathbf{r}))\right]\psi_i = \varepsilon_i\psi_i. \tag{69}$$

It is this relativistic correction (factor $\frac{1}{c}$) that leads to the coupling between spin-space ($\boldsymbol{\sigma}$) and lattice ($\mathbf{E}(\mathbf{r})$).
If we assume that the electric field is derived from a spherically symmetric potential, $V(r)$, (as occurs in the vicinity of an atomic nucleus) we can transform this term

$$-\boldsymbol{\sigma} \cdot (\mathbf{p} \times \mathbf{E}(\mathbf{r})) = \boldsymbol{\sigma} \cdot (\nabla V(r) \times \mathbf{p}) = \frac{1}{r}\frac{dV(r)}{dr}\boldsymbol{\sigma} \cdot (\mathbf{r} \times \mathbf{p}) = \frac{1}{r}\frac{dV(r)}{dr}(\boldsymbol{\sigma} \cdot \mathbf{L}) = \xi\boldsymbol{\sigma} \cdot \mathbf{L}, \tag{70}$$

[1] Although this interaction has the form of a Zeeman term (the interaction of the spin with an external magnetic field), its interpretation is not so straightforward: as compared to a classical interpretation, due to kinematical effects a factor of two arises in the expression. The origin of this effect is called Thomas-precession [37].

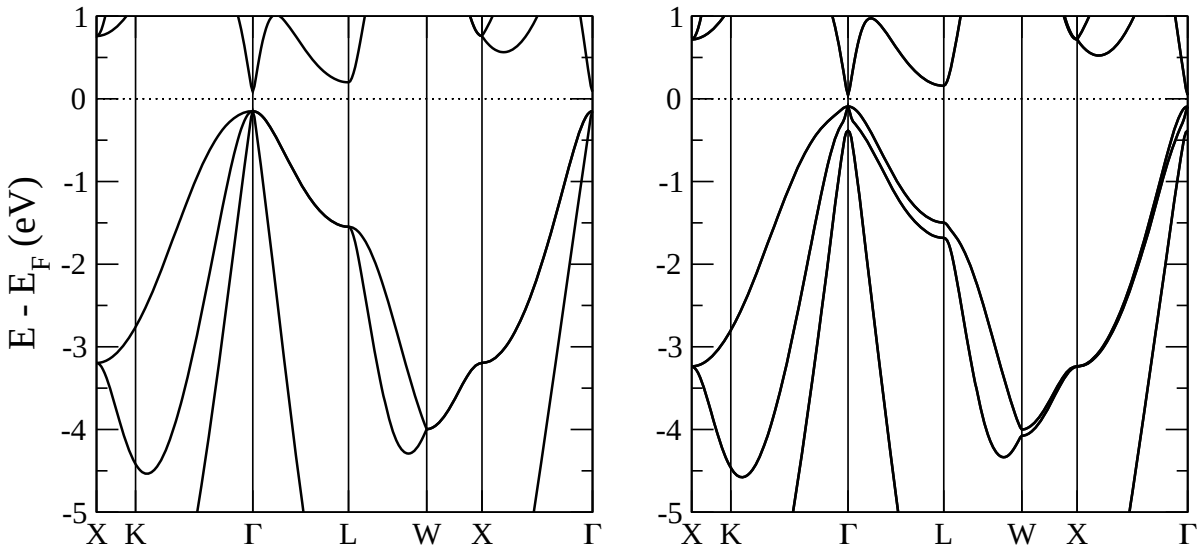


Fig. 7: *Bandstructure of Ge around the Fermi level without spin-orbit coupling (left) and with spin-orbit coupling included (right). Notice, that the three-fold degeneracy of the highest occupied state at the* Γ *point is split by spin-orbit coupling, as well as the doubly degenerate band along the lines* $\overline{\Gamma L}$ *and* $\overline{\Gamma X}$*. The calculation is performed at the experimental lattice constant using the local density approximation to DFT. Note, that the experimentally observed bandgap of* 0.75 *eV closes in a DFT calculation. In the present calculation the gap is opened by applying the LDA+*U *method (cf. subsection 3.3).*

where $\mathbf{L}$ is the orbital momentum operator. This term is called the spin-orbit coupling (SOC) term with the spin-orbit coupling constant ξ. Keep in mind that – although the term $\boldsymbol{\sigma} \cdot \mathbf{L}$ looks like a coupling between a spin- and an orbital moment – the SOC term describes the coupling between the spin and the magnetic field created by the electrons orbital *motion*. Therefore, $\boldsymbol{\sigma} \cdot \mathbf{L}$ enters with a positive sign in Eq.(69), indicating that spin- and orbital moment like to orient antiparallel – giving rise to Hund's third rule.

Since the radial derivative of the potential in a crystal will be largest in the vicinity of a nucleus, we can expect that the major contribution to the spin-orbit interaction will come from this region. For an atom ν then r is the radial part of the vector $\mathbf{r}_\nu = \mathbf{r} - \boldsymbol{\tau}_\nu$. Furthermore, since for small r_ν the potential will be Coulomb-like ($V(r) = -\frac{Z}{r}$), its derivative $\frac{\partial V}{\partial r_\nu}$ is proportional to the nuclear number of the atom, Z_ν. We thus expect that ξ will be large for heavy atoms, but small for lighter ones.

Electrons, that are close to the nucleus (i.e. those of the inner shells) will feel the consequences of this spin-orbit coupling most strongly. As it is well known from free atoms, this term will favor the formation of an orbital momentum, $\mathbf{L}$, which is then coupled to the electrons spin. E.g. the p-electrons can form states with a total orbital momentum $L = 1$, coupling then to the electrons spin. We can classify p-states according to their projections on a selected axis (z) by their magnetic quantum numbers $m_z = -1, 0, 1$. Combined with the electrons spin, this will result in a total angular momentum $J = 3/2$ with projections $m_j = 3/2$ or $1/2$. As a consequence of spin-orbit coupling, this results in a level splitting between the $p_{3/2}$ and $p_{1/2}$ states.

In contrast, the valence electrons in a solid will arrange to optimize the chemical bonding, e.g. in a simple cubic lattice p_x, p_y and p_z states will form. The level splitting is then determined

by the crystal field. Partially, spin-orbit coupling will interfere and lead to additional level splittings as can be observed e.g. in semiconductors at the center of the Brillouin-zone: In Si (cf. figure 4) we see a three-fold degenerate state directly below the Fermi-level that splits into a doubly degenerate and a singly degenerate one. The former one is closest to the Fermi level in turn consists of two bands with different dispersions, the highly dispersive state is called the light-hole band, the other one is termed heavy hole band. The singly degenerate state at Γ forms the spin-orbit split-off band. In a non-relativistic calculation these bands are degenerate in some high symmetry directions, but when spin-orbit coupling is included a splitting can be observed. As expected, this splitting is small in Si, but larger in the isoelectronic but heavier Ge (figure 7).

4.2 The Rashba- and the Dresselhaus effect

In a system without internal or external magnetic field time-reversal symmetry holds, i.e. changing the direction of the arrow of time will not alter the properties of the system. The transformation $t \rightarrow -t$ exchanges a particle moving with momentum $\mathbf{k}$ with a particle moving in $-\mathbf{k}$. Time reversal will also invert the precessional motion of the electron and, therefore, its spin. As a consequence, the energy of a right-moving spin-up particle will equal the energy of a left moving spin-down particle,

$$\varepsilon(\mathbf{k},\uparrow) = \varepsilon(-\mathbf{k},\downarrow)\,. \tag{71}$$

In a crystal with inversion symmetry, additionally $\varepsilon(\mathbf{k}) = \varepsilon(-\mathbf{k})$ holds, both for spin-up and spin-down electrons. This means, that the bandstructure is symmetric around the center of the Brillouin-zone, $\mathbf{k} = 0$, and all bands are doubly degenerate. E.g. in the bandstructures in figure 4 or 7 show this degeneracy.
In contrast, crystals without inversion symmetry the degeneracy of the bands can be lifted as a consequence of spin-orbit coupling and only Eq. (71) holds. This can be understood if we realize that a lack of inversion symmetry, $V(\mathbf{r}) \neq V(-\mathbf{r})$, will result in a non-vanishing potential gradient or electric field, $\mathbf{E}(\mathbf{r})$. As we have seen in the last section an electron moving in an electric field will experience this field Lorentz-transformed as $\mathbf{B}$-field and

$$\varepsilon(\mathbf{k},\uparrow) \neq \varepsilon(\mathbf{k},\downarrow)\,. \tag{72}$$

This will, depending on symmetry, result in different consequences for the bandstructures.
Performing a Taylor expansion of the potential $V(\mathbf{r})$, $V(\mathbf{r}) = V_0 + e\mathbf{E}(\mathbf{r})\cdot\mathbf{r} + \cdots$, in lowest order the inversion asymmetry of the potential $V(\mathbf{r})$ is characterized by an electric field $\mathbf{E}(\mathbf{r})$. When electrons with an effective mass m^* propagate with a velocity $\mathbf{v} = \mathrm{d}\varepsilon/\mathrm{d}\mathbf{p} = \frac{1}{m^*}\mathbf{k}$ in an external electric field $\mathbf{E}$ defined in a global frame of reference, then the relativistic Lorentz transformation gives rise to magnetic field $\mathbf{B} = \frac{1}{c}(\mathbf{v}\times\mathbf{E}) = \frac{1}{m^*c}(\mathbf{k}\times\mathbf{E})$ in local frame of the moving electron. We have seen this term appearing in the spin-orbit coupling Hamiltonian in Eq.(69) in the last chapter. Again, the interaction of the spin with this $\mathbf{B}$ field leads to a coupling term. This term is called Rashba or Bychkov-Rashba Hamiltonian [39, 40]

$$H_\mathrm{R} = \alpha_\mathrm{R}\boldsymbol{\sigma}\cdot(\mathbf{p}\times\mathbf{E}) \qquad \text{or} \qquad H_\mathrm{R} = \alpha_\mathrm{R}\,\boldsymbol{\sigma}\cdot(\mathbf{k}\times\mathbf{E}) \qquad \text{or} \qquad H_\mathrm{R} = \alpha_\mathrm{R}(|\mathbf{E}|)\,\boldsymbol{\sigma}\cdot(\mathbf{k}\times\hat{\mathbf{e}}) \tag{73}$$

describing the Rashba spin-orbit coupling as additional contribution to the kinetic energy. $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices, Eq.(68). The latter two terms are strictly correct only for plane wave eigenstates as, e.g. for a two-dimensional electron gas (2DEG). An important realization of a 2DEGs are electrons in doped semiconductor heterostuctures, that support an

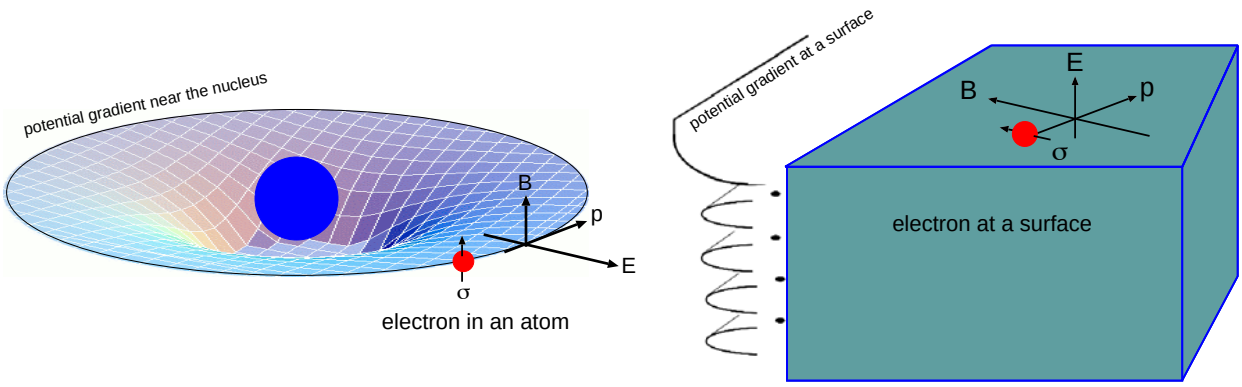


Fig. 8: *Schematic illustration of an electron moving in the potential gradient near a nucleus (left) and a surface (right). In both cases the electric field* $\mathbf{E}$ *resulting from the potential gradient is Lorentz-transformed into a* $\mathbf{B}$ *field by the motion of the electron. In the case of the atomic spin-orbit coupling (left) this is an orbital motion, in the case of the Rashba-effect (right) it is a linear motion. In both cases, the electrons spin,* $\boldsymbol{\sigma}$ *couples to the resulting* $\mathbf{B}$ *field.*

electron gas at the interface between two materials, e.g. (InGa)As and InP [41]. Another possibility to study the Rashba-effect in 2DEGs is shown in figure 8: on surfaces which support a surface state, e.g. in Au(111) [42] the electrons of the surface state move in a potential gradient that is provided by the surface itself (but can also be modified slightly by external electric fields [43]). More examples will follow in other lectures.
The general features of the Rashba-model are studied for the 2DEG in a potential with structural inversion asymmetry (SIA) and the corresponding bandstructure are displayed schematically in figure 9. For electrons propagating in the 2DEG extended in the (x, y) plane subject to an electric field normal to the 2DEG, $\hat{\mathbf{e}}_z = (0, 0, 1)$, the Hamiltonian takes the form

$$H = H_{\mathrm{K}} + H_{\mathrm{R}} = \frac{\mathbf{p}_\parallel^2}{2m^*} + \alpha_{\mathrm{R}} \left(\boldsymbol{\sigma} \times \mathbf{p}_\parallel\right)_{|z} = \frac{\mathbf{p}_\parallel^2}{2m^*} + \alpha_{\mathrm{R}} \left(\sigma_x p_y - \sigma_y p_x\right), \tag{74}$$

which can be solved analytically. For a Bloch vector in the plane of the 2DEG, $\mathbf{k}_\parallel = (k_x, k_y, 0) = k_\parallel(\cos\varphi, \sin\varphi, 0)$, the eigenstates written as a product of plane wave in space and two-component spinor are

$$\psi_{\pm\mathbf{k}_\parallel}(\mathbf{r}_\parallel) = \frac{e^{i\mathbf{k}_\parallel\cdot\mathbf{r}_\parallel}}{2\pi} \frac{1}{\sqrt{2}} \begin{pmatrix} ie^{-i\varphi/2} \\ \pm e^{i\varphi/2} \end{pmatrix} \tag{75}$$

with eigenenergies

$$\varepsilon_\pm(\mathbf{k}_\parallel) = \frac{\mathbf{k}_\parallel^2}{2m^*} + \alpha_{\mathrm{R}} \left(\boldsymbol{\sigma} \times \mathbf{k}_\parallel\right) = \frac{\mathbf{k}_\parallel^2}{2m^*} \pm \alpha_R |\mathbf{k}_\parallel| = \frac{1}{2m^*}(k_\parallel \pm k_{\mathrm{SO}})^2 - \Delta_{\mathrm{SO}}\,, \tag{76}$$

where $\pm$ denotes the spin-up and -down states with respect to a spin orientation axis $\hat{\mathbf{n}}(\mathbf{k}_\parallel)$, local in $\mathbf{k}_\parallel$ space. With the exception of the high-symmetry state $k_\parallel = 0$, we find that the original two-fold degenerate energy paraboloid of the 2DEG in a constant potential is indeed spin-split. This splitting $\varepsilon_+(\mathbf{k}_\parallel) - \varepsilon_-(\mathbf{k}_\parallel) = 2\alpha_{\mathrm{R}} k_\parallel$ is linear in $k_\parallel$. Due to the presence of the SIA potential and the spin-orbit interaction, the origin of the degenerate parabola is shifted by $k_{\mathrm{SO}} = m^*\alpha_{\mathrm{R}}$, but in opposite directions for up- and down-spins with in overall spin-orbit

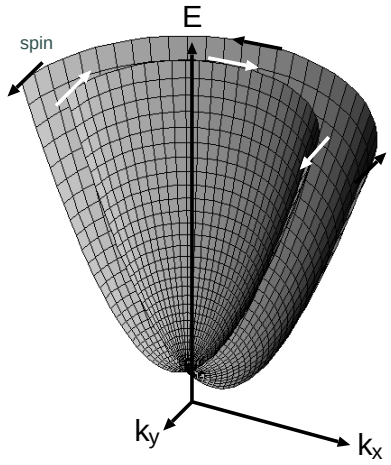


Fig. 9: *Cut through the parabolic energy dispersions of a two-dimensional electron gas in a structure inversion asymmetric (SIA) environment. Indicated are the vector fields of the spin-quantization axes (or the patterns of the spin) at the Fermi surface. As the opposite spins have different energies, the Fermi surface becomes two concentric circles with opposite spins. The effective B-field,* $\mathbf{B}_{\rm eff}$ *is always perpendicular to the propagation direction defined by* $\mathbf{k}_\parallel$.

lowering of $\Delta_{\rm SO} = m^*\alpha_{\rm R}/2$. The orientation axis is given by the expectation value

$$\hat{\mathbf{n}}_\pm(\mathbf{k}_\parallel) = \langle\psi_{\pm\mathbf{k}_\parallel}|\boldsymbol{\sigma}|\psi_{\pm\mathbf{k}_\parallel}|\rangle = \pm\begin{pmatrix}\sin\varphi\\ -\cos\varphi\\ 0\end{pmatrix} \quad \perp \quad \mathbf{k}_\parallel = k_\parallel\begin{pmatrix}\cos\varphi\\ \sin\varphi\\ 0\end{pmatrix}. \qquad (77)$$

We find that the orientation axis is independent of the magnitude $k_\parallel$ and depends only on the direction of the $\mathbf{k}_\parallel$ vector. In fact, it is in the plane of the 2DEG and the orientation axis is perpendicular to the propagation direction of the electron. Considering $\mathbf{k}_\parallel \to -\mathbf{k}_\parallel$, φ changes to $\varphi + \pi$, we find that the spin orientation axis reverses as indicated in figure 9. Thus for $\mathbf{k}_\parallel$ and $-\mathbf{k}_\parallel$ the spin-up and -down states refer to opposite orientations. Defining a global quantization axis along the line $(-\mathbf{k}_\parallel, \mathbf{k}_\parallel)$, e.g. according to $\hat{\mathbf{n}}_\pm(+\mathbf{k}_\parallel)$, then a spin-up state appears as spin-down state if $\mathbf{k}_\parallel$ changes sign. Together with the eigenvalue spectrum given in equation (76) the Kramer degeneracy $\varepsilon_\uparrow(\mathbf{k}_\parallel) = \varepsilon_\downarrow(-\mathbf{k}_\parallel)$ holds. In all, the magnetic moment is zero when averaged over all states $\mathbf{k}_\parallel$. This is consistent with the absence of an $\mathbf{B}$ field.

The Rashba spin-orbit splitting may be observed either by electron photoemission or transport experiments. Transport experiments work typically on a shell of constant energy ε. Here we expect at a given energy two different wave vectors for up- and down-electrons, which will be exploited in the Datta-Das proposal of a spin-transistor [44].

That the Rashba-type spin-orbit coupling may have important consequences for the one-electron energy levels in bulk semiconductors was first emphasized by Dresselhaus *et al*. [45] already in 1955. Unlike the diamond structure of Si and Ge, the zinc blende structure, in which for example the III-V semiconductor crystallize, exhibit a bulk inversion asymmetry (BIA), i.e. this crystal structure lacks a center of inversion, so that we can have a spin splitting of the electron and hole states at nonzero wave vectors $\mathbf{k}$ as for the Rashba effect even if $\mathbf{B} = 0$. Today, this is called the Dresselhaus effect. The corresponding Dresselhaus Hamiltonian

$$H_{\rm D} = \alpha_{\rm D}\left[\sigma_x p_x\left(p_y^2 - p_z^2\right) + \sigma_y p_y\left(p_z^2 - p_x^2\right) + \sigma_z p_z\left(p_x^2 - p_y^2\right)\right] \qquad (78)$$

describes the BIA spin splitting due to the Dresselhaus spin-orbit coupling, which produces spin vector fields quite different from those produced by the SIA splitting. One difference is obviously that the Dresselhaus term produced a spin splitting which is proportional to k^3, $\varepsilon_{\rm D} \propto k^3$, while the spin splitting of the Rashba-term is linear in k, $\varepsilon_{\rm R} \propto k$.

In a magnetic system time-reversal symmetry is broken and there is a shift between the spin-up and spin-down states, e.g. in the Stoner-model for a magnetization M and the intraatomic

exchange integral I we find $\varepsilon(\mathbf{k},\uparrow) = \varepsilon(\mathbf{k},\downarrow) - IM$. In addition to this splitting, the Rashba-type spin-orbit splitting can add a term $\alpha_{\mathrm{R}}k$ to the energy difference between electrons of the same $\mathbf{k}$ but different spin. However, in a magnetic material the exchange interaction tries to align the electron spins in a parallel (collinear) manner and there is a well-defined spin-quantization axis (SQA) that fixes the orientation of the spins in the crystal. Then, only electrons moving in a direction perpendicular to the SQA can be influenced by a Rashba-type effect, for electrons with $\mathbf{k}$ parallel to the SQA the term $\boldsymbol{\sigma} \cdot (\mathbf{E} \times \mathbf{k})$ has to vanish [46].

4.3 Magnetic anisotropy

As mentioned shortly at the end of the last subsection, in a magnetic system the spin-orbit induced splittings in a bandstructure will be influenced by the direction of the spin-quantization axis. This gives a small, but important contribution to the total energy of a magnetic system since in a non-relativistic Hamiltonian there is no term which could give a dependence of the total energy on the direction of the SQA. Generally, a dependence of the total energy of the magnetization-direction (with respect to the crystal axes) is termed a magnetic anisotropy. This anisotropy fixes the magnetization direction w.r.t. the lattice and allows for a stable magnetization direction in a material, which is the basis of almost all magnetic applications.
There are several interactions that can lead to a magnetic anisotropy, e.g. the dipolar interaction [49] which is of substantial importance in bulk materials. In low-dimensional magnets also other interactions that lead to magnetically anisotropic behavior can become dominant, most importantly the spin-orbit coupling. The magnetocrystalline anisotropy energy (MAE) results from the anisotropy of the spin-orbit interaction, i.e. it is the difference of total energies obtained from Hamiltonians including the spin-orbit coupling term with the magnetization pointing in two different directions.
To see how this can happen, remember that out of certain d-levels, only orbital moments pointing in a certain direction can be formed. E.g. a d_{xy} and a $d_{x^2-y^2}$ orbital can only be combined to form an orbital moment in z direction. An orbital moment pointing in x-direction has to be formed from electrons that can move in the (y,z)-plane, and this is impossible within only the d_{xy} and $d_{x^2-y^2}$ orbitals. If now two appropriate orbitals are degenerate and occupied by a single electron (and thus forming the Fermi level), it is rather straightforward to identify the resulting direction of the orbital moment using group theory [47]. In a metal, where several bands are crossing the Fermi level, ϵ_{F}, it is basically the sum of all contributions from bands near ϵ_{F} that determine the orbital moment. In second-order perturbation theory the expectation value of the orbital moment operator $\mathbf{L}$ can be written as:

$$\langle \mathbf{L} \rangle = \sum_{i,j} \frac{\langle \psi_i | \mathbf{L} | \psi_j \rangle \langle \psi_j | H_{\mathrm{so}} | \psi_i \rangle}{\epsilon_i - \epsilon_j} f(\epsilon_i) \left[1 - f(\epsilon_j)\right], \tag{79}$$

where H_{so} is the spin-orbit coupling Hamiltonian and f is the Fermi function ensuring that the wavefunction ψ_i is occupied and ψ_j is unoccupied. Van der Laan [47] has shown, that in the absence of spin-flip terms (i.e. when the majority and minority band are well separated by the exchange interaction), the spin-orbit coupling changes the total energy of a system in second-order perturbation theory as:

$$\delta E = \sum_{i,j} \frac{\langle \psi_i | H_{\mathrm{so}} | \psi_j \rangle \langle \psi_j | H_{\mathrm{so}} | \psi_i \rangle}{\epsilon_i - \epsilon_j} f(\epsilon_i) \left[1 - f(\epsilon_j)\right] \approx -\frac{1}{4} \xi \hat{\mathbf{S}} \cdot \left[\left\langle \mathbf{L}^{\downarrow} \right\rangle - \left\langle \mathbf{L}^{\uparrow} \right\rangle \right] \tag{80}$$

where ξ is the radial part of the spin-orbit Hamiltonian (Eq. (70)), $\hat{\mathbf{S}}$ is the direction of the spin moment, and $\mathbf{L}^{\downarrow}$ and $\mathbf{L}^{\uparrow}$ are the orbital moment vectors of the spin-down and spin-up bands, respectively. If the spin-up band is completely filled, we see that energy change, δE, is proportional to the size of the orbital moment and the magnetocrystalline anisotropy energy (MAE), i.e. the difference of δE for two different magnetization directions, will be proportional to the difference in the orbital moments. This relation between orbital moment anisotropy and MAE was first derived by Bruno [48].

Practically, one starts from a solution Ψ_0 of the Schrödinger equations (possibly including scalar-relativistic corrections, cf. Appendix), and then solves the Hamiltonian including the spin-orbit coupling term with the spin-quantization axis turned into the required direction by means of a spin-rotation matrix U:

$$\begin{aligned}\langle U\Psi_0|H_{\mathrm{S}}+\xi\boldsymbol{\sigma}\cdot\mathbf{L}|U\Psi_0\rangle &= \langle\Psi_0|H_{\mathrm{S}}|\Psi_0\rangle+\xi\,\langle U\Psi_0|\boldsymbol{\sigma}\cdot\mathbf{L}|U\Psi_0\rangle = \\ &= \varepsilon_0+\xi\left\langle\begin{array}{c}\Psi_0^{\uparrow}\\ \Psi_0^{\downarrow}\end{array}\right|U^{\dagger}\begin{pmatrix}L_z & L_x-iL_y\\ L_x+iL_y & -L_z\end{pmatrix}U\left|\begin{array}{c}\Psi_0^{\uparrow}\\ \Psi_0^{\downarrow}\end{array}\right\rangle \quad (81)\end{aligned}$$

If ξ or the orbital moment is small, the last part of Eq. (81) is only a small correction to the energy ε_0 obtained from the Schrödinger equation and the magnetization direction of the solution will point into the direction of the spin-quantization axis described by U.

From the above equations it is clear that both, strong spin-orbit coupling and a sizeable orbital moment, $\mathbf{L}$, are necessary for a large contribution to the magnetic anisotropy. But it is also necessary that the spin-orbit interaction gives different energy contributions for different magnetizations of the sample. In principle there are two possibilities to imagine how this could happen: (i) the orbital moment is fixed to the lattice and its projection on the axis of the spin moment varies with the magnetization direction or (ii) the spin and orbital moments are collinear and depending on the magnetization direction the size of the orbital moment varies. Normally, we observe collinear spin- and orbital moments. The rotation of the orbital moment by an external magnetic field can then lead to structural changes of the crystal. This phenomenon is called magnetostriction and is discussed e.g. in Ref. [49].

The MAE is a typically a small energy, for elemental bulk magnets it is in the order of microelectronvolts (μeV). This is mainly a consequence of the high symmetry in these bulk systems. Low-dimensional systems (thin films, chains and wires) can show much higher MAE's, up to a few milli-electronvolts. Since other sources of magnetic anisotropy can be even smaller in these systems, spin-orbit coupling can get very important in magnetic nanostructures.

Sometimes, in analogy to the Heisenberg Hamiltonian describing the exchange interaction in a crystal, the spin-orbit coupling is cast into a form $\sum_i \xi_i L_i S_i$ where i is a particular atomic site. Then, evidently, another term coupling the spin of a site i to the orbital motion at site j is conceivable: $C_{ij}L_jS_i$. This spin-other orbit interaction is, like the dipole-dipole interaction, derived from the Breit equation. In the Hartree approximation it was included in *ab-initio* calculations but was found to be much weaker than the formerly described spin- (same) orbit interaction [50].

Appendix

The Dirac equation

A relativistic theory for an electron (of charge $-e$) in an external scalar potential V and a vector potential $\mathbf{A}$ can be formulated via the Dirac equation (for clarity, no atomic units are used in this section)

$$\mathcal{H}\Psi = +i\hbar\frac{\partial}{\partial t}\Psi = \varepsilon\Psi; \qquad \mathcal{H} = -eV(\mathbf{r}) + \beta mc^2 + \boldsymbol{\alpha}\cdot(c\mathbf{p} + e\mathbf{A}(\mathbf{r})). \tag{82}$$

Here, $\boldsymbol{\alpha}$ is a vector of 4×4 matrices, that can be written in terms of the Pauli spin-matrices, $\boldsymbol{\sigma}$, while β is a matrix of same rank, expressible in terms of the 2×2 unit matrix I_2:

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The Hamiltonian acts on a four-component wavefunction Ψ that can be written as a 2-vector of the so-called large and small components, ψ and χ. For this components the Hamiltonian of the Dirac equation can be written as

$$(\varepsilon - 2mc^2 + eV(\mathbf{r}))\psi = \boldsymbol{\sigma}\cdot(c\mathbf{p} + e\mathbf{A}(\mathbf{r}))\chi \tag{83}$$
$$(\varepsilon + 2mc^2 + eV(\mathbf{r}))\chi = \boldsymbol{\sigma}\cdot(c\mathbf{p} + e\mathbf{A}(\mathbf{r}))\psi \tag{84}$$

where $\boldsymbol{\sigma}$ is the vector of Pauli matrices. In the non-relativistic limit, these equations reduce to the Schrödinger equation for the large component. Discussions of the Dirac theory are available in most textbooks on quantum mechanics, we follow here the book of Bethe and Salpeter [38]. Substituting Eq.(84) in (83) and retaining only terms up to order $(v/c)^2$, it is possible to formulate an equation (sometimes termed Pauli equation) for the large component only:

$$\left[\varepsilon + eV(\mathbf{r}) + \frac{\hbar^2}{2m}\nabla^2 + \frac{1}{2mc^2}\left(\varepsilon + eV(\mathbf{r})\right)^2 + i\frac{e\hbar}{mc}\mathbf{A}(\mathbf{r})\cdot\nabla - \frac{e^2}{2mc^2}A^2(\mathbf{r}) \quad + \right.$$
$$\left. + i\frac{e\hbar}{(2mc)^2}\mathbf{E}(\mathbf{r})\cdot\mathbf{p} - \frac{e\hbar}{(2mc)^2}\boldsymbol{\sigma}\cdot(\mathbf{E}(\mathbf{r})\times\mathbf{p}) - \frac{e\hbar}{2mc}\boldsymbol{\sigma}\cdot\mathbf{H}(\mathbf{r})\right]\psi = 0 \tag{85}$$

where the gradient of V and the curl of $\mathbf{A}$ have been written explicitly as electric ($\mathbf{E}$) and magnetic ($\mathbf{H}$) fields. In the non-relativistic limit the first three terms give the ordinary Schrödinger equation. The fourth term gives the relativistic correction due to the change of the mass with velocity. In absence of a vector potential (and magnetic field) only two more terms involving the electric field remain. The expression including the triple product $\boldsymbol{\sigma}\cdot(\mathbf{E}(\mathbf{r})\times\mathbf{p})$ is the spin-orbit coupling term, discussed in section 4.1. Ignoring this term and retaining the $\mathbf{E}(\mathbf{r})\cdot\mathbf{p}$ term (which has no classical analogon) gives a spin-free equation, the so-called scalar relativistic version of the Schrödinger equation.

References

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College, Philadelphia, 1976.

[2] V. L. Moruzzi, J. F. Janak, and A. R. Williams, *Calculated Electronic Properties of Metals*, Pergamon, New York, 1978.

[3] M. I. Aroyo, J. M. Perez-Mato, C. Capillas, E. Kroumova, S. Ivantchev, G. Madariaga, A. Kirov, and H. Wondratschek, Bilbao Crystallographic Server I: Databases and crystallographic computing programs, Z. f. Kristallogr. **221**, 15–27 (2006).

[4] J. E. Inglesfield and E. W. Plummer, The Physics of Photoemission, in *Angle resolved photoemission: theory and current applications*, edited by S. D. Kevan, volume 74 of *Studies in surface science and catalysis*, pages 15–61, Amsterdam, 1992, Elsevier.

[5] D. D. Koelling and B. N. Harmon, A technique for relativistic spin-polarised calculations, J. Phys. C: Solid State Phys. **10**, 3107–3114 (1977).

[6] J. Grotendorst, S. Blügel, and D. Marx, editors, *Computational Nanoscience: Do It Yourself!*, volume 31 of *NIC Series*, Jülich, 2006, Research Center Jülich, also available at `http://www.fz-juelich.de/nic-series/volume31/`.

[7] F. Aryasetiawan and O. Gunnarsson, The GW method, Rep. Prog. Phys. **61**, 237–312 (1998).

[8] J. C. Slater, Note on Hartree's method, Phys. Rev. **35**, 210–211 (1929).

[9] J. C. Slater, A Simplification of the Hartree-Fock Method, Phys. Rev. **81**, 385–390 (1951).

[10] L. D. Landau, Theory of the Fermi liquid, Soviet Phys.-JETP **3**, 920 (1956).

[11] P. Hohenberg and W. Kohn, Inhomogeneous Electron Gas, Phys. Rev. **136**, B864–B871 (1964).

[12] W. Lenz, Über die Anwendbarkeit der statistischen Methode auf Ionengitter, Z. Physik **77**, 713–721 (1932).

[13] L. H. Thomas, The calculation of atomic fields, Proc. Cambridge Philos. Soc. **23**, 542–548 (1927).

[14] E. Fermi, Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente, Z. Physik **48**, 73–79 (1928).

[15] P. A. M. Dirac, Note on Exchange Phenomena in the Thomas–Fermi Atom, Proc. Cambridge Philos. Soc. **26**, 376–385 (1930).

[16] J. C. Slater and H. M. Krutter, The Thomas-Fermi Method for Metals, Phys. Rev. **47**, 559–568 (1934).

[17] W. Kohn and L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, Phys. Rev. **140**, A1133–A1138 (1965).

[18] W. Kohn and P. Vashista, General density functional theory, in *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. H. March, pages 79–147, New York, 1983, Plenum.

[19] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, Descriptions of exchange and correlation effects in inhomogeneous electron systems, Phys. Rev. B **20**, 3136–3164 (1979).

[20] O. Gunnarsson and B. I. Lundqvist, Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism, Phys. Rev. B **13**, 4274–4298 (1976).

[21] D. van der Marel and G. A. Sawatzky, Electron-electron interaction and localization in d and f transition metals, Phys. Rev. B **37**, 10674–10684 (1988).

[22] J. C. Slater, The theory of complex spectra, Phys. Rev. **34**, 1293–1322 (1929).

[23] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Band theory and Mott insulators: Hubbard U instead of Stoner I, Phys. Rev. B **44**, 943–954 (1991).

[24] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LSDA+U method, J. Phys.: Condens. Matter **9**, 767–808 (1997).

[25] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Density-functional theory and strong interactions: Orbital ordering in Mott-Hubbard insulators, Phys. Rev. B **52**, R5467–R5470 (1995).

[26] P. Kurz, G. Bihlmayer, and S. Blügel, Magnetism and electronic structure of hcp Gd and the Gd(0001) surface, J. Phys.: Condens. Matter **14**, 6353–6371 (2002).

[27] I. Turek, J. Kudrnovský, M. Diviš, P. Franek, G. Bihlmayer, and S. Blügel, A first-principle study of electronic structure and exchange interactions in bcc europium, Phys. Rev. B **68**, 224431/1–7 (2003b).

[28] N. Binggeli and M. Altarelli, Orbital ordering, Jahn-Teller distortion, and resonant x-ray scattering in $KCuF_3$, Phys. Rev. B **70**, 085117 (2004).

[29] I. Leonov, N. Binggeli, D. Korotin, V. I. Anisimov, N. Stojić, and D. Vollhardt, Structural Relaxation due to Electronic Correlations in the Paramagnetic Insulator $KCuF_3$, Phys. Rev. Lett. **101**, 096405 (2008).

[30] E. Pavarini, E. Koch, and A. I. Lichtenstein, Mechanism for Orbital Ordering in $KCuF_3$, Phys. Rev. Lett. **101**, 266405 (2008).

[31] I. V. Solovyev, P. H. Dederichs, and V. I. Anisimov, Corrected atomic limit in the local-density approximation and the electronic structure of d impurities in Rb, Phys. Rev. B **50**, 16861–16871 (1994).

[32] I. V. Solovyev and M. Imada, Screening of Coulomb interactions in transition metals, Phys. Rev. B **71**, 045103/1–11 (2005).

[33] S. Picozzi, K. Yamauchi, G. Bihlmayer, and S. Blügel, First-principles stabilization of an unconventional collinear magnetic ordering in distorted manganites, Phys. Rev. B **74**, 094402 (2006).

[34] J. G. W. Pratt, Generalization of Band Theory to Include Self-Energy Corrections, Phys. Rev. **118**, 462 (1959).

[35] L. Hedin, New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem, Phys. Rev. **139**, A796–A823 (1965).

[36] G. Onida, L. Reining, and A. Rubio, Electronic excitations: density-functional versus many-body Green's-function approaches, Rev. Mod. Phys. **74**, 601–659 (2002).

[37] J. D. Jackson, *Classical electrodynamics*, Wiley & Sons, 1962.

[38] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Systems*, Plenum, New York, 1977.

[39] Y. A. Bychkov and E. I. Rashba, Oscillatory effects and the magnetic-susceptibility of carriers in inversion-layers, J. Phys. C: Solid State Phys. **17**, 6039 (1984).

[40] Y. A. Bychkov and E. I. Rashba, Properties of a 2D eelctron-gas with lifted spectral degeneracy, Sov. Phys. JETP Lett **39**, 78 (1984).

[41] T. Schäpers, J. Knobbe, and V. A. Guzenko, Effect of Rashba spin-orbit coupling on magnetotransport in InGaAs/InP quantum wire structures, Phys. Rev. B **69**, 235323 (2004).

[42] S. LaShell, B. A. McDougall, and E. Jensen, Spin Splitting of an Au(111) Surface State Band Observed with Angle Resolved Photoelectron Spectroscopy, Phys. Rev. Lett. **77**, 3419 (1996).

[43] G. Bihlmayer, Y. M. Koroteev, P. M. Echenique, E. V. Chulkov, and S. Blügel, The Rashba-effect at metallic surfaces, Surf. Sci. **600**, 3888 (2006).

[44] S. Datta and B. Das, Electronic analog of the electro-optic modulator, Appl. Phys. Lett. **56**, 665 (1990).

[45] G. Dresselhaus, Spin-orbit coupling effects in zinc blende structures, Phys. Rev. **100**, 580 (1955).

[46] O. Krupin, G. Bihlmayer, K. Starke, S. Gorovikov, J. E. Prieto, K. Döbrich, S. Blügel, and G. Kaindl, Rashba effect at magnetic metal surfaces, Phys. Rev. B **71**, 201403 (R) (2005).

[47] G. van der Laan, Microscopic origin of magnetocrystalline anisotropy in transition metal thin films, J.Phys.: Condens. Matter **10**, 3239 (1998).

[48] P. Bruno, Tight-binding approach to the orbital magnetic moment and magnetocrystalline anisotropy of transition-metal monolayers, Phys. Rev. B **39**, 865 (1989).

[49] S. Blügel, *Materie und Material (2)*, chapter Magnetische Anisotropie und Magnetostriktion, Schriften des FZ-Jülich, Jülich, 1999.

[50] M. D. Stiles, S. V. Halilov, R. A. Hyman, and A. Zangwill, Spin-other-orbit interaction and magnetocrystalline anisotropy, Phys. Rev. B **64**, 104430 (2001).

A 2 Electronic Basis of Magnetism

Rudolf Zeller
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Classical and Quantum Theory **2**
2.1 Bohr-van Leeuwen Theorem 2
2.2 Paramagnetism and Diamagnetism 3
2.3 Electron Spin 3

3 Spin Density Functional Theory **5**
3.1 Basic Formalism 5
3.2 Stoner Model for Ferromagnetism 7
3.3 Ferromagnetism of Alloys 12
3.4 Mapping to Model Hamiltonians 14

4 Relativistic Effects **15**

1 Introduction

Although magnetic materials are known since ancient times and much about the relation between electricity and magnetism has been discovered by Ampere, Faraday, Oersted, Maxwell and others in the nineteenth century, the key for understanding of magnetism was only provided by the development of quantum mechanics in the twentieth century. In my lecture I will first show that classical theory for a system of moving electron in an external magnetic field leads to a vanishing magnetization so that the description of magnetism necessarily requires quantum theory. Quantum theory for a single electron is then used to explain the origin of diamagnetism, paramagnetism and spin magnetism. Then I will show how the many-electron system can be treated by spin density functional theory, that this theory provides a powerful tool to study ferromagnetism in elements and alloys and that it can be used to obtain parameters for simpler model Hamiltonians. Finally, I will discuss that for practical applications of magnetism also relativistic effects are important.

2 Classical and Quantum Theory

2.1 Bohr-van Leeuwen Theorem

The magnetization $\underline{m}$ in classical statistical physics can be obtained from the derivative of the free energy $F = -k_B T \ln Z$ with respect to the magnetic field $\underline{B}$ as

$$\underline{m} = -\frac{1}{V}\frac{\partial F}{\partial \underline{B}} = \frac{k_B T}{V}\frac{\partial \ln Z}{\partial \underline{B}} \,, \tag{1}$$

where V is the volume, T the temperature and k_B the Boltzmann constant. For a classical system of N electrons the partition function Z is given by an integral over all momenta $\underline{p}_i$ and coordinates $\underline{r}_i$ as

$$Z = \int \cdots \int \exp\left(-\frac{H}{k_B T}\right) \mathrm{d}\underline{p}_1 \ldots \mathrm{d}\underline{p}_N \mathrm{d}\underline{r}_1 \ldots \mathrm{d}\underline{r}_N \,. \tag{2}$$

In an electric field described by a potential v and a magnetic field $\underline{B} = \underline{\nabla} \times \underline{A}$ described by a vector potential $\underline{A}$ the Hamilton function H can be written as

$$H = \sum_i^N \left\{ \frac{1}{2m_e} \left[\underline{p}_i - e\underline{A}(\underline{r}_i) \right]^2 + v(\underline{r}_i) \right\} + \frac{1}{4\pi\epsilon_0}\frac{e^2}{2} \sum_{i \neq j} \sum \frac{1}{|\underline{r}_i - \underline{r}_j|} \,, \tag{3}$$

where ϵ_0 is the electric constant and m_e and e electron mass and charge. The substitution $\underline{p}'_i = \underline{p}_i - e\underline{A}(\underline{r}_i)$ for all momentum integration variables in (2) leads to the result that the partition function Z cannot depend on $\underline{A}$ and consequently (1) gives $\underline{m} = 0$. This result that in an external magnetic field in thermal equilibrium a classical system of moving electrons has a vanishing magnetization is known as the Bohr-van Leeuwen theorem because it has been derived by Niels Bohr in 1911 and independently by Johanna Hendrika van Leeuwen in 1919 in their doctoral dissertations.

In quantum theory the situation is different. The electron system is described by a Schrödinger equation $\hat{H}\Psi(\underline{r}, t) = \mathrm{i}\hbar\partial_t \Psi(\underline{r}, t)$, where the Hamilton operator

$$\hat{H} = \sum_i^N \left\{ \frac{1}{2m_e} \left[-\mathrm{i}\hbar\underline{\nabla}_i - e\underline{A}(\underline{r}_i) \right]^2 + v(\underline{r}_i) \right\} + \frac{1}{4\pi\epsilon_0}\frac{e^2}{2} \sum_{i \neq j} \sum \frac{1}{|\underline{r}_i - \underline{r}_j|}$$

is obtained from the Hamilton function (3) by replacing the classical momentum variables $\underline{p}_i$ by momentum operators $\hat{\underline{p}}_i = -\mathrm{i}\hbar\underline{\nabla}_i$. Since the momentum operator $\hat{\underline{p}}_i$ does not commute with the coordinate $\underline{r}_i$ and the vector potential $\underline{A}(\underline{r}_i)$, the Bohr-van Leeuwen theorem does not apply with the consequence that magnetism is a pure quantum phenomenon.

2.2 Paramagnetism and Diamagnetism

To simplify the discussion only a single electron will be considered now. Here the stationary Schrödinger equation can be written as

$$\left\{-\frac{\hbar^2}{2m_e}\nabla^2 + \frac{\mathrm{i}e\hbar}{2m_e}\left(\underline{\nabla}\cdot\underline{A}(\underline{r}) + \underline{A}(\underline{r})\cdot\underline{\nabla}\right) + \frac{e^2}{2m_e}A^2(\underline{r}) + v(\underline{r})\right\}\Psi = E\Psi\,. \tag{4}$$

The term linear in $A(\underline{r})$ gives rise to paramagnetism and the term quadratic in $A(\underline{r})$ to diamagnetism. With (31) and (32) derived in the appendix it can be shown that for a constant magnetic field $\underline{B}$ equation (4) simplifies into

$$\left\{-\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e}{2m_e}\underline{L}\cdot\underline{B} + \frac{e^2}{8m_e}r^2B^2(1-\cos^2\vartheta) + v(\underline{r})\right\}\Psi = E\Psi\,, \tag{5}$$

where $r = |\underline{r}|$ and $B = |\underline{B}|$ denote absolute values, ϑ is the angle between $\underline{r}$ and $\underline{B}$ and $\underline{L} = -\mathrm{i}\hbar\underline{r}\times\underline{\nabla}$ is the angular momentum operator. The relative size of the paramagnetic and diamagnetic terms in (5) can be estimated using $r \approx a_0$, where a_0 is the Bohr radius, and $L = |\underline{L}| \approx \hbar$. The ratio of diamagnetic to paramagnetic contribution is then given by

$$\frac{(e^2/8m_e)a_0^2B^2}{(e/2m_e)\hbar B} = \frac{ea_0^2}{4\hbar}B\,.$$

With $e = 1.602\times10^{-19}$ As, $a_0 = 0.529\times10^{-10}$ m and $\hbar = 1.054\times10^{-34}$ Js one obtains a factor $ea_0^2/4\hbar$ which is the order of 10^{-6} T^{-1}. (J = m^2 kg s^{-2} and T = kg s^{-2} A^{-1} are the derived SI units joule and tesla). Thus for fields achievable in laboratories (about 10 T) the diamagnetism of atoms is much smaller than their paramagnetism. For other systems, for instance for metallic electrons, diamagnetism and paramagnetism can be of comparable magnitude as the result $\chi_L = -\frac{1}{3}\chi_P$ for the diamagnetic Landau and paramagnetic Pauli susceptibilities of a free electron gas demonstrates. It is also instructive to compare the paramagnetic term with the Coulomb energy $e^2/4\pi\epsilon_0 a_0$. Their ratio can be estimated as

$$\frac{(e/2m_e)\hbar B}{e^2/4\pi\epsilon_0 a_0} = \frac{2\pi\hbar\epsilon_0 a_0}{em_e}B\,.$$

With $\epsilon_0 = 8.854\times10^{-12}$ Fm^{-1} and $m_e = 9.109\times10^{-31}$ kg one obtains a factor $2\pi\hbar\epsilon_0 a_0/em_e$ which is of the order of 2×10^{-6} T^{-1}. (F = s^4 A^2 kg^{-1} m^{-2} is the derived SI unit farad.) Thus for laboratory conditions magnetic effects arising from the external fields $\underline{A}$ or $\underline{B}$ in (4) and (5) are rather small.

2.3 Electron Spin

In 1922 Stern and Gerlach found that in an inhomogeneous magnetic field a beam of silver atoms splits into two subbeams. This famous result cannot be explained by the Schrödinger

equation given above. The angular momentum operator $\underline{L}$ in (5) splits the beam into $2l+1$ subbeams, for $l=0$ (s state) no splitting is expected and for $l=1$ (p state) three subbeams should appear. Thus the observed splitting cannot be caused by the orbital motion of the electron. The reason for the two subbeams is an internal degree of freedom, the electron spin discovered by Uhlenbeck and Gouldsmith in 1925. In 1927 Pauli proposed to add a term $g\mu_B \underline{S} \cdot \underline{B}$ to (5) and to use a two component wavefunction $\Psi^{\pm}$, where $\underline{S}$ is the spin operator, $\mu_B = e\hbar/2m_e$ the Bohr magneton and $g=-2$ the Landé-g-factor for the electron. One year later the relativistic Dirac equation was discovered and, because the Pauli term naturally follows from the non-relativistic limit of the Dirac equation, for historical reason spin is often considered as a relativistic phenomenon.
However, the Pauli term can also be derived without a relativistic formalism. It is only necessary to linearize the quantum mechanical wave equation. Whereas in a relativistic formalism Lorentz invariance requires a linear relation of space and time variables, which is achieved in the four component Dirac equation, non-relativistically a linear relation only between the three space or momentum variables is required. Obviously the linearization

$$\gamma_0(p_x^2+p_y^2+p_z^2) = (\gamma_1 p_x + \gamma_2 p_y + \gamma_3 p_z)^2 \,. \tag{6}$$

cannot be satisfied if the γ_i are numbers, but it can be achieved with matrices, for instance with 2×2 matrices

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \quad , \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7}$$

if the choice $\gamma_0 = I_2$, $\gamma_1 = \sigma_x$, $\gamma_2 = \sigma_y$ and $\gamma_3 = \sigma_z$ is made. In vector notation (with a vector $\underline{\sigma}$ built from the Pauli matrices σ_x, σ_y, σ_z), equation (6) can be written as

$$I_2 p^2 = (\underline{\sigma} \cdot \underline{p})^2 \,, \tag{8}$$

which is valid as shown in the appendix. By use of (8) one obtains that

$$\begin{pmatrix} EI_2 & \underline{\sigma} \cdot \underline{p} \\ \underline{\sigma} \cdot \underline{p} & 2m_e I_2 \end{pmatrix} \begin{pmatrix} \Phi \\ X \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{9}$$

after elimination of X leads to

$$(\frac{p^2}{2m_e} - E) I_2 \Phi = 0 \,. \tag{10}$$

Thus the four component equation (9), which is linear in p_x, p_y and p_z and which was derived by Levy-Leblond [1] by requiring Galilean invariance instead of Lorentz invariance, is equivalent to the two component Schrödinger equation (10), which is quadratic in p_x, p_y and p_z. If now electric and magnetic field are added to (9) by the usual prescription of gauge invariant minimal coupling one obtains

$$\begin{pmatrix} (E-v)I_2 & \underline{\sigma} \cdot (\underline{p} - e\underline{A}) \\ \underline{\sigma} \cdot (\underline{p} - e\underline{A}) & 2m_e I_2 \end{pmatrix} \begin{pmatrix} \Phi \\ X \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \,, \tag{11}$$

which after elimination of X leads to

$$\left\{ \frac{1}{2m_e} \left[\underline{\sigma} \cdot (\underline{p} - e\underline{A}) \right]^2 - [E - v(\underline{r})] I_2 \right\} \Phi = 0 \,.$$

Substituting $\underline{p} = -\mathrm{i}\hbar\underline{\nabla}$ and using

$$\begin{aligned}[\underline{\sigma}\cdot(\mathrm{i}\hbar\underline{\nabla}+e\underline{A})]^2 &= I_2(\mathrm{i}\hbar\underline{\nabla}+e\underline{A})^2+\mathrm{i}\underline{\sigma}\cdot[(\mathrm{i}\hbar\underline{\nabla}+e\underline{A})\times(\mathrm{i}\hbar\underline{\nabla}+e\underline{A})]\\ &= I_2(\mathrm{i}\hbar\underline{\nabla}+e\underline{A})^2-e\hbar\underline{\sigma}\cdot[\underline{\nabla}\times\underline{A}+\underline{A}\times\underline{\nabla}]\\ &= I_2\left(-\hbar^2\nabla^2+\mathrm{i}e\hbar\,[\underline{\nabla}\cdot\underline{A}+\underline{A}\cdot\underline{\nabla}]+e^2A^2\right)-e\hbar\underline{\sigma}\cdot\underline{B}\end{aligned}$$

which are valid because of (28 - 30) derived in the appendix, leads to

$$\left\{I_2\left[-\frac{\hbar^2}{2m_e}\nabla^2+\frac{\mathrm{i}e\hbar}{2m_e}(\underline{\nabla}\cdot\underline{A}(\underline{r})+\underline{A}(\underline{r})\cdot\underline{\nabla})+\frac{e^2}{2m_e}A^2(\underline{r})+v(\underline{r})-E\right]+g\mu_B\underline{S}\cdot\underline{B}\right\}\Phi=0,$$

where $\mu_B = e\hbar/2m_e = 9.274\times10^{-24}$ JT^{-1} is the Bohr magneton, $\underline{S}=\underline{\sigma}/2$ the spin operator and $g=-2$ the non-relativistic approximation to the electron g-factor $g=-2.002319$.

3 Spin Density Functional Theory

It is well known that certain materials, for instance the elements Fe, Co and Ni, show spontaneous magnetism even in a vanishing external magnetic field. For $\underline{A}=0$ the system is described by the Hamilton operator

$$\hat{H}=\sum_i^N\left[-\frac{\hbar^2}{2m_e}\nabla_i^2+v_{ext}(\underline{r}_i)\right]+\frac{1}{4\pi\epsilon_0}\frac{e^2}{2}\sum_{i\neq j}\sum\frac{1}{|\underline{r}_i-\underline{r}_j|} \tag{12}$$

where $v_{ext}(\underline{r})$ is the external potential provided by the atomic nuclei. Within density functional theory this many-electron system can be treated by effective single-particle equations. By the Hohenberg-Kohn theorem there exists a one-to-one mapping between the external potential and the ground-state density which means that the ground-state density uniquely determines the external potential and the wavefunctions for this potentials and by the wavefunctions all stationary quantum-mechanical observables. Thus for magnetic systems it should be sufficient to calculate the ground-state density $n(\underline{r})$ and then the magnetization $\underline{m}(\underline{r})$ as a functional of $n(\underline{r})$. Unfortunately, the functional dependence $\underline{m}(\underline{r})=\underline{m}[n(\underline{r})](\underline{r})$ is not known.
To circumvent this problem von Barth and Hedin [2] and Rajagopal and Callaway [3] proposed to extend density functional theory into a spin density functional theory by using a Pauli spin term added to (12) and two component wavefunctions. In principle, terms, which are linear and quadratic in $\underline{A}$ can be added also and treated within spin-current density functional theory, but for the present purpose only the spin term will be considered.

3.1 Basic Formalism

In spin density functional theory the basic variables are the density $n(\underline{r})$ and the magnetization $\underline{m}(\underline{r})$. Alternatively, it is possible to use the 2×2 spin density matrix $n^{\alpha\beta}(\underline{r})$, where the spin indices α and β can have two values, either $+$ for spin up (majority spin) or $-$ for spin down (minority spin). The notation $\uparrow$ instead of $+$ and $\downarrow$ instead of $-$ is also often used in the literature. The connection between $n(\underline{r})$, $\underline{m}(\underline{r})$ and $n^{\alpha\beta}(\underline{r})$ is given by

$$n(\underline{r})=\sum_\alpha n^{\alpha\alpha}(\underline{r})\qquad \underline{m}(\underline{r})=\mu_B\sum_{\alpha\beta}\underline{\sigma}^{\alpha\beta}n^{\alpha\beta}(\underline{r})$$

and by

$$n^{\alpha\beta}(\underline{r}) = I_2^{\alpha\beta} n(\underline{r})/2 + \left[\sigma_x^{\alpha\beta} m_x(\underline{r}) + \sigma_y^{\alpha\beta} m_y(\underline{r}) + \sigma_z^{\alpha\beta} m_z(\underline{r})\right]/2\mu_B \,,$$

where the spin indices are explicitly displayed. The spin density functional

$$\begin{aligned} E[n^{\alpha\beta}(\underline{r})] = T_s[n^{\alpha\beta}(\underline{r})] \quad &+ \quad \frac{1}{4\pi\epsilon_0}\frac{e^2}{2}\int\int \frac{n(\underline{r})n(\underline{r}')}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}\mathrm{d}\underline{r}' \\ &+ \quad \sum_{\alpha\beta}\int v_{ext}^{\alpha\beta}(\underline{r}) n^{\alpha\beta}(\underline{r})\mathrm{d}\underline{r} + E_{xc}[n^{\alpha\beta}(\underline{r})] \end{aligned} \tag{13}$$

consists of a sum of the kinetic energy T_s of non-interacting electrons, the electron-electron interaction in the Hartree approximation, the interaction energy with the external potential $v_{ext}^{\alpha\beta}$ and the exchange-correlation energy. The external potential $v_{ext}^{\alpha\beta}$ contains the electrostatic Coulomb potential of the nuclei and the potentials arising from possible external electric and magnetic fields. Because the Hohenberg-Kohn theorem is valid for arbitrary values of the coupling constant e^2, it is also valid for a non-interacting electron system characterized by $e^2 = 0$ so that the spin density matrix $n^{\alpha\beta}(\underline{r})$ of the interacting system uniquely determines a potential $v_s^{\alpha\beta}(\underline{r})$ for the non-interacting system. The energy functional for the non-interacting system is given by

$$E_{e^2=0}[n^{\alpha\beta}(\underline{r})] = T_s[n^{\alpha\beta}(\underline{r})] + \sum_{\alpha\beta}\int v_s^{\alpha\beta}(\underline{r}) n^{\alpha\beta}(\underline{r})\mathrm{d}\underline{r} \tag{14}$$

Subtraction of (13) and (14) leads to an integral equation for $v_s^{\alpha\beta}(\underline{r})$, which by functional derivation with respect to the ground-state spin density matrix $n^{\alpha\beta}(\underline{r})$ gives the result

$$v_s^{\alpha\beta}(\underline{r}) = I_2^{\alpha\beta}\frac{e^2}{4\pi\epsilon_0}\int \frac{n(\underline{r}')}{|\underline{r}-\underline{r}'|}\mathrm{d}\underline{r}' + v_{ext}^{\alpha\beta}(\underline{r}) + v_{xc}^{\alpha\beta}(\underline{r}) \,. \tag{15}$$

Note that both $E[n^{\alpha\beta}(\underline{r})]$ and $E_{e^2=0}[n^{\alpha\beta}(\underline{r})]$ are minimal in the ground state so that their functional derivatives vanish in the derivation of (15). The exchange-correlation potential $v_{xc}^{\alpha\beta}(\underline{r})$ can be obtained by the functional derivative $\delta E_{xc}[n^{\alpha\beta}(\underline{r})]/\delta n^{\alpha\beta}(\underline{r})$. For the non-interacting system the spin density matrix and the kinetic energy can be calculated by single-particle equations

$$n^{\alpha\beta}(\underline{r}) = \sum_i \varphi_i^{\star\alpha}(\underline{r})\varphi_i^{\beta}(\underline{r}) \quad \text{and} \quad T_s[n^{\alpha\beta}(\underline{r})] = \sum_{\alpha i}\int \varphi_i^{\star\alpha}(\underline{r})\left(-\frac{\hbar^2}{2m_e}\nabla^2\varphi_i^{\alpha}(\underline{r})\right)\mathrm{d}\underline{r} \,,$$

where the sum over i includes all occupied orbitals. These orbitals can be determined from the single-particle Kohn-Sham equation

$$-\frac{\hbar^2}{2m_e}\nabla^2\varphi_i^{\alpha}(\underline{r}) + \sum_{\beta} v_s^{\alpha\beta}(\underline{r})\varphi_i^{\beta}(\underline{r}) = \epsilon_i\varphi_i^{\alpha}(\underline{r}) \,, \tag{16}$$

which arises if the energy functional (14) is minimized with respect to the orbitals. Useful approximations for the exchange-correlation energy functional can be given in terms of the eigenvalues $n^{+}(\underline{r})$ and $n^{-}(\underline{r})$ of the matrix $n^{\alpha\beta}(\underline{r})$ which can be diagonalized as

$$\sum_{\alpha'\beta'} U^{\alpha\alpha'}(\underline{r}) n^{\alpha'\beta'}(\underline{r}) U^{\beta\beta'}(\underline{r}) = \delta^{\alpha\beta} n^{\alpha}(\underline{r}) \,, \tag{17}$$

where $U^{\alpha\beta}(\underline{r})$ are spin-1/2 rotation matrices and $n^{\alpha}(\underline{r})$ the eigenvalues. Note that these matrices and eigenvalues generally depend on the position $\underline{r}$. In many applications, for instance in ferromagnetic and antiferromagnetic solids, a common magnetization axis exists for all atoms. The z axis can then be chosen globally along the direction of the magnetic field and the spin-1/2 rotation matrices in (17) are independent of $\underline{r}$. This has the simplifying consequence that the energy and all other physical observables are functionals of the density and of the magnitude of the magnetization $m(\underline{r}) = |\underline{m}(\underline{r})| = \mu_B[n^+(\underline{r}) - n^-(\underline{r})]$ rather than of the vector $\underline{m}(\underline{r})$. In terms of the spin up and spin down orbitals $\varphi_i^+(\underline{r})$ and $\varphi_i^-(\underline{r})$, the spin densities $n^+(\underline{r})$ and $n^-(\underline{r})$ can be represented as

$$n^{\pm}(\underline{r}) = \sum_i |\varphi_i^{\pm}(\underline{r})|^2 \tag{18}$$

and the Kohn-Sham equation can be written as

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_s^{\pm}(\underline{r})\right]\varphi_i^{\pm}(\underline{r}) = \epsilon_i^{\pm}\varphi_i^{\pm}(\underline{r})$$

with the effective potential

$$v_s^{\pm}(\underline{r}) = \frac{e^2}{4\pi\epsilon_0}\int \frac{n(\underline{r}')}{|\underline{r} - \underline{r}'|}d\underline{r}' + v_{ext}^{\pm}(\underline{r}) + v_{xc}^{\pm}(\underline{r})\ .$$

In an external magnetic field B, the external potential $v_{ext}^{\pm}$ contains a field term $-(\pm\mu_B B)$, where the negative sign means that the majority electrons (with spin $+$) are energetically favored compared to the minority electrons (with spin $-$). The exchange-correlation potential $v_{xc}^{\pm}(\underline{r})$ can have different values for the two spin directions even without an external magnetic field. This is, for instance, realized in the ferromagnetic metals Fe, Co and Ni.
Spin density functional theory as presented above is exact in principle, however the functionals E_{xc} and $v_{xc}^{\pm}$, in which all complications of the many-electron system are hidden, are not known and must be approximated. Useful approximation like the local-spin-density approximation (LSDA) [2, 4, 5] and the generalized gradient approximation (GGA) [6, 7] have been developed and have been shown to be rather accurate for many applications.

3.2 Stoner Model for Ferromagnetism

Normally the magnetization $m(\underline{r})$ is a small compared to the density $n(\underline{r})$. Thus the exchange-correlation potential can be expanded in terms of $m(\underline{r})$ and approximated as

$$v_{xc}^{\pm}(\underline{r}) = v_{xc}^{o}(\underline{r}) \mp m(\underline{r})\tilde{v}[n(\underline{r})](\underline{r})\ ,$$

where higher order terms of $m(\underline{r})$ are neglected and v_{xc}^{o} is the non-magnetic exchange-correlation potential. The average value of $\tilde{v}$ is positive such that majority electrons (with spin $+$) feel a more attractive potential and minority electrons (with spin $-$) a less attractive one. In the Stoner model the difference of the potentials is simulated by a constant

$$v_{xc}^{\pm}(\underline{r}) = v_{xc}^{o}(\underline{r}) \mp \frac{1}{2}IM \quad \text{with} \quad M = \frac{1}{\mu_B}\int_{V_{Atom}} m(\underline{r})\mathrm{d}\underline{r} = \int_{V_{Atom}} [n^+(\underline{r}) - n^-(\underline{r})]\mathrm{d}\underline{r}\ .$$

Here I is the exchange integral (Stoner parameter) and M the difference of the number of majority and minority electrons in the atomic unit cell so that $M_{loc} = \mu_B M$ gives the local

atomic moment. In the ferromagnetic elements Fe, Co and Ni all atoms are equivalent (with the same value of I and M) so that $v_{xc}^{\pm}(\underline{r})$ differ by IM, which is constant in space. For a constant shift of the potential the wavefunctions and eigenvalues, which in periodic crystals are characterized by wavevector $\underline{k}$ and band index ν, can be calculated as

$$\varphi_{\underline{k}\nu}^{\pm}(\underline{r}) = \varphi_{\underline{k}\nu}^{o}(\underline{r}) \qquad \text{and} \qquad \epsilon_{\underline{k}\nu}^{\pm} = \epsilon_{\underline{k}\nu}^{o} \mp \frac{1}{2} IM \; .$$

This means the wavefunctions are identical to the non-magnetic ones and the eigenvalues are simply shifted by a constant amount. Consequently, the densities of states $n^{\pm}(E)$ are also shifted compared to the non-magnetic density of states $n^{o}(E)$ as

$$n^{\pm}(E) = \sum_{\nu} \int_{BZ} \delta(E - \epsilon_{\underline{k}\nu}^{\pm}) d\underline{k} = n^{o}(E \pm \frac{1}{2} IM) \; . \tag{19}$$

Here the integral is over the Brillouin zone (BZ). The shift $\pm\frac{1}{2}IM$ for the spin up and down densities of states describes reasonably well the situation which is found in spin density functional calculations for the ferromagnetic elements as can be seen in Fig. 1.

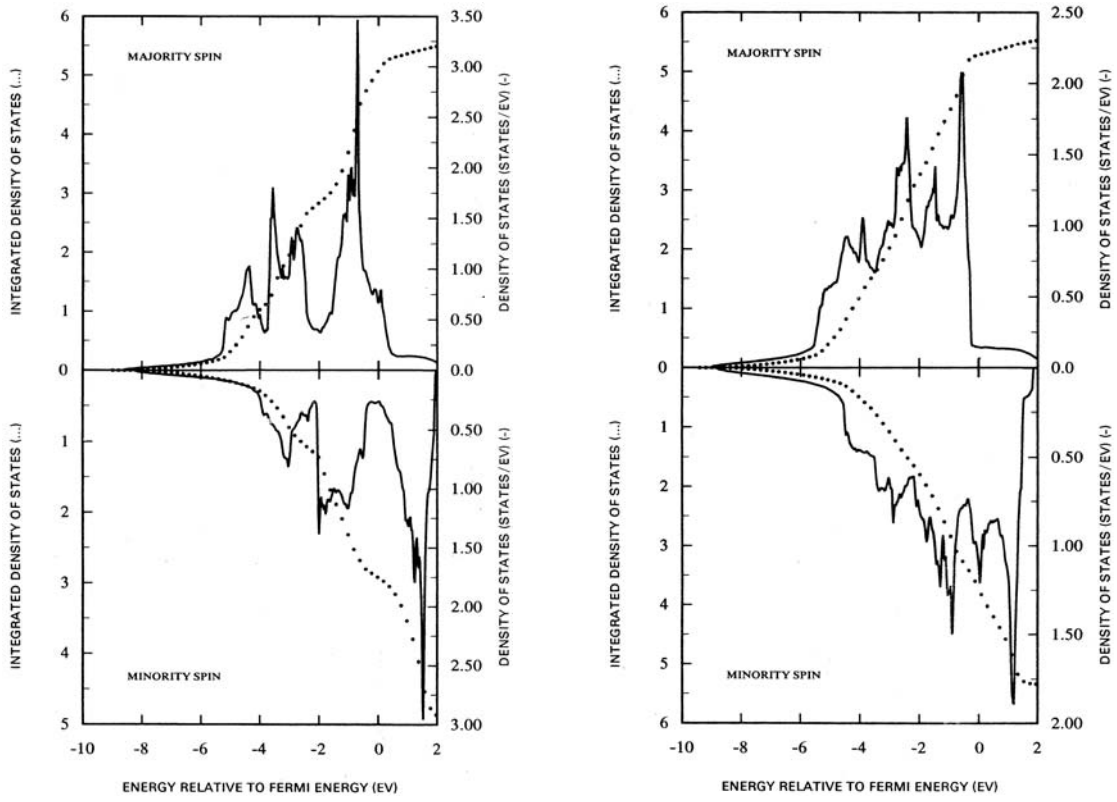


Fig. 1: *Density of states for Fe (left picture) and Co (right picture) from spin-density-functional calculations [8]. The density of states for majority-spin electrons is plotted upwards and for minority-spin electrons downwards. States for negative energies (lower than E_F) are occupied states for positive energies (higher than E_F) are unoccupied. The dotted curves represent energy integrated density of states.*

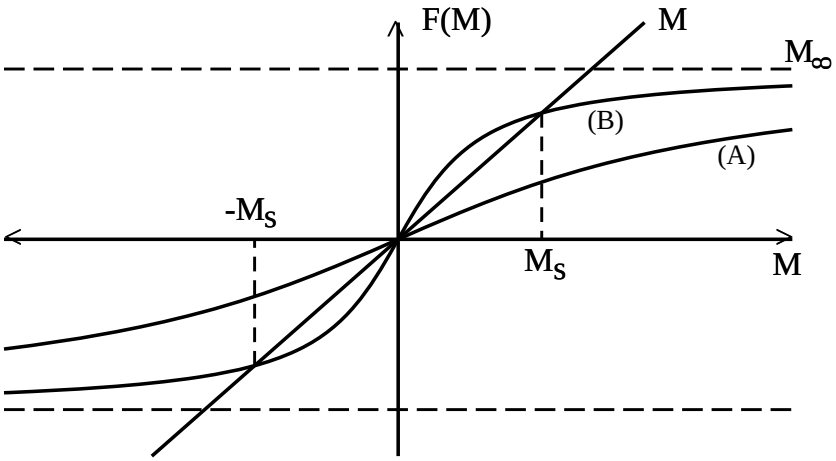


Fig. 2: *Graphical solution for the Stoner model. The intersection of $F(M)$ with the straight line M determines the solution M_s. The intersection at $M = 0$ is always a trivial solution.*

From (19) one obtains by integration over all occupied states

$$M = \int^{E_F} \left[n^+(E) - n^-(E)\right] \mathrm{d}E = \int^{E_F} \left[n^o(E + \frac{1}{2}IM) - n^o(E - \frac{1}{2}IM)\right] \mathrm{d}E \qquad (20)$$

$$N = \int^{E_F} \left[n^+(E) + n^-(E)\right] \mathrm{d}E = \int^{E_F} \left[n^o(E + \frac{1}{2}IM) + n^o(E - \frac{1}{2}IM)\right] \mathrm{d}E\,. \qquad (21)$$

Since $n^o(E)$ is determined by the non-magnetic calculation and N by the condition of charge neutrality, (21) implicitly defines $E_F = E_F(M)$ as function of M and (20) defines a function

$$F(M) = \int^{E_F(M)} \left[n^o(E + \frac{1}{2}IM) - n^o(E - \frac{1}{2}IM)\right] \mathrm{d}E$$

for which the solution $M = F(M)$ must be found self-consistently. The function $F(M)$ satisfies $F(M) = -F(-M)$, $F(0) = 0$, $F(\pm\infty) = \pm M_\infty$ and $F'(M) > 0$, where the last condition arises from $n^o(E) > 0$. Here M_∞ is the spin moment for full spin polarization, when all majority states are occupied and all minority states are empty. This situation corresponds to the atomic limit with maximal spin moment according to Hund's first rule. Two possibilities for $F(M)$ are shown in Fig. 2. For the function denoted by (A) only the trivial non-magnetic solution $M = 0$ exists. For the function denoted by (B) three solutions exist, $M = \pm M_S$ with a finite spontaneous magnetization and $M = 0$. This non-magnetic solution is however unstable. From Fig. 2 it is obvious that a solution with non-zero moment always exists, if the slope $F(M)/M$ at $M = 0$ is larger than one. Thus $F'(0) > 1$ is a sufficient condition for ferromagnetic solutions. From (20) one obtains by differentiation

$$\begin{aligned} F'(M) &= \frac{I}{2}\left[n^o(E_F + \frac{1}{2}IM) + n^o(E_F - \frac{1}{2}IM)\right] \\ &+ \left[n^o(E_F + \frac{1}{2}IM) - n^o(E_F - \frac{1}{2}IM)\right]\frac{dE_F}{dM} \end{aligned} \qquad (22)$$

which for $M = 0$ leads to

$$F'(0) = In^o(E_F).$$

This implies that the Stoner criterion

$$In^o(E_F) > 1$$

is a sufficient condition for ferromagnetic solutions, which are thus favored for large exchange integrals I and large density of states $n^o(E_F)$ at the Fermi energy E_F.
The Stoner model can be extended to include effects of an external magnetic field $\underline{B} = (0, 0, B)$. This allows to determine the spin susceptibility χ , which according to $\mu_B M/V = \chi B$ describes the relation between the magnetization and the magnetic field for small fields. Instead of (19) the relevant potential is given by

$$v^{\pm}(\underline{r}) = v^o(\underline{r}) \mp \frac{1}{2} IM \mp \mu_B B$$

and instead of $M = F(M)$ the equation $M = F(M+2\mu_B B/I)$ must be solved self-consistently. Linearizing around M_0, which would be the magnetic moment without field, leads to

$$\begin{aligned} \Delta M = M - M_0 &= F(M + \frac{2\mu_B B}{I}) - F(M_0) \\ &\approx F'(M_0) \left(\Delta M + \frac{2\mu_B B}{I} \right) \end{aligned}$$

Solving this equation for ΔM, inserting $M_0 = 0$ and $F'(0) = In^o(E_F)$ leads to

$$M = \frac{n^o(E_F)}{1 - In^o(E_F)} 2\mu_B B \qquad \text{and} \qquad \chi = \frac{1}{1 - In^o(E_F)} \chi_P \,,$$

where $\chi_P = 2\mu_B^2 n^o(E_F)/V$ is the Pauli spin susceptibility for non-interacting electrons, which is obtained if the exchange interaction is neglected. The exchange interaction leads to a Stoner enhancement factor $S = [1 - In^o(E_F)]^{-1}$, which diverges for $In^o(E_F) = 1$. For $In^o(E_F) < 1$ the non-magnetic state is stable, whereas for $In^o(E_F) > 1$ the ferromagnetic state is stable.
The density of states shows usually a rather detailed structure. However, in simple approximation it scales inversely to the band width W. For a constant density of states n_l^o for states with quantum number l this relation is exact ($n_l^o = (2l+1)W^{-1}$), because $\int_W n_l^o dE = W n_l^o$ is given by $2l + 1$, the number of states with quantum number l. For delocalized electrons the band width is large and the density of states is small, whereas for more localized electrons the band width is smaller and the density of states larger. In the atomic limit the band width goes to zero, the Stoner criterion is always satisfied and the magnetic moment is maximal according to Hund's first rule. Fig. 3 shows a schematic representation for the band widths of transition metals, rare earth metals and actinides. The 5f electrons of the early actinides and the 3d electrons of the late transition metals from Cr to Ni have a tendency for band magnetism, whereas the late actinides and the rare earth metals show localized magnetism with almost atomic moments in good agreement with Hund's rules.
Quantitative results of non-spin-polarized local density functional calculations [9] for $n^o(E_F)$ and I are given in Table 1 for some selected metals. The results in Table 1 show that Fe and Ni satisfy the Stoner criterion $In^o(E_F) > 1$ and that Co with $In^o(E_F) = 0.97$ almost satisfies this criterion. It is discussed in Ref. [9] that the calculated values for I in Table 1 are lower bounds and that the value $In^o(E_F) = 0.97$ for Co does not contradict that ferromagnetism is observed for this metal. As a matter of fact in spin-polarized calculations [8, 10], which do not

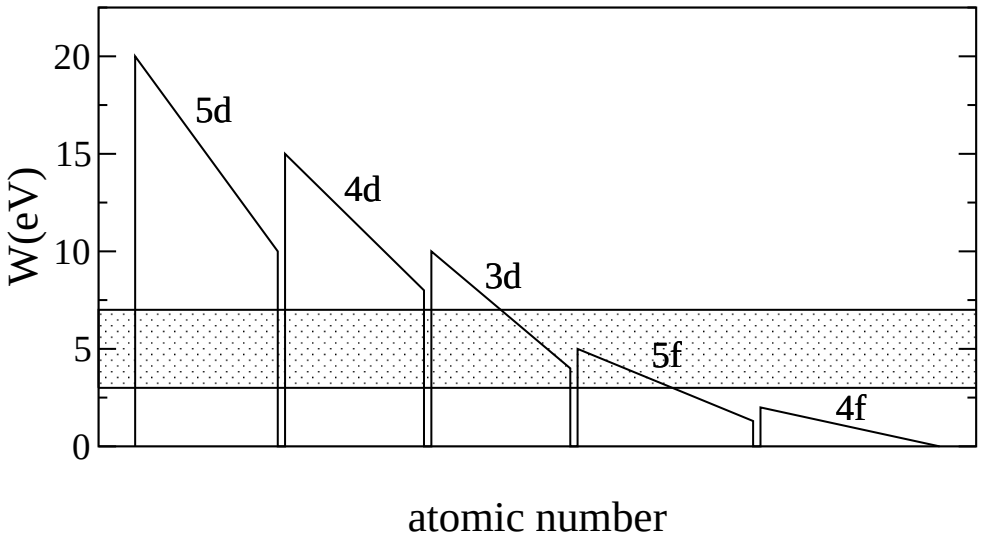


Fig. 3: *Schematic illustration for the band width W of transition metals (3d, 4d and 5d electrons), rare earth metals (4f electrons) and actinides (5f electrons). The shaded rectangle represents the region favorable for band magnetism. Above the rectangle magnetism is suppressed and below the rectangle localized (atomic) magnetism is preferred.*

use the Stoner model and do not rely on an estimate of I, the ferromagnetic state of Co has been found to be more stable than the non-magnetic state. Thus these early spin density functional calculations are consistent with the observed ferromagnetism of Fe, Co and Ni. The metal Pd also has a large Stoner factor and is almost magnetic. The experimental Stoner factor for Pd is even approximately twice as large as the one given in Table 1 in agreement with the fact that the tendency for magnetism is underestimated by the approximation for I used in Ref. [9].

Fig. 1 shows densities of states obtained by spin density functional calculations for Fe and Co, where the bcc structure was used for Fe and the fcc structure for Co. Except for the exchange splitting Fig. 1 shows rather similar densities of states for both spin directions, thus indicating the applicability of the Stoner model for these metals. Ni with fcc structure has a similar density of states as Co, but with a smaller exchange splitting. The majority states for Co and Ni are fully occupied, whereas in the minority states 1.7 electrons are missing for Co and 0.6 electrons for Ni. This leads to moments of 1.7 μ_B and 0.6 μ_B per Co and Ni atom. The calculated moment for Fe is 2.2 μ_B. In contrast to Co and Ni, the majority d states in Fe are not fully occupied. Thus Fe is characterized as a *weak* ferromagnet, whereas Co and Ni are characterized as *strong* ferromagnets. Table 2 shows that the calculated moments agree well with the experimental values.

Remark: While the discussion above was restricted to ferromagnetism, it is equally well applicable for other magnetic states, for instance for antiferromagnetic materials. Here, as Neel pointed out in 1936, the internal magnetic field $\frac{1}{2}IM$ has opposite sign on the two sublattices. In a Stoner like picture the competition between ferromagnetism and antiferromagnetism has been extensively discussed by Heine and Samson [11] who have shown that antiferromagnetism is favored if E_F is in the middle of a band, in particular if the density of states is small at E_F, whereas ferromagnetism is favored if E_F is near to a band edge, in particular if the density of states is large at E_F.

Table 1: *Densities of states $n^o(E_F)$ at the Fermi energy, exchange integrals I, their products and Stoner enhancement factors S obtained in non-spin-polarized density functional calculations [9] for some selected metals.*

metal	$n^o(E_F)[Ry^{-1}]$	$I[Ry]$	$In^o(E_F)$	$S = \chi/\chi_P$
Na	6.2	0.067	0.42	1.71
Al	5.6	0.045	0.25	1.34
Cr	9.5	0.028	0.27	1.36
Mn	21.	0.030	0.63	2.74
Fe	42.	0.034	1.43	-2.34
Co	27.	0.036	0.97	38.2
Ni	55.	0.037	2.04	-0.98
Cu	3.9	0.027	0.11	1.12
Pd	31.	0.025	0.78	4.46

Table 2: *Magnetic moments M_{LSDA} for Fe, Co and Ni obtained in local spin density functional calculations [8] in comparison with the experimental values for the spin-only moments M_{spin} and for the total moments M including the orbital contributions.*

metal	$M_{LSDA}[\mu_B/atom]$	$M_{spin}[\mu_B/atom]$	$M[\mu_B/atom]$
Fe	2.15	2.12	2.22
Co	1.56	1.57	1.71
Ni	0.59	0.55	0.61

3.3 Ferromagnetism of Alloys

Alloys of Fe, Co and Ni and alloys of these elements with other transition metals display a wide variety of magnetic properties since the magnetization of the pure elements can be strongly changed by alloying. The moments of the ferromagnetic atoms can be reduced or enlarged and parallel or antiparallel alignment of moments of the different atoms can occur. A combined representation of the averaged moments of binary magnetic alloys is given by the Slater-Pauling curve which is shown in Fig. 4. The Slater-Pauling curve has two main branches with slopes of 45^o and -45^o which meet in the middle where a maximal moment of about 2.4 μ_B occurs. The left main branch consists of Fe alloys, whereas Co and Ni alloys form the right main branch and the subbranches. The main reason for the two different slopes is a different electronic screening behavior. Alloys on the main branch on the right have a full majority spin band so that the screening of the valence difference introduced by the impurity atoms is provided by minority spin electrons. This leads to a reduced number of minority d electrons which gives increased moments. Alloys on the other branches are characterized by the occurrence of antiparallel moments of the impurities which lead to reduced averaged moments with increasing concentration. Here the screening is mainly provided by the majority spin electrons. The experimental results (upper picture) have been obtained by measurements of the saturation magnetization and contain an orbital contribution of about 5–10 % of the spin contribution. Therefore, the theoretical

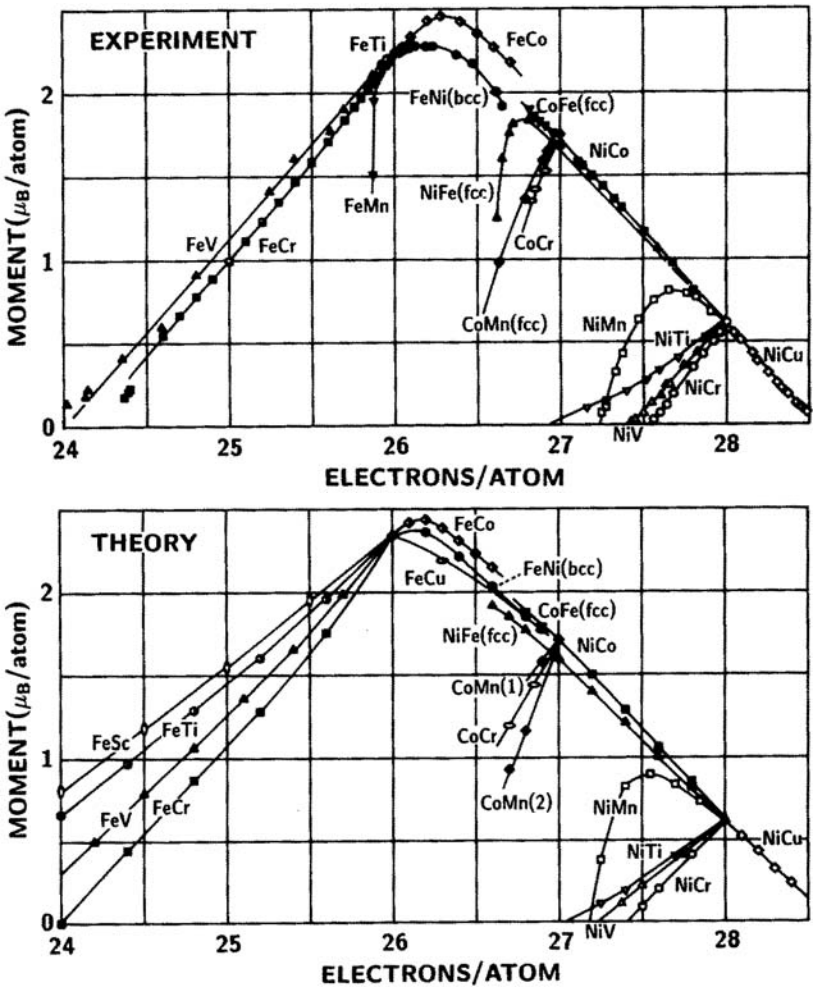


Fig. 4: *Slater-Pauling curve from Ref. [12] for the averaged magnetic moment per atom as function of the averaged number of electrons per atom. Important alloys for applications are Permalloy ($Fe_{20}Ni_{80}$) as a typical soft magnet and Invar ($Fe_{65}Ni_{35}$) because of its small thermal expansion.*

results (lower picture) for the spin moments have been scaled to take into account this effect. For CoMn two solutions have been obtained, CoMn(1) and CoMn(2) with Mn moments parallel and antiparallel to the bulk magnetization.

The theoretical determination of the averaged moments by spin-density functional calculations is not easy for these concentrated and disordered alloys. Statistical averaging over many configurations for many different concentrations requires a substantial amount of computing time [13] which was not available when the results shown in Fig. 4 were calculated in 1991. For these calculations the coherent potential approximation (CPA) was used which is based on an effective medium concept, where the effective medium is obtained by the self-consistent requirement that the averaged scattering of the electrons at A and B atoms in the effective medium vanishes. Although the agreement between experiment and calculation (upper and lower picture in Fig. 4) is not perfect, this figure shows that spin density functional theory is a powerful tool to understand and explain magnetic properties of materials.

3.4 Mapping to Model Hamiltonians

Although spin density functional theory can be used to determine the magnetic ground state, the calculations, particularly for large and complex systems, often require too much computing time. Thus a mapping to simpler model Hamiltonians is of great value, for instance to the Heisenberg Hamiltonian

$$\hat{H} = -\sum_{i,j} J_{ij} \hat{\underline{S}}_i \hat{\underline{S}}_j \,,$$

where J_{ij} is the exchange coupling constant between the spins at sites i and j. Such a mapping can be obtained by constrained density functional theory [14]. Constraints are already used in the formal development of density functional theory, for instance the density is constrained to give the correct number of electrons and the Kohn-Sham orbitals must be normalized as $\sum_\alpha (\varphi_i^\alpha, \varphi_i^\alpha) = 1$, which is guaranteed by the Lagrange parameters ϵ_i in the Kohn-Sham equation (16). The basic idea of constrained density-functional theory is the extension to quite arbitrary constraints. The energy of the lowest state compatible with a constraint can then be found by a modified energy functional. One example considered in Ref. [14] is a modification of the energy functional $E[n(\underline{r})]$ into

$$\tilde{E}[n(\underline{r})] = E[n(\underline{r})] + v \left[N_V - \int_V n(\underline{r}) \mathrm{d}\underline{r} \right] \,, \tag{23}$$

where the constraint, guaranteed by the Lagrange parameter v, describes that the local volume V contains exactly N_V electrons. The minimization of (23) with respect to $n(\underline{r})$ leads to an additional potential v in the Kohn-Sham equations, which is constant in volume V and zero elsewhere. This potential must be adjusted such that the resulting density $n(\underline{r})$ gives exactly N_V electrons in volume V. Instead of calculating the energy from the functional $\tilde{E}[n(\underline{r})]$, it is computationally easier to calculate the energy difference with respect to a reference state, for instance the ground state with N_0 electrons in volume V. This can be done by the Hellmann-Feynman theorem

$$\frac{\mathrm{d}\tilde{E}(N_V)}{\mathrm{d}N_V} = v \qquad \Rightarrow \qquad \Delta\tilde{E}(N_V) = \int_{N_0}^{N_V} v(N') \mathrm{d}N' \,, \tag{24}$$

which only requires the knowledge of the potential $v(N')$. Physically, the potential v can be viewed as the "force" necessary to constrain the system to the desired state and ΔE as the "strain energy" of the system.

An early application of constrained density functional theory is the calculation of interaction energy differences between the ferromagnetic and antiferromagnetic configuration of impurity pairs in metals [15]. In these calculations the local magnetic moment of one of the impurities is constrained to an arbitrary value $M_{loc} = \mu_B M$ and the lowest energy compatible with the constraint is determined by a modified functional

$$\tilde{E}[n(\underline{r}), m(\underline{r})] = E[n(\underline{r}), m(\underline{r})] + B \left[\mu_B M - \int_V m(\underline{r}) \mathrm{d}\underline{r} \right] \,,$$

where the Lagrange parameter B is a constraining magnetic field, which is constant in the cell of one impurity with volume V and zero elsewhere. This field is chosen such that the integral of the magnetization $m(\underline{r})$ over the cell gives the desired value of the moment. Similar to (24)

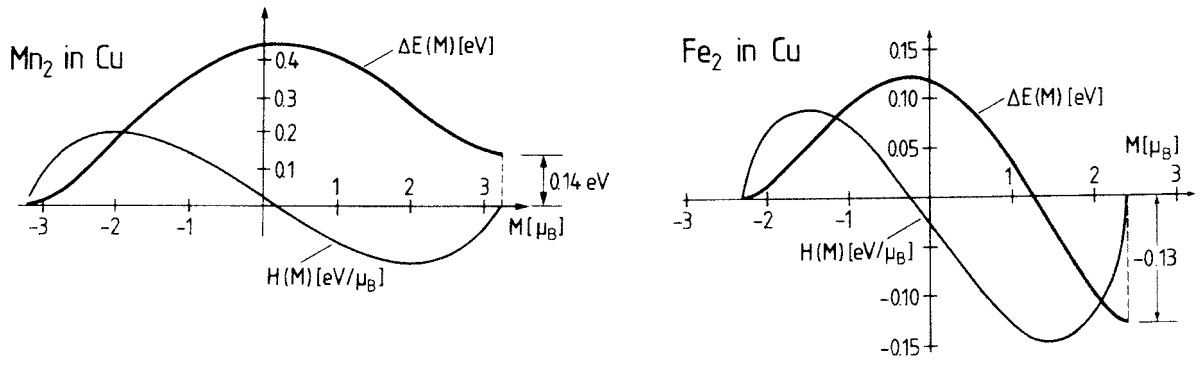


Fig. 5: *Magnetic interaction energy difference and constraining magnetic field for pairs of Mn and Fe impurities on nearest neighbor sites in a Cu crystal (from Ref. [15]). Note that the energies involved in these calculations are several orders of magnitude smaller than the individual energies of the antiferromagnetic and ferromagnetic configurations.*

the energy difference is given by

$$\Delta E(M) = \int_{M_0}^{M} H(M')\mathrm{d}M' \,,$$

where M_0 is the value of M in the reference state. For instance in Fig. 5, the reference state is the antiferromagnetic configuration, for which the moments for the two impurities have opposite sign. This reference state with calculated values $M = -3.22$ and $M = -2.31$ (for Mn and Fe pairs) corresponds to the left minima of the $\Delta E(M)$ curves in Fig. 5. The right minima of the $\Delta E(M)$ curves correspond to the ferromagnetic configuration with calculated values $M = 3.20$ and $M = 2.40$ (for Mn and Fe pairs). Both the antiferromagnetic and the ferromagnetic configurations are stable as the minima of the energies with vanishing constraining field B indicate. The energy differences between the ferromagnetic and antiferromagnetic configuration are 0.14 eV for Mn and -0.13 eV for Fe pairs so that the antiferromagnetic state is more stable for the Mn pair and the ferromagnetic state is more stable for the Fe pair. Note that, whereas energy and field curves depend on the choice of the constraint, for instance the volume V, the resulting energy differences between both configurations are independent of the choice. Instead of constraining the absolute magnitude of the moment also its direction can be constrained. This requires a transversal field perpendicular to the direction of the moment. The calculated energy changes can then be mapped to effective Hamiltonians. For instance, for small deviations from the ground state, the energy difference is quadratic in the changes of the moments

$$\Delta E(\underline{M}_i, \underline{M}_j) \approx -\frac{1}{2}\sum_{i,j} J_{ij}\Delta\underline{M}_i\Delta\underline{M}_j \,,$$

which provides a method to calculate the exchange-coupling constants J_{ij} for the Heisenberg model within density functional theory.

4 Relativistic Effects

In applications of magnetic materials, for instance permanent magnets or magnetic storage media, one usually exploits the fact that the material is more easily magnetized in a certain direction

than in other directions. The energy needed to rotate the magnetization from a direction of low energy (easy axis) to a direction of high energy (hard axis) is of the order of μeV to meV per atom. The small energy difference can be attributed to relavistic effects, the magnetic dipole dipole interaction and the coupling of spin and orbital motion of the electrons. The classical expression for the dipole dipole interaction

$$\frac{\mu_B^2\mu_0}{2}\sum_{i\neq j}\sum\frac{1}{|\underline{r}_i-\underline{r}_j|^3}\left\{\underline{m}_i\cdot\underline{m}_j-3\frac{\left[(\underline{r}_i-\underline{r}_j)\cdot\underline{m}_i\right]\left[(\underline{r}_i-\underline{r}_j)\cdot\underline{m}_j\right]}{|\underline{r}_i-\underline{r}_j|^2}\right\}$$

can be obtained within density functional theory as shown by Jansen [17] from the Hartree energy part of the relativistic Breit equation. An important property of this interaction is its slow decrease with distance like $|\underline{r}-\underline{r}'|^{-3}$. Thus the dipole field experienced at an atom depends significantly on the moments located at the boundary of the sample. This results in the shape anisotropy which is important for small samples of non-spherical shape and for low dimensional systems like magnetic films and wires. In bulk cubic crystals the dipole dipole contribution to the magnetic anisotropy, which is of second order in $\underline{m}(\underline{r})$, is unimportant because cubic symmetry requires that second order terms vanish. Here the main contribution to the magnetic anisotropy comes from spin-orbit coupling, which is a relativistic effect contained in the Dirac equation as can seen, if the Dirac equation is expanded with respect to the inverse of the velocity of light $1/c$.

Phenomenologically the free energy density $f = F/V$ can be expanded into spherical harmonics

$$f(\vartheta,\varphi)/V=\sum_{l\,\mathrm{even}}\sum_{m=-l}^{m=l}\kappa_{lm}(\underline{B})Y_{lm}(\vartheta,\varphi)\,. \tag{25}$$

Here odd terms in l vanish because of time reversal symmetry. Often instead of (25) an expansion in the directional cosines $(\alpha_1,\alpha_2,\alpha_3)=(\sin\vartheta\cos\varphi,\sin\vartheta\sin\varphi,\cos\vartheta)$ is used in the form

$$f(\vartheta,\varphi)=b_0(\underline{B})+\sum_{i,j}b_{ij}(\underline{B})\alpha_i\alpha_j+\sum_{i,j,k,l}b_{ijkl}(\underline{B})\alpha_i\alpha_j\alpha_k\alpha_l+\cdots\,,$$

where again only even terms compatible with time reversal symmetry appear. Due to crystalline symmetry certain relations between coefficients of equal order exist so that for cubic crystals the anisotropy can be written as

$$f(\vartheta,\varphi)=K_0+K_1(\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2)+K_2\alpha_1^2\alpha_2^2\alpha_3^2+\cdots \tag{26}$$

with the coordinate axes taken along the cubic axes and for hexagonal crystals as

$$f(\vartheta,\varphi)=K_0+K_1\sin^2\vartheta+K_2\sin^4\vartheta+\cdots\,, \tag{27}$$

where ϑ is the angle between the magnetization direction and the c axis. (Note that in (26) and (27) the index i of K_i does not describe the order.) Values for the magnetic anisotropy constants for the ferromagnetic elements Fe, Co and Ni are given in Table 3. In general, the anisotropy constants K_1 and particularly K_2 are difficult to measure, since they are strongly temperature dependent and susceptible to crystal imperfections.

Table 3: *Magnetic anisotropy constants for Fe, Co and Ni at 300 K (from Ref. [16]).*

metal lattice	Fe (bcc)	Co (hcp)	Ni (fcc)
K_1 [Jm^{-3}]	$4.8 \cdot 10^4$	$5.3 \cdot 10^5$	$4.5 \cdot 10^3$

Appendix

Useful relations for products of $\underline{\nabla}$ and $\underline{A}$, which do not commute, are

$$\underline{\nabla} \cdot (\underline{A}\Psi) + \underline{A} \cdot \underline{\nabla}\Psi = (\nabla \cdot \underline{A})\Psi + 2\underline{A} \cdot \underline{\nabla}\Psi = 2\underline{A} \cdot \underline{\nabla}\Psi \,, \tag{28}$$

$$\underline{\nabla} \times (\underline{A}\Psi) = (\nabla \times \underline{A})\Psi - \underline{A} \times \underline{\nabla}\Psi = \underline{B}\Psi - \underline{A} \times \underline{\nabla}\Psi \,, \tag{29}$$

where in (28) the Coulomb gauge condition $\underline{\nabla} \cdot \underline{A} = 0$ was used. To derive (8) matrix multiplication can be used

$$\begin{aligned}
(\underline{\sigma} \cdot \underline{u})(\underline{\sigma} \cdot \underline{v}) &= \begin{pmatrix} u_z & u_x - \mathrm{i}u_y \\ u_x + \mathrm{i}u_y & -u_z \end{pmatrix} \begin{pmatrix} v_z & v_x - \mathrm{i}v_y \\ v_x + \mathrm{i}v_y & -v_z \end{pmatrix} \\
&= \begin{pmatrix} u_x v_x + u_y v_y + u_z v_z + \mathrm{i}(u_x v_y - u_y v_x) & -u_x v_z + u_z v_x + \mathrm{i}(u_z v_y - u_y v_z) \\ -u_x v_z + u_z v_x + \mathrm{i}(u_z v_y - u_y v_z) & u_x v_x + u_y v_y + u_z v_z - \mathrm{i}(u_x v_y - u_y v_x) \end{pmatrix} \\
&= I_2(\underline{u} \cdot \underline{v}) + \mathrm{i}\underline{\sigma} \cdot (\underline{u} \times \underline{v}) \,,
\end{aligned}$$

which leads to (8) for the choice $\underline{u} = \underline{v} = \underline{p}$. For a constant magnetic field $\underline{B}$ the vector potential can be written as

$$\underline{A} = -\frac{1}{2}\underline{r} \times \underline{B} = -\frac{1}{2} \begin{pmatrix} yB_z - zB_y \\ zB_x - xB_z \\ xB_y - yB_x \end{pmatrix} \tag{30}$$

which is consistent with $\underline{B} = \underline{\nabla} \times \underline{A}$ because of

$$\underline{\nabla} \times \underline{A} = \begin{pmatrix} \partial_y A_z - \partial_z A_y \\ \partial_z A_x - \partial_x A_z \\ \partial_x A_y - \partial_y A_x \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} -B_x - B_x \\ -B_y - B_y \\ -B_z - B_z \end{pmatrix} = \underline{B} \,.$$

By use of (28) and (30) one obtains

$$\begin{aligned}
\underline{\nabla} \cdot \underline{A} + \underline{A} \cdot \underline{\nabla} &= (B_y z - B_z y)\nabla_x + (B_z x - B_x z)\nabla_y + (B_x y - B_y x)\nabla_z \\
&= (y\nabla_z - z\nabla_y)B_x + (z\nabla_x - x\nabla_z)B_y + (x\nabla_y - y\nabla_x)B_z \\
&= [\underline{r} \times \underline{\nabla}] \cdot \underline{B} = -\frac{1}{\mathrm{i}\hbar}\underline{L} \cdot \underline{B}
\end{aligned} \tag{31}$$

and

$$\begin{aligned}
4A^2 &= (xB_y - yB_x)^2 + (zB_x - xB_z)^2 + (yB_z - zB_y)^2 \\
&= y^2B_x^2 + x^2B_y^2 + x^2B_z^2 + z^2B_x^2 + y^2B_z^2 + z^2B_y^2 - 2xyB_xB_y - 2xzB_xB_z - 2yzB_yB_z \\
&= (x^2 + y^2 + z^2)(B_x^2 + B_y^2 + B_z^2) - (xB_x + yB_y + zB_z)^2 \\
&= r^2B^2 - (\underline{r} \cdot \underline{B})^2 = r^2B^2(1 - \cos^2\vartheta) \,,
\end{aligned} \tag{32}$$

where ϑ is the angle between $\underline{r}$ and $\underline{B}$.

References

[1] J. M. Levy-Leblond, Commun. Math. Phys. **6**, 286 (1967)

[2] U. von Barth and L. Hedin, J. Phys. C **5**, 1629 (1972)

[3] A. K. Rajagopal and J. Callaway, Phys. Rev. B **7**, 1912 (1973)

[4] S. H. Vosko, L. Wilk and M. Nusair, Can. J. Phys. **58**, 1200 (1980)

[5] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981)

[6] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992)

[7] J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)

[8] V. L. Moruzzi, J. F. Janak and A. R. Williams, *Calculated Electronic Properties of Metals* (Pergamon Press, New York, 1978)

[9] J. F. Janak, Phys. Rev. B **16**, 255 (1977)

[10] J. F. Janak, Solid State Commun. **25**, 53 (1978)

[11] V. Heine and J. H. Samson, J. Phys. F **13**, 2155 (1983)

[12] P. H. Dederichs, R. Zeller, H. Akai and H. Ebert, J. Magn. Magn. Mater. **100**, 241 (1991)

[13] P. James, O. Eriksson, B. Johansson and I. A. Abrikosov, Phys. Rev. B **59**, 419 (1999)

[14] P. H. Dederichs, S. Blügel, R. Zeller and H. Akai, Phys. Rev. Lett. **53**, 2512 (1984)

[15] A. Oswald, R. Zeller and P. H. Dederichs, J. Magn. Magn. Mater. **54-57**, 1247 (1986)

[16] E. du Trémolet de Lacheisserie, D. Gignoux and M. Schlenker *Magnetism: Fundamentals* (Springer, New York, 2005)

[17] H. J. F. Jansen, Phys. Rev. B **38**, 8022 (1988)

A 3 Magnetism in Reduced Dimensions

Stefan Blügel
Institut für Festkörperforschung
and
Institute for Advanced Simulation
Forschungszentrum Jülich GmbH, D-52425 Jülich

Contents

1 Introduction **2**

2 Models **4**
2.1 Stoner Model 4
2.2 Role of Coordination Number 6
2.3 Heisenberg Model and Beyond 8
2.4 Critical Temperature 10
2.5 Orbital Moment and Magnetic Anisotropy 12
2.6 Dzyaloshinskii-Moriya Interaction 18

3 The Rashba effect at metallic surfaces **19**
3.1 Nonmagnetic Surfaces 19
3.2 Semimetal Surfaces 22
3.3 Magnetic Surfaces 25

4 Ultrathin Films **27**
4.1 (100) Oriented Monolayers on Nonmagnetic Substrates 29
4.2 (111) Oriented Monolayers on Nonmagnetic Substrates 32
4.3 Magneto-Interlayer Relaxation 35
4.4 Orbital Moment and Magnetic Anisotropy 37
4.5 Spin-Orbit Induced Homochiral Mesoscale Spin Spirals 42

1 Introduction

In this article we focus on two issues: (i) spin related behavior of electrons propagating in the potentials with structure inversion asymmetry (SIA). Owing to the spin-orbit interaction, inversion asymmetric potentials give rise to a Bychkov-Rashba spin-orbit coupling causing a spin-splitting of a spin-degenerate electron gas. In this article we show that the rich spin-orbit driven physics in potentials with SIA is effective also for electrons at metallic surfaces. We illustrate the Rashba spin-splitting of surface electrons at noble-metal surfaces, e.g. Ag(111) and Au(111), at the semimetal surfaces Bi(111) and Bi(110), and the magnetic surfaces Gd(0001) and O/Gd(0001). E.g. on the Bi(110) surface the Rashba spin-splitting is so large that the Fermi surface is considerably altered, so that the scattering of surface electrons becomes fundamentally different. On a magnetic surface, the Rashba splitting depends on the orientation of the surface magnetic moments with respect to the electron wavevector, thus offering a possibility to spectroscopically separate surface from bulk magnetism.
(ii) The second issue is the dimensional aspect of itinerant magnetism, in particular of those systems including d electrons, as relevant for the magnetic ground-state properties of metallic surfaces, interfaces, multilayers, ultrathin films, step edges, or magnetic clusters deposited on surfaces. Considering the vast number of possible systems, a number growing fast with the number of constituent atoms, the surface and interface orientation, the chemical and structural roughness at interfaces, the electronic nature of the substrate (metal, semiconductor, metal), an exhaustive review is unattainable. Instead we discuss chemical trends in order to develop an intuition helpful to understand also new systems or envisage new effects not investigated yet.
The simplest low–dimensional systems are isolated atoms, whose spin moments as function of the the number of d electrons are well described by Hund's first rule: the spins of all electrons are aligned in parallel as long as no quantum number is occupied more than once. Thus nearly all of the 30 transition–metal atoms have magnetic spin moments. The largest possible d moments occur at the center of each series, i.e. 5 $\mu_{\rm B}$ for Cr and Mn in the $3d$ series. On the other hand, it is well-known that only 5 of 30 transition metals remain magnetic in their bulk crystalline phase: Co and Ni are ferromagnetic, Cr is antiferromagnetic, and Mn and Fe are ferromagnetic or antiferromagnetic depending on their crystal structure (cf. Fig. 1). Low–dimensional transition–metals should fall in between these two extremes. Magnetic material may be envisaged, which is nonmagnetic as bulk metal but magnetic as nano-structure. Although these arguments do apply, band narrowing, charge transfer, lift of degeneracies, structural, morphological or thermodynamical changes mire the interpolation and it took about 10 years to settle the "relatively simple" problem of the surface magnetism of Ni(100) [1]). Totally

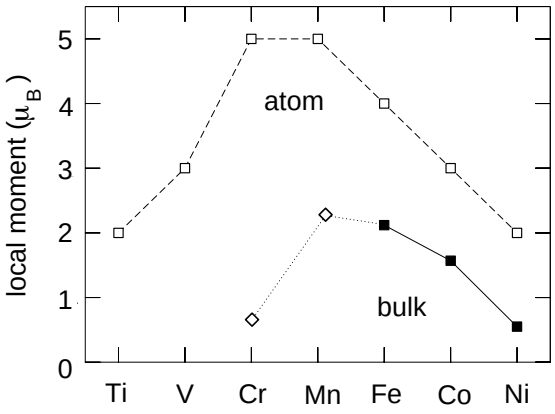


Fig. 1. *Local magnetic moments of isolated $3d$ atoms (empty squares connected by dashed line), ferromagnetic (solid squares connected by solid line) and antiferromagnetic (diamonds connected by dotted line) $3d$ bulk metals. The magnetism of the atom includes only the moment due to the d electrons. For the bulk metals the experimental spin moment are shown.*

Table 1: *Typical ground-state energies E in eV/atom for $3d$ metal films.*

	E (eV/atom)
cohesive energy	5.5
local moment formation	1.0
alloy formation	0.5
magnetic order	0.2
structural relaxation	0.05
magnetic anisotropy	$0.0001 \div 0.002$

unclear is the magnetic coupling between the moments of atoms in systems of reduced dimensions, in particular if the frustration of the magnetic interactions comes into play as for example in exchange-bias systems.

The magnetic ground-state properties may be divided into (i) the formation of local moments of different sizes (ii) the interaction between the local moments responsible for the formation of the magnetic order, the magnetic coupling at interfaces or across spacer layers, and (iii) the magnetic anisotropy energy, which couples the direction of the magnetization to the lattice and determines the easy and hard axes of the magnetization. At this point it may be useful to put the magnetic energies involved in (i)–(iii) into a general perspective by comparing them in Table 1 with the structural and compositional ground-state energies. From the relative importance of the different energies it is evident that the local moment formation has a considerable influence on the stability, alloy formation, atom arrangement and atom relaxation at the interface. Since the local moments may change quite substantially at the interface, materials with new and unknown phases [2], crystal structures and magnetic structures [3] are to be expected. Despite the technological importance and the importance for the finite temperature properties of thin films, the anisotropy energy is a rather small quantity The anisotropy energy depends on all structural and electronic details of an interface, while in turn, with the exception of the magnetostriction, not much influence on structural aspects are expected. In this sense the problem of the magnetic anisotropy can be tackled after the interface is completely determined otherwise.

The issues (i) and (ii) cross in the point that electrons at a SIA invironment of surfaces or interfaces, common to all nanostructures, gives rise not only to the well-known symmetric Heisenberg exchange but in addition also to a less known Dzyaloshinskii-Moriya-type (DM) antisymmetric exchange. Depending on the strength of the DM interaction, we expect in low-dimensional magnets deposited on substrates, such as ultrathin magnetic films, homochiral, i.e. chirality broken two- or three-dimensional magnetic ground-state structures between nanometer and sub-micrometer lateral scale. Little is known about the strength of the DM interactions in low dimenional magnets and this is one of very active research areas.

There are several low-dimensional systems and phenomena which are not covered in this chapter. To these belong the *magnetic chains, wires, clusters* in the gas phase, the *molecular magnets* and the *Kondo-Effekt of adatoms on surfaces*. When the growth of thin films is repeated to form multilayers, in particular those of thin magnetic films separated by non-magnetic spacer layers, an exchange interaction between the films across the spacer layer occurs, which is known as the *interlayer exchange coupling*. The emphases of this lecture is not on thermodynamic properties

of low-dimensional systems, they are only included at minimum. The work which I present are basically predictions, analyses and understanding of the electronic structure, magnetic moment, and magnetic structure as – results obtained from the density functional theory introduced in the lecture of Dr. Bihlmayer and Dr. Zeller.

2 Models

In this section the reader is reminded at the theoretical concepts used to predict and analyze the results. The theories and model have been in part been introduced by previous speaker, e.g. Dr. G. Bihlmayer and Dr. R. Zeller. Further, simple models are discussed to rationalize the results.

2.1 Stoner Model

The occurrence of ferromagnetism can be studied on the basis of the Stoner criterion introduced in the chapter *Density Functional Theory of Magnetism*:

$$I\, n(E_{\rm F}) \;>\; 1\;. \tag{1}$$

The Stoner criterion is an instability condition which expresses the competition between the exchange interaction in terms of the exchange integral I which drives the system into ferromagnetism for large I and the kinetic energy in terms of the nonmagnetic density of states (DOS) $n(E_{\rm F})$ at the Fermi energy $E_{\rm F}$. The kinetic energy rises if the system becomes magnetic, the more the wider the band width or the lower the density of states, respectively. A big exchange integral and a large nonmagnetic DOS at the Fermi energy favors ferromagnetism. When ferromagnetism occurs, the double degeneracy of the energy bands $\varepsilon_{\vec{k}}$ is lifted, majority states $\varepsilon_{k\uparrow}$ and minority states $\varepsilon_{k\downarrow}$ are rigidly shifted in energy by the exchange splitting IM, where M is the value of the local magnetic moment,

$$\varepsilon_{k\uparrow} \;=\; \varepsilon_{\vec{k}} - \frac{1}{2}IM \qquad \text{and} \qquad \varepsilon_{k\downarrow} \;=\; \varepsilon_{\vec{k}} + \frac{1}{2}IM\;. \tag{2}$$

The rigid band shift is a good model if the shift is small as in case of bulk ferromagnets. Deviations can be found for thin films, as the magnetic moments and thus the exchange splitting is large.
The Stoner criterion in equation (1) can be generalized describing the instability against the formation of a frozen spinwave of wave vector $\vec{q}$,

$$I\,\chi_{\vec{q}}(E_{\rm F}) \;>\; 1\;. \tag{3}$$

Obviously the local DOS was replaced by the $\vec{q}$ dependent susceptibility $\chi_{\vec{q}}$, a quantity which is expressed in the Heisenberg model by $J(\vec{q})$. Within equation (3), antiferromagnetism is just a special case. While the DOS at $E_{\rm F}$ is easily accessible by experiment or electronic structure calculations, the static susceptibilities $\chi_{\vec{q}}(E_{\rm F})$ are not. To make use of equation (2) an approximate criterion for antiferromagnetism is derived which makes explicit use of the local DOS. Small magnetic moments with the same magnitude M, but possibly different directions $\widehat{\vec{M}_j}$ at different sites j, induce in linear response theory local moments $\vec{M}_i$ at sites i

$$\vec{M}_i = \sum_j \chi_{ij}(E_{\rm F})\, M\, \widehat{\vec{M}_j}\;. \tag{4}$$

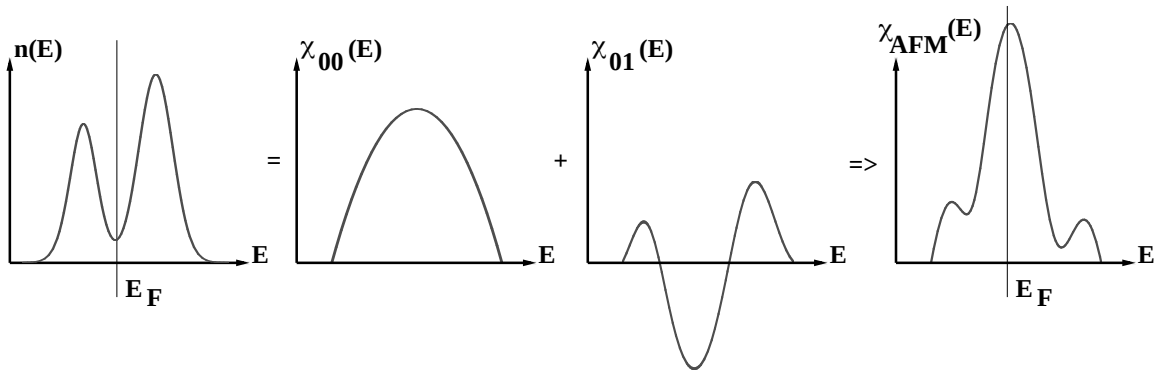


Fig. 2: *Graphical illustration of equation (8) for a DOS typical for transition-metal monolayers on (001) oriented noble metal substrates.*

The staggered susceptibility describing a particular magnetic state (M) is then expressed as

$$\chi_{\mathrm{M}} = \sum_i \chi_{0i}\, \widehat{\vec{M}_0} \cdot \widehat{\vec{M}_i}\ . \tag{5}$$

Particular examples of this staggered susceptibility are the ferromagnetic (χ_{FM})

$$\chi_{\mathrm{FM}} = n = \sum_i \chi_{0i} \tag{6}$$

and the antiferromagnetic (χ_{AFM})

$$\chi_{\mathrm{AFM}} = \sum_i (-1)^{(i)}\, \chi_{0i} \tag{7}$$

susceptibilities. Assuming that for $3d$ metals the nearest-neighbor interaction is the most dominating one, χ_{0i} can be neglected for all sites beyond nearest neighbors ($\chi_{0i} = 0$ for $i > 1$), and χ_{FM} and χ_{AFM} are given approximately by

$$n(E) \approx \chi_{00}(E) + \chi_{01}(E), \quad \text{and} \quad \chi_{\mathrm{AFM}}(E) \approx \chi_{00}(E) - \chi_{01}(E)\ , \tag{8}$$

where $\chi_{00}(E)$ is the local or atomic susceptibility, respectively, at the energy E. The energy dependence of χ_{00} is fairly simple. It follows from atomic Hund's rule-type arguments: The maximum spin M occurs for half band-filling, hence the atomic (local) susceptibility $\chi = \partial M/\partial H$ will also be largest. From equation (8), we can obtain an approximate form for χ_{AFM} using only DOS information. This is illustrated in Figure 2. As function of the d band-filling, from V to Ni, the Fermi energy sweeps from the left to the right through the DOS. If the Fermi energy is positioned at the center of the band as for Cr, and the DOS is low but the antiferromagnetic susceptibility is high, and antiferromagnetism is expected. If the Fermi energy is closer to the end of the band, the antiferromagnetic susceptibility is small but the DOS is large and ferromagnetism is expected as for Fe, Co, and Ni. Mn and Fe are at the edge of both magnetic states, and depending on circumstances different magnetic ground states can be found. Compare also to the calculated DOS, Figure 12, in Sect. 4.1

2.2 Role of Coordination Number

As discussed in Sect. 2.1 the Stoner criterion for ferromagnetism, equation (1), depends (i) on the Stoner parameter I and (ii) the DOS $n(E_{\rm F})$ at the Fermi energy $E_{\rm F}$.
(i) The exchange integral I is an intra-atomic, element specific quantity, and in simplest approximation independent of the local environment, the structure and the site of a given atom, e.g. surface atom or bulk atom. According to Gunnarsson [4] and Janak [5] a global trend

$$I_{3d} > I_{4d} > I_{5d} \tag{9}$$

was found for the exchange integrals of the $3d$, $4d$, and $5d$ transition-metal series.
(ii) Focusing on the d electrons as relevant electrons for itinerant magnetism, the DOS depends on both the coordination number N_{nn} and the hopping matrix elements h_d between the d electrons. This can be understood as follows: The energy integral $\int_W n_\ell(\varepsilon)\, d\varepsilon = 2\ell + 1$ over the band width, W, of the local DOS of angular momentum quantum number $\ell(= 2)$ is normalized to $2\ell + 1$ states. Thus, in simplest approximation possible (e.g. rectangular shaped DOS), one can assume that the local DOS scales inversely proportional to the band width, W,

$$n(E_{\rm F}) \sim \frac{1}{W}\,. \tag{10}$$

At the atomic limit the band width converges to zero, the Stoner criterion is always fulfilled and moments in accordance with Hund's first rule will be found. In general the DOS consists of contributions from electrons in s, p, d, and f states. For transition metals by far the largest contribution comes from the d electrons, and the d–d hybridization determines the shape of the density of states. Therefore, in the following discussion we restrict ourselves to d electrons and write

$$n(E_{\rm F}) \approx n_d(E_{\rm F}) \sim \frac{1}{W_d}\,. \tag{11}$$

The average local band width $\overline{W_d}(\vec{R}_i)$ for an atom i at position $\vec{R}_i$ can be estimated in a nearest neighbor tight-binding model, applicable for the itinerant but tightly bound d electrons of transition-metal atoms, to be

$$W_d \approx \overline{W_d}\,(\vec{R}_i) = 2\sqrt{N_{nn}(\vec{R}_i)}\; h_d(R_{nn})\,. \tag{12}$$

According to equation (12) the band width depends on two quantities: (a) the hopping matrix element h_d of the d electrons and (b) the number of nearest neighbor atoms or coordination number N_{nn}.
(a) The hopping matrix element depends on the overlap of the d wavefunctions. It decreases with increasing lattice constant or distance R_{nn} to the nearest neighbor atom and for a given lattice constant it increases with the extension of the wavefunction or, equivalently, the number of nodes. In Figure 3 the band widths of $3d$, $4d$, and $5d$ bulk transition-metals are schematically shown, together with the band widths of rare earths and actinides. In line with the arguments of increasing number of nodes from $3d$ to $5d$ wavefunctions a clear "macro trend" between the transition-metal series is visible summarized as follows:

$$h_{3d} < h_{4d} < h_{5d} \Longrightarrow W_{3d} < W_{4d} < W_{5d} \Longrightarrow n_{3d} > n_{4d} > n_{5d} \tag{13}$$

Within each transition-metal series there exists an additional "micro trend": due to the incomplete screening of the Coulomb potential of the nucleus by the d electrons, the d wavefunctions

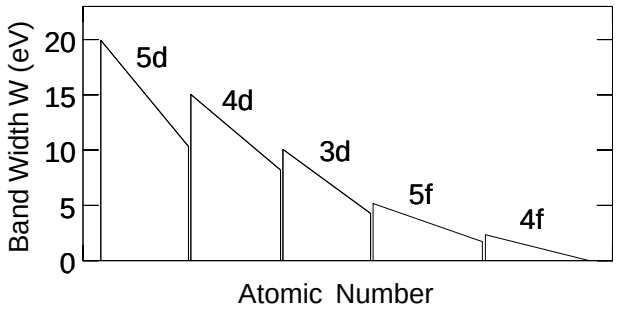


Fig. 3: *Schematic illustration of the band width W of the transition-metals together with rare earths ($4f$) and actinides ($5f$), all in the bulk phase. The $5f$ electrons of the early actinides and the $3d$ electrons of transition-metals from the middle to the end of the $3d$ series (Cr to Ni) show itinerant magnetism, while the magnetism of the late actinides and the rare earths is best described as localized magnetism, and their magnetic properties can in good approximation be explained in terms of Hund's rule.*

at the beginning of the transition-metal series are more extent than at the end of the series, thus the hopping matrix element at the beginning of the series is larger than at the end, with the well-known consequences for the band width W and the DOS $n(E_{\rm F})$.
(b) The smaller the coordination number N_{nn} the smaller the d–d hybridization and the smaller is the band width. Let's consider for example the coordination number of an atom in the environment of a fcc crystal ($N_{\rm fcc} = 12$), of an atom in the (001)–surface of the fcc crystal ($N_{(001)} = 8$), located in a two-dimensional (001) monolayer film ($N_{\rm ML} = 4$) and of an atom in a monoatomic chain ($N_{\rm chain} = 2$), keeping the nearest neighbor distance fixed ($R_{nn} = \text{constant}$) and keeping the bonding strength fixed ($h_d = \text{constant}$). Under these circumstances, one obtains for the ratio of the band widths

$$W_d{}^{\rm chain} \;:\; W_d{}^{\rm ML} \;:\; W_d^{(001)} \;:\; W_d^{\rm fcc} \;=\; 0.41 \;:\; 0.58 \;:\; 0.82 \;:\; 1 \quad ,$$

or the local DOS

$$n_d^{\rm chain} \;: n_d^{\rm ML} \;:\; n_d^{(001)} \;:\; n_d^{\rm fcc} \;=\; 2.45 \;:\; 1.73 \;:\; 1.22 \;:\; 1 \; . \tag{14}$$

Thus, the reduction of the coordination number leads to less d–d hybridization, consequently to band narrowing, and in low-dimensional structures the tendency towards magnetism is considerably boosted. Accordingly, one can expect, that transition-metals, which are nonmagnetic as bulk metals, may become magnetic at surfaces or as ultra-thin films. A nice manifestation of these arguments was recently reported for the size and shape dependence of the local magnetic moments in Fe clusters on the Ni(100) surface [6]. The arguments put forward here for the increased ferromagnetism in reduced dimensions can be carried over directly to the increased antiferromagnetic susceptibility.
The magnetic properties are expected to depend also on the surface or film orientation, because along with a change of the surface orientation goes a change of the coordination number N_{nn} (cf. Table 2) as well as a change of the nearest neighbor distance $R_{\parallel}$ between the surface atoms and $R_{\perp}$ between the surface atoms and the atoms in the next layer. For a fcc lattice, the (111) surface is the most densely packed one, and we expect for it the smallest enhancement of the magnetic moments. Among the three low-index surfaces, with the orientation (001), (011), and (111), the (011) surface leads to the most open surface. For the latter we expect the largest

Table 2: *Coordination number N_{nn}, interlayer distance d, point symmetry S, and packing density ρ (fraction of the area of the surface unit cell, covered by atoms with an atom radius of touching bulk atoms) for a fcc lattice. Only the 3 low-index surfaces, (001), (011), and (111), are considered. a is the lattice parameter of the simple cubic unit cell.*

	N_{nn}	S	d/a	ρ
(111)	9	$C_{3\mathrm{v}}$	0.5774	0.9068
(001)	8	$C_{4\mathrm{v}}$	0.5000	0.7854
(011)	7	$C_{2\mathrm{v}}$	0.3536	0.5554

magnetic moments. At surfaces or ultrathin films of bcc lattice type the trend should be exactly the opposite. The most densely packed surface is the (011) surface for which we expected the smallest enhancements of the magnetic moments. The (111) surface is the most open one. This surface is already close to a stepped one.
The implication of the coordination number, discussed so far is an important aspect in interface magnetism, but it is not the whole story. Further important aspects neglected so far have to be taken into account in order to give a qualitative correct description of the magnetism at interfaces.

2.3 Heisenberg Model and Beyond

To predict the magnetic ground state of a low-dimensional magnetic system can be a highly nontrivial problem. In cases, for example, where competing exchange interactions between neighboring atoms cannot be satisfied, exchange interactions are frustrated giving rise to a multitude of possible spin-structures. In the past, the magnetism of complex spin structures of itinerant magnets has been almost exclusively discussed within the framework of model Hamiltonians, e.g. the classical Heisenberg Hamiltonian,

$$H_{2\text{-spin}} = -\sum_{i,j} J_{ij}\ \vec{S}_i \cdot \vec{S}_j\ . \tag{15}$$

The spins localized on the lattice sites i, j are considered as classical vectors $\vec{S}$, with the assumption that the spins on all lattice sites have the same magnitude S:

$$\vec{S}_i^2 = S^2, \text{ for all } i\ . \tag{16}$$

The exchange interaction between the spins is isotropic and described by the pair interaction J_{ij}. In localized spin systems the J_{ij} can be safely approximated by the ferromagnetic ($J_1 > 0$) or antiferromagnetic ($J_1 < 0$) nearest-neighbor (n.n.) interaction, i.e. $J_{ij} = 0$ for all i, j, except for $J_{n.n.} = J_1$. Also in itinerant magnets J_1 often dominates over the rest of the further distant pairs, however, an attempt to reproduce T_{C} solely from J_1 produces results of limited quality. In many cases interactions between atoms as distant as 20 sites need to be included to give reliable results.
Exchange interactions beyond the classical Heisenberg model can be motivated from a perturbation expansion of the Hubbard model [7]. Expanding the Hubbard model into a spin model, replacing the spin operators by classical spin vectors, a second order perturbation expansion

reproduces the classical Heisenberg model. The fourth order perturbation treatment (the third order is zero in the absence of spin-orbit interaction) yields two additional terms of different form. One is the four-spin exchange interaction (4-spin):

$$H_{4\text{-spin}} = -\sum_{ijkl} K_{ijkl}\left[(\vec{S}_i\vec{S}_j)(\vec{S}_k\vec{S}_l) + (\vec{S}_j\vec{S}_k)(\vec{S}_l\vec{S}_i) - (\vec{S}_i\vec{S}_k)(\vec{S}_j\vec{S}_l)\right] . \tag{17}$$

The 4-spin interaction arises from the hopping of electrons over four sites, i.e. the process $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$. The second term, resulting from the hopping $1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1$, is the biquadratic exchange:

$$H_{\text{biquadr}} = -\sum_{ij} B_{ij}(\vec{S}_i \cdot \vec{S}_j)^2 . \tag{18}$$

The exchange parameters J_{ij}, K_{ijkl}, and B_{ij} depend on the details of the electronic structure and it is known [8] that for transition-metals the sign and magnitude are rapidly varying functions of the d-band filling. In thin films, the nearest neighbor exchange constants scaled by the appropriate power of the magnetic moment, S^4K_1 and S^4B_1, are about one order of magnitude smaller than S^2J_1, which is for example for Mn/Cu(111) about 30 meV [9]. The higher order spin interactions have then the effect, depending on the sign and value, to lift the degeneracy of magnetic states which are degenerate in the Heisenberg model.

In itinerant magnets, the electrons that are responsible for the formation of the magnetic state do participate in the formation of the Fermi-surface and hop across the lattice. Thus, it is by no means clear how far a short-ranged n.n. interaction or even how far the Heisenberg model, and models beyond that, can go in giving a sufficiently good description of the physics of itinerant magnets at surfaces and films. We believe that the interplay of *ab initio* calculations with model Hamiltonians provides a powerful approach to investigate the magnetic structures of complex magnetic systems as low-dimensional magnets and to deal with their thermodynamical properties.

For our purpose here, the value of the Heisenberg model lies in two facts: (i) to construct a zero-temperature phase diagram of relevant spin states as function of the exchange parameters J_{ij} and (ii) that a spin-spiral state, SSDW, with a propagation vector $\vec{q}$ in the first Brillouin zone (BZ) is a fundamental solution of the Heisenberg model for a Bravais lattice. On a Bravais lattice it is convenient to write the spin on lattice sites in terms of their discrete Fourier components $\vec{S}_{\vec{q}}$. The Heisenberg Hamiltonian can then be written in the simple form

$$H_{2\text{-spin}} = -N\sum_{\vec{q}} J(\vec{q})\ \vec{S}_{\vec{q}} \cdot \vec{S}_{-\vec{q}} . \tag{19}$$

The summation goes over the reciprocal lattice vectors $\vec{q}$. N denotes the number of lattice sites in the crystal.

$$J(\vec{q}) = \sum_{i,j} J_{i-j}\ \mathrm{e}^{-i\vec{q}(\vec{R}_j-\vec{R}_i)} = \sum_{\vec{0}-\vec{R}_i} J_{\vec{0}-\vec{R}_i}\ \mathrm{e}^{-i\vec{q}(\vec{0}-\vec{R}_i)} = J(-\vec{q}) = J(\vec{q})^* \tag{20}$$

are the Fourier transformed exchange constants and $\vec{R}_i$ is the real-space coordinate of lattice site i. The lowest energy

$$E(\vec{Q}) = -NS^2J(\vec{Q}) \tag{21}$$

is found for the magnetic ground state $\vec{S}_{\vec{Q}}$ of the SSDW with wavevectors $\pm\vec{Q}$ (as well as symmetry related $\vec{Q}$ vectors) which are obtained by minimizing the energy equation (19) under

the condition equation (16). The corresponding spin structure are helical spin spirals equation is given by

$$\vec{M}_i = M(\cos(\vec{q}\cdot\vec{R}_i)\sin\vartheta, \sin(\vec{q}\cdot\vec{R}_i)\sin\vartheta, \cos\vartheta) \ , \tag{22}$$

for $\vartheta = 90^\circ$ and $\vec{M}_i = -g\mu_{\rm B}\vec{S}_i$. For particular $\vec{Q}$ vectors, e.g. $\vec{Q} = \pm 2\pi/a(0,0,1/2)$ one may find the $uudd$-state as ground state, a collinear bilayer antiferromagnetic state of ferromagnetic double layers, which couple antiferromagnetically. This state, for example, was found in calculations for regime II of fcc-Fe films on Cu(001) [10].
In two dimensions $\vec{Q}$ is typically located at high-symmetry points (lines) of the two-dimensional Brillouin zone, where the energy equation (21) as function of the $\vec{q}$-vector should have an extremum, a maximum, a minimum (or a saddle point), depending on the exchange constants J_{ij}, and the symmetry of the high-symmetry point. In principle, one cannot exclude that the minimum of the energy will be located at any arbitrary point along the high-symmetry lines, representing an incommensurate spiral spin-density wave. In practice, we perform first-principles total energy calculations $E\left[n(\vec{r}), \vec{m}(\vec{r})|\{\vec{q}\}\right]$ for flat spin-spirals along the high symmetry lines to gain an overview of possible minimum energies $E(\vec{Q})$. The role of higher order spin interactions are then investigated carrying out constraint calculations of the total energy $E[n(\vec{r}), \vec{m}(\vec{r})|\{\hat{\vec{e}}\}]$ for particular paths of magnetic configurations. Zero-temperature phase diagrams in the $J_{01}\cdots J_{0i}$ space are very helpful to reduce the relevant phase space of possible spin structures. This recipe had been followed in Sect. 4.1 and 4.2 to explore the magnetic ground state of thin films. The above described mapping of *ab initio* calculations to spin-models relies on the assumption, that the magnetic moment does not depend on the relative difference of the magnetization axis between atoms. For itinerant systems this is not necessarily guaranteed. The change of the moment with respect to the relative quantization axis can be mapped on spin-models introducing also higher order spin interactions.

2.4 Critical Temperature

It is well known that magnetic excitations in itinerant ferromagnets are basically of two different types, namely, the Stoner excitations associated with longitudinal fluctuations of the magnetizations, and the spin-waves or magnons, which correspond to collective transverse fluctuations of the magnetization direction. Near the bottom of the excitation spectrum, the density of states of magnons is considerably larger than that of the Stoner excitations, so that the thermodynamics in the low-temperature regime is completely dominated by magnons. Stoner excitations can be savely ignored. Thus, it seems reasonable to extend this approximation up to the critical temperature, $T_{\rm c}$, to neglect the Stoner excitation systematically, and to describe the transversal fluctuations by the Heisenberg model expressed in equation (15) with exchange parameters determined from first-principles. An overview over the current applications along this line of mapping first-principles results on Heisenberg-type Hamiltonians to study the thermodynamical properties of bulk and low-dimensional magnets can be found in the paper of Turek *et al.* [11].
Below the critical temperature, the so-called Curie temperature $T_{\rm C}$ for ferromagnets or the Néel temperature, $T_{\rm N}$, for magnets with more complex magnetic phases, the spontaneous magnetization remains finite, while it is zero above $T_{\rm C}$. The phase transition is of second order, i.e. the spontaneous magnetization which is the order parameter characterizing the phase transition, vanishes continously at $T_{\rm C}$. A second order phase transition is governed by the principle of universality, where a system close to the phase transition does not depend on details of the system such as its material parameters or the geometry of the sample, but rather on the symmetry of the

underlying model and the dimension of the spin, which is three for the Heisenberg model. In this lecture we are interested in estimating the critical temperatures as these are non-universal quantities, and of great practical importance. It is certainly important to know whether cryogenic, room temperature or elevated temperatures are required to observe particular phenomena.
A first simple estimate of the Néel temperature for a three-dimensional system exhibiting a helical spin-spiral ground state with wave vector $\vec{Q}$ is given by the mean-field approximation (MFA) to the Heisenberg Hamiltonian, which leads to

$$k_{\rm B}T_{\rm N}^{\rm MFA} = \frac{2}{3}S^2 J(\vec{Q}) \qquad \text{and} \qquad k_{\rm B}T_{\rm C}^{\rm MFA}(\text{n.n.}) = \frac{2}{3}S^2 N_{nn} J_1 \ , \tag{23}$$

where $k_{\rm B}$ is the Boltzmann constant. For the ferromagnetic state, $\vec{Q} = (0,0,0)$, the left equation (23) gives the Curie temperature in the MFA, $T_{\rm C}^{\rm MFA}$, expressed explicitely in the right equation in the nearest neighbor approximation to the exchange interaction. N_{nn} is the coordination number of nearest neigbhor atoms and J_1 is the interaction strength as introduced in Subsect. 2.3. The MFA gives the right proportionality of $T_{\rm C}$ with respect to the number of neighbors, but has also a few deficiencies. Besides overestimating $T_{\rm C}$ for three-dimensional systems by typically about 20%, $T_{\rm C}^{\rm MFA}$ does not depend on the lattice structure nor on the dimensionality of the system. These shortcomings are remedied treating the Heisenberg model in the random phase approximation (RPA) [12, 13], which gives for the critical temperatures

$$\frac{1}{k_{\rm B}T_{\rm N}^{\rm RPA}} = \frac{3}{4}\frac{1}{NS^2}\sum_{\vec{q}}\left[\frac{1}{J(\vec{Q})-J(\vec{q})} + \frac{1}{J(\vec{Q})-\frac{1}{2}J(\vec{q}+\vec{Q})-\frac{1}{2}J(\vec{q}-\vec{Q})}\right]$$

$$\text{and} \qquad k_{\rm B}T_{\rm C}^{\rm RPA}(\text{n.n.}) = \frac{2}{3}S^2 N_{nn} J_1 \cdot \begin{cases} 0.660 & \text{sc} \\ 0.718 & \text{bcc} \\ 0.744 & \text{fcc} \end{cases} . \tag{24}$$

The RPA gives weight to the low-energy magnon excitations $E(q) \propto J(\vec{Q}) - J(\vec{q})$ in the summation over all modes. This provides estimates of $T_{\rm C}$ in close vicinity to the numerical analysis using classical Monte-Carlo simulations [14] discussed in detail in the book of Landau and Binder [15].
Both approximations show that the Curie and Néel temperature depend on the number of nearest neighbors and one expects that the critical temperature $T_{\rm C}$ decreases if the dimensionality of the system is reduced. But both approximations show a qualitatively different behavior for low-dimensional magnets. The mean-field approximation overestimates the tendency for long-range order and predicts always a phase transition to ferromagnetic order in the Heisenberg model, no matter whether we have a one, two or three dimensional system, whereas $T_{\rm C}^{\rm RPA} = 0$ already for two-dimensional systems. This is consistent with the theorem of Mermin and Wagner [16], which states that in two dimensions there is no spontaneous long-range ferromagnetic order for isotropic Heisenberg models with short-range interaction ($\sum_j J_{ij} r_{ij}^2 < \infty$) at finite temperature. In thin films the long range order at finite temperature is stabilized by the magnetic anisotropy, which is practically always present. It opens a gap Δ in the excitation spectrum of the spin-waves, $E(q) \propto \Delta + J(\vec{Q}) - J(\vec{q})$, and supresses low-energy long-wavelength fluctuations which occur for low temperatures. According to a renormalization group analysis of Erickson and Mills [17, 18] the transition temperature in two dimensions, $T_{\rm C}^{(2{\rm D})}$, scales with the transition temperature of the three-dimensional Heisenberg model which is renormalized by a logarithmic

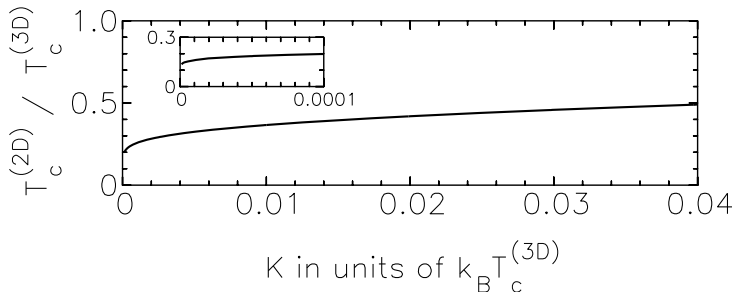


Fig. 4: *Critical temperature of a two-dimensional magnet as function of the uniaxial anisotropy following equation (25). The function startes at zero for $K = 0$. Note its rapid growth in the vicinity of the origin as shown in the inset at magnified scale.*

factor,

$$T_{\rm C}^{(2{\rm D})} = T_{\rm C}^{(3{\rm D})} \frac{2}{\ln\left(\frac{3\pi}{4}\frac{k_{\rm B}T_{\rm C}^{(3{\rm D})}}{K}\right)}\,, \tag{25}$$

which contains the strength of the uniaxial magnetic anisotropy in terms of a constant K (see Subsect. 2.5). This result is displayed in Fig. 4. $T_{\rm C}$ vanishes in the isotropic limit ($K \to 0$) in accordance with the Mermin Wagner theorem. Interestingly, for finite K there is a rapid increase of $T_{\rm C}$ reaching reasonable values, of say 20% of the critical temperature in three dimensional systems, for anisotropy values of less than a percent of the ferromagnetic coupling constant. Consider for example Fe, with a shape anisotropy of 0.140 meV, which corresponds to 1.63 K on the temperature scale. This is only 0.14% of the Curie temperature of Fe, $T_{\rm C}^{(3{\rm D})}({\rm Fe}) = 1183$ K, but causes already a Curie temperature for a Fe film of $T_{\rm C}^{(2{\rm D})}({\rm Fe}) = 0.27 \cdot T_{\rm C}^{(3{\rm D})}({\rm Fe}) = 320$ K. Thus at any finite anisotropy, there is a critical temperature, where the spin degree of freedom is frozen out due to the presence of the anisotropy, i.e. the dimension of the spin is reduced from three for the Heisenberg model to one, spin-up and -down. In terms of universality the Heisenberg model with any finite anisotropy value is in the universality class of the Ising model, and the Ising model shows a phase transition in two-dimensions.
In one dimension even the Ising model does not show long-range order at finite temperatures. Although for quasi-one-dimensional magnetic chains, these are chains of finite size, there is strictly speaking no remanent magnetization or long-range order, but there is a temperature, known as blocking temperature, below which a finite chain seems to a have a spontaneous and remanent magnetization, with long-range order in the chain. In reality, this magnetic order is accompanied by a slow relaxation [19]. The relaxation rate depends on the magnetic anisotropy and can be of macroscopic times, such that a quasi-one-dimensional chain appears as a ferromagnet as it occurs in the experiments of Gambardella et al. [20] in 2002.

2.5 Orbital Moment and Magnetic Anisotropy

A piece of magnetic material is typically magnetically anisotropic. This means, besides the isotropic exchange interaction there are additional interactions, which make the total energy depend on the orientation of the magnetization as measured with respect to the crystal axes and the sample shape. This orientation dependent energy contribution is called the magnetic anisotropy energy (MAE), $E_{\rm MAE}$, given in units of energy per atom throughout this article. Without this

effect of the magnetic anisotropy, magnetism would have been hard to discover and possibly useless. In some way or the other, almost all applications of magnetic materials hinge on the fact that it is easier to magnetize a magnetic material in one direction than another. The magnetic anisotropy is responsible for the occurrence of easy and hard axes, stabilizes magnetic order against thermal fluctuations in dimensions where the exchange interaction alone would not suffice (see Subsect.2.4), and limits the width of a magnetic domain wall. It is for example responsible for the bimodal stability of magnetic domains with uniaxial symmetry, which allows the two possible magnetization directions in space to be interpreted in terms of bit "0" or "1". This makes magnetism very valuable for magnetic storage media. Since the magnetic anisotropy is strongly related to the crystalline symmetry and the shape of the samples, a general expression of E_{MAE} will be a complex function of the orientation of the magnetization relative to the crystal axes. In low-dimensional systems twofold symmetries are the most relevant ones and the magnetic anisotropy is then expressed as

$$H_{\mathrm{MAE}} = \sum_i \vec{S}_i \cdot \vec{\vec{K}}_i \cdot \vec{S}_i \ , \tag{26}$$

where the tensor of single-site anisotropy constants, $\vec{\vec{K}}$, determines the strength of the anisotropy as well as the direction of minimum and maximal energy, named easy and hard axes, respectively. In perfect thin films and wires the presence of a surface holds then responsible for an uniaxial anisotropy energy normal to the surface, i.e. all components of $\vec{\vec{K}}_i$ are zero except $K_i^{zz} = K\delta^{zz}$ for perfect films and $K_i^{xx} = 1/2K\delta^{xx}$ and $K_i^{yy} = 1/2K\delta^{yy}$ for isolated wires. After expressing $\vec{S}_i$ in the form of equation (22), the uniaxial MAE takes the angular angular dependence

$$E_{\mathrm{MAE}}(\theta) = -K\cos^2\theta \ . \tag{27}$$

θ denotes the angle between the magnetization and the film or wire normal and $K = \Delta E_{\mathrm{MAE}} = E_{\mathrm{MAE}}^{(\parallel)} - E_{\mathrm{MAE}}^{(\perp)}$ is the uniaxial anisotropy constant also given in energy per atom. The total MAE, $E_{\mathrm{MAE}}^{(\mathrm{tot})} = N_{\mathrm{A}} E_{\mathrm{MAE}} = V\mathcal{E}_{\mathrm{MAE}}$, of the system depends then on the number of atoms, N_{A}, in it. Frequently, the MAE is also expressed in terms of an energy density $\mathcal{E}_{\mathrm{MAE}}$. By definition, $K > 0$ ($K < 0$) describes the case of a preferred direction of the magnetization perpendicular, $\perp$, (parallel, $\parallel$) to the film plane or wire axis. Additional higher symmetries in plane, for example a fourfold symmetry in a (100) oriented film plane, corresponds to anisotropy contributions which are smaller in energy than the uniaxial anisotropy and are neglected here. The anisotropy constant depends sensitively on the chemical elements involved, structural details, details of the electronic structure and the dimensionality of the system.

The microscopic origins of the magnetic anisotropy are the magnetic dipolar interaction and the spin-orbit interaction. The dipolar interaction is of long range and senses the outer boundaries of the sample. This results in the shape anisotropy. Discussing long range contributions, the underlying atomistic lattice describing the crystallinity of the system can be neglected and the shape anisotropy is described in terms of a continuum theory. Any contribution to the MAE beyond the continuum theory taking explicitely the crystallinity of the system into account is summarized as magnetocrystalline anisotropy energy (MCA). Both the dipolar and the spin-orbit interaction contribute to the MCA and the total anisotropy constant K,

$$K = K_{\mathrm{shape}} + K_{\mathrm{MCA}}^{(\mathrm{dip})} + K_{\mathrm{MCA}}^{(\mathrm{so})} \ , \tag{28}$$

is just a linear superposition of the different contributions.

The shape anisotropy constant, $K_{\rm shape}$ in atomic Rydberg units per atom of a perfectly flat film of infinite extension or an infinitely long perfectly cylindrical wire is given by the local magnetic moment m and the atomic volume V as

$$K_{\rm shape}^{\rm film} = -2\pi\frac{2}{c^2}\frac{m^2}{V} \qquad \text{and} \qquad K_{\rm shape}^{\rm wire} = -\pi\frac{2}{c^2}\frac{m^2}{V}\ , \tag{29}$$

all expressed in atomic units, m in μ_B/atom, V in a.u.^{3} and the speed of light, c, by the inverse of the finestructure constant α, $c = 2/\alpha$. The negative sign denotes that the shape anisotropy pulls the magnetization into the film plane or along the wire axis. For bcc Fe, for instance, with a bulk magnetic moment of $2.215\ \mu_{\rm B}$ per atom and a lattice constant of 5.42 a.u., $K_{\rm shape}^{\rm film}$ is equal to -0.140 meV/atom. The long range interaction senses also the interface or surface roughness which is always present in real films. According to Bruno [21] the roughness gives rise to an effective perpendicular contribution to the shape anisotropy whose order of magnitude depends on the parameters characterizing the roughness. Obviously $K_{\rm shape}^{\rm film}$ and $K_{\rm shape}^{\rm wire}$ is the same for all atoms irrespective of their position in the film and K is thus homogeneous across the film or wire. The same is true for any finite ellipsoidal structure, for any other finite structure, e.g. a nano-pattern structure on a surface, $K_{\rm shape}$ becomes inhomogeneous and becomes typically much smaller at the boundary of the structure. For bulk samples, thick films, patterned nanostructures and wires the shape anisotropy is frequently the most important of the anisotropies.
For thin films and wires of a few atomic layers, the assumption that the magnetization can be treated by a continuous magnetic medium is no longer valid. Instead, the magnetic dipole-dipole energy has to be evaluated explicitely. In transition-metals, the magnetization distribution around the atom is almost spherical and is thus treated to a good approximation as a collection of discrete magnetic dipoles, which are regularly arranged on a crystalline lattice. The dipolar energy $E_{\rm dip}$ per atom experienced by a dipole at site i due to the presence of ferromagnetically aligned dipols on all other sites j can then be expressed as

$$E_{\rm dip}^{(i)}(\theta) = K_{\rm dip}^{(i)}\cos^2\theta = \frac{2}{c^2}\frac{1}{2}\sum_{j(j\neq i)}\frac{m_i m_j}{R_{i,j}^3}\left(1-3\cos^2\theta_{ij}\right)\ . \tag{30}$$

θ_{ij} is the angle between the direction of the magnetic moment m of the dipoles at sites i or j given in units of Bohr magneton and the vector $\vec{R}_{i,j}$ connecting atoms i and j. $R_{i,j}$ denotes the relative distance between these dipoles or atoms, respectively. The θ-dependence expresses explicitly the fact that the dipole-dipole interaction contributes to the magnetic anisotropy. Obviously, in thin films and wires the anisotropy energy depends on the position of the atom i normal to the surface or wire axis, respectively, and as such explicitly on the film thickness or wire diameter (in difference to $K_{\rm shape}$ where all atoms have the same value). For crystalline thin wires and films the sum in equation (30) can be evaluated straight forwardly with fast converging summation techniques [22, 23]. Draaisma *et al.* [24] have worked out in detail the layer dependent dipolar anisotropy $K_{\rm dip}^{(i)}$. In general, the outer atoms experience values of $K_{\rm dip}$ that are appreciably smaller than those of the inner layers which finally approach $K_{\rm shape}$. The inner atoms reach 95% of $K_{\rm shape}$ after about 15Å below the surface. The exact details depend on the crystal structure and surface orientation, e.g. a reduction between 25% and 45% of $K_{\rm shape}$ was reported for a (100) oriented fcc or bcc monolayer, respectively. The deviation of $K_{\rm dip}$ from $K_{\rm shape}$ gives $K_{\rm MCA}^{\rm (dip)}$ in equation (28), the dipolar contribution to the MCA which occurs here due to the presence of a surface or interface and is sometimes also called the surface contribution of the dipolar anisotropy. If the MAE is expressed in terms of energy densities $\mathcal{E}$, this $K_{\rm MCA}^{\rm (dip)}$

is expressed in terms of an areal density. The dipolar energy contributes also to the MCA of bulk systems or thick films or wires, if the underlying lattice structure has a twofold symmetry. For this three-dimensional case more sophisticated summation techniques such as the Ewald summation method [25] is required to obtain reliable results for equation (30).
The spin-orbit interaction, treated typically by a Pauli-type addition to the Hamiltonian as:

$$H_{\mathrm{so}} = \vec{\sigma} \cdot (\vec{E}(\vec{r}) \times \vec{p}) = \vec{\sigma} \cdot (\nabla V(r) \times \vec{p}) \tag{31}$$

provides the essential contribution to the MCA. This Pauli approximation derives naturally from the Dirac equation, and is normally sufficient for treating relativistic effects in transition-metal magnets. For a radially symmetric potential we can rewrite equation (31):

$$H_{\mathrm{so}} = \frac{1}{r}\frac{dV(r)}{dr}\vec{\sigma} \cdot (\vec{r} \times \vec{p}) = \frac{1}{r}\frac{dV(r)}{dr}(\vec{\sigma} \cdot \vec{L}) = \xi(\vec{r})\vec{\sigma} \cdot \vec{L}\ , \tag{32}$$

where $\vec{L}$ is the angular momentum operator. Since the radial derivative of the potential in a crystal will be largest in the vicinity of a nucleus, we can expect that the major contribution to the spin-orbit interaction will come from this region. Furthermore, since for small r the potential will be Coulomb-like ($V = -\frac{Z}{r}$), the radial expectation value of $\xi(r)$ leads to a material-dependent spin-orbit coupling constant ξ, which is roughly proportional to the square of the nuclear number Z, $\xi \propto Z^2$. In low-dimensional systems the MCA dominates over the shape anisotropy. The anisotropy depends crucially on the symmetry of the system.
In a solid, where the symmetry of the states is determined by the crystal field, spin-orbit coupling can now introduce orbital moments and magnetocrystalline anisotropies by coupling states that carry no orbital momentum, e.g. a d_{xy} and a $d_{x^2-y^2}$ orbital, such that the combination form an orbital moment in z direction. In second-order perturbation theory the expectation value of the orbital moment operator $\mu_{\mathrm{B}}\vec{L}$ can be written as:

$$m_{\mathrm{l}} = \mu_{\mathrm{B}}\langle\vec{L}\rangle = \mu_{\mathrm{B}} \sum_{i,j} \frac{\langle\psi_i|\vec{L}|\psi_j\rangle\langle\psi_j|H_{\mathrm{so}}|\psi_i\rangle}{\varepsilon_i - \varepsilon_j} f(\varepsilon_i)\left[1 - f(\varepsilon_j)\right]\ , \tag{33}$$

where f is the Fermi function ensuring that the wavefunction ψ_i is occupied and ψ_j is unoccupied. In a metal, where several bands are crossing the Fermi level, E_{F}, it is basically the sum of all contributions from bands near E_{F} that determine the orbital moment. Van der Laan [26] in 1998 has shown, that in the absence of spin-flip terms (i.e. when the majority and minority band are well separated by the exchange interaction), the spin-orbit coupling changes the total energy of a system in second-order perturbation theory as:

$$\delta E = \sum_{i,j} \frac{\langle\psi_i|H_{\mathrm{so}}|\psi_j\rangle\langle\psi_j|H_{\mathrm{so}}|\psi_i\rangle}{\varepsilon_i - \varepsilon_j} f(\varepsilon_i)\left[1 - f(\varepsilon_j)\right] \approx -\frac{\xi}{4\mu_{\mathrm{B}}}\widehat{m}_{\mathrm{s}} \cdot \left[\vec{m}_{\mathrm{l}}^{\downarrow} - \vec{m}_{\mathrm{l}}^{\uparrow}\right]\ , \tag{34}$$

where $\widehat{m}_{\mathrm{s}}$ is the direction of the spin moment, and $\vec{m}_{\mathrm{l}}^{\downarrow}$ and $\vec{m}_{\mathrm{l}}^{\uparrow}$ are the orbital moment vectors of the spin-down and spin-up bands, respectively. If the spin-up band is completely filled, we see that energy change, δE, is proportional to the size of the orbital moment and the magnetocrystalline anisotropy energy (MCA), i.e. the difference of δE for two different magnetization directions, will be proportional to the difference in the orbital moments. This relation between orbital moment anisotropy and MCA was first derived by Bruno [27].
We have discussed that the reduced coordination number in low-dimensional systems favors the increase of the spin moment. But it also enables the formation of large orbital moments, as can

be seen from most atoms. Also in the case of the orbital moment the hybridization with some neighboring orbitals "locks" the electrons in place and quenches the orbital moment. Imagine a Sc atom with only one d-electron: as an atom, according to Hund's rules, the orbital moment will be maximized and antiparallel to the spin moment. But when Sc atoms are assembled in a square lattice, orbitals with $m = -2$ and $m = +2$ will form linear combinations to build d_{xy} and $d_{x^2-y^2}$ orbitals of which the latter one will be occupied. The more these two levels are split in energy, the more difficult it will be for the electron to "circle" around the atom and, therefore, to form an orbital moment.

Table 3: *Local spin ($m_{\rm s}$) and orbital ($m_{\rm l}$) magnetic moments in units of $\mu_{\rm B}$ of Fe, Co and Ni atoms in bulk materials (n=3), unsupported thin films (n=2), wires (n=1), and as isolated atoms (n=0). For the bulk crystals the variation of the orbital moment with the direction is small, but for films and wires the orbital moments parallel ($\parallel$) and perpendicular ($\perp$) to film-plane or wire-axis are given. The geometry is chosen as if the film or wire would have been grown epitaxial on a Pt(111) substrate. The column $|K_{\rm MCA}|$ indicates the order of magnitude of the magnetocrystalline anisotropy energy for different dimensions. The results were obtained in the generalized gradient approximation to the density functional theory.*

	Fe			Co			Ni			$\lvert K_{\rm MCA}\rvert$
n	$m_{\rm s}$	$m_{\rm l}$		$m_{\rm s}$	$m_{\rm l}$		$m_{\rm s}$	$m_{\rm l}$		$\left[\frac{\rm meV}{\rm atom}\right]$
		$\parallel$	$\perp$		$\parallel$	$\perp$		$\parallel$	$\perp$	
3	2.05	0.05		1.59	0.08		0.62	0.05		0.01
2	3.07	0.07	0.10	2.09	0.20	0.19	0.94	0.18	0.14	1.00
1	3.22	0.72	0.27	2.32	0.98	0.77	1.18	0.84	0.44	10.00
0	4	2		3	3		2	3		–

In Table 3 some representative values of spin and orbital momentum have been collected. These calculations yield very small orbital moments: 0.05 $\mu_{\rm B}$, 0.08 $\mu_{\rm B}$ and 0.05 $\mu_{\rm B}$ for bcc Fe, hcp Co, and fcc Ni (and about twice the value if the orbital polarization (OP) proposed by Brooks [28] is included). It is well-known that the orbital moments are quenched in the bulk due to the strong hybridization with neighboring atoms. Larger orbital moments are obtained for the (111) oriented unsupported $3d$ monolayers. For Fe, Co and Ni the values are 2–3 times larger than the corresponding bulk values. Thus, in monolayer films the quenching of the orbital moments is less pronounced due to the reduced hybridization. However, it is important to realize that these enhanced orbital moments are still an order of magnitude smaller than the corresponding free atom values, as given by Hund's second rule (last row in Table 3). Consequently, we expect for atomic scale magnetic structures such as wires, small clusters and adatoms strong changes in the orbital moment and, in turn, large values of the magnetocrystalline anisotropy energy. In practice, these films are deposited on substrates. That will once more quench the values, especially for the orbital moments. But the spin-polarization of the substrate can lead to additional large contributions to the magnetocrystalline anisotropy energy in particular for substrates with large Z, such as Pt or Ir.

Typically, first-principles calculations based on the LSDA or GGA underestimate the orbital moments. In the literature several methods have been discussed how this deficiency can be overcome [28, 29, 30]. For example, the orbital moments of the bulk magnets are about twice

the value if Brooks' orbital polarization is applied [31, 32]. The effect of OP is much more drastic in low dimensions [33]. A systematic comparison of LSDA results for Pt supported and unsupported Fe and Co magnets in various dimensions can be found in the work of Komelj *et al.* [34] and Ederer *et al.* [35].

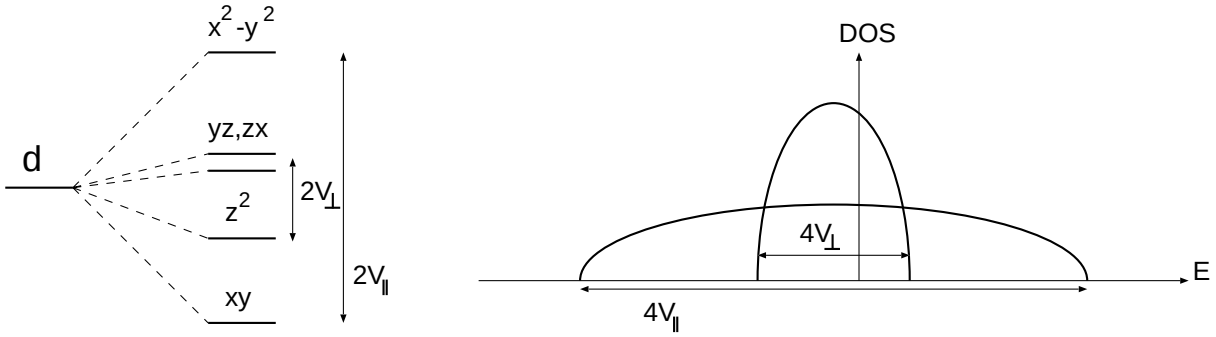


Fig. 5: *d-level splittings shown in the left figure at a given* **k***-point due to a crystal field in a square monolayer results in a density of states, shown on the right.*

In order to interpret *ab initio* results on thin films we discuss the case of an unsupported, (100) oriented d-metal monolayer, in terms of a simple model following Stöhr [36]. Assume that the d-band is substantially exchange split and more than half filled, so that we only have to consider the (partially filled) minority band. The d orbitals at each atom site experience in the monolayer plane a crystal field V, that leads to a splitting of these levels: if the surface normal is assumed to be in z-direction, the d_{xy} and $d_{x^2-y^2}$ levels will experience a stronger field than the out-of-plane directed d_{zx}, d_{yz} and d_{z^2} orbitals. The crystal field leads to a splitting of $2V_\parallel$ for the in-plane oriented orbitals and $2V_\perp$ for the out-of-plane oriented ones. In a band-picture, these splittings can be translated into bandwidths W, which will then be twice as large (cf. Figure 5). Normally, $V_\parallel$ will be larger than $V_\perp$, so that $R = V_\perp/V_\parallel < 1$. (If, however, the monolayer is sandwiched between two slabs of nonmagnetic material the situation could be changed.)
Assume that – like in the case of Co – the minority band is half filled; the d_{xy} and $d_{x^2-y^2}$ states will split symmetrically by $\pm V_\parallel$ around the Fermi level, the (d_{zx}, d_{yz}) and d_{z^2} states by $\pm V_\perp$. In a band-picture, these splittings will of course depend on the considered $\vec{k}_\parallel$ point. Now we can use perturbation theory equation (33) to calculate the orbital moments. The result [36]

$$m_l^\parallel = \frac{\xi\mu_B}{2V_\parallel}\left(\frac{3}{R} + \frac{2}{R+1}\right) \quad \text{and} \quad m_l^\perp = 4\frac{\xi\mu_B}{2V_\parallel} \tag{35}$$

shows, that only the in-plane orbital moment, $m_l^\parallel$ depends on the splitting of the out-of-plane oriented states, while the out-of-plane orbital moment is only quenched by the in-plane crystal field. This is intuitively clear, since $m_l^\perp$ corresponds to an in-plane motion of the electron, i.e. a hopping between the d_{xy} and $d_{x^2-y^2}$ states that are separated by $V_\parallel$. For the calculation of the magnetocrystalline anisotropy energy we can use equation (34), that gives:

$$K_{\mathrm{MCA}} = E_{\mathrm{MCA}}^\parallel - E_{\mathrm{MCA}}^\perp = -\frac{\xi}{4\mu_B}(m_l^\parallel - m_l^\perp) = -\frac{\xi^2}{8V_\parallel}\left(\frac{3}{R} + \frac{2}{R+1} - 4\right) . \tag{36}$$

From this equation we see that, as long as $R < 1$, an in-plane magnetization is obtained, while for $R > 1$ an out-of-plane easy axis is possible. Indeed it is observed that Co-monolayers on a

weakly interacting substrate (like Cu(001)) have an in-plane easy axis, while a Co layer sandwiched in Pt has a perpendicular magnetization. Taking typical values for $3d$-metal monolayers, a spin-orbit coupling strength $\xi \approx 75$ meV and bandwidths $W^{\parallel} \approx 3$ eV and $W^{\perp} \approx 2$ eV, one arrives at orbital moments of $m_{\rm l}^{\parallel} = 0.285\ \mu_{\rm B}$ and $m_{\rm l}^{\perp} = 0.200\ \mu_{\rm B}$ and the magnetocrystalline anisotropy energy per atom of $K_{\rm MCA} = 1.6$ meV, values in the ballpark of the *ab initio* results given in Table 3.

2.6 Dzyaloshinskii-Moriya Interaction

Magnets in low-dimensions face frequently a structure inversion asymmetric environment. Consider for example a thin magnetic film on a substrate with the vacuum potential on one side and the potential to the substrate on the other side. This inversion asymmetry leads to a gradient of the potential that can be interpreted in first approximation as an electric field normal to the film surface. In the rest frame of moving electrons, the electric field $\vec{E}$ appears by Lorentz transformation as a magnetic field $\vec{B} \propto \vec{p} \times \vec{E}$, which interacts then with the spin $\vec{\sigma}$ of the electron, giving rise to an additional term in the Hamiltonian, which was already encountered in equation (31) in the context of spin-orbit coupling. Here, instead of an orbital motion, a linear motion of an electron with momentum $\vec{k}$ in an electric field oriented along $\vec{e}_z$ is considered. This can be described by a Hamiltonian $H = \alpha_R \vec{\sigma} \cdot (\vec{k} \times \vec{e}_z)$, known as Rashba-term [37]. The strength described by the Rashba-parameter, α_R, is determined e.g. by the asymmetry of the wavefunction due to the asymmetry of the potential or the electric field, respectively, and the spin-orbit interaction of the electrons involved.

The magnetic interaction between the spin $\vec{S}_i$ at lattice site i and $\vec{S}_j$ at lattice site j is caused by electrons which hop from site i to site j and back. Electrons in a magnetic film propagate in an exchange field $\pm 1/2 IM$ (cf. equation (2)), the bands are exchange split and the time-inversion symmetry is lost. Due to the spin-orbit interaction caused by the Rashba term, electrons experience a kinetic energy with an additional weak spin-dependent potential, which depends on the propagation direction $\vec{p}$ of the electrons. Thus, the motion from site i to j and the back motion from j to i is slightly different. The same is true for time inverse hopping process, the electron hopping first from site j to i and then back. At first sight, both processes look identical and indeed both contribute equally to the isotropic Heisenberg exchange equation (15). But due to the presence of the spin-orbit interaction, the reflection asymmetric environment and the lack of time inversion symmetry, the interference of both processes does not cancel out completely. Instead, it gives rise to an additional antisymmetric exchange interaction between these sites, known as the Dzyaloshinsky-Moriya (DM) [38, 39] interaction

$$H_{\rm DM} = \sum_{i,j} \vec{D}_{ij} \cdot \left(\vec{S}_i \times \vec{S}_j \right) , \tag{37}$$

where $\vec{D}$ is a constant vector, which depends on the symmetry of the system and on the real space direction given by two sites i and j. For example, for typical (100) and (110) low index surfaces of metals, $\vec{D}$ lies in the film plane and points perpendicular to the direction (i, j) connecting two surface atoms, if the two surface atoms are placed along high a symmetry line. The DM interaction arises as the first-order perturbation in the spin-orbit interaction, and might for this reason be stronger than the magnetocrystalline anisotropy. This chiral interaction tends to orient the spin S_i and S_j orthogonal to each other and to $\vec{D}$, destabilizing a uniform ferro- or antiferromagnetic order and can cause, depending on the strength D, a canting of the magnetization at different atoms, a helical or cycloidal spin-wave. The sign of $\vec{D}$ defines the chirality

of the canting. The DM interaction is practically unknown in metallic bulk magnets, since most metals crystallize in structures with centro-symmetric symmetries. Surprisingly, after 20 years of research on low-dimensional magnetism, the magnitude of $\vec{D}$ has not been established so far and there is currently active research going on to clarify its relevance for the magnetic order in nanomagnets.

3 The Rashba effect at metallic surfaces

3.1 Nonmagnetic Surfaces

A surface state can be considered as a particular realization of a two dimensional electron gas. Since the surface always breaks spatial inversion symmetry, the effective potential which acts on the surface state will generally have a finite gradient along the surface normal, i.e. there is an electric field in this direction. The physical manifestation of this field is the workfunction. Like in the semiconductor heterostructures discussed in the lecture of Dr. Bringer, due to the

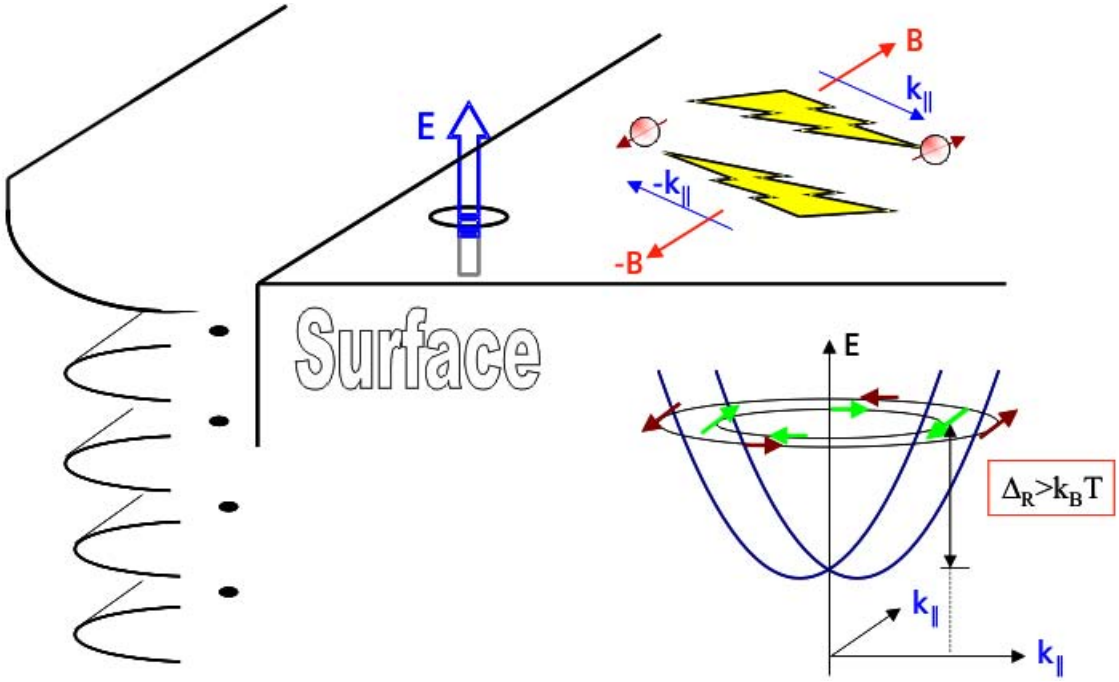


Fig. 6: *Electrons moving with in-plane wavevectors* $\mathbf{k}_{\|}$ *and* $-\mathbf{k}_{\|}$ *in the potential gradient of a surface (indicated on the left). The resulting electric field,* $\mathbf{E}$*, is seen in the restframe of the moving electrons as a magnetic field* $\mathbf{B}$ *or* $-\mathbf{B}$ *which couples to the spin of the electron. This interaction modifies the bandstructure of a spin-degenerate* (s, p_z)*-like surface state as shown in the lower right picture. The degeneracy is lifted and the electrons at the Fermi level have spin directions perpendicular to their propagation directions as indicated by the small arrows (cf. also preceding section).*

movement of an electron with wavevector $\mathbf{k}_{\|}$ in the surface plane, this electric field appears in the rest frame of the electron as a magnetic field which couples to the spin of the electron. The situation is schematically depicted in Figure 6 for two electrons traveling on the surface in

opposite directions. For a non-spinpolarized surface state (e.g. on a nonmagnetic surface) this gives rise to a term in the Hamiltonian

$$H_{\rm soc} = \frac{\hbar}{4m_{\rm e}^2c^2}\boldsymbol{\sigma}\cdot(\nabla V(r)\times\mathbf{p}) \tag{38}$$

which leads to a k-dependent splitting of the dispersion curves. When we simply use the nearly free electron gas (NFEG) model and substitute $\mathbf{p}$ by the $\mathbf{k}$-vector, for usual workfunctions we would expect this splitting to be very small, in the order of 10^{-6} eV. This would be far too small to observe directly with angle resolved photoemission spectroscopy (ARPES). So it came rather as a surprise, when in 1996 LaShell and coworkers [40] discovered a splitting of the surface state of the Au(111) surface, which was not only k-dependent, but also in the order of 0.1 eV at the Fermi level. They correctly interpreted this splitting as a spin-orbit coupling effect, which obviously was influenced by the strong atomic spin-orbit effects in the heavy Au atom. Spin resolved ARPES experiments finally also analysed the spin distribution of this surface state [41] and found it to be in quite good agreement with the NFEG model (cf. Figure 6), as it was also predicted theoretically [42].

While this effect was observed in different studies for the Au(111) surface, on other surfaces which show a similar Shockley state, e.g. Ag(111) or Cu(111), no such splitting was discovered experimentally and in calculations [43] based on density functional theory (DFT). From the calculations it was concluded, that the k-dependent splitting on Ag(111) is by a factor 20 smaller than on Au(111) (cf. also Figure 7). This can neither be explained by the difference in atomic spin-orbit coupling of Au ($Z = 79$) and Ag ($Z = 47$) alone, nor by the potential gradients at the surface. Also the amount of p-character in the sp-surface state is larger for Ag than for Au, so that in principle spin-orbit effects should be more prominent in silver. So what is responsible for the size of the effect?

To resolve this issue, we did calculations based on DFT with the full potential linearized augmented planewave method [44] as implemented in the FLEUR code [45]. Our calculations include spin-orbit coupling (SOC) in a self-consistent manner [46] in the muffin-tin (MT) spheres. For the present discussion it might be interesting to note, that actually only the spherically symmetric part of the potential is included in the calculations, which might seem inconsistent with the above discussion which claims that the potential gradients at the surface are responsible for the effect we want to describe. But we will see, that in all considered cases the agreement with experimental data is fine, suggesting that the theoretical approach includes the dominant terms leading to the Rashba-type splitting in question.

In the calculations we can choose the region where to include SOC: in specific spheres around the atoms, i.e. in certain layers of the film, or we can also vary the size of the sphere, where we want to include spin-orbit coupling. In this way, it is possible to show that a bit less 60% of the k-dependent splitting of the Au(111) surface state comes from the surface layer and the contribution in deeper layers decays more or less like the weight of the surface state in these layers [47]. Moreover, this effect is extremely localized in the core region, where the *radial* potential gradient is largest. For Au(111), more that 90% of the effect originate from a sphere with radius 0.25 a.u. around the nucleus. In this region the potential is almost perfectly spherically symmetric, so that our above mentioned approximation, to include only the $l = 0$ part of the potential, is probably well justified. The potential gradient at the surface enters actually only indirectly, via the asymmetry of the wavefunction in the core region. In a tight-binding model, Petersen and Hedegård showed that the size of the Rashba-type splitting is determined by the product of the atomic spin-orbit coupling parameter and a measure for the

asymmetry of the wavefunction under consideration [48].
A measure for the asymmetry of the wavefunction of a surface state can be found by analysing the l-like character of the state, i.e. to determine how much s, p or d character a surface state shows at a certain $\mathbf{k}_{\|}$-point, in our case the $\overline{\Gamma}$-point. E.g. a surface state of pure p_z character is inversion symmetric and will – in absence of an electric field – show no Rashba-type splitting. The potential gradient or electric field at the surface will distort the wavefunction, so that some s or d_{z^2} contributions to the surface state will arise. The ratio of l- to $l \pm 1$-type character of a surface state (for a given m, e.g. $m = 0$) will therefore give a measure for the asymmetry of this state. In the case of Ag(111), we find that the surface state is predominantly of p_z-type, with a small d_{z^2} admixture ($p : d$ ratio of 9.5) while in Au the surface state has much stronger d_{z^2} character ($p : d$ = 3.3). The fact that the Au d band lies much higher in energy than the Ag d band leads to a stronger d character of the Au surface state and thus to larger asymmetry of the wavefunction [47]. This determines the stronger k-dependent splitting in the Au(111) surface bandstructure.

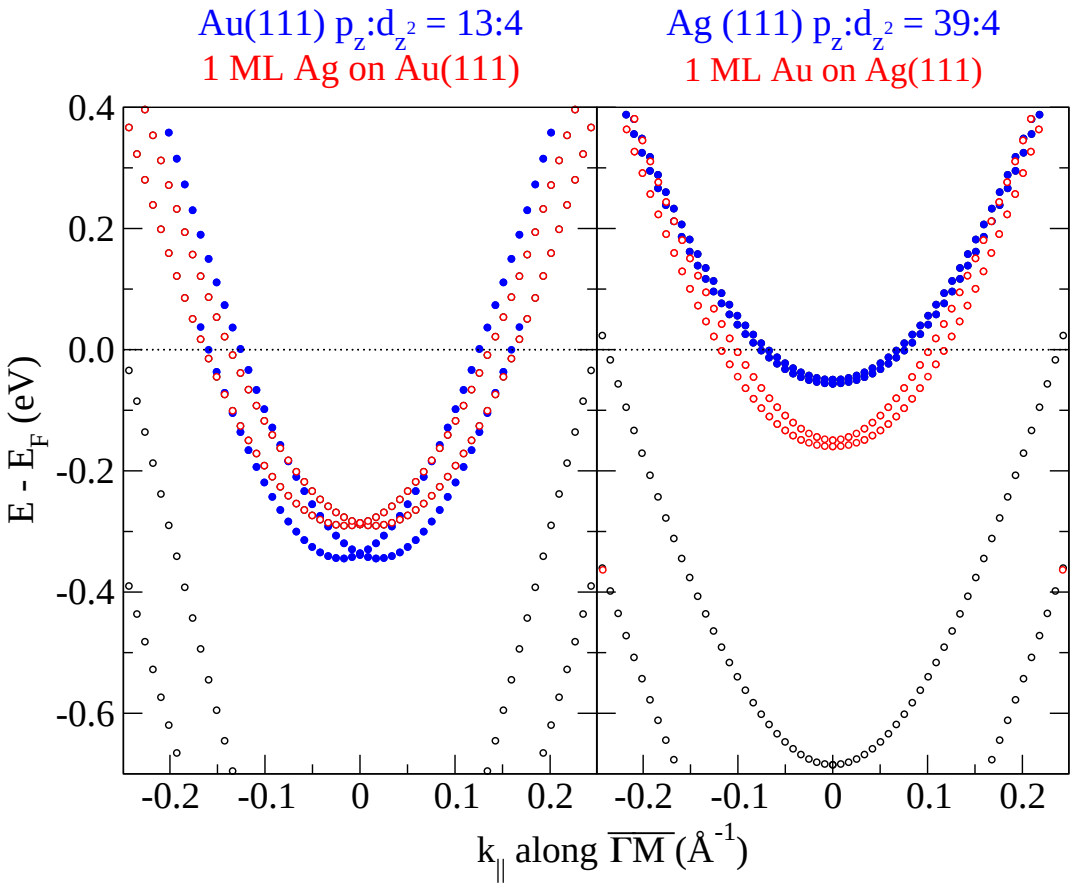


Fig. 7: *Bandstructure of Au(111) (left, full circles) and one monolayer Ag on Au(111) (left, open circles) as compared to a Ag(111) film (right, full symbols) and a single Au monolayer on Ag(111) (right, open symbols).*

This effect can be further demonstrated, when we compare a single monolayer of Ag on Au(111) with a Au monolayer on Ag(111). Just from the point of view of the atomic SOC, we would expect that the Rashba-type splitting of the Au monolayer of Ag(111) is larger than that of the Ag/Au(111) system, since more than 50% of the effect comes from the surface layer. But since

the gold d states of the subsurface layers can induce a larger d character of the Ag surface state in Ag/Au(111) while the Au surface state of Au/Ag(111) has less d character than the one of pure Au(111), finally the Rashba-type spin-orbit splitting is larger in Ag/Au(111) (cf. Figure 7). Other examples, how the asymmetry of a surface state influences the strength of the Rashba-type splitting can be found on lanthanide surfaces (e.g. Lu(0001) [47]), and a particular case will be presented in subsection 3.3.

3.2 Semimetal Surfaces

Up to now we have discussed examples, where the Rashba-type spin-orbit splitting was in the order of 10 to 100 meV (up to 120 meV for Au(111)), so that experimentally it is not so easy to detect in ARPES experiments. Now, we turn to another extreme, where the splitting is so big, that it was not a-priori clear, whether the two experimentally observed features were spin-split partners of the same state or two different surface states: the low-index surfaces of Bi. Bismuth is a non-magnetic, rather heavy metal ($Z = 83$) with semimetallic properties, i.e. the Fermi surface consists only of two tiny pockets, so that the density of states (DOS) at the Fermi level ($E_{\rm F}$) is almost zero. In the surface projected bulk-bandstructure extended gaps are observed around $E_{\rm F}$, in which surface states can be localized.
ARPES measurements on the Bi(110) surface [49] showed the existence of two spectroscopic features in the gap, which could be interpreted as to two surface states. Bismuth has a rhombohedral crystal structure and the (110) surface consists of unreconstructed pseudocubic bilayers [50], where dangling bonds can give rise to surface states. Similarly, on Bi(111) two states were identified spectroscopically [51]. The (111) surface has closed-packed layers and again shows a bilayer structure, but without dangling bonds and with a much larger separation of the bilayers [52]. In both cases of course only the occupied part of the surface bandstructure could be observed spectroscopically. Using DFT calculations, we have the possibility to access also the unoccupied part of the spectrum. It can be seen that the observed spectroscopic features are actually a Rashba-type spin-split pair of a surface state which forms – at least for the (110) and (100) surface – a band through the whole surface Brillouin zone [53, 54]. That these surface state is actually split by spin-orbit coupling can be demonstrated by comparison of a scalar-relativistic calculation without inclusion of SOC and with the inclusion of SOC [55] (cf. also left of Figure 8). In this cases, the splittings are very large (in the order of 300 meV) and, since the surface states extend throughout the Brillouin zone, they are also no longer linear in k, except in the vicinity of high symmetry points.
It is not only of academic interest, whether two surface states are a spin-split pair or two spin-degenerate surface states. For example, on the Bi(111) surface the Fermi surface forms a small hexagon around the $\overline{\Gamma}$ point, which led to speculations about the formation of a charge density wave on this surface [56]. If the Fermi surface were indeed formed by spin-degenerate surface states, this would be possible. If, on the other hand, Rashba-type spin-split bands form this part of the Fermi surface, the electrons at $+\mathbf{k}_{\|}$ and $-\mathbf{k}_{\|}$ were of opposite spin and instead of a peak in the (spin) diagonal part of the susceptibility χ, we would expect a large contribution to the spin off-diagonal part, $\chi^{\pm}$, leading to a modulation of the spin-structure. Since the surface is of course still nonmagnetic, these modulations have to cancel and a direct observation is difficult. When magnetic atoms were present at the surface, their interaction would be modified and this effect could be detected. We will show in a later chapter, that this is actually possible.
Using scanning tunneling microscopy (STM) techniques, consequences of the spin polarization of the surface states have indeed been observed for another Bi surface [53]. If a scanning

tunneling spectrum (STS) is recorded for a dense mesh of positions on a surface, this STS map can be Fourier transformed for a given energy within this spectrum. The Fourier transformed (FT) STS map gives then a picture of the energy dispersion in reciprocal space, i.e. a two dimensional cut through the function $\varepsilon(\mathbf{k}_{\|})$, but with doubled length of the $\mathbf{k}$-vectors, since the STS maps the scattering between two states of different $\mathbf{k}$ but at the same E. In particular, for $E = E_{\mathrm{F}}$, this yields an image of the Fermi surface. It is easily seen, that a surface state with a Fermi surface of a wavevector $\pm\mathbf{k}_{\mathrm{F}}$ will give rise to standing waves with $2\mathbf{k}_{\mathrm{F}}$ which can be seen in the STS map. This correspondence between FT-STS and Fermi surface has been used extensively to study the electronic properties of high-temperature superconductors. A Rashba splitting will not change this picture, since for one spin channel the Fermi vectors are changed to $\pm\mathbf{k}_{\mathrm{F}}+\Delta\mathbf{k}$, while for the other spin we get $\pm\mathbf{k}_{\mathrm{F}}-\Delta\mathbf{k}$, so that both spin channels will lead to a contribution of $\pm 2\mathbf{k}_{\mathrm{F}}$ in the STS map, i.e. the picture is indistinguishable from the one without Rashba splitting [48]. But if the Fermi surface is more complex, like in the case of Bi(110), the fact, that the surface states are spin polarized can be seen the FT-STS clearly.

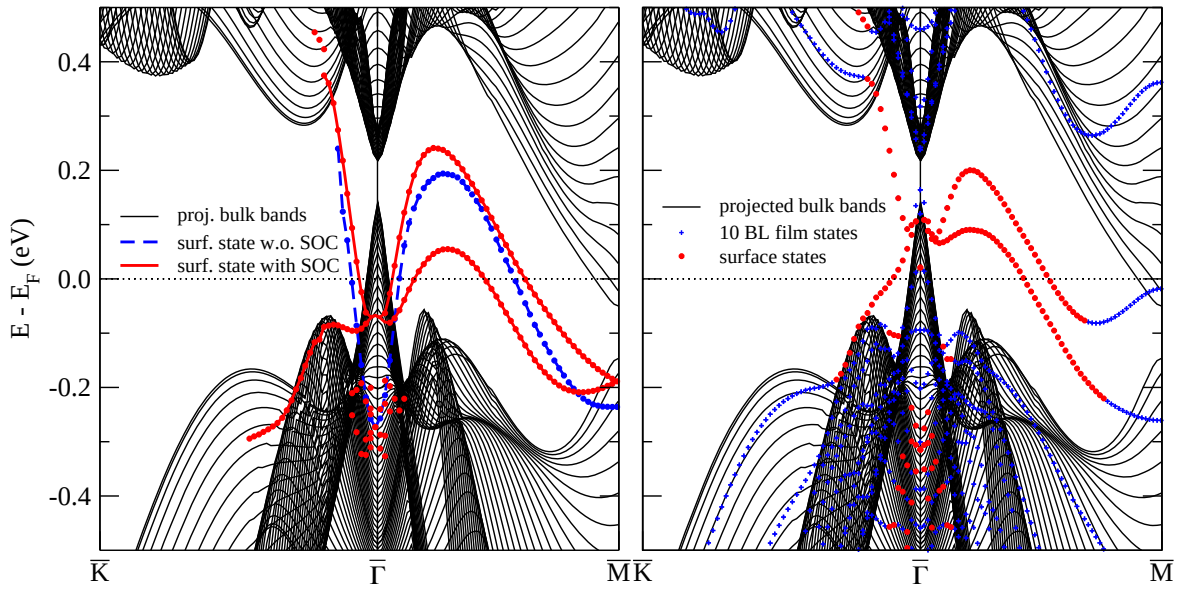


Fig. 8: *Bulk-projected bandstructure of Bi and surface bandstructure of a 22 layer Bi(111) film with H termination on one side (left) with (full line) and without (broken line) spin-orbit coupling included in the calculation. A similar calculation with SOC included for a symmetric 20 layer film without H termination is shown on the right.*

Consider a simple one-dimensional example: along the line $\overline{\Gamma\mathrm{M}}$ in Bi(111) we can see in Figure 8 (left, broken line) a surface state obtained without inclusion of SOC. This state originates at -0.3 eV at $\overline{\Gamma}$, crosses the Fermi level at a wavevector we denote k_a, disperses down again and crosses E_{F} once more at k_b and reaches $\overline{\mathrm{M}}$ at -0.22 eV. Surface states at E_{F} can scatter between k_a and k_b and give rise to standing waves with wavelength $2k_a, 2k_b, (k_a + k_b)$ and $(k_a - k_b)$, if the state is spin-degenerate. Now, consider that spin-orbit coupling splits this degeneracy and gives rise to spin-up states at $k_a + \Delta k$ and $k_b - \Delta k$, while spin-down states cross the Fermi level at $k_a - \Delta k$ and $k_b + \Delta k$. In this case, spin conserving scattering events will again give rise to oscillations with wavelength $2k_a, 2k_b$, but also $(k_a + k_b) \pm 2\Delta k$ and $(k_a - k_b) \pm 2\Delta k$. Here, the effect of spin is clearly visible. On the Bi(110) surface, this effect was also verified

experimentally in a two dimensional case [53].

The occurrence of spin-polarized surface states of course suggests, that this could be utilized in some way for spintronic applications. In the case of Ag(111), where the surface state contributes very little to the density of states at the Fermi level, this might not seem very promising, but in the case of a semimetal surface, where the DOS at $E_{\rm F}$ originates almost exclusively from surface states, this might be more realistic. Alternatively, the surface of a thin film on insulating or semiconducting substrates could be interesting, since in this case the relative contribution of the surface state to the conduction electrons is also increased. This works of course only, if the thin film still supports the same surface state as the semiinfinite crystal, i.e. localized Tamm states of d-character as they occur on lanthanide (0001) surfaces will be more suitable for very thin films than the extended s,p-derived Shockley states of the closed packed coinage metal surfaces.

Another effect, that can disturb the surface states in thin films, is the interaction between the two surfaces of the film. If, like in Bi, the screening is very weak, surface states at the upper and lower surface of a symmetric film interact to form even and odd linear combinations. This of course interferes with the concept of broken inversion symmetry at the surface. On the other hand, in our theoretical calculations for Au and Ag surfaces, we always used symmetrical films where a tiny interaction between upper and lower surface cannot be avoided, even in thicker films. For the bandstructures of Figure 7 we used 23 layer films and especially in the case of Ag(111), a finite splitting of the surface state parabolas at the $\overline{\Gamma}$ point can be seen. At the first glance it might seem surprising, that the two different splittings, the even-odd and the Rashba-type splitting result in only two dispersion curves. Without the interactions that lead to the splittings, we can think of having two states (spin up, $\uparrow$ and down, $\downarrow$) on each surface. The spin-orbit coupling leads for the spin up states of the upper surface ($\uparrow_u$) to the same shift in energy as for the spin down states of the lower surface ($\downarrow_l$) (since the potential gradient is reversed there) and they will have an energy ε^+. In the same way of course $\varepsilon(\downarrow_u) = \varepsilon(\uparrow_d) = \varepsilon^-$. A hybridization of $\uparrow_u$ and $\downarrow_u$ leads to energies $\varepsilon^+ + \varepsilon^s$ and $\varepsilon^- - \varepsilon^s$, respectively, but in the same way the two downspin states, $\downarrow_u$ and $\downarrow_l$ will be shifted to energy values $\varepsilon^- - \varepsilon^s$ and $\varepsilon^+ + \varepsilon^s$. The stronger the interaction across the film, the more each state will be localized at both sides of the film so that finally the spin-polarization for a given energy and $\mathbf{k}_\parallel$ gets reduced.

A case, where this scenario has been actually observed in experiment are thin Bi films grown on a Si substrate [57]. The interaction with the substrate is very weak, since the Bi film is deposited on a seeding layer of Bi atoms and can adopt (for more than a few bilayers) the structure of Bi(111). Angle resolved photoemission has shown that near the zone center the electronic structure of these Bi films is not so different from what has been observed on single crystal surfaces. But when the $\mathbf{k}_\parallel$ vector approaches the zone boundary at $\overline{\rm M}$, the crossing of the two spin-split states is no longer observed. Instead, quantum well states (QWS) are formed when the surface state gets near to the bulk continuum at $\overline{\rm M}$ [57]. The energy levels of these states agree nicely with those obtained by the calculation of symmetric films of the same thickness (cf. right of Figure 8). As the surface state character is lost, also the spin-polarization of these states vanishes. The very bad screening of Bi makes this QWS disappear only for very thick films (more than 40 bilayers). Therefore, when we simulate Bi single crystal surfaces, we have to terminate one side of the film with H atoms to saturate the dangling bonds and explicitly remove the inversion symmetry of the film, even if it is 22 layers thick.

3.3 Magnetic Surfaces

Let us finally consider the case of a surface of a magnetic metal, like Gd(0001). On this closed packed surface a bulk projected bandgap around $\overline{\Gamma}$ contains a surface state of d_{z^2} character, like it can be found also on other lanthanide surfaces. Exchange interaction splits this surface state into an occupied majority spin state and an unoccupied minority state. This splitting is mainly controlled by the $4f$ electrons of Gd and amounts to about 0.8 eV, which is large as compared to spin-orbit effects in this system. No matter how SOC affects the electrons of the surface state, their spin will remain more or less parallel to the exchange field, which is oriented in plane in the directions of nearest neighbor atoms by the magnetic anisotropy.
An electron traveling on the surface in a direction *perpendicular* to its spin quantization axis, will experience the potential gradient at the surface as a magnetic field parallel to its spin. Therefore, a magnetic coupling can arise and the dispersion curves will split more or less similar to what is observed on a nonmagnetic surface. If, on the other hand, the propagation direction of the electron is parallel to its spin quantization axis, the field arising from SOC cannot couple to the electron's spin and no Rashba-like splitting can be observed. Schematically, this situation is shown in Figure 9. In contrast to the surface state on the nonmagnetic surface, where the spin of the electron is always oriented perpendicular to the propagation direction and the surface normal, $\mathbf{e}_z$, (with some deviation, depending on the shape of the potential [42]), on

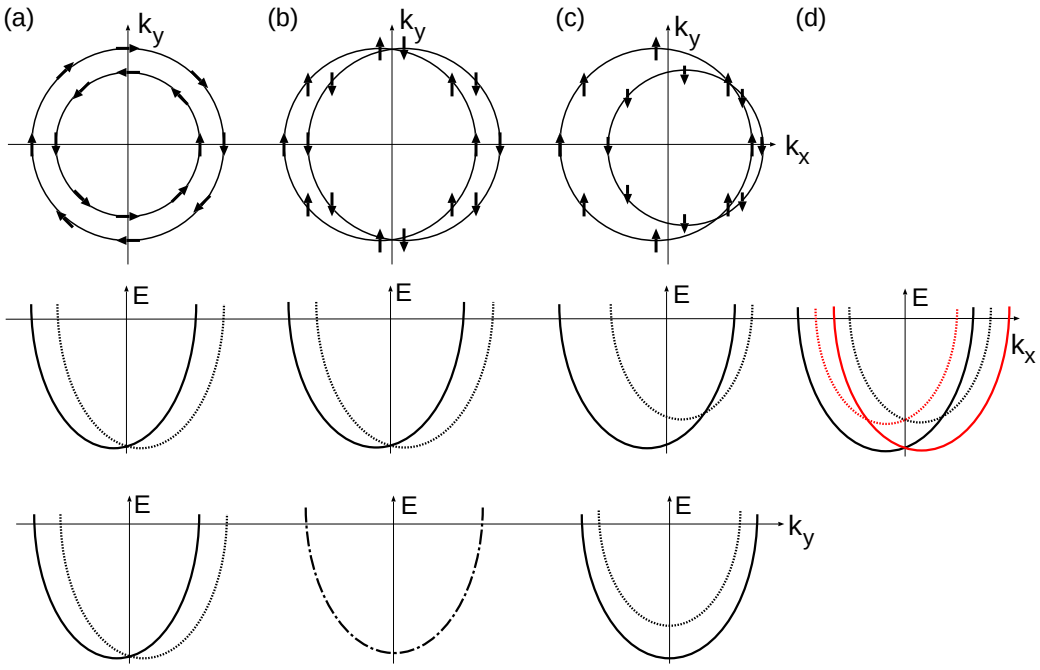


Fig. 9: *(a): Rashba splitting on a non-magnetic surface: the top panel shows the Fermi surface and the spin-polarization of the states at the Fermi level. In the middle and lower panel the bandstructures along two orthogonal directions in reciprocal space are shown. (b): The same relations as in (a) are shown, but now for the case where the spin-quantization axis of the electrons has been aligned in a particular direction. For electrons propagating in this direction, the Rashba splitting vanishes. (c): Effect of an additional exchange splitting on the situation as described in (b). (d): Surface bandstructure arising from the calculation of a symmetric film, where the surface states from the lower surface are shown in red.*

the spin-polarized surface, the spins are more or less collinear. This changes the shape of the Fermi surface significantly, especially if exchange splitting is considered (Figure 9 (c)). If the exchange splitting is large, this leads to a Fermi surface consisting of a single circle shifted away from the zone center. The consequences for the bandstructure are simple: along a certain direction in reciprocal space SOC will have no particular effect. In a direction orthogonal to this one, the dispersion curves for majority and minority spin will be shifted in opposite directions. For the eigenvalues this results in an expression

$$\varepsilon_{\downarrow(\uparrow)}(\mathbf{k}) = \varepsilon(\mathbf{k}) \pm IM \pm \alpha_{\mathrm{R}}(\mathbf{k} \times \mathbf{e}_z) \cdot \hat{\mathbf{M}} \tag{39}$$

where $\mathbf{M} = M\hat{\mathbf{M}}$ is the magnetization and IM represents the exchange splitting of the bands. Of course in a calculation of a symmetric film, again on the lower surface ($-\mathbf{e}_z$) the directions of the spin-orbit induced shifts will be exactly opposite to the shifts on the upper surface, so that in total a picture as shown in Figure 9 (d) is obtained. From this picture the two splittings for the spin-up and the spin-down surface state can be determined directly as $\Delta\varepsilon(\mathbf{k}) = \varepsilon(\mathbf{k}, \mathbf{M}) - \varepsilon(\mathbf{k}, -\mathbf{M})$.

Experimentally, for a single crystal surface, it is possible to measure with ARPES two spectra of the same surface, but rotated by 180 degrees. In the case of an in-plane anisotropy, as for Gd(0001), this rotation reverses the spin and leads, therefore, also to a picture as Figure 9 (d). A comparison of these two spectra allowed to determine the Rashba splitting in Gd(0001), even though its magnitude is rather small [58]. A particular advantage of magnetic surfaces is, that the measurements allow the determination of the sign of the Rashba parameter, α_R, even without the need of spin-analysis via a Mott detector.

Modification of the Gd surface also alters the characteristics of the surface state: if (atomic) oxygen is adsorbed in the surface, the surface state shifts down in energy and both, minority and majority spin states become occupied. Moreover, the dispersion of the surface state changes from almost flat on Gd(0001) to parabolic for O/Gd(0001). Both surface states were observed experimentally, and DFT calculations show, that these states are actually interface states residing between the topmost Gd/O layer and the underlying Gd bulk [58].

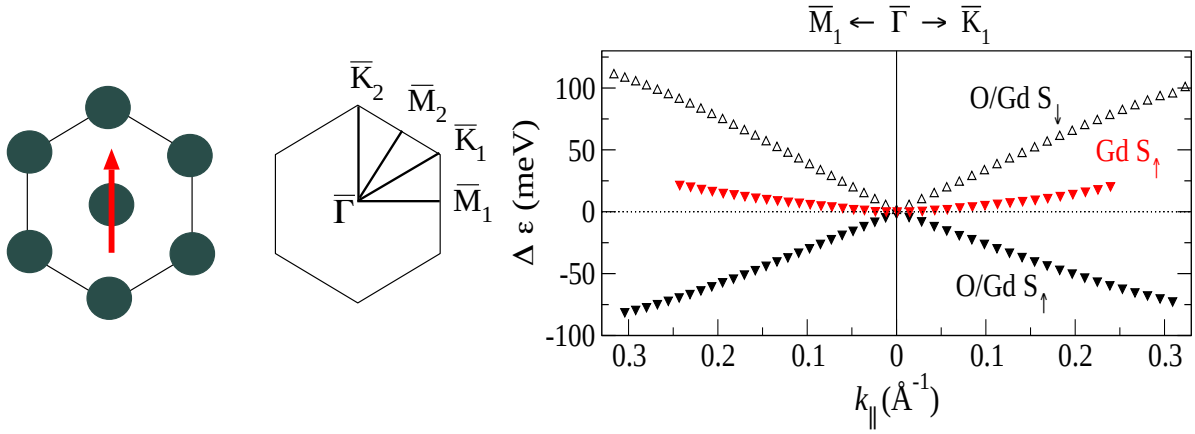


Fig. 10: *Left: Magnetization direction on the Gd(0001) surface indicated by the red arrow and surface Brillouin zone and labeling of the high symmetry points. Right: Rashba-type splitting of the surface state of Gd(0001) and O/Gd(0001) in the directions* $\overline{\Gamma\mathrm{K}_1}$ *and* $\overline{\Gamma\mathrm{M}_1}$.

Let us now focus on the Rashba-type splitting on O/Gd(0001). In Figure 10 we plotted the splitting as a function of $k_{\|}$ for the directions $\overline{\Gamma \mathrm{M}_1}$ and $\overline{\Gamma \mathrm{K}_1}$ as indicated in the Figure. Since the magnetization is directed towards nearest neighbours (a-direction), the largest effects on the surface state dispersion should be observed in $\overline{\Gamma \mathrm{M}_1}$ direction while no splitting should be visible in the direction $\overline{\Gamma \mathrm{K}_2}$. A closer look at Figure 10 reveals, that the splitting, $\Delta\varepsilon$ is indeed smaller in $\overline{\Gamma \mathrm{K}_1}$ than in $\overline{\Gamma \mathrm{M}_1}$. Furthermore, we observe that $\Delta\varepsilon$ for the majority spin state ($\mathrm{S}_\uparrow$) is not only of opposite sign as compared to $\Delta\varepsilon$ for the minority state ($\mathrm{S}_\downarrow$), but also their absolute values differ. This is also observed experimentally, and can be explained again by the different positions of the states in the bulk-projected bandgap and the different asymmetry of the wavefunctions. One should note here, that also the effective masses of the $\mathrm{S}_\uparrow$ and $\mathrm{S}_\downarrow$ states differ and this shows, that spin is not the only difference of these states.
Even more drastic is the difference of the $\mathrm{S}_\uparrow$ surface state of O/Gd(0001) to the $\mathrm{S}_\uparrow$ state of Gd(0001). We can see from the right of Figure 10 that not only the magnitude of the splitting is a factor 3 to 4 smaller, even the sign is different. Since this reversal of sign cannot be attributed to the spin, it must result from a different admixture of p_z-character to the d_{z^2} surface state. A reversal of the gradient of the wavefunction at the position of the Gd nucleus can be interpreted as the result of hybridization with p_z-type wavefunctions of different signs. In some sense we can say, that we see the sign of the wavefunction here in the sign of the Rashba-parameter.
Spin-orbit coupling effects on surfaces can be very diverse and recently this field expanded in various directions: Studies of the spin-orbit splitting of surface states in Bi or Pb surface alloys have shown more complex spin-orientation patterns than what is expected from the Rashba-model [59]. In thin films of Bi on Si(111) substrates, a gradual decrease of the Rashba-type spin splitting has been observed as the orbitals change their character from surface states to quantum well states as a function of momentum [60]. On the other hand, in Pb films on Si(111) even quantum well states can be spin split due to their stronger interaction with the substrate [61]. Thin Bi films turn out to be quite attractive, since Bi can undergo a transition from its semimetallic to a semiconducting state at low thickness [62]. Edge states of these semiconducting films can provide a non-trivial band topology that can support the so called quantum spin-Hall effect. A two-dimensional variant of this effect has been recently confirmed on BiSb alloy surfaces [63].

4 Ultrathin Films

The transition-metal monolayers on noble-metal substrates are the classical systems exhibiting two-dimensional (2D) magnetism. Because of the reduced coordination number of nearest neighbor atoms in a monolayer film the d-band width in two-dimensions is considerably smaller and correspondingly the LDOS at the Fermi energy is considerably larger than in the bulk situation. Thus the magnetic instability should occur for a much wider variety of transition-metal elements. Following this line of argument it is clear that the strength of the d–d hybridization between monolayer and substrate is an additional parameter which controls the d-band width of the monolayer. For instance large band-gap material, e.g. MgO(100), as substrate allows the formation of two-dimensional monolayer bands within the band gap of the substrate material. In this case the impact on the magnetization of the monolayer due to the substrate is expected to be small. The same is true for noble-metal substrates, which have d bands well below the Fermi energy. The width of the monolayer d band is not significantly broadened by the monolayer-substrate d–d interaction, and magnetism is restricted to the monolayer. Increasing the d–d

Table 4: *Local magnetic moments in μ_B/atom for $3d$ transition–metal atoms as ferromagnetic (F) and antiferromagnetic (AF) $3d$ monolayers (ML) on Ag(001) [64], Pd(001) [65], W(110) and on Cu(001) [2, 66]; compared with results for $3d$ monolayers as interlayers (IL) in Cu(001) [2], unsupported (001) monolayers (UL) in the lattice constant of Cu(111) and Ag(001) [67], and with results for ferromagnetic $3d$ monolayers on Cu(111) and Ag(111) [68]. "–" indicates that no calculation was performed for this system. "0" indicates that the calculated moment was smaller than the numerical accuracy estimated to be about 0.02 μ_B/atom. "?" indicates a system, for which the calculation was not finished up to self-consistency, but result is approximately correct.*

				Ti	V	Cr	Mn	Fe	Co	Ni
Ag	ML	on Ag(001)	F	0.34	2.09	3.78	4.04	3.01	2.03	0.65
			AF	0	2.08	3.57	4.11	3.06	?	0
	UL	– Ag(001)	F	1.72	2.87	4.50	4.32	3.29	2.20	1.02
			AF	0	2.59	4.09	4.32	3.32	2.10	0
	ML	on Ag(111)	F	0	1.39	3.43	3.91	2.95	1.93	0.51
Pd	ML	on Pd(001)	F	0	0.51	3.87	4.11	3.19	2.12	0.89
			AF	0	1.39	3.46	4.05	3.20	1.99	0.59
W	ML	on W(001)	F	–	0.00	–	2.97	2.37	1.14	0.00
			AF	–	0.00	2.52	3.32	–	–	0.00
Cu	ML	on Cu(001)	F	–	0	0	2.97	2.61	1.76	0.33
			AF	–	0	2.52	2.92	2.35	?	0
	IL	in Cu(001)	F	–	0	0	2.01	2.39	1.51	0
			AF	–	0	1.84	2.15	–	–	–
	ML	on Cu(111)	F	–	0	0	3.05	2.69	–	
	UL	– Cu(111)	F	–	0	0	3.06	2.75	–	–

hybridization by choosing appropriate nonmagnetic transition metal substrates, e.g. Pd(100) or W(110), will lead to a considerable broadening of the monolayer bands and introduce a significant spin-polarization of the substrate until we have changed from the two-dimensional limit to the semi–infinite regime. Choosing a magnetic substrate an additional complexity arises due to the competition of the magnetic coupling in the monolayer and between monolayer and substrate.

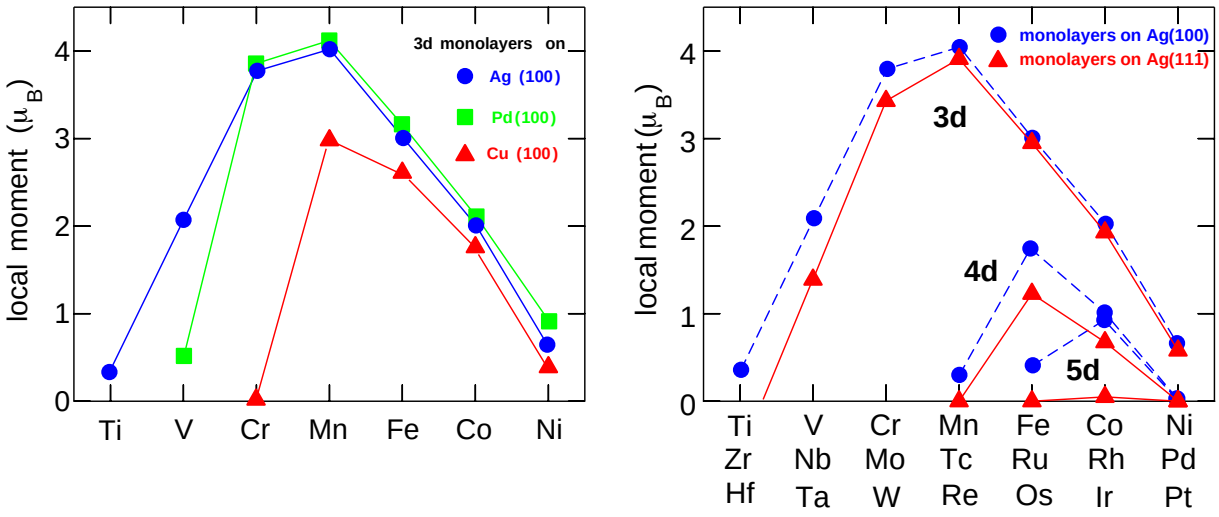


Fig. 11: *Local magnetic moments as calculated for ferromagnetic (left figure) $3d$ metal monolayers on Ag(100) [64] (dots), Pd(100) [65] (squares), and Cu(001) [2] (triangles), and (right figure) $3d$, $4d$ [69], and $5d$ [70] monolayers on Ag(001) (dots) and Ag(111) [68] (triangles)*

4.1 (100) Oriented Monolayers on Nonmagnetic Substrates

Ferromagnetic Monolayers

A systematic investigation of the magnetism of all possible $3d$, $4d$, and $5d$ transition-metals monolayers on Ag(001) are collected in Fig. 11 and in Table 4. One finds that all $3d$ metal monolayers (Ti, V, Cr, Mn, Fe, Co, Ni) on Ag(001) substrate show ferromagnetic solutions. Tc, Ru, and Rh are ferromagnetic among the $4d$-metals, and Os and Ir are ferromagnetic among the $5d$-metals on Ag(001). The local magnetic moments are partly very large, not only for the $3d$ monolayers, but surprisingly also for the $4d$ and $5d$ ones. In the $3d$ series the overall trend of the local moments follows Hund's first rule. The largest local moment of about 4 μ_B was found for Mn and from Mn to Ni the magnetic moment decreases in steps of 1 μ_B. The latter is a consequence of the strong ferromagnetism in these monolayers. The magnetic moments of Ti, V, and Cr monolayers show a pronounced dependence on the substrate: Ti is magnetic on Ag, but nonmagnetic on Pd; the magnetic moment of V is reduced by more than 1.5 μ_B when changing the substrate from Ag to Pd; and for Cr the magnetic moment changes from 3.8 μ_B as an adlayer on Ag or Pd to zero as an adlayer on Cu. Although not as dramatic, the reduction is also visible for Mn. We attribute the drastic reductions of the monolayer moments to the reduction of the lattice constants in the sequence Ag to Pd to Cu.

When comparing the results of the local moments between $3d$, $4d$, and $5d$ monolayers on Ag(001) an interesting trend is observed: The element with the largest magnetic moment among each transition metal series is shifted from Mn to Ru (isoelectronic to Fe) and at last to Ir (isoelectronic to Co), respectively. Following these trends we do not expect ferromagnetism for any other $4d$ or $5d$ metal on noble metal (001) substrates, and indeed Mo and Re remained nonmagnetic. The overall picture of monolayers on Ag and Au is the same, but the different substrate interactions cause Tc and Os on Au to be nonmagnetic and lead to a slightly larger moment for Rh. Pd and Pt are predicted to be nonmagnetic. With the exception of Ru, for which a rather small magnetic moment of 0.2μ_B was calculated, no monolayer magnetism was found for $4d$ metals on Pd(100). Investigations [71] including the spin-orbit interaction have shown that the spin-orbit interactions reduces significantly the magnetic spin moment of the $5d$ metal

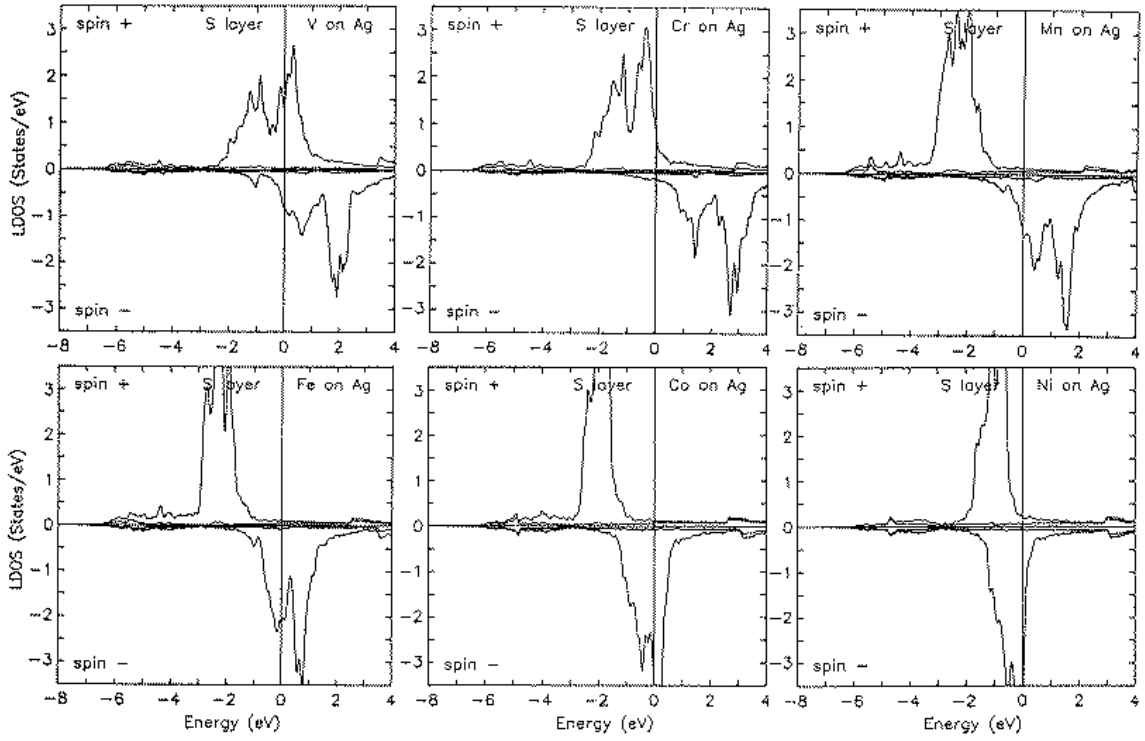


Fig. 12: *Local density of states (LDOS) of ferromagnetic $3d$ metal monolayers on Ag(100). The Fermi energy defines the origin of the energy scale, separating occupied (at negative energies) from unoccupied states (at positive energies). Majority (minority) states are indicated by positive (negative) values of LDOS.*

monolayers and depending on the interlayer relaxation the spin moment might be suppressed.

Antiferromagnetic Monolayers

It is by no means clear whether the ferromagnetic state is actually the magnetic ground state. Looking at the LDOS of the $3d$ monolayers in Fig. 12 and considering the analysis of the antiferromagnetic susceptibility (8) we expect an antiferromagnetic phase for Cr and possibly also for V and Mn monolayers. In reality, various antiferromagnetic states as well as non-collinear spin configurations could be anticipated. Studying an Heisenberg model (15) for a square lattice as formed by the (001) monolayers up to the second nearest-neighbor interaction (J_1, J_2) the situation becomes relatively simple. As long as the nearest-neighbor interaction is the dominating one, there are only two phases to be considered: the ferromagnetic p(1×1) structure ($J_1 > 0$) discussed in the previous section and the antiferromagnetic c(2×2) superstructure ($J_1 < 0$, a checkerboard arrangement of up and down spins with moments of identical size on both sublattices). The c(2×2) structure corresponds to the $\overline{\mathrm{M}}$-point in the 2DBZ of the square lattice. If the next-nearest neighbor interaction is antiferromagnetic, $J_2 < 0$, and sufficiently strong, $|J_1| < 2|J_2|$, then the magnetic structure with a 2D $\vec{Q}_{\parallel}$ vector of the $\overline{\mathrm{X}}$-point in the 2DBZ, corresponding an antiferromagnetic p(2×1) or p(1×2) structure (ferromagnetic rows of atoms along the [100] or [010] direction coupling antiferromagnetically from row to row) becomes the magnetic ground state.

Figure 13 shows the local moments for the ferromagnetic and c(2×2) antiferromagnetic phase

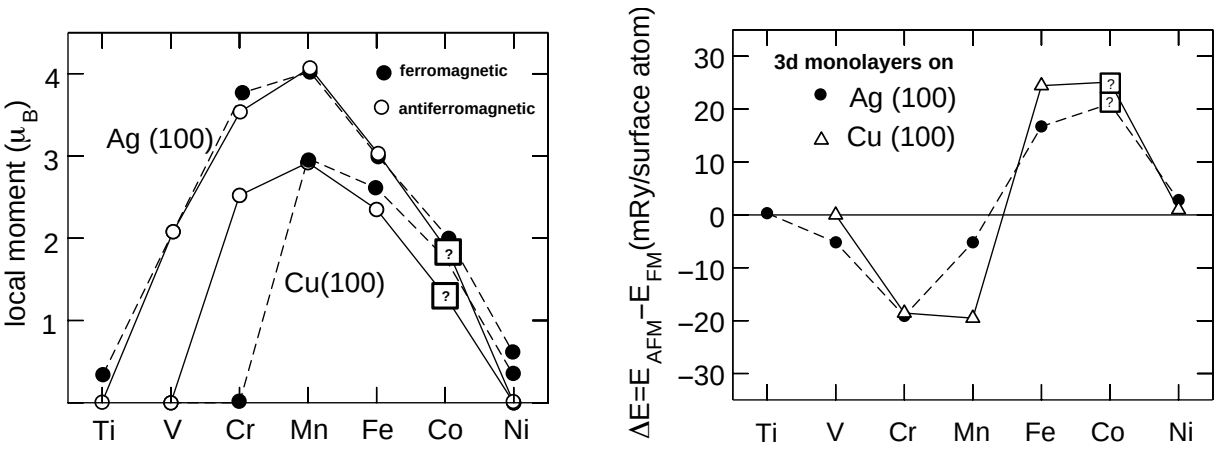


Fig. 13: *Left figure: Local magnetic moments of $3d$ monolayers on Cu(100) [2] and Ag(100) [64] calculated for the p(1×1) ferro– (solid circles connected by dashed line) and the c(2×2) antiferromagnetic configuration (open circles connected by solid line). Right figure: Total energy difference $\Delta E = E_{AFM} - E_{FM}$ per $3d$ atom between the c(2×2) antiferromagnetic and p(1×1) ferromagnetic phase for $3d$ monolayers on Cu(100) (triangle connected by full line) and Ag(001) (solid circles connected by dashed line). $\Delta E > 0$ (< 0) means, the ferromagnetic (antiferromagnetic) configuration is the most stable one. "?" indicates an result which is not fully converged*

of $3d$ monolayers on Cu(001). It becomes evident that, for many systems (see also Table 4) both configurations exist with moments of similar values. Depending on the inplane lattice constant, differences in the local moments for the two magnetic phases develop for earlier transition metals, e.g. for Cr on Cu(001,) for V on Pd(001) or for Ti on Ag(001). Figure 13 shows also the energy differences $\Delta E = E_{AFM} - E_{FM}$ per atom between the c(2×2) antiferromagnetic and the ferromagnetic configuration for $3d$ metal monolayers on Cu(001) and Ag(001). A clear trend emerges: The Ni, Co, and Fe overlayers ($\Delta E > 0$) prefer the ferromagnetic configuration and the Mn, Cr, and V ones favor the antiferromagnetic one. From the strong similarities of the monolayer trends for these two substrates we conclude, that this is a general trend: Fe, Co, and Ni favor the p(1×1) ferromagnetism on the (001) surfaces of Pd, Pt and the noble metals Cu, Ag and Au [72] whereas V, Cr, and Mn monolayers prefer the c(2×2) antiferromagnetic configuration. The same trend was recently found for monolayers on W(110) [73], and is expected for Al substrates although V and Ni might then be nonmagnetic. Since $\Delta E \approx 8S^2 J_1$, ΔE reflects basically the change of J_1 as function of the band filling (number of d electrons) or how $E_{\rm F}$ moves through the LDOS in Fig. 12. For Mn on Ag(001), where ΔE or J_1, respectively, is relatively small, the J's between more distant pairs may determine the picture. We investigated by total energy calculations the stability of the possible p(2×1) structure and found that the c(2×2) structure is indeed the magnetic ground state.

The c(2×2) antiferromagnetic phase was first predicted by theory. After the prediction several experiments indicated that the c(2×2) state may indeed exist: no ferromagnetic long range order was found at low temperatures for a V monolayer on Ag(100) [74], but a local exchange splitting was found for V, Cr, and Mn monolayers on Ag(100) [75]. More than 10 years after the theoretical prediction a direct proof of the c(2×2) antiferromagnetic state became for the first time possible by using the spin-polarized scanning tunneling microscopy in the constant-current mode [76, 77]. The experiments were carried out for a Mn monolayer on W(110).

4.2 (111) Oriented Monolayers on Nonmagnetic Substrates

Ferromagnetic Monolayers

The (0001) surface of an hcp crystal and the (111) surface of a fcc crystal establish a triangular lattice. Compared to the (100) surface the coordination number changes from 4 to 6, and the symmetry changes from fourfold to threefold or sixfold, respectively. Moreover, the differences in the magnetic properties between films on a square lattice and on a triangular lattice gives an estimate of the importance of the pseudomorphic growth condition for the magnetism of the films.
Figure 11 exhibits the general trend that the magnetic moments of the sixfold coordinated monolayers on Ag(111) are smaller in magnitude than those of the fourfold coordinated ones on Ag(001). On the Ag(111) surface we found magnetism for all $3d$ metals with the exception of Ti, which was very small anyway. There is nearly no difference between the monolayer moments of Mn, Fe, Co, and Ni on the differently oriented Ag substrates. A comparatively larger reduction of the magnetic moments is found at the beginning of the $3d$ series where the wavefunction is more extended than at the end of the series. Thus changing the coordination number from 4 to 6 changes the local moments not significantly. One consequence of this result is that for monolayers which do not grow pseudomorphically on any substrate, but keep an average distance between monolayer atoms similar to the pseudomorphic films, no dramatic difference in the formation of large local moments are expected.
With the exception of Ru (1.23 $\mu_{\rm B}$), and Rh (0.67 $\mu_{\rm B}$) and a tiny moment for Ir (0.05 $\mu_{\rm B}$) among the $5d$ metals, no ferromagnetism was found for any other $4d$ and $5d$ monolayers on Ag(111). For the $4d$ metal monolayers Ru and Rh, the moments are reduced to about 70% of the (001) values and for the $5d$ metal Ir only a tiny magnetic moment of 0.05 $\mu_{\rm B}$, about 15% of the (001) value, remains. Obviously the degree of the reduction of the magnetic moments due to the increase of the hybridization with the increase of the coordination number from 4 to 6, follows simply the increasing degree of delocalization of the d wavefunction when moving from the $3d$ to the $4d$ and $5d$ transition–metal wavefunctions.

Monolayers with Complex Spin Structures

Antiferromagnetic interactions on a triangular lattice are the origin of frustrated spin systems. In recent years the epitaxial growth of such ultra-thin films has been studied intensively by various experimental techniques. In particular, pseudo-hexagonal c(8×2)Mn films on Cu(100) [78], Mn films on the (111) surfaces of fcc Pd [79], Ir [80], Cu [81, 82, 83], and MgO [84] and on the (0001) surface of Ru [85] and Co [86] have been prepared and analyzed. But also other ultra-thin hexagonal films, e.g. Cr and V on Pt(111) and Ru(0001) [87, 88, 89], have been investigated.
To obtain an overview of all relevant spin-structures we develop first a zero-temperature phase diagram in the context of the Heisenberg model. As discussed in Sect. 2.3 the magnetic ground states are SSDWs, most likely with a commensurate propagation vector $\vec{q}_{\|}$ located at the high-symmetry points in the first 2DBZ of a 2D Bravais lattice. For the 2DBZ of the triangular (hexagonal) lattice, displayed in Fig. 14 (Left), the high-symmetry points are the corner points $\overline{\Gamma}$, $\overline{\rm K}$, and $\overline{\rm M}$ of the irreducible wedge of the 2DBZ (I2DBZ). The $\overline{\Gamma}$-point corresponds to the ferromagnetic solution. The $\overline{\rm K}$-point corresponds to a 120° Néel state (Fig. 14 (Center)), a 2D coplanar spin structure with three atoms in a ($\sqrt{3} \times \sqrt{3}$) R30° unit cell for which the relative angle between the spins at the different sites is always 120°. The $\overline{\rm M}$-point corresponds to row-

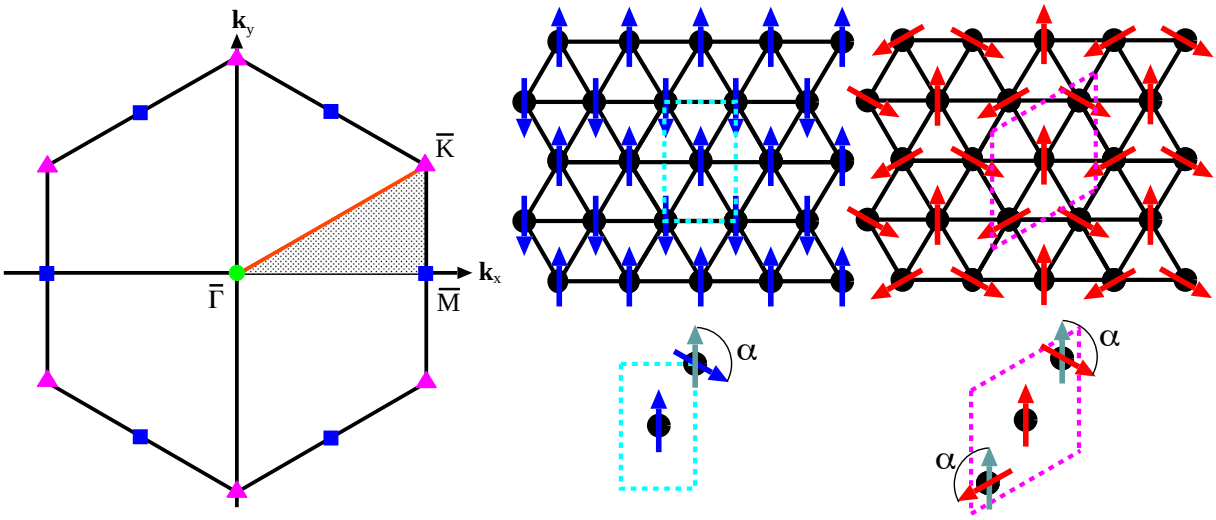
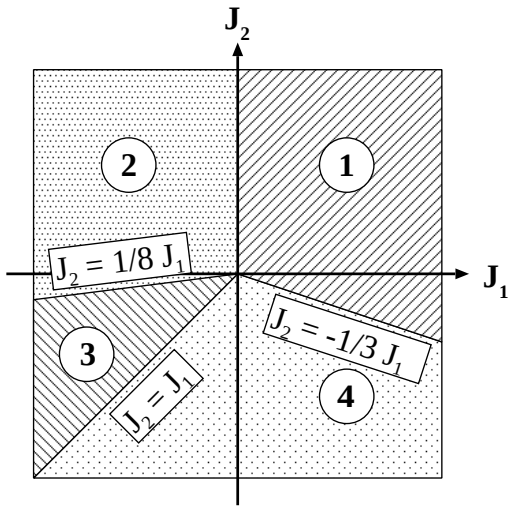


Fig. 14: *(Color) (Left:) The hexagon shows the first BZ of the 2D hexagonal Bravais lattice. The gray-shaded area indicates the irreducible part. (Center:) The RW-AFM structure. (Right:) the coplanar non-collinear Néel (120°) structure. Indicated are the corresponding two- and three-atom unit cells and the continous paths, which connect the corresponding magnetic structure to the FM state*



1 Ferromagnetic
2 Néel State
3 Row-wise AFM
4 Spin-Spiral (0,k_y)

Fig. 15. *Zero temperature phase-diagram in the (J_1, J_2) space for the triangular lattice indicating the regions of the four possible magnetic states*

wise antiferromagnetic (RW-AFM) configuration (Fig. 14 (Right)), which can be described by a rectangular unit cell with two antiferromagnetically aligned atoms. Magnetic ground states with incommensurate $\vec{q}_{\parallel}$-vectors are also possible preferentially with $\vec{q}_{\parallel}$-vectors from the connecting high-symmetry lines $\overline{\mathrm{M}}$-$\overline{\Gamma}$-$\overline{\mathrm{K}}$-$\overline{\mathrm{M}}$.

Along the line $\overline{\mathrm{M}}$-$\overline{\Gamma}$-$\overline{\mathrm{K}}$-$\overline{\mathrm{M}}$ we investigated the energetics within the Heisenberg model up the second nearest-neighbor interaction, i.e. including the exchange constants J_1, J_2. The results are summarized in Fig. 15 in terms of a zero-temperature phase diagram. Depending on the signs and values of J_1, and J_2 four kinds of possible magnetic ground states exist: FM, RW-AFM, 120°, and the SSDW. If J_2 is zero or positive (ferromagnetic) than there are only two possible magnetic ground states, determined by the sign of J_1, the FM and the Néel state. But small values of J_2 are already sufficient to change the magnetic ground state and an infinite number of magnetic states becomes possible, the RW-AFM state or the incommensurable SSDW at any

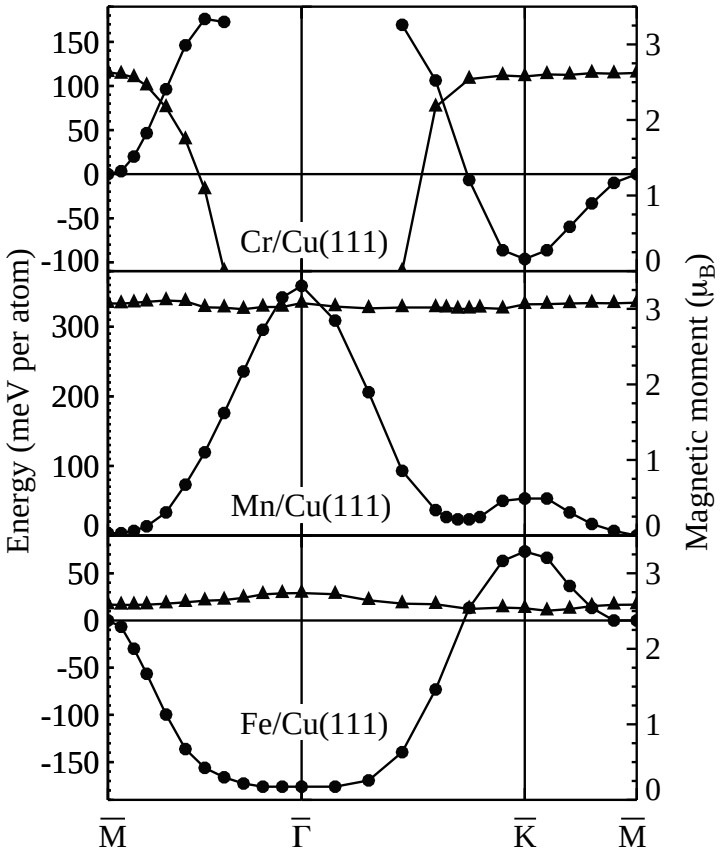


Fig. 16. *Calculated total energies (circles, left scale) and magnetic moments (triangles, right scale) for spin-spiral states in $3d$-UMLs with the Cu(111) geometry as function of the 2D wave vector $\vec{Q}_\parallel$ along the high symmetry lines of the 2DBZ. The energy is shown relative to the energy of the RW-AFM state*

possible wave-vector $\vec{q}_\parallel$ at the high-symmetry line $\overline{\Gamma}$-$\overline{\mathrm{M}}$. Extending the model by including also J_3, a magnetic state with a $\vec{q}_\parallel$ at any high-symmetry line can become ground state.

Since the J's are rapidly varying functions of the number of d electrons, ab-initio calculations are carried out to determine the element specific ground states. Since the calculations are very time comsuming, the full overview has been worked out only for unsupported, free-standing monolayers (UML). Fig. 16 shows for the UMLs with the Cu lattice constant the total energy $E(\vec{Q}_\parallel)$ and the magnetic moments $M(\vec{Q}_\parallel)$ calculated for a discrete set of the spin-spiral $\vec{Q}_\parallel$ vectors along the high-symmetry lines. Among all the SSDWs calculated, the high-symmetry points have the lowest energies: the 120° Néel state ($\overline{\mathrm{K}}$-point) for Cr(111), the RW-AFM state ($\overline{\mathrm{M}}$-point) for Mn(111), and the FM state ($\overline{\Gamma}$-point) for Fe(111). For Fe and Mn, the $M(\vec{Q}_\parallel)$ are nearly a constant, but the Cr moments change drastically, as no ferromagnetic solution could be found for Cr(111). One more interesting observation is the local minimum of $E(\vec{Q}_\parallel)$ for Mn on the line $\overline{\Gamma}$-$\overline{\mathrm{K}}$, which is only 21 meV higher in energy than the RW-AFM state. We expect that a small change in the d-band filling, e.g. due to alloying with Fe, may change the energetics and an incommensurate SSDW may become the magnetic ground state.

For Mn, the lowest energy magnetic state found so far is the RW-AFM state, which corresponds to the commensurate SSDW state with one single $\vec{Q}_\parallel$-vector at the $\overline{\mathrm{M}}$-point of the 2DBZ, and the RW-AFM is also called single-$\vec{Q}_\parallel$ (1Q) state. In the 2DBZ there are three $\overline{\mathrm{M}}$-points corresponding to the three possible directions of the long axis of the RW-AFM unit cell on a triangular lattice. They are equivalent in symmetry but are different to each other with $\vec{Q}_\parallel$-vectors, $\vec{Q}_\parallel^{(k)}$, for $k = 1, 2, 3$. Within the Heisenberg model the energy of each SSDW denoted by one of the three wave vectors $\vec{Q}_\parallel^{(k)}$ or any SSDW being an orthogonalized linear combination of those are degenerate. Higher order spin interactions (17) and (18) may lift this degeneracy and a so-

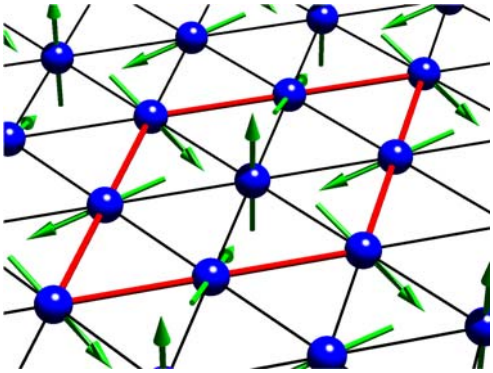


Fig. 17. *(Color) An image of the magnetic 3Q-structure, with spins pointing in all three directions of the spin-space. Note that, due to the neglect of the spin-orbit interaction only the relative orientation of the moments is specified*

called triple-$\vec{Q}_{\parallel}$ (3Q)-state, may become lower in energy. The 3Q-state is a three-dimensional non-collinear spin-structure on a 2D lattice (see Fig. 17) with four chemically identical atoms per surface unit-cell, where the relative angle between all nearest-neighbor spins is given by the tetrahedron angle of $109.47°$. The 3Q-state is formed as a linear combination of the three RW-AFM (1Q) structures orthogonal in spin-space, each having one of the three $\vec{Q}_{\parallel}^{(k)}$-vectors of the $\overline{\mathrm{M}}$-points:

$$\vec{m}(\vec{r}+\vec{R}_i) = m(\vec{r}) \times \frac{1}{\sqrt{3}} \sum_{k=1}^{3} e^{i\vec{Q}_{\parallel}^{(k)}\,\vec{R}_i}\,\hat{\mathbf{e}}^{(k)}, \tag{40}$$

where the $\hat{\mathbf{e}}^{(k)}$ are orthogonal unit vectors in spin space. We see that in the nearest-neigbhor approximation to the higher order exchange contributions the sign of K_1 and B_1 determine the sign of the energy difference $\Delta E = E_{3Q} - E_{1Q} = 16/3S^4(2K_1 + B_1)$ and thus whether the 3Q or the 1Q state becomes the magnetic ground state. From the ab-initio calculations for the Mn UML in the geometry of Cu(111) we [3] found that the 3Q-state is 15 meV/atom lower in energy than the 1Q-state.

Calculations including the Cu(111) substrate show that the energy differences between different magnetic states change due to the present of the substrate, but the magnetic ground state remains unaltered: Cr/Cu(111) exhibits the 120° Néel state (2.35 μ_{B}), Mn/Cu(111) the 3Q-structure (2.74 μ_{B}), which is 17 meV lower in energy than the 1Q-state (3.00 μ_{B}), and Fe/Cu(111) is ferromagnetic (2.63 μ_{B}). On the Ag(111) substrate [90] the overall picture is the same, but two differences were noticed: V/Ag(111) is magnetic (2.19 μ_{B}) and exhibits as Cr/Ag(111) (3.65 μ_{B}) the 120° Néel state and the magnetic gound state of Mn/Ag(111) is the RW-AFM state (3.91 μ_{B}) and not the 3Q-state (3.88 μ_{B}). Fe/Ag(111) is ferromagnetic (3.02 μ_{B}). We believe that the complex spin-structures presented here, can be resolved using the spin-polarized scanning tunneling microscope in the constant-current mode [91, 90].

4.3 Magneto-Interlayer Relaxation

In order to give the reader an impression (i) how strongly the formation of large monolayer moments may affect the interlayer relaxation and (ii) what is the influence of the magnetic order on the interlayer distance, total energy calculations as function of the interlayer distances are presented for two selected systems: Mn/Ag(001), and Mn/Cu(001). Prior to these calculations we determined the in-plane lattice constants which are taken to be the bulk lattice constants of the substrate; we found a value of $\mathrm{a}_0^{\mathrm{Cu}} = 6.65$ a.u. for Cu and $\mathrm{a}_0^{\mathrm{Ag}} = 7.58$ a.u. for Ag. Clearly, the Mn monolayers show the largest magnetic moments on any substrate and the magneto-volume effects should be most substantial.

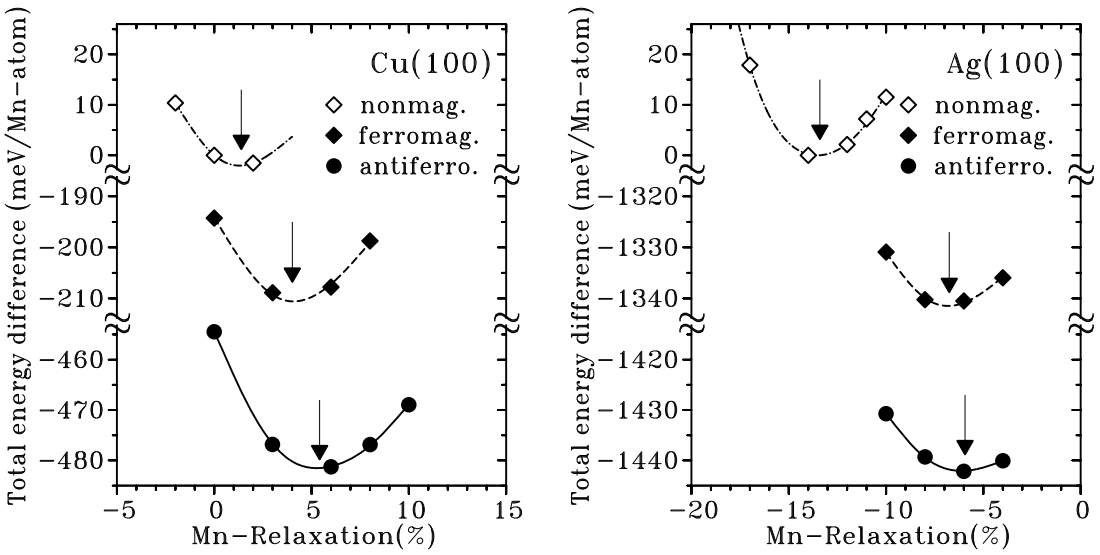


Fig. 18: *Total energies as function of the interlayer relaxation for nonmagnetic (open diamonds), ferromagnetic (solid diamonds), and c(2×2) antiferromagnetic (solid circles) Mn monolayers on Cu(001) and Ag(001). The energy of the nonmagnetic monolayer at 0% relaxation was chosen as the origin of the total energy scale. The interlayer relaxation is given in relative units with respect to the interlayer distance of the substrate. The vertical arrows indicate the minimum energy interlayer relaxation*

Fig. 18 shows the total energy as function of the interlayer distance for a Mn monolayer on Cu(001) and Ag(100) for three different magnetic states: nonmagnetic, ferromagnetic and c(2× 2) antiferromagnetic. We find, as already discussed in Sect. 4.1 that the nonmagnetic solution is the highest in energy and the antiferromagnetic one is the lowest energy magnetic state. Second, we find a substantial change of the minimum energy interlayer distances with change of the magnetic state. On Cu(100) the most contracted minimum energy distance was found for the nonmagnetic solution with $\Delta z_{\rm N} = 1.39\%$. For the ferromagnetic state a relaxation of $\Delta z_{\rm F} = 4.02\%$ and for the antiferromagnetic state a relaxation of $\Delta z_{\rm AF} = 5.41\%$ was determined. We find that the effect of the long range magnetic ground state on the relaxation is equally important as the formation of moments itself: the formation of a magnetic moment expands the interlayer distance by about 2.6% and the change in the magnetic state changes the interlayer distance by 1.4%. This coincides with the energy differences between the ferromagnetic state and the nonmagnetic state which is comparable to the energy difference between the antiferromagnetic state and the ferromagnetic one.

On Ag(001), the interlayer relaxations for the nonmagnetic, ferromagnetic, and antiferromagnetic Mn monolayers are determined to $\Delta z_{\rm N} = -13.4\%$, $\Delta z_{\rm F} = -6.75\%$, and $\Delta z_{\rm AF} = -5.94\%$, respectively. The lattice constants of Ag is 14% larger than the lattice constant of Cu. Consequently the Mn atoms relax inwards on these substrates. Due to the large Mn moments, around 4 $\mu_{\rm B}$ on these substrates (recall the moment of Mn on Cu is slightly below 3 $\mu_{\rm B}$), the magneto-volume effect is very large. The ferromagnetic Mn monolayers experience a large expansion of their minimum energy interlayer distance of about 7%, much larger than for Cu and the magnetic configuration modifies this expansion by an other 1% to 2%. The impact of the magnetic order on the interlayer distance is within about 2%, but the magneto–volume effect due to the formation of large magnetic moments is much larger for Mn on Ag than for Mn on Cu. This is in line with the arguments based on energy differences. The energy difference between the

antiferromagnetic state and the ferromagnetic state is for all Mn systems in the same ballpark of about 300 meV/Mn atom (cf. Table 1), while the formation energy of local moments is at large difference: about 200 meV for Mn on Cu but 1300 meV for Mn on Ag. This explains the large difference in the magneto-volume effects between Mn on Cu and Mn on Ag. In all cases the relaxations stabilize the ferromagnetic and antiferromagnetic phases, respectively.

Concluding, the atomic volume depends on the magnetism, mostly on the size of the moment and to a smaller extent on the magnetic state. An extreme example of this is the experimentally observed unusually large atomic buckling of the c(2×2)MnCu/Cu(001) [92] and c(2×2)MnNi/Ni(001) [92] surface alloys. In these alloys a buckling of the surface atoms of 0.30 Å (MnCu) [92] and 0.25 Å (MnNi) [92] was found. Although the atomic radii of Pd and Au are much larger than for Mn, the buckling of the c(2×2)CuPd/Cu(001) and c(2×2)CuAu/Cu(001) atoms was observed to just 0.02 Å[93] and 0.10 Å[94], respectively. It was shown that this buckling was a consequence of the magnetovolume effect, due to the large moments of Mn (3.75 $\mu_{\rm B}$) in Cu [95] and Ni (3.55 $\mu_{\rm B}$) [96].

4.4 Orbital Moment and Magnetic Anisotropy

Trends in Unsupported (100) Monolayers

The orbital magnetic moments ($m_{\rm l}$) and the magnetocrystalline anisotropy (MCA) are fairly small quantities as compared to spin moments and exchange energies. This holds at least for $3d$ transition metals. These quantities depend on fine details of the electronic structure which alters with lattice constant, film thickness, choice of substrate and surface orientation. Although it is important to know the fact values of the $m_{\rm l}$ and the MCA for particular systems, in this subsection we try to provide insight and intuition into the behavior of these quantities by studying the chemical trend of these properties across the transition-metal series. We focus on (100) oriented unsupported $3d$, $4d$ and $5d$ transition-metal monolayers in the lattice constant of Ag(100) ($a_0/\sqrt{2} = 5.459$ a.u.). In order to proceed with a fine scale analysis of these properties as function of the electronic structure or the band filling, respectively, the relativistic density functional calculations are carried out for films of hypothetical atoms with non-integer nuclear numbers. The rational behind this modus operandi is the idea that due to the required charge neutrality, the nuclear number and the number of electrons are the same. Thus, a fine change of the nuclear number is followed by a fine change of the number of electrons, caused by the adjustment of the Fermi energy. This facilitates a fine scan of $m_{\rm l}$ and the MCA as function of the band filling across the transition-metal series. For example, for the $3d$ monolayers we have varied the nuclear number Z from the beginning to the end of the transition-metal series, e.g. from $Z = 21$ with $N_V = 3$ valence electrons till Cu, $Z = 29$ with $N_V = 11$ valence electrons, in steps of approximately $\Delta Z = 0.10 \sim 0.15$.

The results are summarized in Figure 19. For each Z, two self-consistent relativistic *ab initio* calculations have been carried out, one with the magnetization direction perpendicular to the film plane, $\widehat{m} = \uparrow$, and one with the magnetization parallel to the film plane along the $[100]$ direction, $\widehat{m} = \rightarrow$. As results one obtains the orientation dependent spin moments, $m_{\rm s}(\uparrow)$ and $m_{\rm s}(\rightarrow)$, orbital moments, $m_{\rm l}(\uparrow)$ and $m_{\rm l}(\rightarrow)$, and electronic total energies, $E(\uparrow)$ and $E(\rightarrow)$. Here we focus on the ferromagnetic phase.

The magnetic spin moments exhibited in Figure 19(a), follow the trend discussed in Subsect. 4.1: The $3d$ monolayers behave according to Hund's first rule with a maximum moment of more than 4 $\mu_{\rm B}$ in the center of the series. Also $4d$- and $5d$-metal monolayers are magnetic

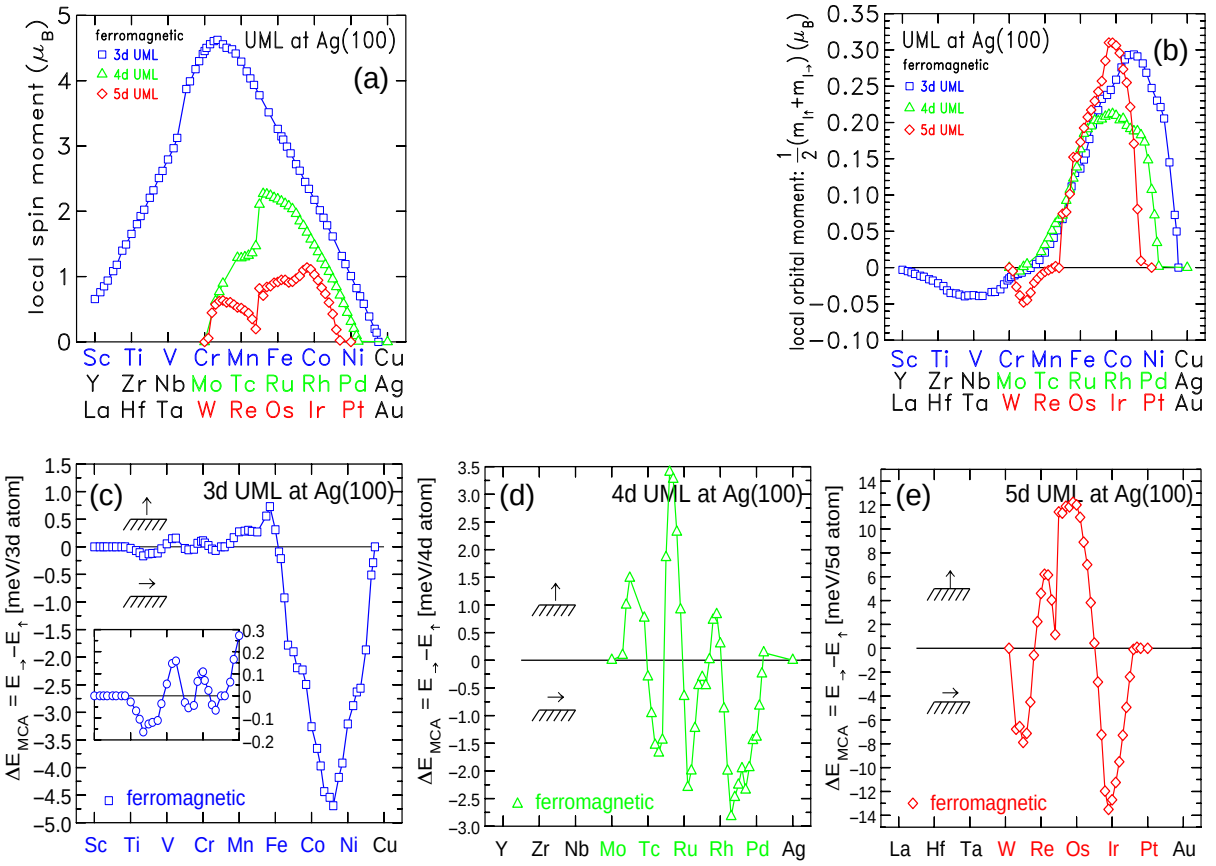


Fig. 19: *Local magnetic spin moments (a) and orbital moments (b), magnetocrystalline anisotropy constant $K_{\rm MCA} = \Delta E_{\rm MCA}$ calculated as energy difference between between two magnetization directions, the magnetization in the film plane ($\rightarrow$) and out-of the film plane ($\uparrow$), calculated for ferromagnetic, freestanding, unsupported $3d$- (squares) (c), $4d$- (triangles) (d), and $5d$- (diamonds) (e) metal monolayers (UML) in the (100) surface orientation and in the lateral lattice constant of the Ag(100) substrate [97]. In (b) the average orbital moments between those of the out-of-plane, $m_{l\uparrow}$, and inplane, $m_{l\rightarrow}$, magnetization directions are shown. The difference of the spin moments on the magnetization direction is difficult to distinguish on the scale of (a) and is not shown. Positive energies in figures (c)–(e) means that the out-of-plane magnetization is energetically preferred.*

for elements between Mo till close to Pd and between W till close to Pt. The magnetic moments decrease from the $3d$ to the $4d$ and $5d$ series and at the same time the element with the maximum magnetic moment in each series shifts to the right in the series. The anisotropy of the spin moments, $\Delta m_{\rm s} = m_{\rm s}(\uparrow) - m_{\rm s}(\rightarrow)$, is very small, e.g. for an Ir monolayer one yields $m_{\rm s}(\uparrow) = 1.044\ \mu_{\rm B}$ and $m_{\rm s}(\rightarrow) = 1.012\ \mu_{\rm B}$, and is therefore not further considered. However, for $5d$ elements, relativistic calculations have an impact on the size of the spin moments. For example, the spin-orbit interaction reduces the magnetic moment of Ir by 0.5 $\mu_{\rm B}$ to about 1 $\mu_{\rm B}$. According to equation (30), $E_{\rm dip}$ is proportional to $m_{\rm s}^2$, and inversely proportional to third power of the lattice constant, a. Since the lattice constants of all systems are fixed to the one of Ag, Figure 19(a) mirrors the functional behavior of the dipole energy $E_{\rm dip}$ with respect to the band filling. Since the dipolar anisotropy or shape anisotropy in the continuum limit, respectively, $K_{\rm dip} = \Delta E_{\rm dip} = E_{\rm dip}(\rightarrow) - E_{\rm dip}(\uparrow)$, favors always a magnetization in the film plane, the shape anisotropy is strictly negative according to our sign convention for the magnetic anisotropy.

The largest value is obtained for a film with elements between Cr and Mn and the dipolar anisotropy amounts to $K_{\rm dip} = -0.32$ meV/atom. This is about 30% of the value of $K_{\rm shape}$ as calculated according to equation (29) using continuum theory, in good agreement with the results of Draaisma *et al.*[24].

Also the orbital moments, collected in Figure 19(b) exhibit a clear trend: The orbital moments are negative in the first half of the transition-metal series, e.g. between Sc and Mn, and positive in the second half of the transition-metal series. The change in sign is reminiscent to Hund's third rule which governs the coupling of the spin and orbital moment. A negative (positive) sign of the orbital moment means that the orbital moment couples opposite (parallel) to the spin moment. The orbital moments of elements in the second half of the transition-metal series are an order of magnitude larger than in the first one. It is surprising that the orbital moments in films made of the $3d$, $4d$, or $5d$ elements in the second half of the corresponding transition-metal series show very similar values although the spin moments are substantially different. We recall from equation (35) that for uniaxial symmetry as present in thin films, the orbital moment scales as $m_{\rm l} \propto \xi \propto Z^2$, which explains the increase of $m_{\rm l}$ switching the transition-metal series. The anisotropy of the orbital moments, $\Delta m_{\rm l}$, has maximum values of about $\pm 0.07\ \mu_{\rm B}$ in each series and cannot be neglected. Therefore, in Figure 19(b) the average moments are shown. $\Delta m_{\rm l}$ is a rapidly varying function with respect to the band-filling and relates according the equation (36) to the corresponding rapid oscillation of the magnetocrystalline anisotropy.

The uniaxial magnetocrystalline anisotropy constants $K_{\rm MCA} = \Delta E_{\rm MCA} = E(\rightarrow) - E(\uparrow)$, calculated as total energy differences for magnetizations in- and out-of the film plane are exhibited in Figs. 19(c)–(e). Indeed, results show a continuous and very rapidly varying behavior as function of the bandfilling. We focus first on the $3d$-metal monolayers (Figure 19(c)). Between Fe and Cu the magnetocrystalline anisotropy energy is negative and the magnetization lies in the film plane. Between Mn and Fe the magnetization normal to the film plane is energetically most favorable. A closer look reveals several changes of sign as function of the band filling. Surprising is the large variation of the value of K as function of bandfilling from 4.75 meV for a bandfilling between Co and Ni ($N_V = 9.5$) and 0.15 meV for V. Non-integer bandfilling are not only of theoretical interest but have a concrete meaning in the spirit of the virtual crystal approximation. For example, we calculated an ordered c(2×2) CoNi UML film, which has also a bandfilling of $N_V = 9.5$ per atom. K is practically on the spot of the curve Figure 19(c). Adding $K_{\rm dip}$ on top of $K_{\rm MCA}$ one finds that with the exception of a small interval between Mn and Fe, where the positive $K_{\rm MCA}$ exceeds the negative $K_{\rm dip}$, the magnetization is energetically most favorable to be in the film plane. Thus, among the $3d$-metal monolayers (integer nuclear number) only the Fe(100) UML has a magnetization direction out-of-plane.

Comparing the $K_{\rm MCA}$ between the $3d$, $4d$ and $5d$ monolayers remarkable, results are observed. The most spectacular results are the gigantic $K_{\rm MCA}$ values for the $5d$ UMLs which reach values of 12.32 meV for Os and -13.50 meV for Ir. Although the maximum magnetic spin moment within each transition-metal series drops from 4.6 μ_B in the $3d$ series to 2.3 μ_B and 1.1 μ_B in the $4d$ and $5d$ series, respectively, and the orbital moments are roughly the same between the transition-metal series, the variation of $K_{\rm MCA}$ changes from -4.69 meV to 0.73 meV in the $3d$ series, and from 3.40 meV to 2.82 meV in the $4d$ series and to these truly gigantic values of -13.50 meV to 12.32 meV in the $5d$ series. One further notices that the latter is accompanied by a rapid change of $K_{\rm MCA}$ of about 25 meV when going from Os to its chemical neighbor Ir. One further notices that the functional characteristics of $K_{\rm MCA}$ shows a much more oscillatory behavior in the $4d$ and $5d$ series, both exhibiting two maxima and three minima, than in the $3d$ one with one minimum, one maximum and then several small rapidly oscillating peaks.

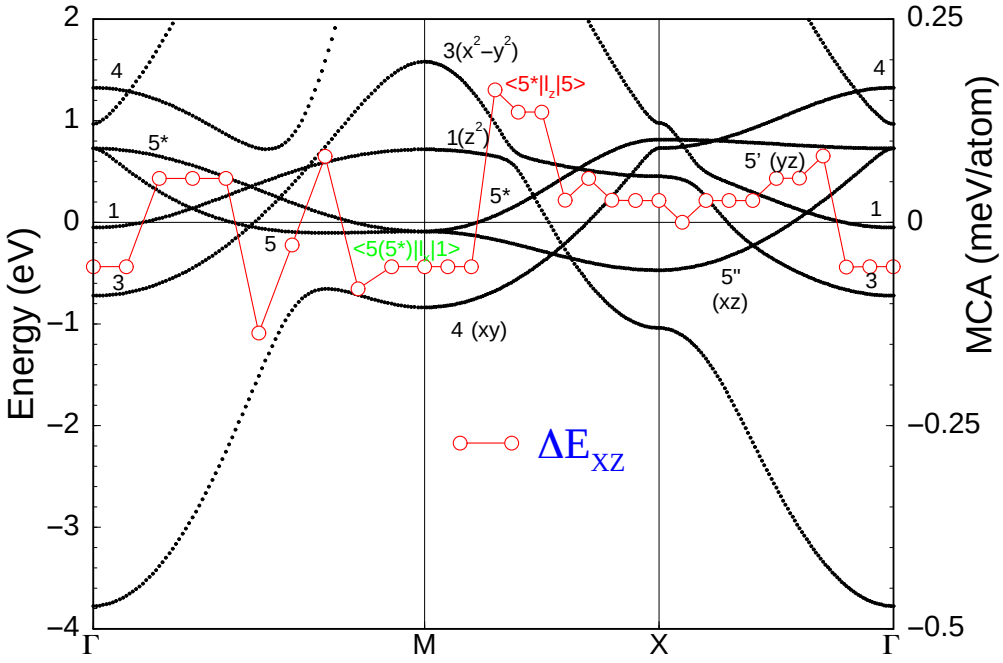


Fig. 20: *Minority spin bandstructure (black) and statewise contribution to* K_{MCA} (red circles) *calculated by perturbation theory for Fe(100) UML in the lattice constant of Ag. The symmetry of the bands is indicated as well as some matrix elements which contribute to* K_{MCA} *(see text).*

These results can be interpreted on the basis of the corresponding bandstructures and the second-order perturbation theory of the MCA as presented in section 2.5. As an example, the Fe monolayer in the Ag(100) lattice constant is analyzed: As can be inferred from Figure 12, the majority d-band of Fe on Ag(100) is filled, so that these states will not contribute to equation (34), where only pairs of occupied and unoccupied states near the Fermi level can contribute significantly. Therefore, we can focus on the minority states, and the corresponding bandstructure is shown in Figure 20. The matrix elements in equation (34) depend on the symmetry of the states, ψ, and the spin-orbit operator [26]. The spatial part of H_{so} has the symmetry of the orbital moment operator, e.g. l_x or l_z. Therefore it is possible to find out which pairs of states can lead to nonzero matrix elements in equation (34), depending of course on the magnetization direction. For an estimate of the MCA, at each $\vec{k}$-point $K_{\mathrm{MCA}} = \delta E(l_x) - \delta E(l_z) = \Delta E_{xz}$ can be calculated individually, as shown in Figure 20. E.g. near the M-point, states of $d_{xz}, d_{yz}(5, 5^*)$ symmetry are just below the Fermi level and states of $d_{z^2}(1)$ symmetry above. These states are coupled by the l_x operator, therefore the bands in this region will contibute more to $\delta E(l_x)$ than to $\delta E(l_z)$, favoring an in-plane magnetization. Between M and X, a coupling of states with 5 and 5^* symmetry favors an out-of-plane magnetization. After summation over the whole Brillouin-zone, the latter contributions dominate and, in accordance with Figure 19(c), Fe/Ag(100) has an out-of-plane magnetization.

If the Fermi level is shifted to higher energies (or the bandfilling increases), the band with 5^* symmetry gets more occupied and the coupling of the 5 and 5^* states is no longer possible. The contribution of the $\langle 5(5^*)|l_x|1\rangle$ matrix elements near M gets stronger. Finally, K_{MCA} changes sign (see Figure 19(c)) and the Co UML is in-plane magnetized. As can be inferred from Figure 20, the contributions to the MCA oscillate strongly in $\vec{k}$-space and for an accurate summation a fine resolution in reciprocal space, i.e. a fine $\vec{k}$-point mesh, is necessary. It should

also be noticed, that the above used decoupling of majority and minority bands can only be applied for $3d$ metals. In $4d$ or $5d$ monolayers, the exchange splitting is much smaller and both spin channels give contributions to the MCA. Therefore, a much more complex behavior of the anisotropy as function of the bandfilling can be seen in Figure 19(d) and (e).
These results give a very excellent overview of the trends of the uniaxial anisotropy K of magnetic monolayers. However, the substrate plays also an important role. For weakly magnetically polarizing substrates, e.g. Cu and Ag, the same trend is expected although the actual values will change. Substrates with large nuclear numbers and thus large spin-orbit interactions, which have in addition a large Stoner enhanced susceptibility, e.g. W or Pt, and can thus be easily magnetically polarized, may at the end determine the magnetic anisotropy of these systems. For example, in the light of the experimental results of ultrathin Co films in contact with other metal films, as sandwich or as multilayers, e.g. Co/Pd(100) [98] or Co/Pt(100) [99] the large negative K_{MCA} value of the unsupported Co monolayer, which is in accordance with results of Bruno [27] based on perturbation theory, is a fairly surprising result. In order to get a better understanding of the influence of the substrate on the magnetocrystalline anisotropy, we compare the uniaxial K_{MCA} for a Co monolayer with and without substrate and found, $K_{\mathrm{MCA}} = -4.75$ meV for an UML(100) in the lattice constant of Ag and $K_{\mathrm{MCA}} = -1.39$ meV on Ag(100), $K_{\mathrm{MCA}} = -1.33$ meV for an UML(100) in the lattice constant of Cu and $K_{\mathrm{MCA}} = -0.32$ meV on Cu(100). This can be understood in the sprit of the model of Stöhr [36], introduced in section 2.5, realizing that the presence of a substrate quenches predominantly the in-plane orbital moment. Therefore, we observe a clear reduction of K_{MCA} due to the presence of the substrate, but the general trend across the transtion-metal series will still hold. At arbitrary substrates $4d$- and $5d$-metal monolayers will be nonmagnetic. But thin films show a significant Stoner enhanced susceptibility. In contrast to $3d$ transition-metals they may develop a magnetic moment, an electronic structure and an uniaxial K_{MCA} comparable to the isolated monolayers. In turn, strongly spinpolarized substrates with large nuclear number change even the sign of the MCA from inplane to out-of-plane.

Magnetic Reorientation Transition: Ni/Cu(100)

If more than one or two layers of magnetic material are deposited as thin film, the layers that are not forming an interface (to the vacuum or the substrate) will show more bulk-like properties. It is common to separate the volume-like contributions to the effective anisotropy constants, K^V (energy per unit volume), from the surface term K^S and interface term K^I (energy unit per area). This yields for the effective magnetic anisotropy K of a magnetic layer of thickness t

$$K^{\mathrm{eff}} = K^V + (K^S + K^I)/t \; . \tag{41}$$

All three anisotropy constants K contain contributions of the dipolar and and the spin-orbit derived anisotropy. On phenomenological grounds the dipolar anisotropy is also split into a volume term, K^V_{shape}, namely the shape anisotropy due to the average dipolar energy as obtained by the continuum theory and the contribution due to the reduction of the dipole anisotropy field experienced by the atoms in the surface and interface region, $K^{S\,(\mathrm{dip})}_{\mathrm{MCA}} + K^{I\,(\mathrm{dip})}_{\mathrm{MCA}}$. Thus, we can write for the volume term $K^V = K^V_{\mathrm{shape}} + K^V_{\mathrm{MCA}}$, and for the surface term $K^S = K^{S\,(\mathrm{dip})}_{\mathrm{MCA}} + K^{S\,(\mathrm{so})}_{\mathrm{MCA}}$ and analogously for the interface term. While for smaller thicknesses K^S_{MCA} and K^I_{MCA} can dominate, for thick films the negative shape anisotropy which has a constant value per atom and thus increases with thickness of the film can determine the easy axis.

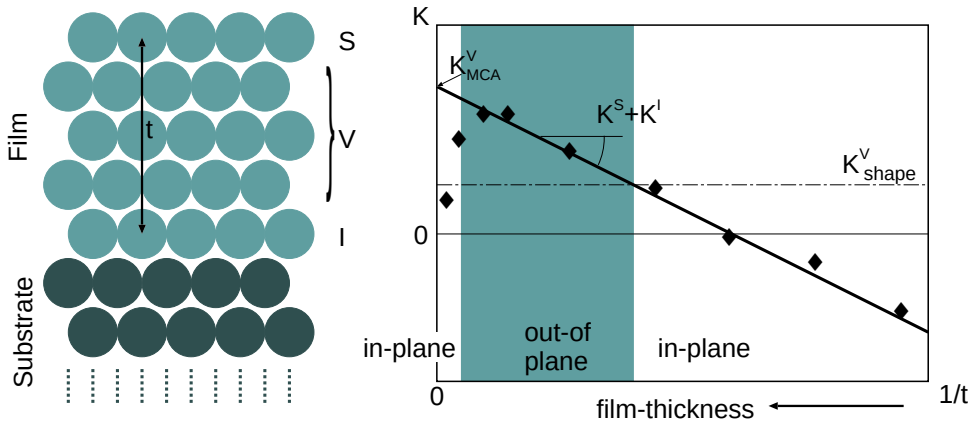


Fig. 21: *Schematic illustration of a five monolayer Ni film on a Cu substrate; the bulk-like, surface and interface regions are marked as V, S, and I, respectively. Right: Schematic plot of the anisotropy (K) versus inverse film-thickness (t): the volume-like contribution, K^V, is given by the intersection with the ordinate, the surface- and interface-like term, $K^S + K^I$, can be deduced from the slope of the curve. Together with the (volume) shape anisotropy, $K^V_{\rm shape}$, these values determine the in-plane / out-of-plane / in-plane transition of the easy axis for Ni films on Cu(001).*

If these terms have different signs, a reorientation transition of the easy axis can occur. A well investigated example is the system Ni/Cu(001) [100]: For up to seven layers Ni, an in-plane easy axis of the Ni film is found. Then a reorientation to perpendicular magnetization sets in and only very thick films (more than 50 monolayers) show again in-plane magnetization. Here K^S is negative, about -85μeV/atom at room temperature, while $K^V_{\rm MCA}$ is positive, approximately 30μeV/atom. So we expect a reorientation between 5 and 6 monolayers, but actually the shape anisotropy contributes another -10μeV/atom to K^V (the shape anisotropy in thin films always favors in-plane magnetization). Therefore, the transition sets in after 7 monolayer thicknesses. But we have to realize, that the value of K^V is much larger than the bulk value of fcc Ni. In fact, LEED measurements demonstrated that Ni grown on Cu(001) is actually strained, the in-plane lattice constant is 1.6% larger than in fcc Ni. To compensate this strain, the spacing between the Ni layers is smaller than in the bulk. From the arguments of the last paragraph we would now suppose that $V_\perp > V_\parallel$, therefore $R > 1$ and, indeed, K^V favors perpendicular magnetization (although the band-filling of Ni does not correspond to the assumptions underlying equation (36)). In very thick films, the structure of Ni relaxes back to fcc and the size of K^V decreases until the influence of the shape anisotropy once more brings the easy axis back in-plane.

4.5 Spin-Orbit Induced Homochiral Mesoscale Spin Spirals

At surfaces, and in other geometries with broken inversion symmetry, magnetic structures are subject to an antisymmetric exchange interaction, Eq. (37). This Dzyaloshinskii-Moriya interaction favors spiral magnetic structures of a specific handedness. In this chapter, we illustrate the impact of this interaction on the mesoscale magnetic structure of a monolayer of Mn atoms deposited on the W(110) surface. In this system, the Dzyaloshinskii-Moriya interaction is indeed strong enough to induce a spiraling magnetic ground state.

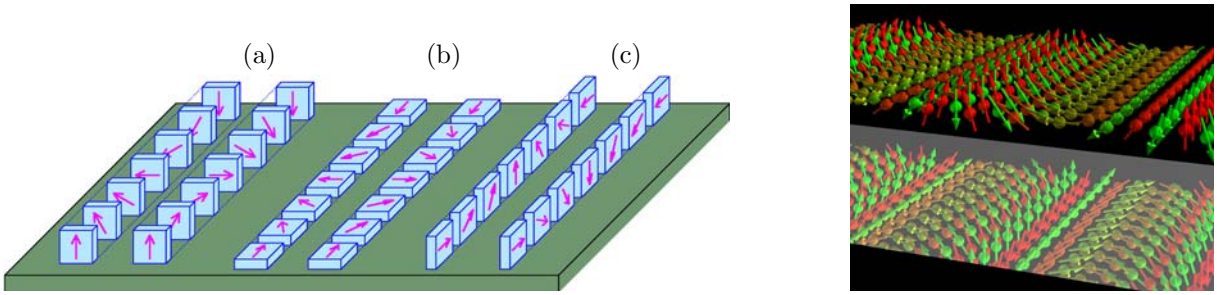


Fig. 22: *Spin spirals with different rotation axes on a symmetric surface. The left panel shows a right- and a left-handed spiral for each rotation axis. For (a) and (b), the right- and left-handed spirals are mirror images of each other. In the case (c), however, the surface breaks the mirror symmetry. Therefore, the two spirals in (c) are not equivalent to each other and may differ in energy. The right panel illustrates a spiral and its mirror image for Mn on W(110) a material with local antiferromagnetic order. The top picture shows a left rotating cycloidal spiral, which was found in nature. The bottom picture shows the mirror image, a right rotating spiral, which does not exist.*

In a wide class of magnetic materials, the magnetic structure changes on a mesoscopic length scale and this changes can be described by a continuous vector field $\vec{m}(\vec{r})$ with $|\vec{m}| = \text{const.}$. In the simplest case, $\vec{m}$ varies only along one spatial coordinate x and the energy of a magnetic configuration can be described by a simple Landau-Lifshitz functional of the form

$$E_0[\vec{m}] = \int \mathrm{d}x \left(A \left(\tfrac{\mathrm{d}\,\vec{m}(x)}{\mathrm{d}\,x}\right)^2 + \vec{m}(x)^\dagger \cdot \vec{\vec{K}} \cdot \vec{m}(x) \right) . \tag{42}$$

Thereby, the spin stiffness A represents the exchange interactions that favor collinear spin alignment, and the symmetric anisotropy tensor $\vec{\vec{K}}$ accounts for the preferred orientation of the magnetization with respect to the crystal lattice, cf. Eq. (26). The exchange term $A\,(\frac{\mathrm{d}\,\vec{m}}{\mathrm{d}\,x})^2$ is symmetric with respect to the rotational direction, i.e. the energy does not depend on the sign of $\frac{\mathrm{d}\,\vec{m}}{\mathrm{d}\,x}$. The exchange term reflects the contribution of the Heisenberg Hamiltonian (15) in the continuous model. In some systems, however, the Dzyaloshinskii-Moriya term (37) has to be taken into account. This term implies, that the energies for right- and left-handed spin rotations differ, and in the continuous approximation it is described by

$$E_{\mathrm{DM}}[\vec{m}] = \int \mathrm{d}x \left(\vec{D} \cdot \left(\vec{m}(x) \times \tfrac{\mathrm{d}\,\vec{m}(x)}{\mathrm{d}\,x} \right) \right) . \tag{43}$$

As already pointed out in Chapter 2.6, $\vec{D}$ is nonzero only if the underlying crystal structure does not posses inversion symmetry. Prominent examples for such structures are surfaces and interfaces [101]. Fig. 22 illustrates how a surface can break the symmetry between right- and left-handed spiral magnetic structures.

The DM term competes with the symmetric exchange and the anisotropy energy: The latter terms favor collinear spin alignment, whereas the DM term favors a spatially spiraling magnetic structure of a specific rotational direction. The resulting ground state was already discussed in 1965 for a simplified energy functional [102], but the actual values of the corresponding parameters (in particular the size of $\vec{D}$) are still unknown in most cases.

One can expect, that the DM term is of particular importance in atomically thin magnetic films that are deposited on non-magnetic surfaces: In such systems the local environment of all mag-

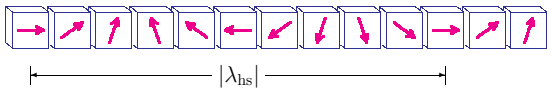


Fig. 23. *Homogeneous spin spiral with period length $|\lambda_{\rm hs}|$. The angle between the magnetization directions of two adjacent lattice sites is constant.*

netic atoms is inversion asymmetric. On the other hand, such low-dimensional structures usually show a large magnetocrystalline anisotropy that favors collinear spin alignment. We want to find out, to what extent the DM interaction can modify the the magnetic ground state of such systems. For the exemplary system Mn/W(110), we estimate the model parameters (A, $\vec{D}$, $\vec{\vec{K}}$) from first-principles, i.e. by electronic-structure calculations in the framework of the density functional theory.

We deduce the values of the spin stiffness A and the size and sign of the $\vec{D}$-vector from the electronic energies of homogeneous spin spirals. Such spirals are characterized by a fixed rotation axis and $(\frac{\mathrm{d}\,\vec{m}}{\mathrm{d}\,x})^2 = \mathrm{const.}$ (cf. Fig. 23). For these spirals Eqs. (42-43) simplify to

$$\frac{E_0 + E_{\rm DM}}{|\lambda_{\rm hs}|} = 4\,\pi\,A\,\lambda_{\rm hs}^{-2} + 2\,\pi\,D\,\lambda_{\rm hs}^{-1} + \mathrm{const.}\,, \tag{44}$$

where the integration is performed over one period length $|\lambda_{\rm hs}|$, the sign of $\lambda_{\rm hs}$ distinguishes between right- and left-handed spirals, and D depends on the size and orientation of $\vec{D}$. The computational scheme that allows us to deal with these large magnetic unit cells is described in Ref. [103].

In the following, we discuss the exemplary system of a monolayer of Mn atoms deposited on

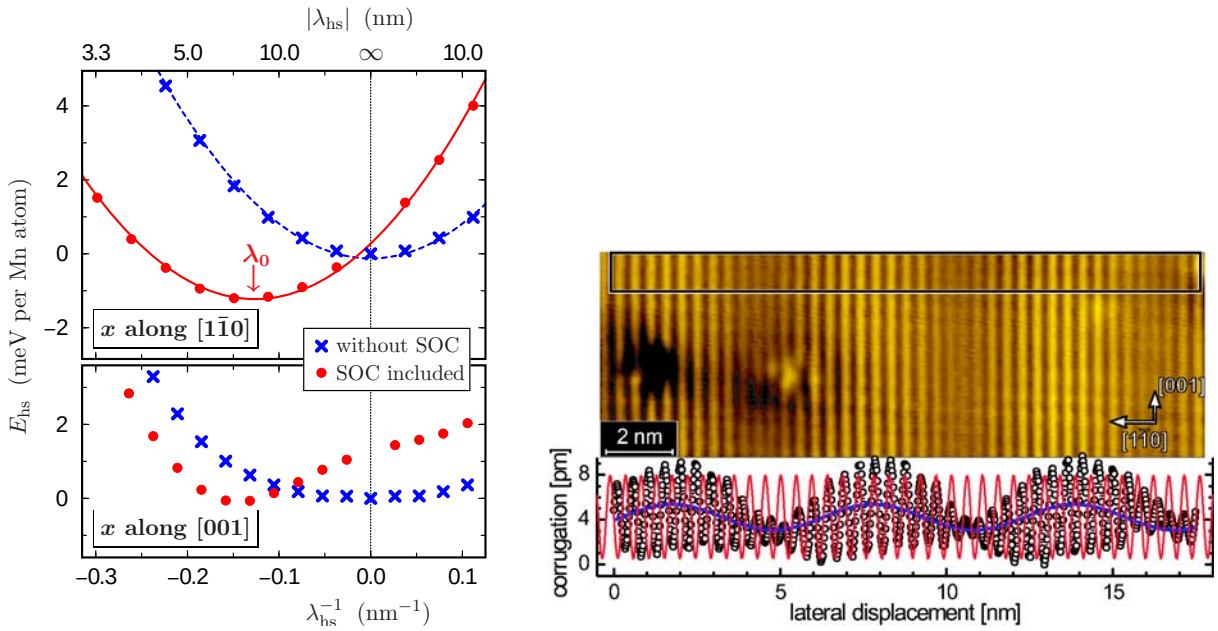


Fig. 24: *(from Ref. [104]) Left panel: Electronic energy of a homogeneous spin spiral depending on the period length $|\lambda_{\rm hs}|$. The sign of $\lambda_{\rm hs}$ depends on the rotational direction. The data points show the results obtained from the electronic-structure calculations and the lines indicate the fits with the terms of Eq. (44). Right panel: Magnetic structure of Mn/W(110) imaged with spin-polarized STM. The stripes on the nanometer scale represent the local antiferromagnetic structure. On a larger length scale, however, the image shows a spiral structure that is driven by the DM interaction.*

the W(110)-surface [104]. The magnetic structure of this system is well studied experimentally by spin-polarized STM. Locally it shows a $\mathrm{c}(2\times2)$-antiferromagnetic structure [76] (this does not conflict in with our ansatz, since the vector field $\vec{m}$ in Eqs. (42-43) represents only the spatial modulations on larger length scales). In Fig. 24, the calculated energies are shown for spin spirals propagating along the high-symmetry lines and the rotation axes aligned parallel to the corresponding $\vec{D}$-vectors. When the spin-orbit operator, Eq. (32), is neglected the curves represent the symmetric exchange, whereas the DM interaction and the anisotropy terms are a consequence of the spin-orbit coupling. From the figure, we can directly identify the homogeneous spiral of lowest energy (indicated with λ_0). In order to permit other magnetic structures, we obtain the parameters of Eq. (44) from fits to the calculated data and insert these values in the functionals (42-43). This way, we predict for the system Mn/W(110) a DM-driven left-handed spin spiral propagating along the $[1\bar{1}0]$-direction with a period $|\lambda| = 7.9\,\mathrm{nm}$. Since we are describing an antiferromagnet, the mesoscale magnetic structure shows a period of $\frac{1}{2}\,|\lambda|$. We find nice agreement with the experimental result that is presented in Fig. 24.
The studied system nicely illustrates the impact, that the DM interaction can have on magnetic structures of low symmetry. Here, it is indeed strong enough to compete with the spin stiffness and anisotropy and induces a so called Dzyaloshinskii spiral on the mesoscopic length scale.

Acknowlegements. I gratefully acknowledge the contributions of G. Bihlmayer, P.H. Dederichs, M. Heide, K. Hirai, J. Kudrnovský, P. Kurz, and I. Turek to the results described in this work.

References

[1] M. Donath, Surf. Sci. Rep. **20**, 251 (1994).

[2] S. Blügel, Appl. Phys. A **63**, 595 (1996).

[3] P. Kurz, G. Bihlmayer, K. Hirai, and S. Blügel, Phys. Rev. Lett **86**, 1106 (2001).

[4] O. Gunnarsson, J. Phys. F: Met. Phys. **6**, 587 (1976).

[5] J. F. Janak, Phys. Rev. B **16**, 255 (1977).

[6] P. Mavropoulos, S. Lounis, R. Zeller, and S. Blügel, Appl. Phys. A **82**, 103 (2006).

[7] M. Takahashi, J. Phys. C **10**, 1289 (1977).

[8] K. Terakura, N. Hamada, T. Oguchi, and T. Asada, J. Phys. F: Met. Phys. **12**, 1661 (1982).

[9] P. Kurz, G. Bihlmayer, K. Hirai, and S. Blügel, Phase Trans. **75**, 101 (2002).

[10] T. Asada and S. Blügel, Phys. Rev. Lett **79**, 507 (1997).

[11] I. Turek, J. Kudrnovský, V. Drchal, and P. Bruno, Phil. Mag. **86**, 1713 (2006).

[12] S. V. Tyablikov, *Methods of Quantum Theory of Magnetism* (Plenum Press, New York, 1967).

[13] I. Turek, J. Kudrnovský, M. Diviš, M. Franek, G. Bihlmayer, and S. Blügel, Phys. Rev. B **68**, 224431 (2003).

[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[15] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).

[16] N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966).

[17] R. P. Erickson and D. L. Mills, Phys. Rev. B **43**, 11527 (1991).

[18] J. A. C. Bland and B. Heinrich, eds., *Thermodynamic Properties of Ultrathin Ferromagnetic Films* (Springer–Verlag, Berlin/Heidelberg, 1994), vol. I, chap. D. L. Mills, p. 91.

[19] R. J. Glauber, J. Math. Phys. **4**, 294 (1963).

[20] P. Gambardella, A. Dallmeyer, K. Maiti, M. C. Malagoli, W. Eberhardt, K. Kern, and C. Carbone, Nature **416**, 301 (2002).

[21] P. Bruno, J. Appl. Phys. **64**, 3153 (1988).

[22] J. Topping, Proc. Roy. Soc (London) **A 114**, 67 (1927).

[23] L. Szunyogh, B. Újfalussy, and P. Weinberger, Phys. Rev. B **51**, 9552 (1995).

[24] H. J. G. Draaisma and W. J. M. de Jonge, J. Appl. Phys. **64**, 3610 (1988).

[25] P. P. Ewald, Ann. Physik **64**, 253 (1921).

[26] G. van der Laan, J. Phys.: Condens. Matter **10**, 3239 (1998).

[27] P. Bruno, Phys. Rev. B **39**, 865 (1989).

[28] M. S. S. Brooks, Physica B **130**, 6 (1985).

[29] I. V. Solovyev, A. I. Liechtenstein, and K. Terakura, Phys. Rev. Lett. **80**, 5758 (1988).

[30] I. V. Solovyev, Phys. Rev. Lett. **95**, 267205 (2005).

[31] O. Eriksson, M. S. S. Brooks, and B. Johansson, Phys. Rev. B **41**, 7311 (1991).

[32] O. Hjortstam, J. Trygg, J. M. Wills, B. Johansson, and O. Eriksson, Phys. Rev. B **53**, 9204 (1996).

[33] B. Nonas, I. Cabria, R. Zeller, P. H. Dederichs, T. Huhne, and H. Ebert, Phys. Rev. Lett. **86**, 2146 (2001).

[34] M. Komelj, C. Ederer, J. W. Davenport, and M. Fähnle, Phys. Rev. B **66**, 140407 (2002).

[35] C. Ederer, M. Komelj, and M. Fähnle, Phys. Rev. B **68**, 052402 (2003).

[36] J. Stöhr, J. Magn. Magn. Mater. **200**, 470 (1999).

[37] Y. A. Bychkov and E. I. Rashba, J. Phys. C: Solid State Phys. **17**, 6039 (1984).

[38] I. Dzialoshinski, J. Phys. Chem. Solids **4**, 241 (1958).

[39] T. Moriya, Phys. Rev. **120**, 91 (1960).

[40] S. LaShell, B. A. McDougall, and E. Jensen, Phys. Rev. Lett. **77**, 3419 (1996).

[41] M. Hoesch, M. Muntwiler, V. N. Petrov, M. Hengsberger, L. Patthey, M. Shi, M. Falub, T. Greber, and J. Osterwalder, Phys. Rev. B **69**, 241401(R) (2004).

[42] J. Henk, A. Ernst, and P. Bruno, Phys. Rev. B **68**, 165416 (2003).

[43] G. Nicolay, F. Reinert, S. Hüfner, and P. Blaha, Phys. Rev. B **65**, 033407 (2001).

[44] E. Wimmer, H. Krakauer, M. Weinert, and A. J. Freeman, Phys. Rev. B **24**, 864 (1981).

[45] For program description see *http://www.flapw.de*.

[46] C. Li, A. J. Freeman, and C. L. Fu, J. Magn. Magn. Mater. **83**, 51 (1990).

[47] G. Bihlmayer, Y. M. Koroteev, P. M. Echenique, E. V. Chulkov, and S. Blügel, Surf. Sci. **600**, 3888 (2006).

[48] L. Petersen and P. Hedegård, Surf. Sci. **459**, 49 (2000).

[49] S. Agergaard, C. Søndergaard, H. Li, M. B. Nielsen, S. V. Hoffmann, Z. Li, and P. Hofmann, New J. Phys. **3**, 15 (2001).

[50] J. Sun, A. Mikkelsen, M. F. Jensen, Y. M. Koroteev, G. Bihlmayer, E. V. Chulkov, D. L. Adams, P. Hofmann, and K. Pohl, Phys. Rev. B **74**, accepted (2006).

[51] C. R. Ast and H. Höchst, Phys. Rev. Lett. **87**, 177602 (2001).

[52] H. Mönig, J. Sun, Y. M. Koroteev, G. Bihlmayer, E. V. Chulkov, K. Pohl, and P. Hofmann, Phys. Rev. B **72**, 085410 (2005).

[53] J. I. Pascual, G. Bihlmayer, Y. M. Koroteev, H.-P. Rust, G. Ceballos, M. Hansmann, K. Horn, E. V. Chulkov, S. Blügel, P. M. Echenique, and P. Hofmann, Phys. Rev. Lett. **93**, 196802 (2004).

[54] P. Hofmann, J. E. Gayone, G. Bihlmayer, Y. M. Koroteev, and E. V. Chulkov, Phys. Rev. B **71**, 195413 (2005).

[55] Y. M. Koroteev, G. Bihlmayer, J. E. Gayone, E. V. Chulkov, S. Blügel, P. M. Echenique, and P. Hofmann, Phys. Rev. Lett. **93**, 046403 (2004).

[56] C. R. Ast and H. Höchst, Phys. Rev. Lett. **90**, 016403 (2003).

[57] T. Hirahara, T. Nagao, I. Matsuda, G. Bihlmayer, E. V. Chulkov, Y. M. Koroteev, P. M. Echenique, M. Saito, and S. Hasegawa, Phys. Rev. Lett. **97**, 146803 (2006).

[58] O. Krupin, G. Bihlmayer, K. Starke, S. Gorovikov, J. E. Prieto, K. Döbrich, S. Blügel, and G. Kaindl, Phys. Rev. B **71**, 201403 (R) (2005).

[59] G. Bihlmayer, E. V. Chulkov, and S. Blügel, Phys. Rev. B **75**, 195414 (2007).

[60] T. Hirahara, K. Miyamoto, A. Kimura, Y. Niinuma, G. Bihlmayer, E. V. Chulkov, T. Nagao, I. Matsuda, S. Qiao, K. Shimada, H. Namatame, M. Taniguchi, *et al.*, New J. Phys. **10**, 083038 (2008).

[61] H. Dil, F. Meier, J. Lobo-Checa, L. Patthey, G. Bihlmayer, and J. Osterwalder, Phys. Rev. Lett **101**, 266802 (2008).

[62] Y. M. Koroteev, G. Bihlmayer, E. V. Chulkov, and S. Blügel, Phys. Rev. B **77**, 045428 (2008).

[63] D. Hsieh, Y. Xia, L. Wray, D. Qian, A. Pal, J. H. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C. L. Kane, Y. S. Hor, R. J. Cava, *et al.*, Science **323**, 919 (2009).

[64] S. Blügel and P. H. Dederichs, Europhys. Lett. **9**, 597 (1989).

[65] S. Blügel, Europhys. Lett. **7**, 743 (1988).

[66] T. Asada and S. Blügel, Physica B **237-238**, 359 (1997).

[67] S. Blügel, D. Drittler, R. Zeller, and P. H. Dederichs, Appl. Phys. A **49**, 547 (1989).

[68] J. Redinger, S. Blügel, and R. Podloucky, Phys. Rev. B **51**, 13852 (1995).

[69] S. Blügel, Europhys. Lett. **18**, 257 (1992).

[70] S. Blügel, Phys. Rev. Lett. **68**, 851 (1992).

[71] B. Újfalussy, L. Szunyogh, and P. Weinberger, Phys. Rev. B **51**, 12836 (1995).

[72] A. J. Freeman and C. L. Fu, J. Appl. Phys. **61**, 3356 (1987).

[73] X. Nie, S. Heinze, G. Bihlmayer, and S. Blügel, *Ferromagnetism and Antiferromagnetism of 3d Transition Metal Monolayers on W(110).*

[74] M. Stampanoni, A. Vaterlaus, D. Pescia, M. Aeschlimann, F. Meier, W. Dürr, and S. Blügel, Phys. Rev. B **37**, 10380 (1988).

[75] J. E. Ortega and F. J. Himpsel, Phys. Rev. B **47**, 16441 (1993).

[76] S. Heinze, M. Bode, A. Kubetzka, O. Pietzsch, X. Nie, S. Blügel, and R. Wiesendanger, Science **288**, 1805 (2000).

[77] M. Bode, S. Heinze, A. Kubetzka, O. Pietzsch, M. Hennefarth, M. Getzlaff, R. Wiesendanger, X. Nie, G. Bihlmayer, and S. Blügel, Phys. Rev. B **66**, 014425 (2002).

[78] T. Flores, M. Hansen, and M. Wuttig, Surf. Sci. **279**, 251 (1992).

[79] D. Tian, H. Li, S. Wu, F. Jona, and P. Marcus, Phys. Rev. B **45**, 3749 (1992).

[80] S. Andrieu, H. Fischer, M.Piecuch, A. Traverse, and J. Mimault, Phys. Rev. B **54**, 2822 (1996).

[81] D. Tian, A. Begley, and F. Jona, Surf. Sci. Lett. **273**, 393 (1992).

[82] I. Grigorov and J. Walker, J. Appl. Phys. **81**, 3907 (1997).

[83] I. Grigorov, J. Walker, M. Hawley, G. B. adn M. Lütt, and M. Fitzsimmons, J. Appl. Phys. **83**, 7010 (1998).

[84] I. Grigorov, I.-L. Siu, M. Fitzsimmons, and J. Walker, Phys. Rev. Lett. **82**, 5309 (1999).

[85] A. Arrott, B. Heinrich, S. Purcell, J. Cochran, and L. Urquhart, J. Appl. Phys. **61**, 3721 (1987).

[86] K. Ounadjela, P. Vennegues, Y. Henry, A. Michel, V. Pierron-Bohnes, and J. Arabski, Phys. Rev. B **49**, 8561 (1994).

[87] L. Zhang, M. Kuhn, and U. Diebold, Surf. Sci. **371**, 223 (1997).

[88] M. Albrecht, J. Pohl, H. Wider, E. Malang, J. Kohler, K. Friemelt, and E. Bucher, Surf. Sci. **397**, 354 (1998).

[89] M. Sambi and G. Granozzi, Surf. Sci. **426**, 235 (1999).

[90] S. Heinze, P. Kurz, D. Wortmann, G. Bihlmayer, and S. Blügel, Appl. Phys. A **75**, 25 (2002).

[91] D. Wortmann, S. Heinze, P. Kurz, G. Bihlmayer, and S. Blügel, Phys. Rev. Lett **86**, 4132 (2001).

[92] M. Wuttig, C. C. Knight, T. Flores, and Y. Gauthier, Surf. Sci. **292**, 189 (1993).

[93] S. C. Wu, S. H. Lu, Z. Q. Wang, C. K. C. Lok, J. Quinn, Y. S. Li, D. Tian, F. Jona, and P. M. Marcus, Phys. Rev. B **38**, 5363 (1988).

[94] Z. Q. Wang, Y. S. Li, C. K. C. Lok, J. Quinn, F. Jona, and P. M. Marcus, Solid State Commun. **62**, 181 (1987).

[95] M. Wuttig, Y. Gauthier, and S. Blügel, Phys. Rev. Lett. **70**, 3619 (1993).

[96] O. Rader, W. Gudat, C. Carbone, E. Vescovo, S. Blügel, R. Kläsges, W. Eberhardt, M. Wuttig, J. Redinger, and F. J. Himpsel, Phys. Rev. B **55**, 5404 (1997).

[97] X. Nie, G. Bihlmayer, and S. Blügel, *Relativistic total energy magnetic anisotropy calculations for transition metals: free-standing monolayers*.

[98] F. J. A. den Broeder, D. Kuiper, H. C. Donkersloot, and W. Hoving, Appl. Phys. A **49**, 507 (1989).

[99] C. J. Lin, G. L. Gorman, C. H. Lee, R. F. C. Farrow, E. E. Marinero, H. V. Do, H. Notarys, and C. J. Chien, J. Magn. Magn. Mater. **93**, 194 (1991).

[100] K. Baberschke, Appl. Phys. A **62**, 417 (1996).

[101] A. Fert, Materials Science Forum **59-60**, 439 (1990).

[102] I. E. Dzyaloshinskiĭ, Sov. Phys. JETP **20**, 665 (1965).

[103] M. Heide, G. Bihlmayer, and S. Blügel, Physica B, submitted (2009).

[104] M. Bode, M. Heide, K. von Bergmann, P. Ferriani, S. Heinze, G. Bihlmayer, A. Kubetzka, O. Pietzsch, S. Blügel, and R. Wiesendanger, Nature (London) **447**, 190 (2007).

A 4 Highly Correlated Electron Materials: Dynamical Mean Field Theory

Ansgar Liebsch
Institut für Festkörperforschung
Forschungszentrum Jülich

Contents

1 Introduction **2**

2 Dynamical Mean Field Theory **4**

3 Multi-Band Exact Diagonalization **6**

4 $Ca_{2-x}Sr_xRuO_4$ **9**

5 $SrVO_3$ and $LaTiO_3$ **12**

6 $La_{1-x}Sr_xTiO_3$ **14**

7 V_2O_3 **15**

8 $Na_{0.3}CoO_2$ **16**

9 DMFT for Heterostructures **18**

10 Cluster DMFT: Organic Salts **20**

11 Summary **23**

1 Introduction

One of the hallmarks of strongly correlated electron materials is their extreme sensitivity to small changes of parameters such as temperature, pressure, structural distortion, or impurity concentration [1]. This sensitivity is the result of a striking competition between kinetic energy associated with electron hopping in narrow bands and intra-atomic Coulomb repulsion. It gives rise to a fascinating range of electronic and magnetic phenomena whose understanding has been at the focus of fundamental research in condensed matter physics for many years. In addition, strongly correlated electron systems, in particular, transition metal oxides, are presently of great interest because of the possibility of utilizing them for technological applications.

An example which illustrates the range of phenomena that are observed due to small changes of external parameters is $Ca_{2-x}Sr_xRuO_4$. It has the same layer perovskite structure as the famous high-T_c superconductors, with transition metal ions at the center of oxygen octahedra (see Fig. 1, left panel). Although the replacement of Sr via Ca is iso-electronic, the structural distortions caused by the smaller size of Ca indicated in the right panel give rise to a remarkably rich phase diagram (see Fig. 2, left panel): The pure Sr ruthenate ($x = 2$) is superconducting with $T_c = 1.5$ K, whereas the pure Ca compound ($x = 0$) is a paramagnetic or anti-ferromagnetic insulator. At finite $x < 0.5$, other magnetic phases are observed, while the range $x > 0.5$ is paramagnetic.

Density functional theory (DFT) in the local density approximation (LDA) predicts Ca_2RuO_4 to be metallic rather than insulating. This failure is quite typical for many transition metal oxides, including V_2O_3 and $LaTiO_3$. It is caused by the inadequate description of the Coulomb interaction within the partially filled d electron shell. This interaction is comparable or larger

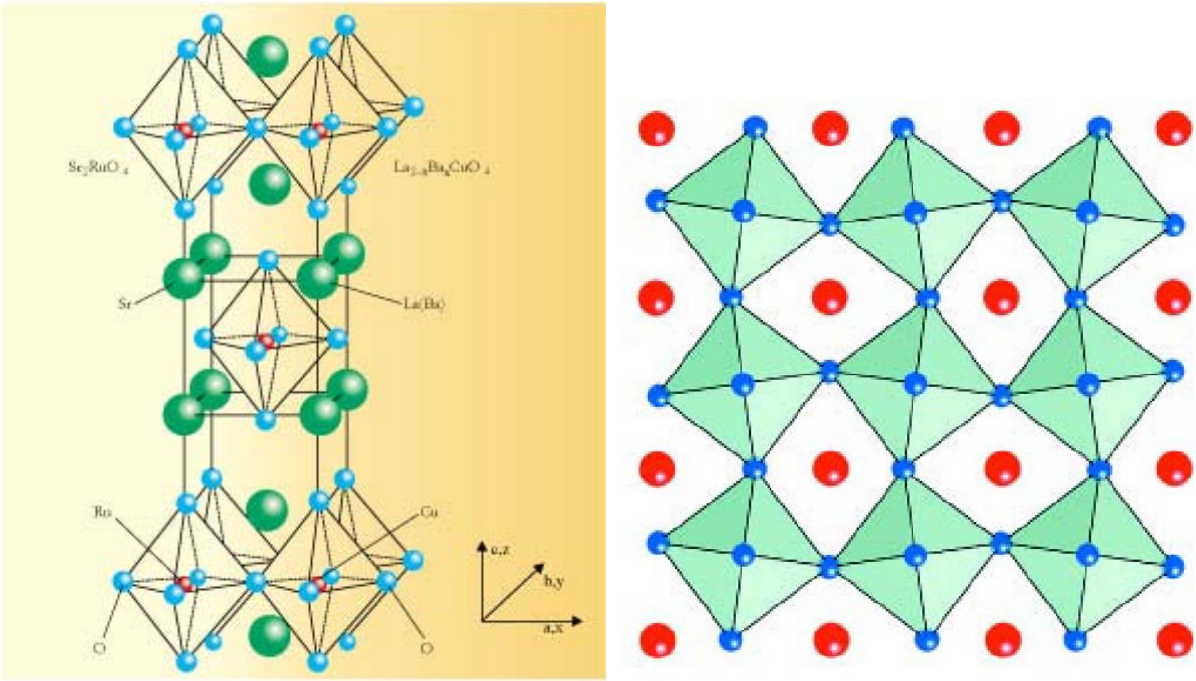


Fig. 1: *Left: Layer perovskite structure of quasi-two-dimensional ruthenate Sr_2RuO_4 or cuprate Ba_2CuO_4. Right: structural distortions in x/y plane of ruthenate due to substitution of Sr via Ca; blue dots: O^{2-} ions, red dots: Ca/Sr^{2+} ions, Ru^{4+} ions are at the center of O octahedra.*

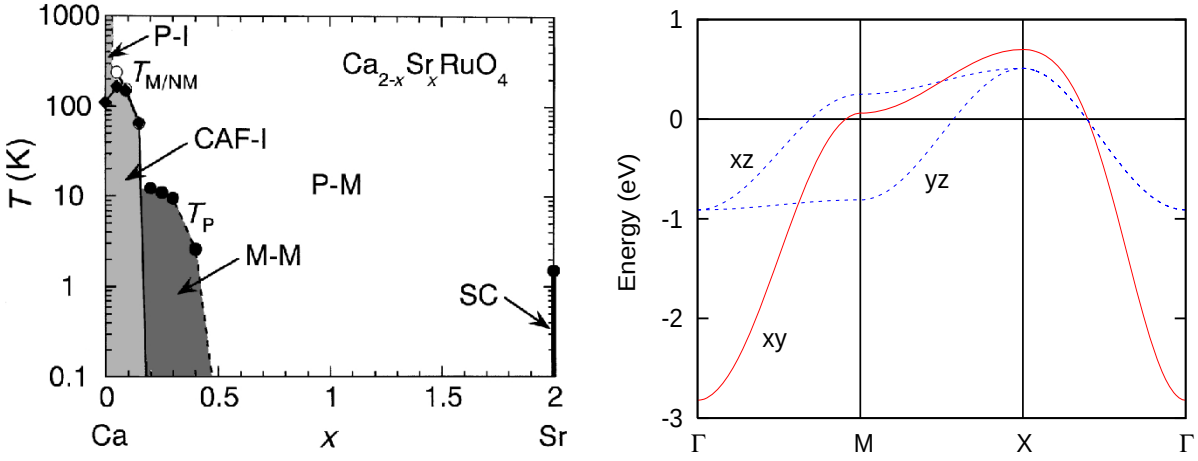


Fig. 2: *Left: Iso-electronic phase diagram of $Ca_{2-x}Sr_xRuO_4$, with superconducting (SC) phase for $x = 2$, paramagnetic and canted anti-ferromagnetic insulating phases (PI and CAFI) for $x \to 0$ [2]. Right: wide d_{xy} and narrow $d_{xz,yz}$ Ru t_{2g} bands of Sr_2RuO_4 [3].*

than the hopping interaction between atoms and cannot be simply expressed in terms of the local electron density. In some systems, the combination of LDA with an onsite Coulomb interaction – the so-called LDA+U method [4] – can explain the insulating behavior observed in long-range anti-ferromagnetic phases (see lectures by G. Bihlmayer). In general, however, the transition from metal to insulator requires a more refined formalism that accounts explicitly for quantum mechanical fluctuations within electron shells. The possibility of a purely correlation driven transition between paramagnetic metallic and insulating phases was first discussed by Mott about 60 years ago [5]. A theoretical formulation of such a transition in realistic materials, where metallic and insulating properties are treated on the same footing, can be achieved via the Dynamical Mean Field Theory (DMFT) [6, 7, 8] which is the subject of this lecture.

The unit cells of strongly correlated materials tend to have complex shapes, with several different types of atoms (see Fig. 1). An important first step towards a theoretical description is therefore the detailed one-electron calculation of the electronic properties within the LDA. There exist nowadays various codes which provide sophisticated information on many of these properties. Since many-body calculations become exponentially more time-consuming with increasing number of orbitals, the second step consists in identifying those partially filled bands near the Fermi level E_F in which Coulomb interactions lead to the most dramatic effects. In the case of Sr_2RuO_4, these are the three Ru $4d$ t_{2g} bands where, because of the layer structure, d_{xy} is a nearly two-dimensional wide band and $d_{xz,yz}$ are nearly one-dimensional narrow bands, all of which are about 2/3 filled (total filling is $n = 4$), see Fig. 2 right panel. Because of the octahedral crystal field, the e_g bands ($d_{x^2-y^2}$ and d_{z^2}) are empty. Although the O $2p$ bands are filled, it is important to keep in mind that hopping between Ru $4d$ orbitals takes place indirectly via O ions which are located between Ru ions. Sr $5s$ orbitals form empty bands far above E_F. The bands close to E_F can therefore be described in terms of an effective single-particle Hamiltonian $H(\mathbf{k})$ (in the present case a 3×3 matrix) with band energies $\epsilon_m(\mathbf{k})$, where m is the band index.

The main changes of the independent-particle band structure due to Coulomb interactions are shifts of band energies giving rise to band narrowing, life-time broadening, and new excited states (satellites or Hubbard bands) that do not occur within the single-electron picture. Thus,

correlations cause a transfer of spectral weight from low to high energies. Another important consequence of Coulomb interactions occurring in multi-band materials is that orbital occupancies of interacting bands can differ from those of non-interacting bands. As discussed in [9], in Sr_2RuO_4 this effect arises since the narrow $d_{xz,yz}$ subbands are more strongly correlated than the wide d_{xy} band. Of course, the total number of electrons does not change due to Coulomb effects, but how these electrons are distributed over the various conduction bands depends on how they interact. Thus, Coulomb correlations can lead to inter-orbital charge transfer, i.e., to an increase or decrease of orbital polarization. In the following sections we discuss this kind of correlation induced internal charge transfer in a variety of transition metal oxides.

The phenomena discussed above can be theoretically formulated in terms of a complex self-energy $\Sigma(\mathbf{k},\omega)$. The Green's function describing the electron motion through the crystal is

$$G(\mathbf{k},\omega) \;=\; \frac{1}{\omega+\mu-H(\mathbf{k})-\Sigma(\mathbf{k},\omega)}, \tag{1}$$

where μ is the chemical potential insuring the correct total electron count. Neglecting matrix element effects associated with the frequency and polarization of the incoming light, the imaginary part of this Green's function is proportional to the photoemission intensity for the point $\mathbf{k}$ of the Brillouin Zone. For clarity we omit here orbital indices. Thus all quantities are assumed to be matrices in orbital space. The spin index is also dropped since we consider only paramagnetic systems. Because of the translational symmetry of the crystal, the lattice Fourier transform of G is defined as

$$G_{ij}(\omega) \;=\; \sum_{\mathbf{k}} e^{i\mathbf{k}(\mathbf{R}_i-\mathbf{R}_j)}\frac{1}{\omega+\mu-H(\mathbf{k})-\Sigma(\mathbf{k},\omega)}, \tag{2}$$

so that the so-called 'local' Green's function is given by the expression

$$G(\omega) \;\equiv\; G_{00}(\omega) \;=\; \sum_{\mathbf{k}} \frac{1}{\omega+\mu-H(\mathbf{k})-\Sigma(\mathbf{k},\omega)}. \tag{3}$$

Analogous equations can be written down for the self-energy. Since $H(\mathbf{k})$ is known, the main task is now to find some (approximate) scheme for the evaluation of $\Sigma(\mathbf{k},\omega)$.

2 Dynamical Mean Field Theory

The great appeal of DMFT is that it is applicable at weak and strong Coulomb interactions, and that it describes metallic and insulating behavior in a consistent manner. Let us consider the Hubbard model with on-site Coulomb interactions:

$$\begin{aligned} H \;=\; & -\sum_{ijmn\sigma} t_{imjn}c^{+}_{im\sigma}c_{jn\sigma} + \sum_{im} U n_{im\uparrow}n_{im\downarrow} + \frac{1}{2}\sum_{im\neq n\sigma\sigma'}(U'-J\delta_{\sigma\sigma'})n_{im\sigma}n_{in\sigma'} \\ & -\sum_{im\neq m'} J[\,c^{+}_{im\uparrow}c_{im\downarrow}c^{+}_{im'\downarrow}c_{im'\uparrow} + c^{+}_{im\uparrow}c^{+}_{im\downarrow}c_{im'\uparrow}c_{im'\downarrow}], \end{aligned} \tag{4}$$

where m,n are orbital indices, i,j denote lattice sites $\mathbf{R}_{i,j}$, and t_{imjn} is the lattice Fourier transform of $H(\mathbf{k})_{mn}$. U and U' are intra- and inter-orbital Coulomb matrix elements and J is the Hund exchange integral. Below we consider mainly systems involving t_{2g} orbitals.

The key conceptual feature of DMFT is that, instead of solving the hopelessly complicated Hamiltonian defined above, a simpler version is considered where Coulomb interactions are retained only at one site (say at $i = 0$) while at all other sites they are replaced by the local self-energy. Thus the interacting lattice is approximated by a single interacting site which is surrounded by a lattice with an extra complex potential given by $\Sigma(\omega)$. The appropriate Green's function G' for this 'impurity' problem can be derived by using the Dyson equation which removes $\Sigma(\omega)$ from the origin. Thus,

$$G'_{ij}(\omega) \;=\; G_{ij}(\omega) - G_{i0}(\omega)\Sigma_{00}(\omega)G'_{0j}(\omega). \tag{5}$$

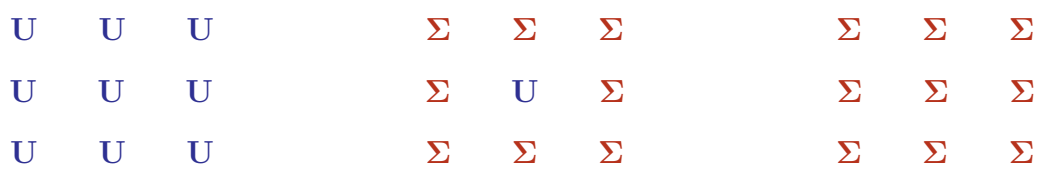
For the special case $i = j = 0$, and with the abbreviations $\Sigma(\omega) \equiv \Sigma_{00}(\omega)$, $G(\omega) \equiv G_{00}(\omega)$, and $G_0(\omega) \equiv G'_{00}(\omega)$, we find the important relationships

$$G_0(\omega) \;=\; [G^{-1}(\omega) + \Sigma(\omega)]^{-1}, \tag{6}$$

$$\Sigma(\omega) \;=\; G_0^{-1}(\omega) - G^{-1}(\omega). \tag{7}$$

The physical meaning of $G_0(\omega)$ is the following: It describes electronic motion from site $i = 0$ through the actual crystal and returning to $i = 0$, with one-electron hopping specified by the coefficients t_{imjn}. This motion, however, takes place within an extra complex potential given by $\Sigma(\omega)$ at all sites except $i = 0$.

Graphically we can represent the above approximation as follows:



The true interacting lattice on the left is simulated via the single-site interacting impurity at the center. The spectral information to be compared with photoemission data then follows from the lattice on the right, where all sites have a complex local self-energy. The picture at the center demonstrates the importance of removing the local self-energy from the origin before starting the many-body impurity calculation. If $G(\omega)$ were used in this step rather than $G_0(\omega)$, one would add the Coulomb terms involving U to the local $\Sigma(\omega)$, which would amount to a severe double-counting of Coulomb interactions.

How do we find the so far unknown self-energy for the impurity calculation at the center? This is done iteratively, by starting with some reasonable guess, or by putting $\Sigma(\omega) = 0$. We then proceed via the following steps:

1. calculate $G(\omega)$ via Eq. (3)
2. calculate $G_0(\omega)$ via Eq. (6)
3. calculate new $G(\omega)$ via impurity solver (see below)
4. calculate new $\Sigma(\omega)$ via Eq. (7) and return to step 1.

Typically 10 to 20 iterations are required to achieve convergence, except close to metal insulator transitions where convergence tends to be slower.

For the quantum impurity calculation a variety of methods is available, such as quantum Monte Carlo (QMC), exact diagonalization (ED), numerical renormalization group (NRG), density

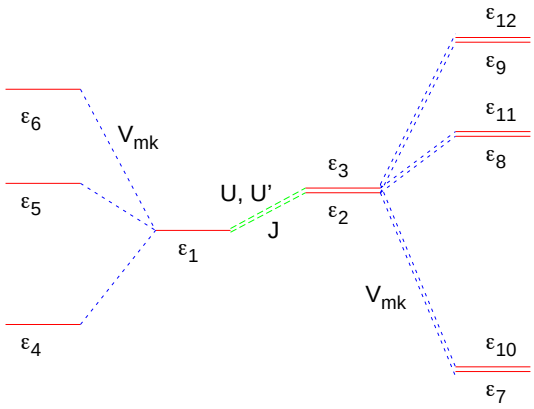


Fig. 3: *Level scheme for exact diagonalization [11]. $\varepsilon_{1,2,3}$ denote the t_{2g} orbitals at the impurity site which interact via Coulomb and exchange interactions. Each of these levels hybridizes with its own set of bath levels. Levels 2 and 3 and their baths are assumed to be degenerate.*

matrix renormalization group (DMRG), iterated perturbation theory (IPT), non-crossing approximation (NCA), etc. The first two are the most accurate and versatile ones for realistic materials. They are complementary in the sense that their range of applicability and/or computational tractability differs somewhat. Where they overlap they have been shown to be in excellent numerical agreement. Here we focus on exact diagonalization [10] which during the recent years has been demonstrated to be a highly useful and efficient method for a variety of multi-band systems [11]. Compared to QMC it does not suffer from sign problems. Thus, it is applicable at rather low temperatures, large Coulomb energies, and for full Hund exchange.

In the procedure outlined above we have effectively replaced the true lattice self-energy $\Sigma(\mathbf{k},\omega)$ by its local version $\Sigma(\omega)$ which may be viewed as lowest order term in a lattice site expansion of $\Sigma(\mathbf{k},\omega)$. To go beyond this 'single-site' or 'local' approximation one needs to retain Coulomb interactions explicitly in a (small) cluster of sites rather than only at $i = 0$. Although the many-body problem then becomes computationally much more involved, there is currently considerable interest in cluster extensions of DMFT since the momentum dependence of the self-energy can have a significant influence on the nature of the metal insulator transition (see Section 10).

3 Multi-Band Exact Diagonalization

To solve the single-site quantum impurity problem within ED/DMFT, the true lattice environment of the impurity at $i = 0$ is simulated via a discrete set of non-interacting 'bath' levels. Thus, instead of H defined in Eq. (4) we consider the finite 'cluster' sketched in Fig. 3:

$$H^{cl} = \sum_{m\sigma} \varepsilon_m n_{m\sigma} + \sum_{k\sigma} \varepsilon_k n_{k\sigma} + \sum_{mk\sigma} V_{mk}[c^+_{m\sigma} c_{k\sigma} + \text{H.c.}] + \sum_m U n_{m\uparrow} n_{m\downarrow}$$
$$+ \frac{1}{2} \sum_{m\neq m'\sigma\sigma'} (U' - J\delta_{\sigma\sigma'}) n_{m\sigma} n_{m'\sigma'} - \sum_{m\neq m'} J[\, c^+_{m\uparrow} c_{m\downarrow} c^+_{m'\downarrow} c_{m'\uparrow} + c^+_{m\uparrow} c^+_{m\downarrow} c_{m'\uparrow} c_{m'\downarrow}]. \quad (8)$$

The levels representing the impurity orbitals are ε_m, the levels of the surrounding bath are ε_k, and the V_{mk} specify the hybridization interactions between impurity and bath levels. The

remaining Coulomb and exchange interactions at the impurity site are identical to those of the original lattice Hamiltonian in Eq. (4). The site index $i = 0$ is dropped for convenience. The total number of cluster levels is denoted as n_s. (The term 'cluster' refers here in ED to impurity + bath and should not be confused with cluster extensions of DMFT.)

The non-interacting cluster Green's function is given by

$$G_0^{cl} = (i\omega_n + \mu - H_0^{cl})^{-1}, \tag{9}$$

where H_0^{cl} represents the first three terms of H^{cl}. Since we are interested in correlation effects at finite temperatures we evaluate Green's functions and self-energies at Matsubara frequencies $\omega_n = (2n+1)\pi/\beta$, where $n \geq 0$ and $\beta = 1/T$ [12]. If G_0^{cl} is diagonal in orbital space this expression can be easily reduced to

$$G_{0,m}^{cl}(i\omega_n) = \Big(i\omega_n + \mu - \varepsilon_m - \sum_k \frac{|V_{mk}|^2}{i\omega_n + \mu - \varepsilon_k}\Big)^{-1}. \tag{10}$$

The energy levels and hopping terms appearing in the cluster Hamiltonian Eq. (8) do not have any physical meaning. Their sole purpose is to achieve an accurate 'cluster' representation of the corresponding lattice impurity Green's function $G_{0,m}(i\omega_n)$. A standard conjugate gradient fitting routine can be used to find the cluster parameters, such that $G_{0,m}(i\omega_n) \approx G_{0,m}^{cl}(i\omega_n)$, where we assume again diagonality in orbital space.

Once the non-interacting cluster parameters are found, the eigenvalues E_ν and eigenstates $|\nu\rangle$ of the many-body cluster Hamiltonian are evaluated via exact diagonalization. The finite temperature interacting cluster Green's function is given by [12]

$$\begin{aligned} G_m^{cl}(i\omega_n) &= \frac{1}{Z}\sum_{\nu\mu} \frac{|\langle\mu|c_{m\sigma}^+|\nu\rangle|^2}{E_\nu - E_\mu + i\omega_n}[e^{-\beta E_\nu} + e^{-\beta E_\mu}] \\ &= \frac{1}{Z}\sum_\nu e^{-\beta E_\nu}\Big(\sum_\mu \frac{|\langle\mu|c_{m\sigma}^+|\nu\rangle|^2}{(E_\nu - E_\mu) + i\omega_n} + \sum_\mu \frac{|\langle\mu|c_{m\sigma}|\nu\rangle|^2}{(E_\mu - E_\nu) + i\omega_n}\Big), \end{aligned} \tag{11}$$

where $Z = \sum_\nu e^{-\beta E_\nu}$ is the partition function. Applying Eq. (7) to the cluster, we find

$$\Sigma_m^{cl}(i\omega_n) = G_{0,m}^{cl}(i\omega_n)^{-1} - G_m^{cl}(i\omega_n)^{-1}. \tag{12}$$

The key assumption is now that this 'cluster' self-energy is a physically reasonable representation of the lattice self-energy, $\Sigma_m^{cl}(i\omega_n) \approx \Sigma_m(i\omega_n)$, which can then be inserted in $G_m(i\omega_n)$. Evidently, at each iteration step two projections are carried out: (i) The lattice impurity Green's function G_0 is projected onto the corresponding cluster Green's function G_0^{cl}, and (ii) the cluster self-energy Σ^{cl} is projected onto the lattice self-energy Σ. In ED, each iteration therefore involves the following steps:

$$\mathbf{\Sigma \to G \to G_0 \to G_0^{cl} \to G^{cl} \to \Sigma^{cl} \to \Sigma}. \tag{13}$$

To illustrate the quality of the projection $G_0 \to G_0^{cl}$ we show in Fig. 4 $\mathrm{Im}\, G_{0,m}(i\omega_n)$ obtained from

$$G_{0,m}(i\omega_n) = \int d\omega\ \rho_m(\omega)/(i\omega_n - \omega), \tag{14}$$

which corresponds to Eq. (6) in the non-interacting limit $\Sigma = 0$ for diagonal G. [Note that $\mathrm{Im}\, G_{0,m}(i\omega_n \to 0) \to -i\pi\rho_m(0)$ and $\mathrm{Im}\, G_{0,m}(i\omega_n \to \infty) \to 1/i\omega_n$.] As density of states

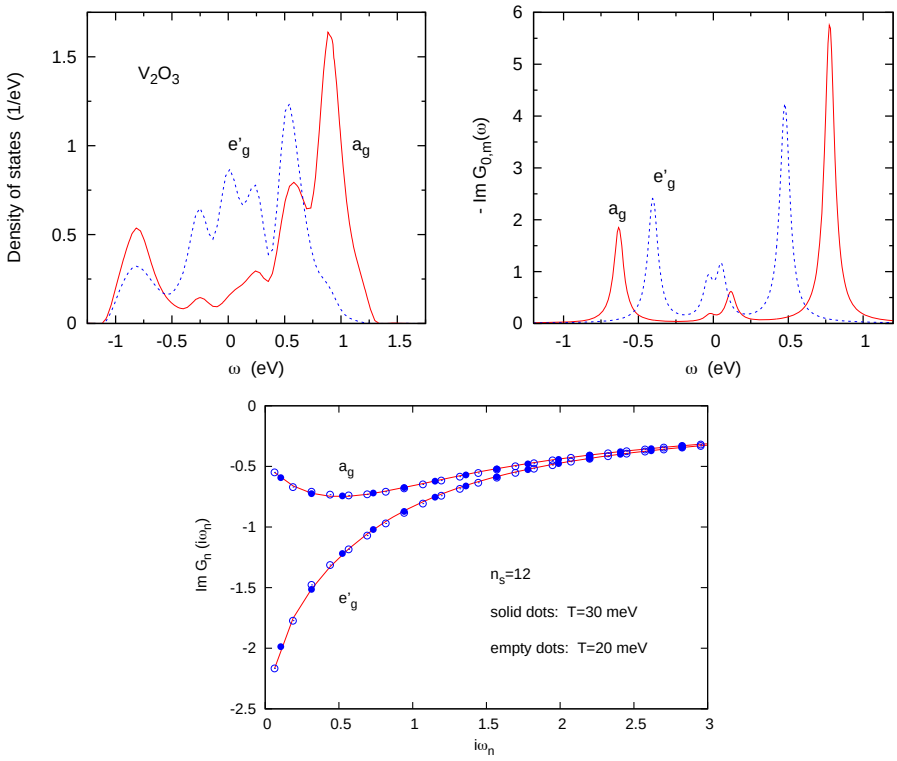


Fig. 4: *Upper left:* a_g *and* e'_g *density of states components of* V_2O_3 *[13]; upper right: spectral distributions at real* ω *derived from Eq. (10) after fit along Matsubara axis. Bottom: comparison of Green's functions at imaginary frequencies for two temperatures. Solid red curves:* $\mathrm{Im}\, G_{0,m}(i\omega_n)$*, Eq. (14); blue dots:* $\mathrm{Im}\, G^{cl}_{0,m}(i\omega_n)$*, Eq. (10).*

$\rho_m(\omega)$ we take the a_g and e'_g components corresponding to V_2O_3 [13] (see Section 7). These lattice impurity Green's functions $G_{0,m}(i\omega_n)$ are compared with fits achieved via Eq. (10), where each impurity orbital is assumed to hybridize with three bath levels, as sketched in Fig. 3. At real frequencies $G_{0,m}$ has a continuous spectral distribution given by the density of states component $\rho_m(\omega)$, whereas $G^{cl}_{0,m}$ is discrete by construction (a small artificial broadening is included for illustrative purpose). Nevertheless, along the Matsubara axis both are seen to be in excellent agreement. This comparison demonstrates that, for finite clusters, the representation at Matsubara frequencies is not unique. In fact, within a certain accuracy, $G_{0,m}(i\omega_n)$ derived from a continuous real-ω spectrum can be represented via an infinite number of discrete spectra corresponding to different cluster sizes n_s. Because of the projections $lattice \rightarrow impurity \rightarrow lattice$ indicated in Eq. 13, continuous lattice and discrete cluster versions of Green's function and self-energies are assumed to exist at each iteration of the ED/DMFT procedure (see also Sections 4 and 8).

Because of the Boltzmann factor in Eq. (11), at low T only a small number of eigenstates of H^{cl} are needed. Moreover, H^{cl} is block-diagonal in spin sectors, consisting of extremely sparse submatrices. The eigenstates can therefore be efficiently evaluated by using the Arnoldi algorithm [14]. Basis vectors are of the form $|n_{1\uparrow},\ldots,n_{n_s\uparrow},n_{1\downarrow},\ldots,n_{n_s\downarrow}\rangle$, where $n_{i\sigma}=0$ or 1. For $n_s=12$, the largest spin sector corresponds to $n_\uparrow=n_\downarrow=6$, yielding matrix dimension $N=d(n_\uparrow)\times d(n_\downarrow)=924^2=853776$, where $d(n_\sigma)=n_s!/[n_\sigma!(n_s-n_\sigma)!]$. The sums over

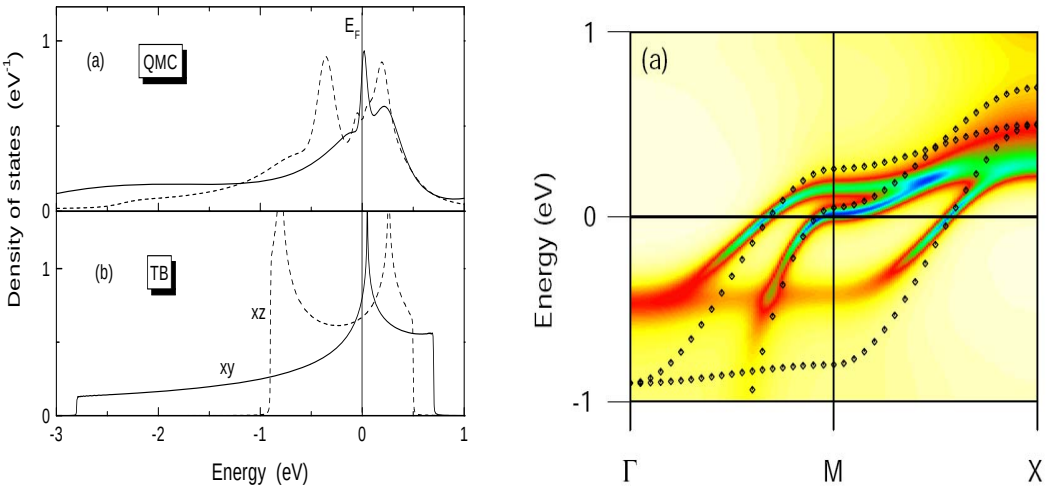


Fig. 5: *Left: Comparison of Sr_2RuO_4 quasi-particle spectra derived using QMC/DMFT with tight-binding density of states. Solid (dashed) curves: d_{xy} ($d_{xz,yz}$) distributions. Right: comparison of quasi-particle band structure with tight-binding bands (symbols) [9].*

μ in Eq. (11) can be readily obtained via a Lanczos procedure after applying $c_{m\sigma}$ and $c^+_{m\sigma}$ to the excited states $|\nu\rangle$. Most of the examples discussed in the subsequent sections are for t_{2g} bands, where each orbital couples to only two bath levels, giving $n_s = 9$ with $N = 126^2$. Since the baths of the impurity levels interact indirectly via the on-site Coulomb interaction, the level spacing of excited states is very small. Thus, convergence with cluster size is achieved far more quickly than in the single-band case which requires $3 \ldots 5$ bath levels. As a result, ED/DMFT can be used to investigate realistic multi-band materials. For $n_s = 9$, one iteration then takes only a few minutes. Additional details concerning the multi-band ED/DMFT approach can be found in Ref. [11].

4 $Ca_{2-x}Sr_xRuO_4$

Fig. 5 illustrates the typical modifications of the one-electron bands of Sr_2RuO_4 due to Coulomb interactions. There is an overall narrowing of bands by about a factor of two which is consistent with data obtained using angle-resolved photoemission spectroscopy (ARPES) [15]. Also, due to their finite lifetime the quasi-particle states acquire intrinsic broadening. Moreover, orbital occupancies differ from those of the bare tight-binding bands. As discussed in Ref. [9], charge is transferred from the more strongly correlated narrow $d_{xz,yz}$ bands to the wider d_{xy} band, implying correlation induced enhancement of orbital polarization as a result of the planar geometry. The d_{xy} van Hove singularity at M is therefore shifted very close to E_F. While in LDA it is about 50 meV above E_F [3], correlations reduce this value to about 10 meV. Thus, the topology of the Fermi surface remains the same as predicted in LDA and confirmed in de Haas-van Alphen (dHvA) measurements. If the van Hove singularity would sink below E_F, the d_{xy} Fermi surface would change from electron-like to hole-like, in contradiction to the dHvA data. Similar Coulomb driven internal charge redistributions are found in the systems discussed in the following sections.

The topology of the Fermi surface of Sr_2RuO_4 was controversial for several years since early photoemission data had observed the d_{xy} van Hove singularity below E_F. This conflict was resolved when it was demonstrated [16] that the surface layer of Sr_2RuO_4 exhibits a reconstruction similar to the distortion shown in the right panel of Fig. 1. This causes a downward shift of the d_{xy} band relative to the $d_{xz,yz}$ bands which is detected in surface sensitive ARPES measurements. A similar reconstruction takes place when Sr is replaced by Ca at $x = 0.5$, with the d_{xy} van Hove singularity below E_F, as observed in ARPES [17].

As pointed out in the Introduction, Sr_2RuO_4 is metallic, but iso-electronic Ca_2RuO_4 is insulating. To understand this striking difference, it is necessary to take into account the structural distortions that occur when Sr ions are replaced by the smaller Ca ions. Similar to the surface reconstruction mentioned above, this substitution induces rotations of Oxygen octahedra, and for $x < 0.5$ also tilting and flattening of octahedra. As discussed in Ref. [18], these deformations cause hybridization among the Ru t_{2g} bands, an effective narrowing of the main d_{xy} component, and, most importantly, a lowering of the d_{xy} band due to inter-orbital charge transfer from $d_{xz,yz}$ to d_{xy}. This structure induced splitting is denoted here as $\Delta = \varepsilon_{xz,yz} - \varepsilon_{xy}$. Thus, $\Delta = 0$ for the pure Sr compound ($x = 2$) and $\Delta \approx 0.4 \ldots 0.5$ eV for Ca_2RuO_4. Note that the Ca induced splitting enhances the crystal field splitting $\Delta_0 \approx 0.26$ eV present already at $x = 0$ as a result of the planar geometry [9].

Fig. 6 illustrates the enhancement of Ru t_{2g} orbital polarization due to Coulomb correlations as calculated within ED/DMFT for Hund exchange $J = U/4$ and $U' = U - 2J$ [19]. $\Delta = 0$ yields only mild polarization enhancement. Thus, even at sizable Coulomb energies Sr_2RuO_4 remains metallic (see QMC/DMFT spectra in Fig. 5, left panel; ED/DMFT spectra support this conclusion). For $\Delta > 0$, however, orbital polarization increases strongly, until for $\Delta \approx 0.4 \ldots 0.5$ eV and $U \approx 3 \ldots 4$ eV, it becomes complete: $n_{xy} \to 1$ and $n_{xz,yz} \to 0.5$ (per spin), i.e., the d_{xy} band is pushed below E_F, while the $d_{xz,yz}$ bands are half-filled. The spectra shown in the lower panels suggest that at $U \approx 3$ eV Ca_2RuO_4 is metallic, whereas at $U \approx 4.5$ eV it is insulating, with a filled d_{xy} band and $d_{xz,yz}$ bands split into lower and upper Hubbard bands. Thus, because of the structural anisotropy of this layer material, Coulomb interactions modify the Ru t_{2g} valence bands in qualitatively different ways.

Since the insulating phase shown in the lower right panel of Fig. 6 is paramagnetic, it corresponds to the high-temperature PI phase for $x \to 0$ in the phase diagram presented in Fig. 2. For $T < 100$ K, a (canted) anti-ferromagnetic phase is observed which has been analyzed within the LDA+U method [20].

These results show that the metal insulator transition in multi-band systems can be quite complex. Similar combinations of filling or emptying of subbands, with other subbands becoming half-filled, are found in other transition metal oxides, such as $LaTiO_3$ [21] and V_2O_3 [13, 22] which are discussed in subsequent sections. Thus, in $Ca_{2-x}Sr_xRuO_4$ there does not appear to occur a so-called orbital-selective Mott transition, as had been proposed in [20]. In the presence of narrow and wide bands, it is conceivable that these bands undergo separate Mott transitions, at different critical Coulomb energies U_c. As a result of correlation driven charge transfer between subbands, however, enhancement of orbital polarization appears to dominate the metal insulator transition in $Ca_{2-x}Sr_xRuO_4$. On the other hand, in the special case of half-filled particle-hole symmetric bands, orbital polarization is precluded. Narrow and wide subbands then indeed exhibit sequential, 'orbital selective' transitions, implying an intermediate phase in which bad-metallic behavior in the wide band coexists with insulating behavior in the narrow band [23, 24]. Depending on the nature of the Coulomb interaction between these bands, both

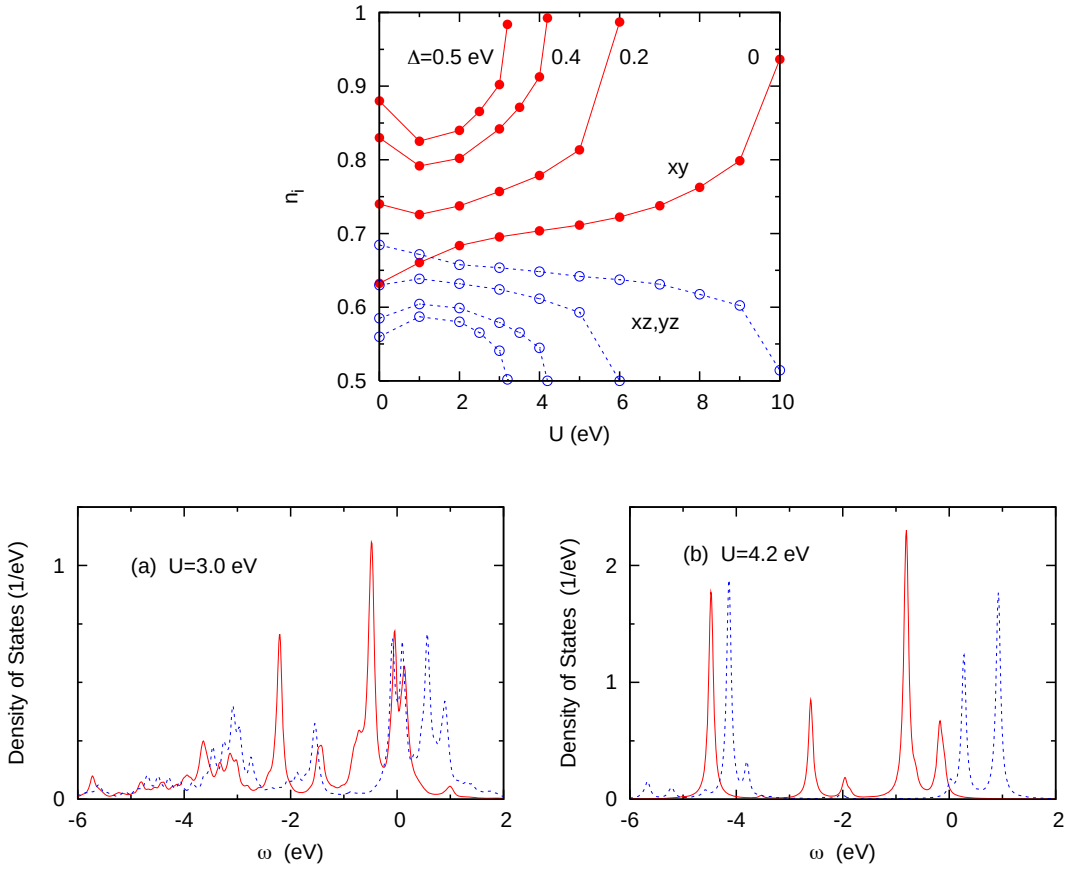


Fig. 6: *Top panel: Correlation induced change of orbital occupancies for $Ca_{2-x}Sr_xRuO_4$ [19]. Δ denotes the splitting between $d_{xz,yz}$ and d_{xy} levels due to structural distortions when Sr is replaced by Ca. Bottom panels: ED/DMFT spectral distributions for $\Delta = 0.4$ eV; solid (dashed) curves: d_{xy} ($d_{xz,yz}$) components; left: metallic phase; right: insulating phase.*

transitions can be first-order, or a combination can occur of first-order transition for the narrow band and a $T = 0$ quantum phase transition for the wide band [25].

The spectral distributions shown in Fig. 6 are obtained from the interacting cluster Green's function, Eq. (11), at real ω. (For illustrative purposes, a small artificial broadening is included.) To distinguish metallic from insulating phases, the inspection of cluster spectra is sufficient. To compare with photoemission data, it would be desirable to generate the equivalent lattice spectra via Eq. (3). This can be achieved via analytical continuation of $G_m(i\omega_n) \approx G_m^{cl}(i\omega_n)$ or $\Sigma_m(i\omega_n) \approx \Sigma_m^{cl}(i\omega_n)$ to real ω. The latter is preferable since one can then use $\Sigma_m(\omega)$ directly in Eq. (3), thereby avoiding the back-transformation of single-particle features stemming from $H(\mathbf{k})$. As discussed in Section 3, the extrapolation from Matsubara frequencies to real ω is not unique. Nevertheless, in principle, one could derive from the discrete cluster spectra shown in Fig. 6 the equivalent continuous spectra, in analogy to the non-interacting example given in Fig. 4. In the case of QMC/DMFT, real-ω spectra are usually generated via the maximum entropy method [26] which accounts for statistical uncertainties of the calculated results. In ED/DMFT inaccuracies originate from the small finite size of the bath surrounding the quantum impurity. Extrapolations of $\Sigma_m(i\omega_n)$ or $G_m(i\omega_n)$ to real ω can be done, for example, using the routine '*ratint*' [27], as discussed in [11] for Na_xCoO_2 (see Section 8).

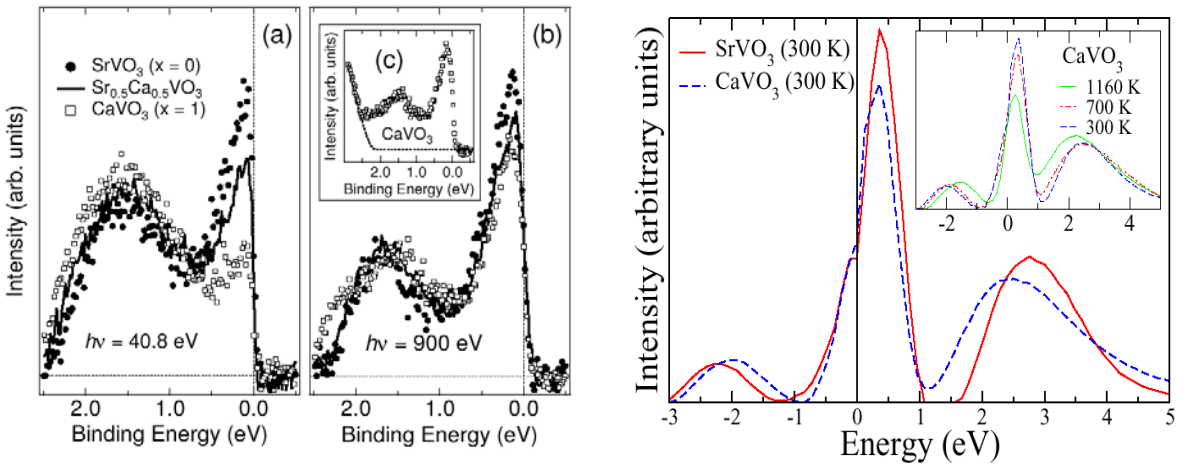


Fig. 7: *Left: Comparison of photoemission spectra for $SrVO_3$ and $CaVO_3$ at low and high photon energies [28]. Right: QMC/DMFT spectra for $SrVO_3$ and $CaVO_3$ [29].*

5 $SrVO_3$ and $LaTiO_3$

In the preceding section we have seen that Sr_2RuO_4 and Ca_2RuO_4 have fundamentally different electronic properties, even though both have four electrons in the Ru $4d$ t_{2g} valence shell. Evidently small differences of single-particle interactions due to structural distortions can lead to qualitatively different results once strong local Coulomb interactions are taken into account, with metallic (superconducting) behavior in one limit and paramagnetic or anti-ferromagnetic insulating behavior in the opposite limit.

Analogous qualitative differences exist in $3d^1$ materials, such as $SrVO_3$ and $LaTiO_3$ [21]. The former system is a cubic perovskite. Thus all six V t_{2g} spin bands are perfectly degenerate, with occupancy 1/6. Coulomb correlations do not affect this degeneracy, i.e., orbital polarization remains zero. Thus, although the on-site Coulomb energy of V ions is much larger than the single-particle band width ($U \approx 5.55$ eV, $J \approx 1.0$ eV, compared to $W \approx 2.5$ eV), the only effects that occur due to correlations are band narrowing of the quasi-particle peak near E_F, broadening of quasi-particle states due to finite lifetime, and satellite formation associated with lower and upper Hubbard bands [29].

Fig.7 shows these effects for $SrVO_3$ and $CaVO_3$. The latter exhibits slight deviations from cubic symmetry due to the smaller size of Ca ions compared to Sr. The occupied part of the spectrum can be compared with the photoemission data shown in the left panel. The unoccupied range can be compared to inverse photoemission spectra. A crucial point here is to be aware of the surface sensitivity of photoemission data. As the experimental data indicate, bulk-like spectra taken at high photon energies tend to be less correlated than data at low energies which contain more surface contributions due to the shorter electronic escape depth. As discussed in [30] the local density of states at the surface of $SrVO_3$ is effectively narrowed compared to the bulk, as a result of the reduced coordination. (This band narrowing affects the $d_{xz,yz}$ bands more than the mainly intra-planar d_{xy} band.) Thus, surface quasi-particle distributions calculated within QMC/DMFT [30] exhibit more pronounced narrowing of the main peak near E_F and stronger Hubbard bands, in agreement with the experimental spectra shown in Fig. 7. A detailed discussion of correlation effects at $SrVO_3$ surfaces is given in [31].

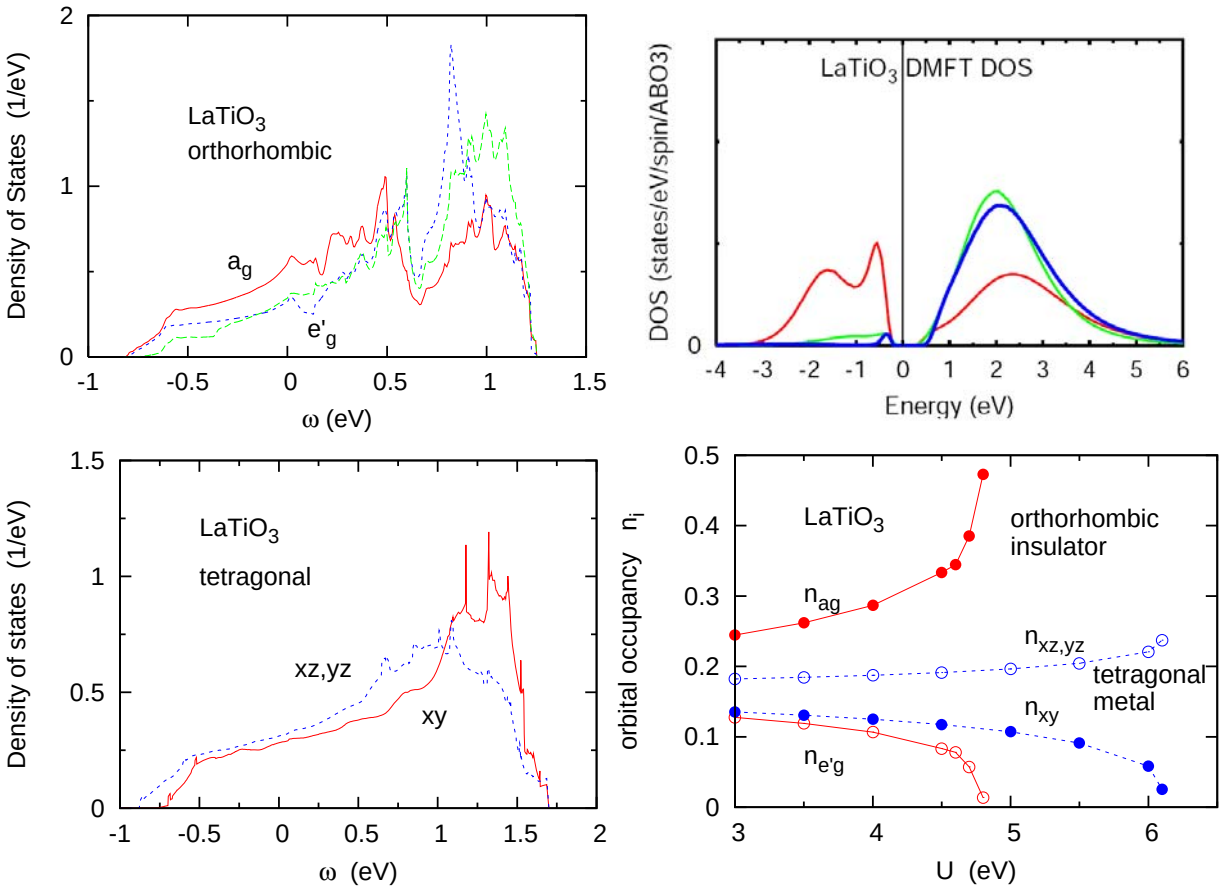


Fig. 8: *Upper row: Orthorhombic $LaTiO_3$ density of states components and QMC/DMFT quasi-particle spectra [21]. Lower left: Density of states components for tetragonal $LaTiO_3$; right: orbital polarization for orthorhombic and tetragonal $LaTiO_3$, calculated within ED/DMFT [32]. For $U^{\rm Ti} \approx 5 \ldots 6$ eV orthorhombic (tetragonal) $LaTiO_3$ is insulating (metallic).*

In contrast to $SrVO_3$ and $CaVO_3$, $LaTiO_3$ exhibits sizable orthorhombic distortions, giving rise to substantially different t_{2g} subband occupancies [21]. Fig. 8 shows the density of states components, indicating larger a_g than e'_g occupancy. QMC/DMFT quasi-particle spectra for this structure demonstrate that Coulomb correlations greatly enhance this orbital polarization. In fact, for $U = 5$ eV, $J = 0.7$ eV, the e'_g bands are nearly empty, whereas the a_g band is half-filled, with a Mott gap between lower and upper Hubbard bands. Effectively, therefore the t_{2g} subband degeneracy is reduced from three to one. A similar orbital polarization is obtained for $YTiO_3$ which exhibits even stronger non-cubic distortions than $LaTiO_3$ [21].

In view of this striking enhancement of orbital polarization due to correlations it is interesting to inquire what happens when thin layers of a material such as $LaTiO_3$ are placed in artificial environments. In fact, heterostructures consisting of $LaTiO_3$ and $SrTiO_3$ layers have recently been observed to be metallic [33] although both systems in their bulk forms are insulating. ($SrTiO_3$ with $3d^0$ is a band insulator.) The natural explanation of the observed metallicity appears to be the interface layer of Ti ions with $3d^{0.5}$ occupancy. On the other hand, it is likely that the first few layers of $LaTiO_3$ grow in a tretragonal fashion, with the a/b plane dictated by the cubic $SrTiO_3$ substrate. The lower left panel of Fig. 8 shows that the crystal field splitting of a hypothetical tetragonal $LaTiO_3$ structure has the opposite sign compared to the usual orthorhombic

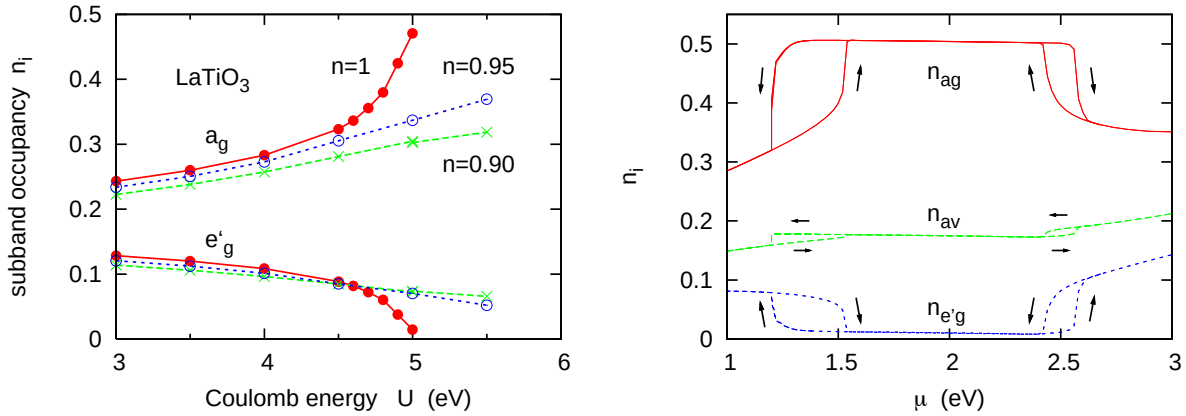


Fig. 9: *Left: Orbital polarization of $La_{1-x}Sr_xTiO_3$ for different hole dopings calculated within ED/DMFT ($n = 1 - x$). Right: Subband and average occupancies (per spin) as functions of chemical potential. Arrows mark hysteresis behavior for increasing/decreasing μ [34].*

version. Moreover, the band width is appreciably larger. ED/DMFT calculations then indicate that correlations enhance this reversed orbital polarization, with a Mott transition at considerably larger $U_c > 6$ eV within the quarter-filled doubly degenerate $d_{xz,yz}$ bands and the d_{xy} band empty [32]. Although the true structure of interlayer $LaTiO_3$ has not yet been determined, these results suggest that the tetragonal compound is a strongly correlated metal rather than a Mott insulator. Thus, the observed metallicity of $LaTiO_3/SrTiO_3$ heterostructures should arise not only from the interface Ti $3d^{0.5}$ ions but from the entire $LaTiO_3$ layer.

6 $La_{1-x}Sr_xTiO_3$

The remarkable orbital polarization close to the Mott transition of Ca_2RuO_4 and $LaTiO_3$ discussed in Sections 4 and 5, in particular, its sensitivity to structural distortions, raises the questions: How robust are these metal insulator transitions against doping, i.e., deviations from integer occupancy? How do the different t_{2g} subbands participate in the doping process? Of course, if all bands are identical as in cubic $SrVO_3$, doping affects all subbands in the same way. Fig. 9 suggests a fundamentally different picture in the case of $La_{1-x}Sr_xTiO_3$. The nearly complete orbital polarization of the Mott phase is greatly diminished, giving rise to a large flow of charge from the half-filled a_g band to the empty e'_g bands. For instance, at 5 % hole doping ($n = 0.95$), $n_{a_g} \approx 0.32$ and $n_{e'_g} \approx 0.08$ per spin band, increasing the total e'_g occupancy from near zero to 0.31 and decreasing the total a_g occupancy from near unity to 0.64. Thus, the internal charge flow is six times larger than the external charge transfer. Evidently external hole doping takes place via simultaneous electron *and* hole doping of t_{2g} subbands. Analogous results are found for electron doping [34].

This behavior may be understood by analyzing the variation of the subband occupancies with chemical potential. As shown in the right panel of Fig. 9, the subband charge compressibilities $\kappa_m = \partial n_m/\partial\mu$ have opposite signs and their magnitudes are much larger than the average charge compressibility. Near the Mott transition the κ_m become singular, permitting large internal charge rearrangements. These result indicate that the combined effect of charge and orbital degrees of freedom leads to a non-trivial generalization of the one-band picture close to half-filling and of the multi-band picture involving identical orbitals.

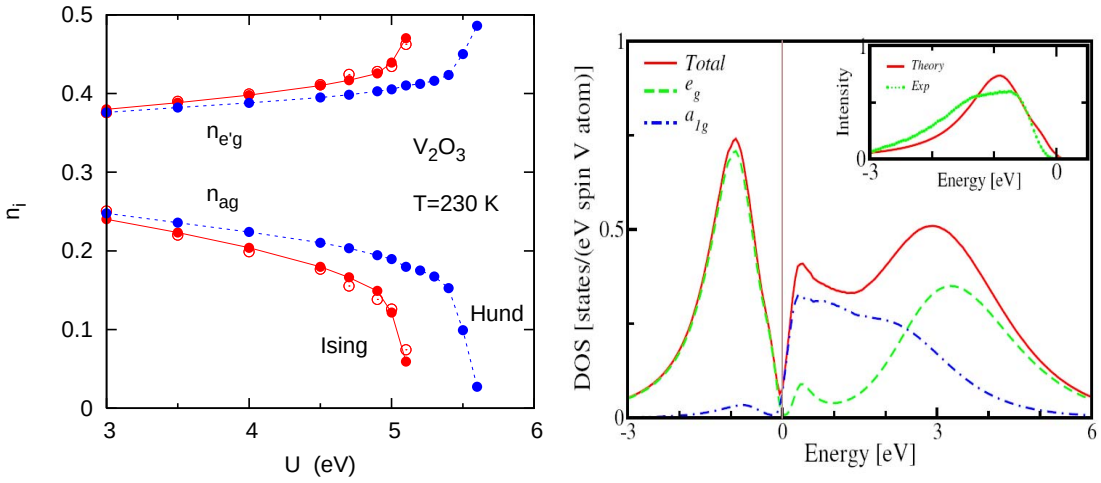


Fig. 10: *Left: V_2O_3 t_{2g} subband occupancies as functions of U, derived within ED/DMFT [35] for t_{2g} density of states shown in Fig. 4. Blue dots: Hund exchange, solid red dots: Ising exchange; empty red dots: QMC/DMFT results for Ising exchange [13]. Right: blue and green curves: a_g and e'_g spectral distributions, red curve: total density, in the insulating phase of V_2O_3, calculated within QMC/DMFT [22].*

7 V_2O_3

The transition between paramagnetic metallic and insulating phases of V_2O_3 observed at $T \approx 150 \ldots 400$ K [36] has been studied using DMFT by several groups [13, 22]. As in the case of $SrVO_3$, the partially occupied valence bands correspond to V $3d$ t_{2g} orbitals. Because of the corundum lattice structure, these orbitals are split into singly-degenerate a_g and doubly-degenerate e'_g components, whose densities of states are shown in Fig. 4 [13]. Within LDA, the subband occupancies are $n_{a_g} = 0.275$ and $n_{e'_g} = 0.362$ per spin band (total occupancy $n = 2$). Fig. 10 shows that in the presence of local Coulomb interactions, this orbital polarization is amplified, until, in the range $U \approx 5.1 \ldots 5.6$ eV (for $J = 0.7$ eV), it becomes complete: The e'_g bands are half-filled and the a_g band is empty. As shown by the spectral distributions on the right [22], the e'_g bands exhibit lower and upper Hubbard bands, while the a_g band is pushed above E_F.

The V_2O_3 subband occupations as functions of U reveal slight differences between full Hund exchange in the quantum impurity calculation and the more approximate Ising-like exchange. In particular, the critical U_c is about 10 % smaller for Ising exchange. The latter amounts to the neglect of spin-flip and pair-exchange terms in the Hamiltonians H and H^{cl}, i.e., the last terms in Eqs. (4) and (8). Thus, only density-density Coulomb and exchange interactions are included. This approximation is usually made in QMC/DMFT calculations to avoid sign problems. In ED/DMFT these problems do not arise. Thus, both Hund and Ising-like exchange can be included. The solid and red dots in the left panel of Fig. 10 are for identical LDA density of states input and for the same Ising-like exchange, demonstrating excellent agreement between ED/DMFT [35] and QMC/DMFT [13].

The correlation driven enhancement of orbital polarization in V_2O_3, Ca_2RuO_4 and $LaTiO_3$ shown in Figs. 10, 6 and 8, can be illustrated schematically as indicated in Fig. 11. The occupied part of the uncorrelated density of states of these transition metal oxides has contri-

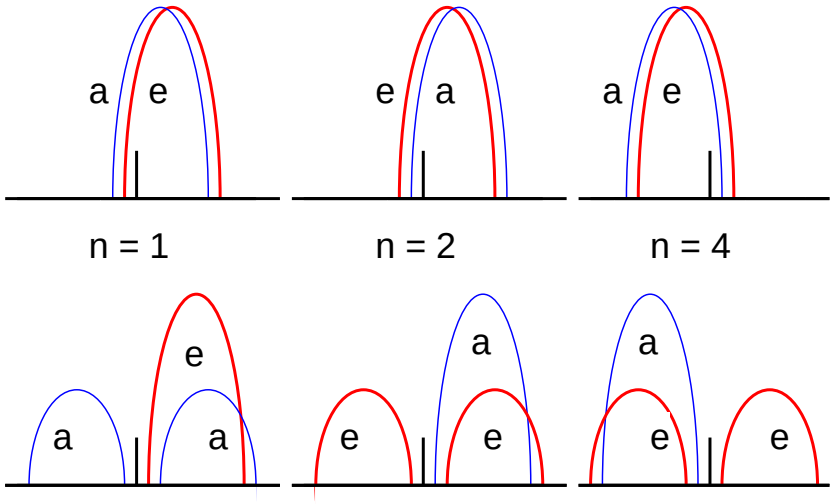


Fig. 11: *Schematic illustration of correlation driven enhancement of orbital polarization. Upper row: crystal field split t_{2g} LDA densities of states for occupancies $n = 1, 2, 4$, corresponding to $LaTiO_3$, V_2O_3 and Ca_2RuO_4, respectively. Blue curves: singly-degenerate a_g band, red curves: doubly-degenerate e'_g bands. In the case of Ca_2RuO_4, a refers to d_{xy}, e to $d_{xz,yz}$. The vertical bars denote the Fermi level. Lower row: orbitally polarized Mott phase. $n = 1$: empty e'_g bands, lower and upper Hubbard peaks of half-filled a_g band; $n = 2$: empty a_g band, lower and upper Hubbard peaks of half-filled e'_g bands; $n = 4$: filled a_g band, lower and upper Hubbard peaks of half-filled e'_g bands.*

butions from all t_{2g} components. In the strongly correlated metallic phase, this remains true, except that orbital polarization is increased. In the insulating Mott phase, however, some subbands are completely empty or filled, while the remaining ones are half-filled and split into lower and upper Hubbard bands. The Mott gap therefore involves transitions between states of opposite symmetry character. Note, however, that other materials can exhibit a different behavior. The hypothetical tetragonal structure of $LaTiO_3$ shown in Fig. 8 reveals a Mott phase with $n_{xz,yz} \to 1/4$ and $n_{xy} \to 0$ [32]. Moreover, orbital polarization in $BaVS_3$ was shown to decrease with increasing local Coulomb interaction [37]. Also, the Mott transition in $LaVO_3$ and YVO_3 occurs before orbital polarization is complete [38]. Finally, in Section 4 we pointed out the possibility of orbital selective Mott transitions. Thus, 60 years after Mott first discussed the paramagnetic, correlation induced metal insulator transition, multi-band DMFT treatments reveal that these transitions in realistic transition metal oxides can be highly complex.

8 $Na_{0.3}CoO_2$

In Section 4 we have seen that Coulomb correlations in Sr_2RuO_4 give rise to a charge transfer between t_{2g} orbitals, so that the d_{xy} van Hove singularity at the M point of the Brillouin Zone is pushed very close to the Fermi level. A slight reduction of the band width, as induced via Sr $\to$ Ca substitution, together with moderate Coulomb interactions, shifts this singularity below E_F, so that the topology of the d_{xy} Fermi surface sheet changes from electron-like to hole-like. The Fermi surface of the layer compound Na_xCoO_2 has remained controversial for several years,

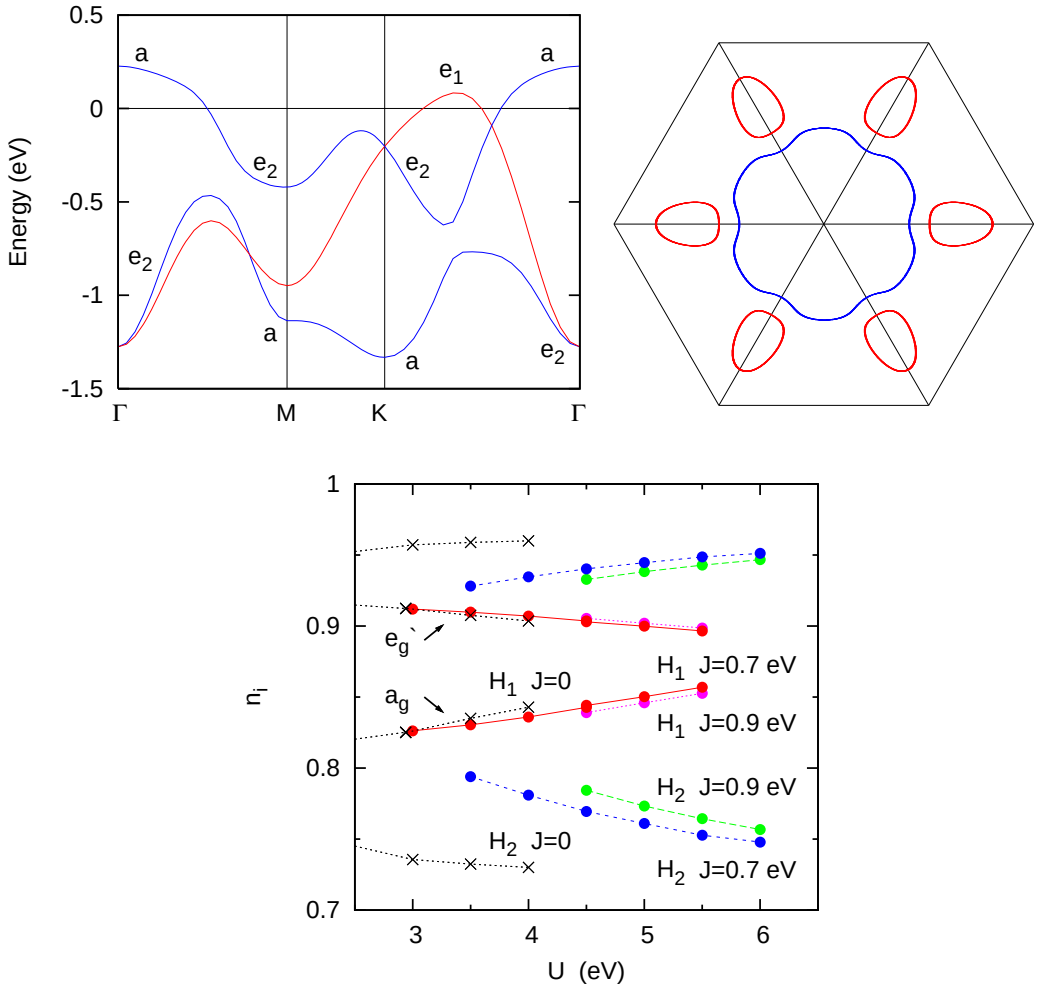


Fig. 12: *Upper panels: LDA t_{2g} bands of $Na_{0.3}CoO_2$ and Fermi surface (schematic), revealing six small e_g' hole pockets [39]. Lower panel: Subband occupancies as functions of U for several fixed values of J, derived within ED/DMFT. H_1 and H_2 denote tight-binding Hamiltonians with slightly different t_{2g} crystal field splittings: $\Delta_1 \approx -130$ meV , $\Delta_2 \approx -10$ meV [42].*

raising the question whether Coulomb correlations might be the origin of the discrepancies between LDA predictions [39] and photoemission data [40]. As shown in Fig. 12, because of the hexagonal structure of this material, the LDA Fermi surface exhibits six small hole pockets arising from the partially filled Co e_g' subbands which have not yet been observed using ARPES. Na_xCoO_2, with $3d^{5+x}$ occupancy of the Co $3d$ bands, as a function of Na doping reveals a remarkably rich phase diagram, ranging from Mott insulator at $x = 0$, superconductor at $x = 0.3$ (if hydrated), charge disproportionation at $x = 0.5$, pronounced Curie-Weiss behavior near $x = 0.7$, and band insulator at $x = 1$. The presence of the hole pockets is believed to have a strong influence on the nature of the superconductivity at $x = 0.3$. Thus, it is clearly important to understand the topology of the Fermi surface.

Early attempts to reproduce these measurements by taking into account strong local Coulomb interactions within the Co $3d$ shells failed since QMC/DMFT results revealed stabilization of the e_g' pockets, rather than their disappearance [41]. An extensive analysis within ED/DMFT demonstrated that orbital polarization between Co t_{2g} subbands may increase or decrease, de-

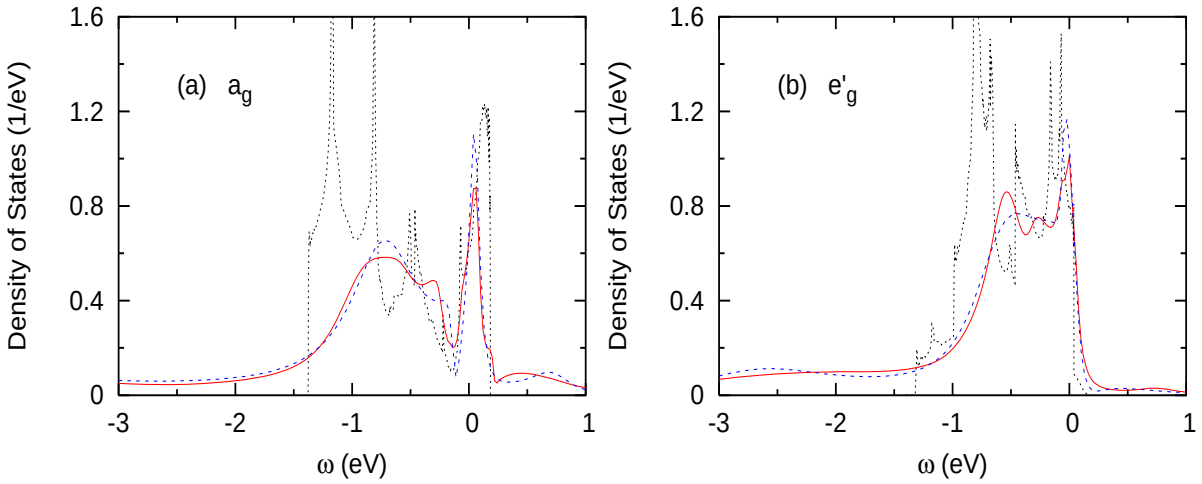


Fig. 13: *Quasi-particle spectra of $Na_{0.3}CoO_2$ calculated within ED/DMFT [11]. Left panel: a_g states, right panel: e'_g states. Solid red curves: spectra derived from $G_m(\omega)$ after extrapolating $\Sigma_m(i\omega_m)$ to real ω; dashed blue curves: spectra derived by extrapolating $G_m(i\omega_m)$ to real ω; dotted curves: bare density of states.*

pending on the details of the input single-particle Hamiltonian [42], as shown in the lower panel of Fig. 12. A crucial parameter is the t_{2g} crystal field splitting, $\Delta = \varepsilon_{a_g} - \varepsilon_{e'_g}$. As long as $\Delta \leq -0.1$ eV as predicted within LDA, Coulomb correlations within single-site DMFT consistently yield e'_g pockets, whereas $\Delta > 0.1$ eV pushes the e'_g bands sufficiently down so that the pockets disappear [43]. Thus, the fate of the e'_g pockets is more strongly influenced by single-particle effects governing the sign and magnitude of Δ than by correlation effects. Recent quantum chemical calculations, with special focus on the strong Co $3d$ – O $2p$ covalency, yield $\Delta \approx 0.3$ eV [44], suggesting filled e'_g pockets. On the other hand, surface effects might also shift the e'_g bands down [45], which could explain surface sensitive ARPES data. More theoretical and experimental work is needed to clarify the Fermi surface of $Na_{0.3}CoO_2$.

We close this section by discussing the analytical continuation from Matsubara frequencies to the real-ω axis. As pointed out in Section 3, quasi-particle spectra at real frequencies can be derived by transforming the solid Green's function $G_m(i\omega_n)$ to real ω, or by first transforming $\Sigma_m(i\omega_n)$ and then applying Eq. (3) at real ω. The comparison shown in Fig. 13 proves that both methods are consistent, and that the latter scheme retains finer spectral details originating from the single-particle Hamiltonian. For instance, the e'_g spectrum obtained via Eq. (3) and $\Sigma_m(\omega)$ shows two peaks below E_F which evidently are the shifted and broadened density of states features near 0.4 and 0.8 eV below the Fermi level. Also, the peak close to E_F exhibits some of the fine structure of the single-particle density of states. These details are lost if the spectrum is instead derived via extrapolation of $G_m(i\omega_n)$ to real ω.

9 DMFT for Heterostructures

So far we have discussed bulk properties of strongly correlated transition metal oxides. The analysis of these properties using photoemission is nevertheless non-trivial because of surface induced changes of the electronic structure, such as band narrowing due to reduced coordination, surface crystal field splitting of t_{2g} orbitals, lateral surface reconstruction, etc. Frequently,

surface contributions to photoemission spectra tend to be more correlated, with stronger Hubbard satellites than the bulk components. In addition to surfaces, heterostructures made out of thin layers of transition metal oxides have recently attracted a lot of interest as promising candidates for electron-correlation-based devices. In Section 5, we mentioned the example of the Mott insulator $LaTiO_3$ which exhibits metallicity when it is combined in heterostructures with the band insulator $SrTiO_3$ [33].

Inhomogeneous correlated layered systems have been studied within DMFT by several groups. Here we briefly discuss a new embedding approach [46] which permits an efficient application of DMFT to semi-infinite surfaces and heterostructures. The semi-infinite substrate leads connected to both sides of the central region of interest are represented via complex, energy-dependent embedding potentials that incorporate one-electron as well as many-body effects within the substrates. As a result, the number of layers which must be treated explicitly in the layer-coupled DMFT equation is greatly reduced compared to previous schemes. The interface region is assumed to include the first few surface layers of the actual substrates. Both the central region and the substrates may exhibit strong correlation effects.

Let us consider for simplicity a simple cubic lattice with nearest neighbor hopping t. The interface region has layer index $1 \leq i \leq N$, the left substrate $i \leq 0$ and the right substrate $i \geq N+1$. The DMFT calculation for this heterostructure then consists of three steps: First, the self-energies $\Sigma_\alpha(i\omega_n)$ of the infinite substrate bulk materials are calculated, using the formalism outlined in Sections 2 and 3. The index $\alpha = L,\ R$ denotes the left or right substrate. Second, the embedding potentials for the surface layers $i = 1$ and $i = N$ of the interface region are derived from the expression

$$S_\alpha(\mathbf{k}, i\omega_n) = (w - \sqrt{w^2 - 4t^2})/2 \tag{15}$$

$$w = i\omega_n + \mu - \varepsilon(\mathbf{k}) - \Sigma_\alpha(i\omega_n), \tag{16}$$

where $\varepsilon(\mathbf{k}) = -2t[\cos(k_x) + \cos(k_y)]$. The effective Hamiltonian for the interface region is given by the $N \times N$ matrix

$$H_{ij}(\mathbf{k}, i\omega_n) = -t_{ij} + \delta_{ij}[\varepsilon(\mathbf{k}) + \delta_{1i} S_L(\mathbf{k}, i\omega_n) + \delta_{iN} S_R(\mathbf{k}, i\omega_n)]. \tag{17}$$

The local lattice Green's function and the corresponding impurity Green's function are

$$G_i(i\omega_n) = \sum_{\mathbf{k}} [i\omega_n + \mu - H(\mathbf{k}, i\omega_n) - \Sigma(i\omega_n)]^{-1}_{ii} \tag{18}$$

$$G_{0,i}(i\omega_n) = [G_i(i\omega_n)^{-1} + \Sigma_i(i\omega_n)]^{-1}, \tag{19}$$

where $H(\mathbf{k}, i\omega_n)$ and $\Sigma(i\omega_n)$ denote the matrices $H_{ij}(\mathbf{k}, i\omega_n)$ and $\delta_{ij}\Sigma_i(i\omega_n)$. Finally, the $G_{0,i}(i\omega_n)$ are used as input for the ED/DMFT calculations for each interface layer, providing the new self-energies $\Sigma_i(i\omega_n)$.

Fig. 14 illustrates the layer variation of the quasi-particle weight $Z_i \approx 1/(1 - \mathrm{Im}\,\Sigma_i(i\omega_0)/\omega_0)$ for a model heterostructure consisting of a simple cubic lattice, with local Coulomb interaction $U = 6$ for the weakly correlated metallic substrates and $U = 10,\ 12$ for the more strongly correlated interface region $i = 1 \ldots 4$ (energy unit is $t = 1$). The calculation is carried out using $N = 8$ embedded layers, which comprise the central 4-layer film and the two outermost layers of the substrates. At each layer, the impurity level is surrounded by 7 bath levels in the ED/DMFT calculation. The quasi-particle weight of the substrate surface layers, $Z_{2,7}$, is seen

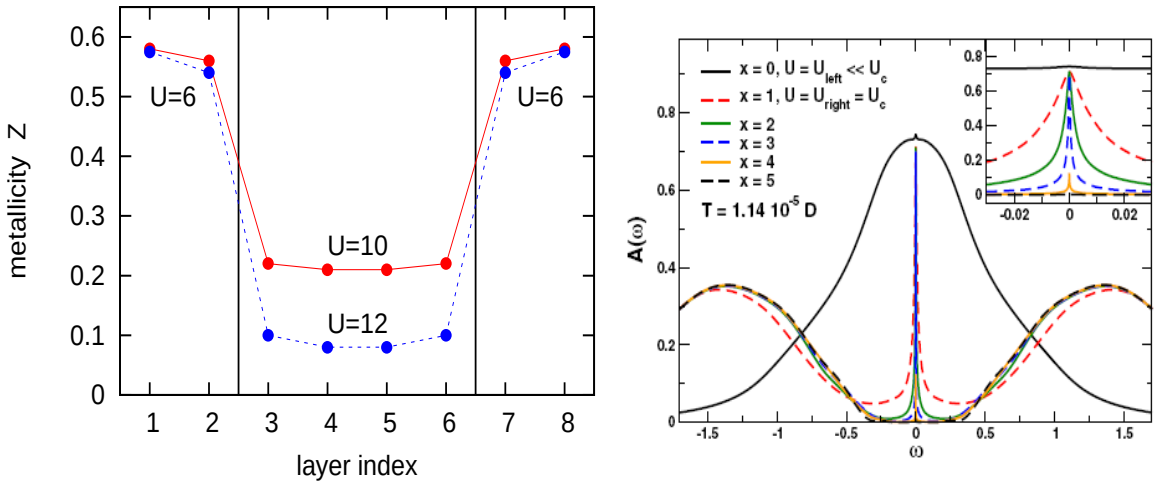


Fig. 14: *Left panel: Quasi-particle weight Z_i of the heterostructure consisting of a 4-layer film ($i = 3 \ldots 6$) sandwiched between two semi-infinite metals, calculated using ED/DMFT [46]. Temperature $T = 0.02$. Right panel: Layer variation of spectral distribution at metal insulator interface, calculated using NRG/DMFT [47]. $x = 0$: metallic surface layer; $x \geq 1$: successive layers in insulator. Inset: low-frequency region.*

to be reduced whereas at the surfaces of the poor metal it is enhanced. Evidently, the good or bad metallic character of one metal spills over into the neighboring metal [46].

Analogous layer-dependent DMFT calculations were recently performed for thick slabs using NRG as impurity solver [47]. The right panel of Fig. 14 shows the layer variation of the spectral distribution at a metal insulator interface. Metallic states near E_F are seen to form exponential tails within the Mott gap of the insulator. With increasing U in the insulator, the gap gets progressively wider and the penetration depth of metallic tails in the insulator decreases rapidly.

10 Cluster DMFT: Organic Salts

The influence of spatial quantum fluctuations on the nature of the Mott transition in strongly correlated systems is currently of great interest. To address this problem within DMFT it is necessary to go beyond the single-site or local approximation discussed in Sections 2 and 3 and include Coulomb interactions within clusters rather than single atoms [48, 49]. This extension allows one to study the momentum variation of the self-energy and examine, for instance, whether the Mott gap opens uniformly across the Fermi surface, or whether it appears first at the so-called 'hot spots' (strongly correlated points or regions of the Brillouin Zone) and only subsequently (at larger U) at 'cold spots' (weakly correlated regions).

A class of materials in which spatial fluctuations can be studied in detail are the layered charge transfer salts of the κ-(BEDT-TTF)$_2X$ family, where X denotes an inorganic anion. The electronic properties of these compounds have been shown to be highly sensitive functions of hydrostatic pressure [50, 51]. As a result, the temperature versus pressure phase diagram is remarkably rich, exhibiting Fermi-liquid and bad-metallic behavior, superconductivity, as well as paramagnetic and anti-ferromagnetic insulating phases, with striking analogies to the phase diagrams of transition metal oxides obtained via chemical doping [1].

A feature of particular interest in these salts is magnetic frustration. Since their structure cor-

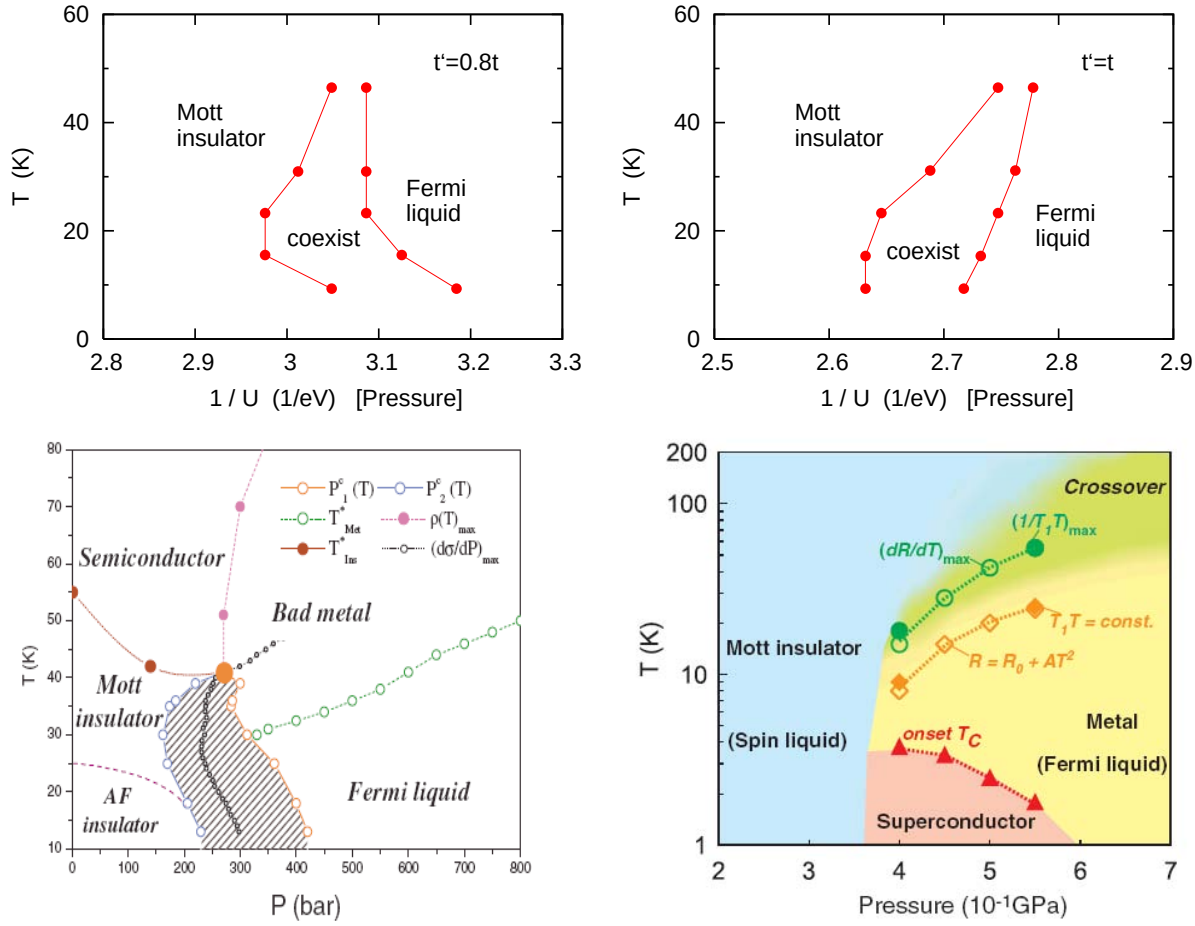


Fig. 15: *Upper panels: Phase diagrams of Hubbard model for anisotropic and isotropic triangular lattices ($t' = 0.8t$ and $t' = t$, respectively), evaluated within cluster ED/DMFT for $t = 0.04$ eV [53]. Plotted are the first-order metal-insulator phase boundaries as functions of inverse local Coulomb energy U. In the experimental setup increasing hydrostatic pressure P implies increasing band width W or decreasing U. Lower panels: Temperature / pressure phase diagrams for organic salts, κ-Cl [50] (left) and κ-CN [51] (right). The reentrant behavior observed for κ-Cl is absent for κ-CN.*

responds to an anisotropic triangular lattice, with inequivalent nearest neighbor hopping interactions t and t', long-range magnetic ordering becomes increasingly frustrated for $t' \to t$, giving rise to a spin-liquid phase in the absence of symmetry breaking [52]. Such a spin-liquid phase appears to be realized in κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ (denoted here as κ-CN) with $t' \approx 1.06t$, whereas κ-(BEDT-TTF)$_2$Cu[N(CN)$_2$]Cl (denoted as κ-Cl) with $t' \approx 0.75t$ is an anti-ferromagnetic insulator.

The minimal model Hamiltonian that captures the interplay between geometrical frustration and strong Coulomb interaction present in the conducting layers of organic salts is

$$H = -\sum_{ij\sigma} t_{ij}(c^+_{i\sigma}c_{j\sigma} + \text{H.c.}) + U\sum_i n_{i\uparrow}n_{i\downarrow} - \mu\sum_{i\sigma} c^+_{i\sigma}c_{i\sigma}, \tag{20}$$

where the sum in the first term is limited to nearest neighbor sites. The hopping integrals in a unit cell consisting of three sites are $t_{13} = t_{23} = t$ and $t_{12} = t'$. The chemical potential μ is

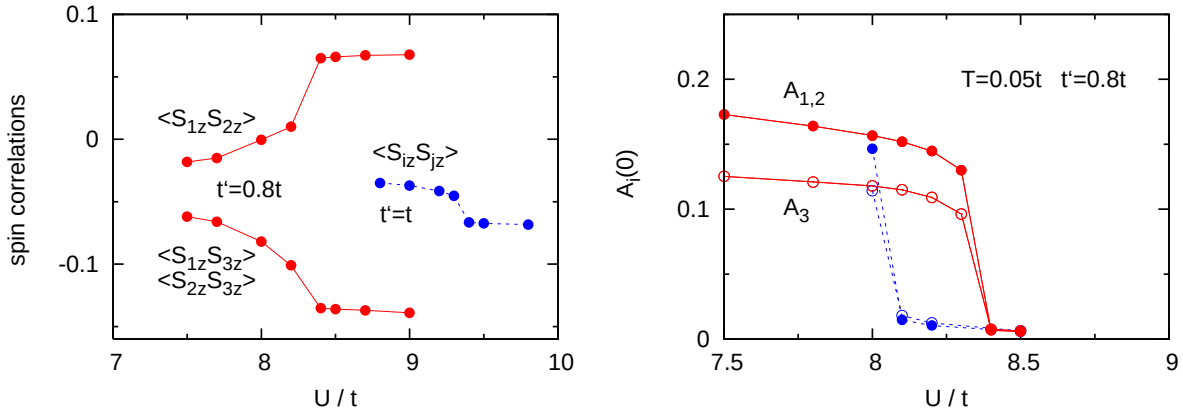


Fig. 16: *Left panel: Nearest-neighbor spin correlations in isotropic and anisotropic triangular lattices for $T = 0.05t$, $t' = t_{12}$, $t = t_{13} = t_{23}$. Strong enhancement of spin correlations occurs for moderate deviations from the isotropic limit. Right panel: Hysteresis behavior of spectral weights $A_i(0)$ of cluster sites at E_F for anisotropic triangular lattice. Red (blue) curves: increasing (decreasing) U [53].*

fixed at half-filling. Within cluster DMFT the lattice Green's function is defined as

$$G_{ij}(i\omega_n) = \sum_{\mathbf{k}} [i\omega_n + \mu - t(\mathbf{k}) - \Sigma(i\omega_n)]^{-1}_{ij} \,, \tag{21}$$

where $\mathbf{k}$ extends over the reduced Brillouin Zone. $t(\mathbf{k})$ denotes the hopping matrix for the superlattice and $\Sigma(i\omega_n)$ represents the non-diagonal cluster self-energy matrix.

Fig. 15 shows the ED/DMFT phase diagrams for the anisotropic and isotropic triangular lattices. The critical temperatures, $T_c \approx 50$ K are consistent with the measured values for κ-Cl and κ-CN. For $t' = 0.8t$, the phase boundaries separating the Fermi liquid from the Mott insulator show the same kind of reentrant behavior as measured for κ-Cl. For instance, at $U = 1/3$ eV and $T \approx 50$ K the system is a Mott insulator which turns into a Fermi liquid when T is lowered to about 20 K. Further reduction of T reverts the system to a Mott insulator, just as seen in the data. This reentrant behavior is absent in the case of the isotropic triangular lattice. This striking difference can be understood by analyzing the magnetic correlations $< S_{iz}S_{jz} >$ shown in Fig. 16. The results demonstrate that spin correlations are strongly enhanced as the geometrical frustration is suppressed. Thus, $t' = 0.8t$ induces a stronger tendency towards magnetic order than $t' = t$. At low T, therefore, the electron entropy is suppressed for $t' = 0.8t$ as compared to $t' = t$. As T is increased for $t' = 0.8t$, the system lowers its free energy by transforming to a metal since the entropy of the metal exceeds that of the ordered insulator. At even higher temperatures the system gains entropy of $\log(2)$ by transforming back into a paramagnetic insulator. In the isotropic lattice magnetic ordering is suppressed and the reentrant behavior disappears.

Finally, to illustrate the first-order nature of the metal-insulator transition we show in the right panel of Fig. 16 the spectral weights at $E_F = 0$ for the three cluster sites as functions of U. These quantities exhibit hysteresis for increasing and decreasing U, indicating coexistence of metallic and insulating solutions.

11 Summary

DMFT provides a combination of (i) high-quality single-electron Hamiltonians for complex materials consisting of many electrons per unit cell and (ii) accurate many-body formalisms for the study of complex quantum impurities. Together with increasingly powerful computational resources, this combination today allows detailed studies of the electronic properties of strongly correlated systems, such as transition metal oxides and organic salts. As a result, the nature of the Mott transition in multi-orbital compounds and its remarkable sensitivity to key parameters, like doping, temperature, pressure, crystal symmetry, etc., are now much better understood than only a decade ago.

The examples discussed in this lecture demonstrate that finite temperature exact diagonalization has emerged as a versatile and efficient tool for the study of highly correlated materials. The correlation induced charge transfer between valence orbitals is now a well established phenomenon found in a variety of systems. In the future, it should become feasible, via generalizations of DMFT for inhomogeneous systems, to investigate these kinds of effects at neutral and charged heterostructures consisting of realistic transition metal oxides. Further studies of spatial fluctuations, by combining exact diagonalization with cluster DMFT, will also be of great interest.

References

[1] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1999).

[2] S. Nakatsuji and Y. Maeno, Phys. Rev. Lett. **84**, 2666 (2000).

[3] I. I. Mazin and D.J. Singh, Phys. Rev. Lett. **79**, 733 (1997).

[4] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

[5] N. F. Mott, *Metal Insulator Transition* (Taylor and Francis, London, 1990).

[6] Key early papers:
W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989);
A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992);
M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992).

[7] Excellent review of early work:
A. Georges, G. Kotliar, W. Krauth and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).

[8] Recent reviews:
G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006);
K. Held, Adv. in Physics, **56**, 829 (2007).

[9] A. Liebsch and A. I. Lichtenstein, Phys. Rev. Lett. **84**, 1591 (2000).

[10] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994).

[11] C. A. Perroni, H. Ishida, and A. Liebsch, Phys. Rev. B **75**, 045125 (2007).
See also: A. Liebsch and T. A. Costi, Eur. Phys. J. B **51**, 523 (2006).

[12] A. A. Abrikosov, L. P. Gorkov, and I. Y. Dzyaloshinskii, *Quantum Field Theory in Statistical Physics* (Pergamon, Oxford, 1965).

[13] G. Keller, K. Held, V. Eyert, D. Vollhardt, and V. I. Anisimov, Phys. Rev. B **70**, 205116 (2004).

[14] R. B. Lehoucq, D. C. Sorensen, and C. Yang, ARPACK *Users' Guide* (1997).

[15] See also: Z. V. Pchelkina, I. A. Nekrasov, Th. Pruschke, A. Sekiyama, S. Suga, V. I. Anisimov, and D. Vollhardt, Phys. Rev. B **75**, 035122 (2007).

[16] R. Matzdorf *et al.*, Science **289**, 746 (2000).

[17] S.-C. Wang *et al.*, Phys. Rev. Lett. **93**, 177007 (2004).

[18] Z. Fang, N. Nagaosa, and K. Terakura, Phys. Rev. B **69**, 045116 (2004).

[19] A. Liebsch and H. Ishida, Phys. Rev. Lett. **98**, 216403 (2007).

[20] V. I. Anisimov, I. A. Nekrasov, D. E. Kondakov, T. M. Rice and M. Sigrist, Eur. Phys. J. B **25**, 191 (2002).

[21] E. Pavarini, S. Biermann, A. Poteryaev, A. I. Lichtenstein, A. Georges, and O. K. Andersen, Phys. Rev. Lett. **92**, 176403 (2004).

[22] A. I. Poteryaev, J. M. Tomczak, S. Biermann, A. Georges, A. I. Lichtenstein, A. N. Rubtsov, T. Saha-Dasgupta, and O. K. Andersen, Phys. Rev. B **76**, 085127 (2007).

[23] A. Koga, N. Kawakami, T. M. Rice and M. Sigrist, Phys. Rev. Lett. **92**, 216402 (2004).

[24] A. Liebsch, Phys. Rev. B **70**, 165103 (2004); Phys. Rev. Lett. **95**, 116402 (2005).

[25] T. A. Costi and A. Liebsch, Phys. Rev. Lett. **99**, 236404 (2007).

[26] M. Jarrell and J. E. Gubernatis, Phys. Rep. **269**, 133 (1996).

[27] *Numerical Recipes in Fortran 77* (Cambridge University Press, Cambridge, 1992), p. 106.

[28] A. Sekiyama *et al.*, Phys. Rev. Lett. **93**, 156402 (2007);
see also K. Maiti *et al.*, Eur. Phys. Lett. **58**, 246 (2001).

[29] I. A. Nekrasov *et al.*, Phys. Rev. B **73**, 155112 (2006).

[30] A. Liebsch, Phys. Rev. Lett. **90**, 096401 (2003).

[31] H. Ishida, D. Wortmann, and A. Liebsch, Phys. Rev. B **73**, 245421 (2006).

[32] H. Ishida and A. Liebsch, Phys. Rev. B **77**, 115350 (2008).

[33] A. Ohtomo *et al.*, Nature **419**, 378 (2002).

[34] A. Liebsch, Phys. Rev. B **77**, 115115 (2008).

[35] A. Liebsch, H. Ishida and J. Merino, Phys. Rev. B **78**, 165123 (2008).

[36] D. B. McWhan *et al.*, Phys. Rev. B **7**, 1920 (1973).

[37] F. Lechermann, S. Biermann, and A. Georges, Phys. Rev. Lett. **94**, 166402 (2005).

[38] M. De Raychaudhury, E. Pavarini, and O. K. Andersen, Phys. Rev. Lett. **99**, 126402 (2007).

[39] D. J. Singh, Phys. Rev. B **61**, 13397 (2000).

[40] See: H.-B. Yang *et al.*, Phys. Rev. Lett. **95**, 146401 (2005) and references herein.

[41] H. Ishida, M. D. Johannes, and A. Liebsch, Phys. Rev. Lett. **94**, 196401 (2005).

[42] A. Liebsch and H. Ishida, Eur. Phys. J. B **61**, 405 (2008).

[43] C. A. Marianetti, K. Haule, and O. Parkollet, Phys. Rev. Lett. **99**, 246404 (2007).

[44] A. Bourgeois, A. A. Aligia, and M. J. Rozenberg, arXiv:0901.1834.

[45] D. Pillay, M. D. Johannes, and I. I. Mazin, Phys. Rev. Lett. **101**, 246808 (2008).

[46] H. Ishida and A. Liebsch, Phys. Rev. B **79**, 045130 (2009).

[47] R. W. Helmes, T. A. Costi, and A. Rosch, Phys. Rev. Lett **101**, 066802 (2008).

[48] G. Kotliar, S. Y. Savrasov, G. Palsson, and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001).

[49] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005).

[50] P. Limelette *et al.*, Phys. Rev. Lett. **91**, 016401 (2003),
S. Lefebvre *et al.*, Phys. Rev. Lett. **85**, 5420 (2000).

[51] Y. Kurosaki *et al.*, Phys. Rev. Lett. **95**, 177001 (2005).

[52] P. W. Anderson, Mater. Res. Bull. **8**, 153 (1973).

[53] A. Liebsch, H. Ishida, and J. Merino, submitted to Phys. Rev. Lett.

A 5 Electronic structures of transition-metal oxides

Fumiyuki Ishii
Institut für Festkörperforschung,
Forschungszentrum Jülich GmbH
and
School of Mathematics and Physics,
Kanazawa University, Japan

Contents

1 Prelude **2**

2 Atomic orbitals in crystal **2**
2.1 Crystal field effect and chemical bonding 2
2.2 Spin polarization 4

3 Exchange interactions **4**
3.1 Superexchange 4
3.2 Double exchange and antisymmetric exchange 6
3.3 Exchange strictions 7
3.4 First-principles calculation of exchange coupling parameters 7

4 The case of perovskite manganite **8**
4.1 $R\mathrm{MnO_3}$(R: rare earth) 8
4.2 $\mathrm{La_{1-x}Sr_xMnO_3}$ 11

1 Prelude

Transition metal oxides (TMO) exhibit a variety of useful properties, such as ferromagnetism, ferroelectricity, half-metallicity, high-temperature superconductivity, colossal magnetoresistance (CMR), and multiferroelectricity. In this Lecture, we present important ingredients for the electronic structure of TMO and magnetic-related properties.

2 Atomic orbitals in crystal

2.1 Crystal field effect and chemical bonding

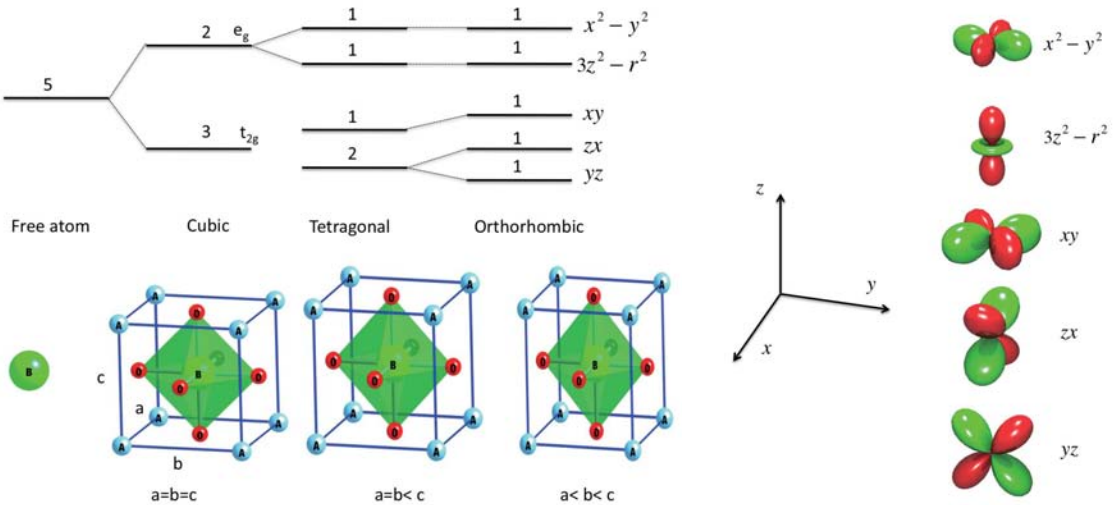


Fig. 1: *Possible d-orbital splitting by crystal field effect in ABO_3 with perovskite structure. Crystal structure and atomic orbitals are presented.*

The one-electron wavefunctions of atoms or ions are expressed in terms of a product of a radial wave function R_{nl} and angular wave functions, the spherical harmonics Y_{lm}.

$$\Psi_{nlm}(\mathbf{r}) = R_{nl}(r)Y_{lm}(\theta,\phi) \tag{1}$$

where n is principal quantum number, l is angular quantum number and m is magnetic quantum number. Transition metals have partially filled d (l=2, m=-2, -1 , 0, 1, 2) states. For simplicity, we can obtain real angular functions by linear combination of $\Psi_{nlm}(\mathbf{r})$ as follows,

$$\Phi_{yz}(\mathbf{r}) = \frac{i}{\sqrt{2}}(\Psi_{n21} + \Psi_{n2-1}) = \sqrt{\frac{15}{4\pi}}\frac{yz}{r^2}R_{n2}(r) \tag{2}$$

$$\Phi_{zx}(\mathbf{r}) = -\frac{1}{\sqrt{2}}(\Psi_{n21} - \Psi_{n2-1}) = \sqrt{\frac{15}{4\pi}}\frac{xz}{r^2}R_{n2}(r) \tag{3}$$

$$\Phi_{xy}(\mathbf{r}) = -\frac{i}{\sqrt{2}}(\Psi_{n22} - \Psi_{n2-2}) = \sqrt{\frac{15}{4\pi}}\frac{xy}{r^2}R_{n2}(r) \tag{4}$$

$$\Phi_{3z^2-r^2}(\mathbf{r}) = \Psi_{n20} = \sqrt{\frac{5}{16\pi}}\frac{3z^2 - r^2}{r^2}R_{n2}(r) \tag{5}$$

$$\Phi_{x^2-y^2}(\mathbf{r}) = \frac{1}{\sqrt{2}}(\Psi_{n22} + \Psi_{n2-2}) = \sqrt{\frac{15}{16\pi}}\frac{x^2 - y^2}{r^2}R_{n2}(r) \tag{6}$$

$$\tag{7}$$

There are two types of angular functions, t_{2g} and e_g. The t_{2g} (xy, yz and zx) orbitals have large amplitudes between the x, y and z axes, while e_g ($3z^2 - r^2$ and $x^2 - y^2$) orbitals have large amplitudes along these axes. The angular parts of these five wavefunctions are plotted in Fig. 1. In a free atom or ion, five d-states are degenerate in the energy. However in the crystal, such degeneracy is lifted up by the electric field derived from surrounding ions, crystal field. The energy diagram of octahedral crystal field in the perovskite oxides is shown in Fig. 1. We can intuitively understand that an electron at a transition-metal site can feel different repulsive potentials due to the oxygen ions.

It is known that the above description of the energy level splitting by a crystal field can explain experimentally observed results (photoemission spectroscopy etc.) though it is based on a point charge approximation. However, these results only explain *anti-bonding states* of transition-metal d states and oxygen p states. The energy levels reverse if we discuss the *bonding states*. Figure 2 shows a schematic diagram of chemical bonding.

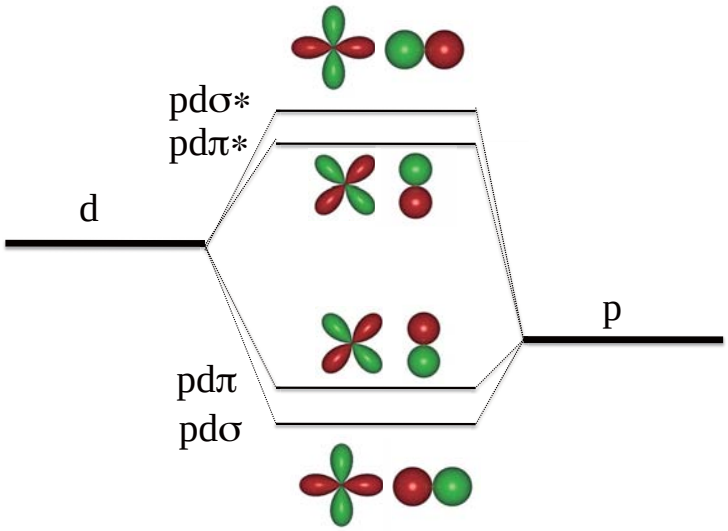


Fig. 2: *Schematic diagram of energy level and chemical bonding.*

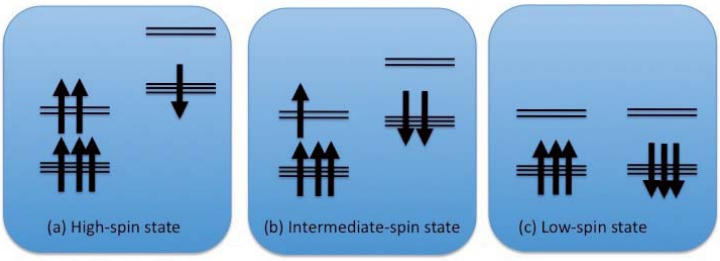


Fig. 3: *Possible spin configurations in Co^{3+} ion in cubic crystal field.*

2.2 Spin polarization

The ground state of an ion is explained by the well-known Hund's rule as follows [1].

(i) Arrange the electronic wave function to maximize S.

(ii) Arrange the electronic wave function to maximize L.

The first rule minimizes the Coulomb energy because of Pauli exclusion principle. The second rule minimizes also the Coulomb energy, because electrons with the same rotation direction avoid each other. These Hund's rules are not always applicable to ions in a crystal. The reason lies in the lift of the degeneracy by the crystal field and chemical bonding. For example, the $Co^{3+}(d^6)$ ion has the possibility of three spin arrangements in perovskite $LaCoO_3$, (i) High spin $S = 2$, (ii) Intermediate spin $S = 1$, (iii) Low spin $S = 0$ shown in Fig. 3. The energy levels are different for up-spin and down-spin as well as for high-spin and intermediate-spin case. These differences in energy levels for different spins are called spin polarization. The difference in the number of up-spin and down-spin electrons becomes the local spin moment of an ion. Figure 4 shows the schematic diagram of the electronic structure and oxygen coordination of different Co^{3+} sites in $SrCo_6O_{11}$. The octahedral site is the same as in figure 1. However, in the trigonal bipyramid case, the symmetry is not octahedral and the notation used is different from e_g-t_{2g} notation. These differences in the crystal field effect in this system gives rise to a novel *half-metallic antiferromagnetic* electronic structure.

3 Exchange interactions

3.1 Superexchange

Many of the transition-metal oxides have a ground state which has a ferromagnetic or antiferromagnetic order. To discuss the magnetic stability, the interaction between each pair of ions (i and j) is often represented by a Heisenberg model.

$$E = -\frac{1}{2}\sum_{ij} J_{ij}\mathbf{S}_i \cdot \mathbf{S}_j, \tag{8}$$

where the exchange parameter J_{ij} is positive for ferromagnetic, and negative for antiferromagnetic interactions. The microscopic origin of J_{ij} is often derived by a perturbation approximation to the Hubbard model given approximately by

$$J_{ij} \propto -\frac{t_{eff}^2}{U}, \tag{9}$$

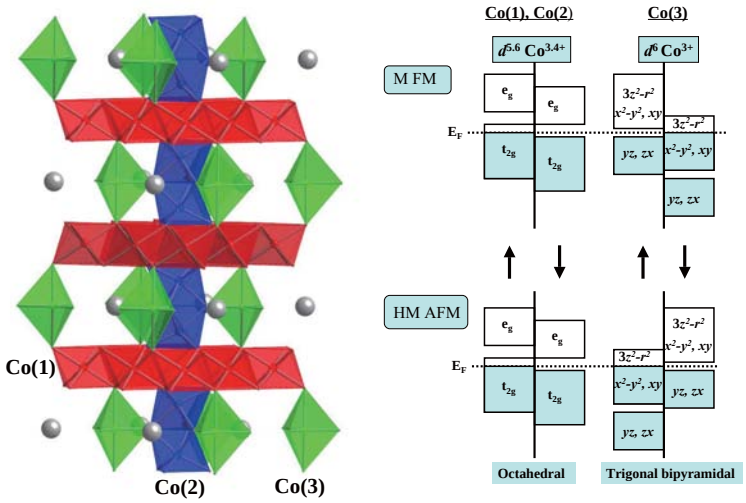


Fig. 4: *Crystal structure of $SrCo_6O_{11}$ and schematic view of electronic structure for ferromagnetic metal and antiferromagnetic half-metal.*

where t_{eff} is an effective hopping integral between transition-metal ions, U is on-site Coulomb interaction. In the transition-metal oxides, the exchange interaction between magnetic metal ions is mediated by a non-magnetic oxygen. This origin of exchange interaction is called *superexchange*. The effective transfer integral t_{eff} is expressed by the energy difference between the transition-metal d orbital and oxygen p orbital, Δ_{pd}, and the hopping integral between transition metal and oxygen t_{pd} as follows,

$$t_{eff} \propto \frac{t_{pd}^2}{\Delta_{pd}}. \tag{10}$$

Figure 2 shows that there are two types of overlap between p orbitals and d orbitals, $pd\sigma$ and $pd\pi$ which mainly contribute to hopping integrals. The sign and magnitude of the superexchange coupling parameter depend on the electronic structure and the geometry of magnetic ions M and oxygen ions. There are rules developed by Goodenough [2, 3] and Kanamori [4],

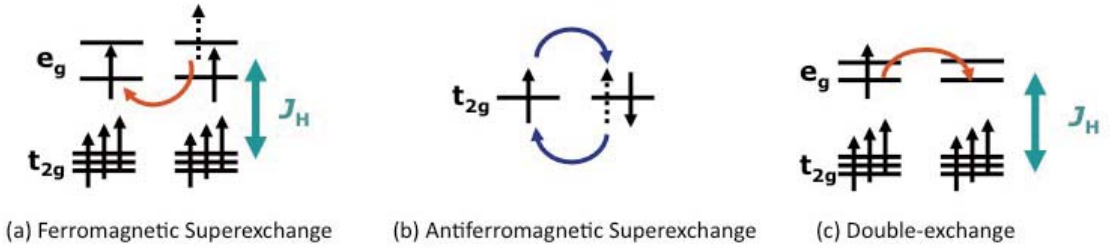


Fig. 5: *Schematic diagram of exchange interactions.*

so called *Goodenough-Kanamori rules* [5].

Rule 1: The hopping of electrons between ions with the same d orbitals via oxygen atom give antiferromagnetic exchange.
(i) $pd\sigma$ type orbitals with 180° M-O-M angle gives the largest superexchange case.
(ii) $pd\pi$ type orbitals with 180° M-O-M angle gives antiferromagnetic coupling.
(iii) $pd\sigma$ and $pd\pi$ case with 90° M-O-M angle gives strong antiferromagnetic coupling.

Rule 2: Combination of different atomic orbitals gives ferromagnetic exchange.
(iv) $pd\sigma$ and $pd\pi$ with 180° M-O-M angle, there is strong ferromagnetic coupling.

3.2 Double exchange and antisymmetric exchange

There are ferromagnetic exchange interactions which are induced by the carrier in the metallic compounds of mixed valency. This exchange interaction called *double exchange*(DE) [6]. The magnitude of the exchange coupling constant depends on t_{ij}, the transfer integral of carriers and the carrier concentration. When the localized spins $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ are aligned parallel, t_{ij} is maximum and is expressed by t^0_{ij}. When $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ are antiparallel, t_{ij}=0. More generally, when $\boldsymbol{S}_i$ shares an angle θ_{ij} with $\boldsymbol{S}_j$, the transfer integral for carriers of spin 1/2 is θ_{ij},

$$t_{ij} = t^0_{ij} \cos(\theta_{ij}/2) \tag{11}$$

The ferromagnetic coupling between carriers and the local magnetic moment is known as Hund's coupling J_H as mentioned in section 1.2 and shown in Fig. 5(c). The mechanism of double exchange is explained by kinetic energy gain, because of the relative spin angle of local moment. For example the canted antiferromagnetism in $La_{1-x}Sr_xMnO_3$ (LSMO) is explained by the double exchange mechanism. We also discuss the canted antiferromagnetism of LSMO from first principles in the Section 3.2. The energy of the double exchange which contributes to total energy is given as follows,

$$E_d = -\frac{x}{2} \sum_{ij} t^0_{ij} \cos(\theta_{ij}/2) \tag{12}$$

where x is carrier concentration. The angular dependence of the relative local magnetic moments differ between double exchange (DE) and superexchange (SE) terms.
Besides the double exchange mechanism, there is other exchange interactions which give raise to canted antiferromagnetic or weak ferromagnetic ground states. The exchange interaction is known as a *antisymmetric exchange*, or also as the *Dzyaloshinskii-Moriya interaction* and *anisotropic exchange* [7, 8]. The antisymmetric exchange is derived by virtual processes subject to the spin-orbit interaction $-\lambda \boldsymbol{l} \cdot \boldsymbol{S}$ and the Coulomb interaction. It is known that the energy gain due to antisymmetric exchange has following expression,

$$E_{DM} = \boldsymbol{D}_{ij} \cdot \mathbf{S}_i \times \mathbf{S}_j. \tag{13}$$

The vector $\boldsymbol{D}$ vanishes when the crystal field has an inversion symmetry with respect to the center between the two magnetic ions.

3.3 Exchange strictions

The exchange interaction depends on the atomic structure, bond distance, bond angle etc. Accordingly, the system may have energy gain if the atomic structure changes. Every mechanism of exchange has the possibility to contribute to an exchange striction mechanism for magnetoelectric or magnetoelastic properties.
Here we demonstrate a simple one-dimensional case of *superexchange striction* [9, 10].

$$E_{tot} = -J_{12}(r)\boldsymbol{S}_1 \cdot \boldsymbol{S}_2 + \frac{\kappa}{2}(r - r_0)^2 \tag{14}$$

$$\frac{dE_{tot}}{dr} = -J'(r)\boldsymbol{S}_1 \cdot \boldsymbol{S}_2 + \kappa(r - r_0) \tag{15}$$

$$\text{with} \quad J_{12}(r) \simeq J_{12}(r_0) + J'(r_0)(r - r_0) \tag{16}$$

where $J_{12} > 0$ for ferromagnetic coupling and $J_{12} < 0$ for antiferromagnetic coupling, κ is the force constant of the harmonic potential, r the is bond length of magnetic ions. From Eqs. (14), (15) and (16) we obtain an atomic displacement by exchange striction as follows,

$$\delta r \equiv r - r_0 = \frac{1}{\kappa} J'(r_0)\boldsymbol{S}_1 \cdot \boldsymbol{S}_2 \tag{17}$$

Clearly $J'(r_0) > 0$ for antiferromagnetic configuration and $J'(r_0) < 0$ for ferromagnetic configuration. *The ground state gives always shorter bond length than excited states.* The total energy for a small atomic displacement δr is given by,

$$E_{tot} \simeq -J_{12}(r_0)\boldsymbol{S}_1 \cdot \boldsymbol{S}_2 - \frac{1}{2\kappa} J'(r_0)^2 (\boldsymbol{S}_1 \cdot \boldsymbol{S}_2)^2 \tag{18}$$

Above discussion can be extended to higher dimensions and the bond length r can be replaced by a generalized coordinate. The microscopic origin of improper ferroelectricity in collinear multiferroics may originate from this superexchange striction.

3.4 First-principles calculation of exchange coupling parameters

Using the Green function technique [11, 12], we have calculated parameters of the inter-atomic exchange interaction following the expression,

$$J_{ij} = \frac{1}{\pi} \int^{\varepsilon_{\rm F}} d\varepsilon \, \mathrm{Im} \, \mathrm{Tr}_{\rm L} \left(\hat{G}^{\uparrow}_{ij} \hat{V}_j \hat{G}^{\downarrow}_{ji} \hat{V}_i \right) \tag{19}$$

where $G^{\uparrow,\downarrow}_{ij}$ in Eq. (19) is a block of real-space Green function with the site indices ij; $V_i = V_i^{\uparrow} - V_i^{\downarrow}$, Tr runs over the localized orbital indices. The method are implemented in OPENMX code [13]. To discuss the filling dependence of exchange interactions, we have calculated exchange coupling parameters for different Fermi energies as discussed by Belhadji *et al.* [14]. Figure 8 shows the density of states of $LaMnO_3$ and the filling-dependence of the exchange-coupling constants based on the rigid band approximation. These results can explain qualitatively the magnetic ground state of isostructural orthorhombic $LaMO_3$(M=Ti-Fe).

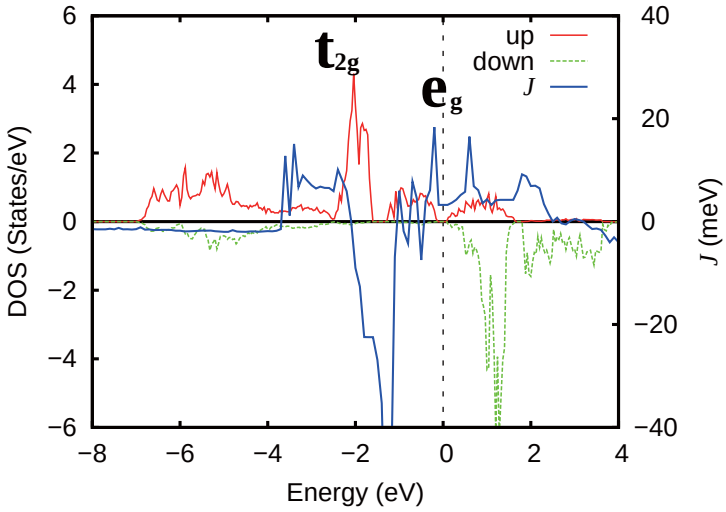


Fig. 6: *Partial density of states of Mn atom and energy dependence of nearest neighbor exchange coupling constant J_{ij} in ab plane of LaMnO$_3$.*

4 The case of perovskite manganite

4.1 RMnO$_3$(R: rare earth)

*R*MnO$_3$ systems exhibit a distorted perovskite structure of GdFe$_0$3-type. With changing from La with a large ionic radius to the small ionic radius of Er, the bond angles of Mn-O-Mn decrease and the GdFeO$_3$-type distortions increase. It was reported on ground of experiments [15], that the magnetic ground state changes from A-type antiferromagnetic (AFM-A) (Fig. 7(a)) to incommensurate magnetic (spiral) state (Fig. 7(d)) to E-type antiferromagnetic (AFM-E) (Fig. 7(c)).

In this section, to explain theoretically the correlation between the lattice distortions and the magnetic structure, we investigate the exchange interaction between the nearest-neighbor Mn sites. We perform first-principles collinear and non-collinear density-functional calculation for perovskite manganites *R*MnO$_3$ (*R* = rare-earth elements) varying *R*. To explore the mechanism for the appearance of the non-collinear magnetic structure, we calculate the exchange interactions in the nearest-neighbor and next nearest-neighbor Mn sites.

First, we predicted theoretically the transition of the magnetic ground state by focusing on the exchange interactions between the nearest-neighbor Mn sites. In *R*MnO$_3$, the double degeneracy of the e_g state of Mn^{3+} ion is lifted by the Jahn-Teller effect. The electron preferentially occupies the $3x^2 - r^2(3y^2 - r^2)$ orbitals. The magnetic ordering is caused by the superexchange interaction between Mn sites. Zhou and Goodenough proposed that the magnetic phase transition was induced by change in the nearest-neighbor exchange interaction [16]. According to their paper, $3x^2 - r^2(3y^2 - r^2)$ and $y^2 - z^2(x^2 - z^2)$ orbitals mix with increasing GdFeO$_3$ distortions and the ferromagnetic interaction for e_g orbital decrease. There is the competition between the antiferromagnetic interaction for t_{2g} orbital and ferromagnetic interaction for e_g orbital.

If we include the GdFeO$_3$-type distortion, the occupied and unoccupied wave function com-

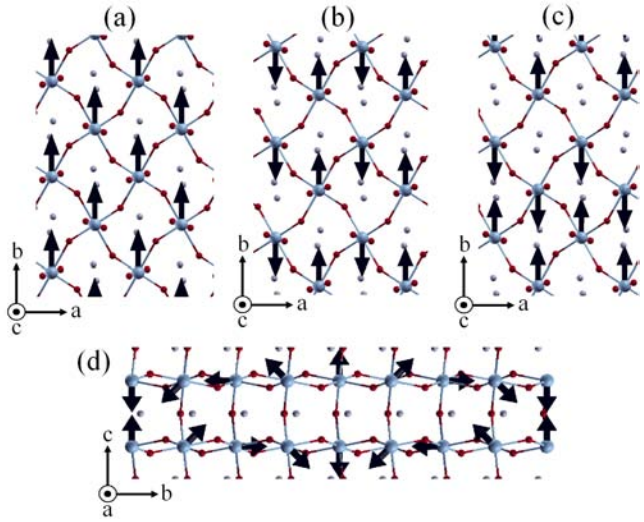


Fig. 7: *(Color). The magnetic structure of $RMnO_3$. Black arrows and large gray spheres denote the spin and Mn atoms, respectively. Small gray and red spheres denote the rare earth and O atoms, respectively. All magnetic structures in this figure consist of AFM coupling along the c-axis. (a), (b) and (c) are AFM-A, AFM-G and AFM-E, respectively. (d) describes the spiral magnetic structure. The rotational angle is the 45° in this structure.*

posed of e_g orbitals in the Mn-1 site and Mn-2 site are expressed as follows,

$$|\psi_1^{\rm o}\rangle = \cos\frac{\theta}{2}\left|3x_1^2 - r_1^2\right\rangle + \sin\frac{\theta}{2}\left|y_1^2 - z_1^2\right\rangle \tag{20}$$

$$|\psi_2^{\rm u}\rangle = \cos\frac{\theta}{2}\left|x_2^2 - z_2^2\right\rangle - \sin\frac{\theta}{2}\left|3y_2^2 - r_2^2\right\rangle \tag{21}$$

θ describes the material depended lattice distortions. For the non-distorted lattice θ=0. The ferromagnetic superexchange for e_g orbital is expressed as follows,

$$\begin{aligned} J_{\rm FM}^{\sigma} &= \left|\cos^2\frac{\theta}{2}\left\langle 3x_1^2 - r_1^2\left|H\right|x_2^2 - z_2^2\right\rangle - \sin^2\frac{\theta}{2}\left\langle y_1^2 - z_1^2\left|H\right|3y_2^2 - r_2^2\right\rangle\right|^2\frac{1}{\Delta_{\rm JT}} \\ &= \frac{t_{eff}^2}{\Delta_{\rm JT}}. \end{aligned} \tag{22}$$

$\Delta_{\rm JT}$ indicates the energy gap between $3x^2 - r^2(3y^2 - r^2)$ and $y^2 - z^2(x^2 - z^2)$ orbitals induced by the Jahn-Teller effect. The antiferromagnetic superexchange for the e_g orbital is expressed as follows,

$$J_{\rm AF}^{\sigma} = \frac{1}{2}\left(\frac{t_{eff}^2}{\Delta_{\rm ex} + \Delta_{\rm JT}\delta_{\theta 0}} + \frac{t_{eff}^2}{\Delta_{\rm ex} + \Delta_{\rm JT}}\right). \tag{23}$$

$\Delta_{\rm ex}$ indicates the exchange splitting. Schematic diagrams of the superexchange between the nearest-neighbors are shown in Fig. 9. The exchange interactions for $\rm e_g$ in the nearest-neighbor is expressed as follows,

$$J_{\sigma} = J_{\rm FM}^{\sigma} - J_{\rm AF}^{\sigma}. \tag{24}$$

t_{eff} in Eq. (22) decreases with increasing distortions, while $\Delta_{\rm JT}$ increases, because the Jahn-Teller effect increases with decreasing ionic radius. As a result, $J_{\rm FM}^{\sigma}$ decreases with increasing distortions. $J_{\rm AF}^{\sigma}$ becomes stronger with orbital mixing induced by distortions. The effective

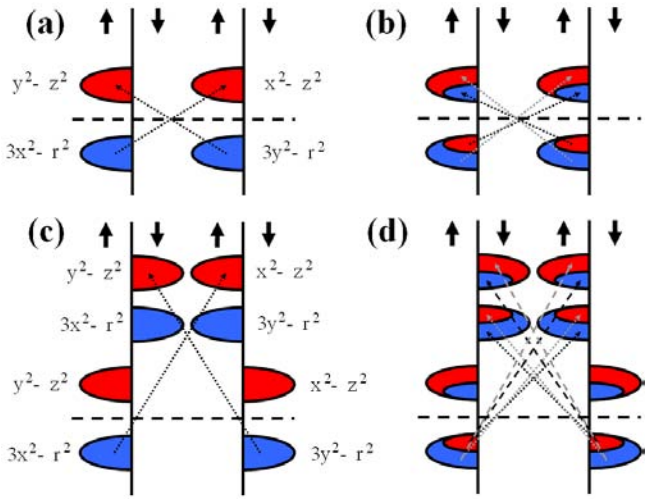


Fig. 8: *(Color). Schematic density of states (DOS) to explain the exchange-interaction mechanism between nearest-neighbor Mn sites due to e_g orbitals including Jahn-Teller distortion. Blue and red DOS are corresponding to $3x^2 - r^2(3y^2 - r^2)$ and $y^2 - z^2(x^2 - z^2)$ states, respectively. (a) and (b) denote ferromagnetic interactions. (c) and (d) denote antiferromagnetic interactions. (a) and (c) denote the DOS excluding the $GdFeO_3$-type distortion, (b) and (d) denote the DOS taking the $GdFeO_3$-type distortion into account.*

exchange interaction between nearest neighbors for e_g and t_{2g} orbitals is known according to Zhou and Goodenough [16]: According to Eq. (24), J_1 decreases with increasing distortions and the ferromagnetic interaction becomes weak. Since non-collinear magnetism can arize due to the competition between exchange interactions between ions at different neighbors, and since J_1 decreases with increasing distortions, it is predicted that non-collinear magnetism appears by the spin frustration with the competition between J_1 and J_2. Thus, to explore the appearance of non-collinear magnetism, we considered also the next nearest-neighbor interaction J_2. When AFM-A is the magnetic ground state, the contribution of J_1 is large. However, because of the fact that the antiferromagnetic J_2 contributes more than J_1 with increasing distortions, a competition between J_1 and J_2 is induced, and hence long-period noncollinear magnetic structure (spiral state) can be stabilized. We performed first-principles calculation for RMnO$_3$ (R = La$\sim$Er) and compared the total-energy differences among different magnetic structures. ϕ is the Mn-O-Mn bond angle, which decreases from La to Er. The magnetic structures were restricted to the AFM-A, AFM-G, AFM-E, and the spiral state (SP) shown in Fig. 7. Although we calculated the ferromagnetic (FM) and AFM-C state, both magnetic structures were not stable at any range of ϕ. Figure 9 shows the total-energy difference per Mn atom for the AFM-A order as a function of ϕ. According to the experiment by T. Kimura *et al.* [15], AFM-A is ground state in the range of $148^\circ < \phi < 156^\circ$, magnetically incommensurate in the range of $145.5^\circ < \phi < 148^\circ$, and AFM-E is ground-state in the range of $\phi < 145.5^\circ$. Even though we use a La pseudopotential for all R in RMnO$_3$ with the respective experimental crystal structures, the transition of magnetic ground state is consistent with experimental results. This means that the lattice structures play a crucial role in determining the magnetic ground state. AFM-A is stable in the range of $153^\circ < \phi < 155^\circ$, Sp-45° is stable in the range of $148.5^\circ < \phi < 153^\circ$, and AFM-E is stable in the range of $\phi < 148.5^\circ$. According to the experiment by T. Arima *et al.* [17], it was reported that a SP state of about 45 degree appears. Our first-principles results are consistent with the experimental result.

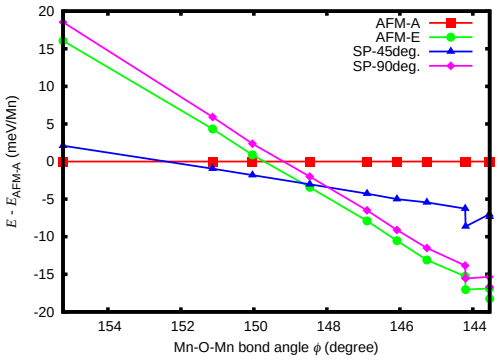


Fig. 9: *(Color). The total-energy difference (in meV/Mn) from the AFM-A as a function of ϕ. Red squares, green circles, blue triangles and pink diamonds denote the AFM-A, AFM-E, SP-45° and SP-90°, respectively. The inset shows the total energy difference between AFM-E and SP-90°. The lines are guide to the eye.*

4.2 $La_{1-x}Sr_xMnO_3$

We show our first-principles results for the double exchange mechanism in carrier-doped perovskite manganite [18]. Perovskite manganites $La_{1-x}Sr_xMnO_3$ (LSMO) exhibit novel physical properties such as colossal magnetoresistance [19] and half-metallicity [20]. These novel physical properties originate from a variety of magnetic configurations in LSMO such as ferromagnetic (FM), A-type (inter-plane antiferromagnetic (AFM) and intra-plane FM orders), C-type (inter-plane FM and intra-plane AFM orders) and G-type (inter-plane AFM and intra-plane AFM orders) AFM states (see Fig. 10 (a)). These magnetic states are controlled by the carrier concentrations and lattice distortions in LSMO [21, 22]. An experimental study revealed that the magnetic phase changes as AFM-A $\rightarrow$ FM $\rightarrow$ AFM-A $\rightarrow$ AFM-C $\rightarrow$ AFM-G states with increasing carrier concentrations in LSMO [23].

The magnetic states around the phase boundary in carrier-doped manganite have been extensively discussed on the basis of the long-range ordered *noncollinear* spin-canting magnetic states or the coexistence of AFM and FM states, i.e., phase separation. While the spin-canting magnetic state was suggested as a possible ground state with competition between magnetic interactions [6, 24, 25], the FM–AFM phase separation was also suggested as a stable phase by an experimental study and model calculations [26, 27, 28]. Despite the extensive studies on the magnetic state around the phase boundary, the issue remains unresolved. Similar problems are encountered in the interface of artificial superlattices [29, 30, 31] such as $(LaMnO_3)_m/(SrMnO_3)_n$ where inhomogeneous carriers are introduced. To design magnetic states in the artificial superlattice, a detailed systematic study of the carrier-dependence of magnetism in LSMO is of great importance.

We have performed first-principles calculations on $La_{1-x}Sr_xMnO_3$ ($0.0 \leq x \leq 1.0$) by the noncollinear density functional theory (DFT) [32, 33]. A generalized gradient approximation (GGA) [34] is adopted to determine the exchange correlation potential after the diagonalization of the noncollinear spin-density matrix. The norm-conserving pseudopotential method [35] with a partial core correction [36] is used and wavefunctions are expanded by a linear combination of multiple pseudo-atomic orbitals (LCPAO) [37, 38]. We neglected the spin-orbit interactions in all calculations. The calculations were done for a four-formula unit cell, i.e.,

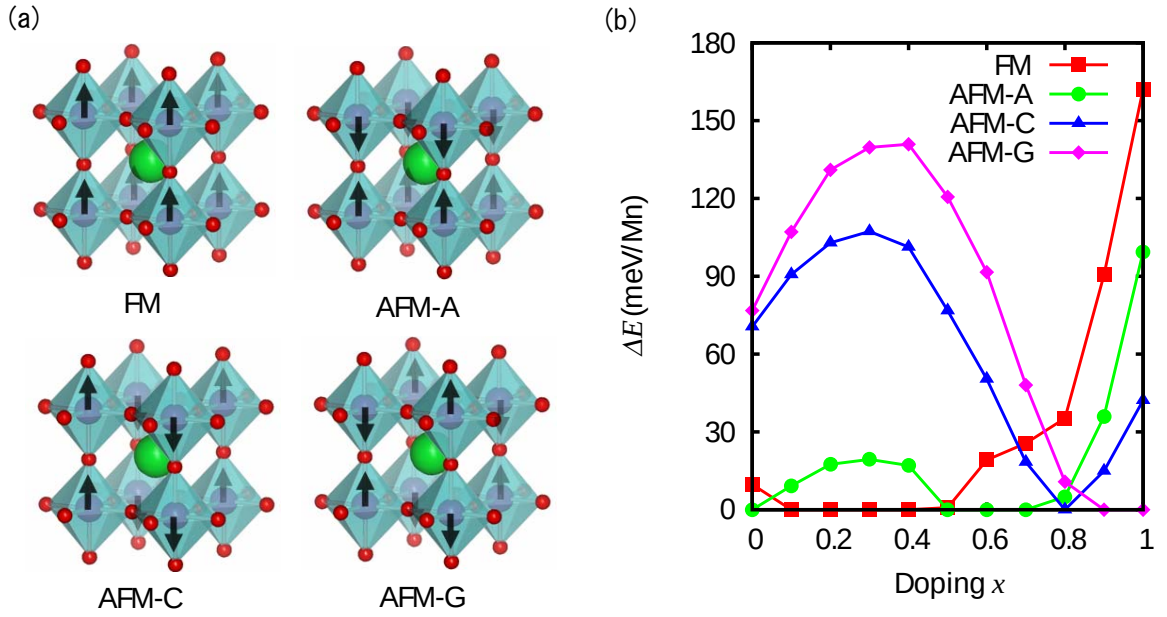


Fig. 10: *(a) The collinear magnetic structures of $La_{1-x}Sr_xMnO_3$. (b) The total energy difference per Mn atom from the stable state as a function of x. The red squares, green circles, blue triangles and pink diamonds denote the FM, AFM-A, AFM-C and AFM-G states, respectively. The lines are guide to the eyes [18].*

20 atoms in the unit cell. The hole carrier doping x is performed by a shift in the Fermi level and a uniform background charge is introduced to balance the charge neutrality of the system. Noncollinear spin orientations are fixed by using constrained DFT, where the penalty functions are introduced in the total-energy functional [39, 40]. All the above methods are implemented in the OPENMX code [13]. We use the atomic coordinates of orthorhombic $LaMnO_3$ ($x = 0.0$) and the cubic $SrMnO_3$ ($x = 1.0$) determined by experimental studies [41, 42]. In the region of $0.0 < x < 1.0$, we assumed that the lattice structure is continuously changed from $LaMnO_3$ to $SrMnO_3$.

We study the stability of the collinear magnetic states in LSMO. The calculated collinear magnetic states are the FM, AFM-A, AFM-C and AFM-G states (Fig. 10 (a)). Figure 10 (b) shows the total energy difference per Mn atom from the stable state as a function of x. When $x = 0.0$ and $0.5 \leq x < 0.8$, the AFM-A state becomes stable. The FM state becomes stable in the region of $0.1 \leq x < 0.5$. The AFM-C state becomes stable around $x = 0.8$. The AFM-G state becomes stable in the region of $0.8 < x \leq 1.0$. The AFM order becomes favorable with increasing x in the region of $0.5 \leq x \leq 1.0$. This result is consistent with the previous theoretical and experimental studies [22, 23].

We extended the calculation of magnetic states for noncollinear configurations, as shown in Fig. 11 (a). In Fig. 10, $x = 0.5$ is the carrier concentration at which the total energies of the FM and AFM-A states are nearly degenerate within 1.0 meV/Mn. Figure 11 (b) indicates the total energy difference per Mn atom from the stable state as a function of θ. θ is the inter-plane spin-canting angle, i.e., $\theta = 0^{\circ}$ and $\theta = 180^{\circ}$ correspond to the FM and AFM-A states, respectively. We found that the spin-canting magnetic state ($\theta = 105^{\circ}$) is stable for $x = 0.5$.

We also investigated the carrier dependence of the noncollinearity in the spin-canting magnetic

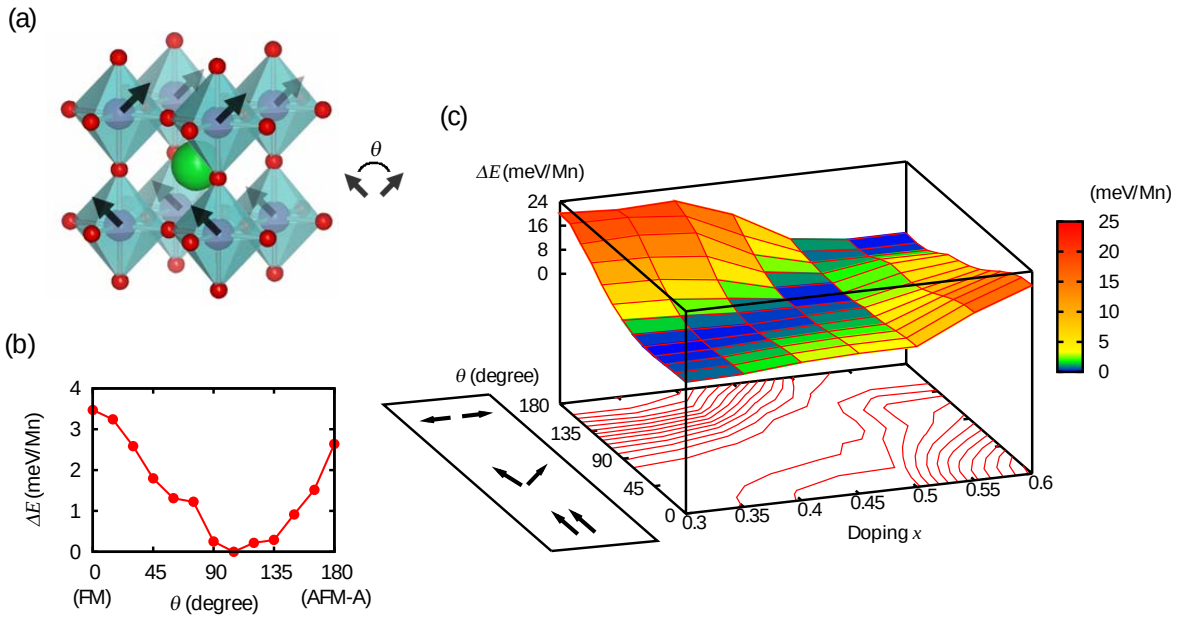


Fig. 11: *(a) The spin-canting magnetic structure of $La_{1-x}Sr_xMnO_3$ (LSMO). θ is the spin-canting angle between the inter-plane Mn sites. (b) The total energy difference per Mn atom from the stable state as a function of θ in LSMO ($x = 0.5$). (c) The total energy difference per Mn atom from the stable state in the parameter space as a function of hole doping x ($0.3 \leq x \leq 0.6$) and θ (degree). The change from blue to red in the color bar on the right-hand side represents the increase in the total energy difference. The lines serve as a visual reference [18].*

state of LSMO ($0.3 \leq x \leq 0.6$). Figure 10 (c) shows the magnetic phase stability as a function of x and θ. ΔE denotes the total energy difference per Mn atom from the stable state. The spin-canting magnetic state is stable in the region of $0.3 \leq x \leq 0.6$. With increasing x, stable θ continuously increases from the FM ($\theta = 0^{\circ}$) state and AFM-A ($\theta = 180^{\circ}$) state.

We discuss the stability of noncollinear magnetism around the magnetic phase boundary. It is understood that the carrier-induced magnetism in perovskite manganites is governed by the double exchange (DE) interaction [5, 6, 43]. The spin-canting magnetic state has been explained by de Gennes [6] in terms of the DE mechanism. According to his theory, in addition to the AFM superexchange (SE) interaction, the FM interaction is caused by electron hopping from a half-filled e_g state to an empty e_g state with Hund's coupling. Then, the spin-canting magnetic states are stable because of the competition between the FM DE and AFM SE interactions in LSMO. Although de Gennes restricted his discussions to low carrier concentrations, Solovyev and Terakura extended de Gennes's theory to a wide range of carrier concentrations [24] and predicted that the spin-canting magnetic state may be stable around the half-doped concentration (x = 0.5). Our first-principles results are consistent with this prediction, and the spin-canting magnetic state is stable in the region of $0.3 \leq x \leq 0.6$. We suggest that a noncollinear magnetic state may appear in a wide range of hole-doped perovskite manganites.

We discuss the effect of lattice distortions at $x = 0.5$. We have performed a calculation of the cubic LSMO with an averaged lattice constant. The total energy difference between the FM and AFM-A states is 9.8 meV/Mn at $x = 0.5$. The corresponding energy difference in the orthorhombic structure is 0.8 meV/Mn. We predict that the spin-canting magnetic state is more stable in the orthorhombic structure than in the cubic structure. We also discuss why the total

energy difference is large in the cubic structure. We attribute this difference to a decrease in the AFM SE interaction at the inter-plane. The average lattice constant in the cubic structure (3.873 Å) is larger than that in the orthorhombic structure (3.826 Å). The larger lattice constant leads to a decrease in the overlap between the wavefunctions of the inter-plane Mn atoms, i.e., the AFM SE interaction decreases at the inter-plane. Therefore, the lattice distortion may affect the stability of the noncollinear magnetic phase. We propose that the control of the noncollinear magnetic states is possible by the superlattice composition.

References

[1] S. Blundell, *Magnetism in Condensed Matter* (Oxford University Press, 2003).

[2] J. B. Goodenough, Phys. Rev. **100**, 564 (1955).

[3] J. B. Goodenough, J. Phys. Chem. Solids **6**, 287 (1958).

[4] J. Kanamori, J. Phys. Chem. Solids **10**, 87 (1959).

[5] P. W. Anderson, in *Magnetism*, edited by G. T. Rado and H. Suhl (Academic Press, New York, 1963), pp. 25–83.

[6] P. G. de Gennes, Phys. Rev. **118**, 141 (1960).

[7] I. Dzyaloshinsky, J. Phys. Chem. Solids **4**, 241 (1958).

[8] T. Moriya, Phys. Rev. **120**, 91 (1960).

[9] E. Harris, J. Phys. C: Solid State Phys. **5**, 338 (1972).

[10] T. Strässle, F. Juranyi, M. Schneider, S. Janssen, A. Furrer, K. W. Krämer, and H. U. Güdel, Phys. Rev. Lett. **92**, 257202 (2004).

[11] A. I. Liechtenstein, M. I. Katsnelson, V. P. Antropov, and V. A. Gubanov, J. Magn. Magn. Mater. **67**, 65 (1987).

[12] M. J. Han, T. Ozaki, and J. Yu, Phys. Rev. B **70**, 184421 (2004).

[13] T. Ozaki, H. Kino, J. Yu, M. J. Han, N. Kobayashi, M. Ohfuti, F. Ishii, T. Ohwaki, H. Weng, and K. Terakura, http://www.openmx-square.org/.

[14] B. Belhadji, L. Bergqvist, R. Zeller, P. H. Dederichs, K. Sato, and H. Katayama-Yoshida, J. Phys.: Condens. Matter **19**, 436227 (2007).

[15] T. Kimura, S. Ishihara, H. Shintani, T. Arima, K. T. Takahashi, K. Ishizaka, and Y. Tokura, Phys. Rev. B **68**, 060403 (2003).

[16] J. Zhou and J. Goodenough, Phys. Rev. Lett. **96**, 247202 (2006).

[17] T. Arima, A. Tokunaga, T. Goto, H. Kimura, Y. Noda, and Y. Tokura, Phys. Rev. Lett. **96**, 97202 (2006).

[18] K. Sawada and F. Ishii, J. Phys.: Condens. Matter **21**, 064246 (4pp) (2009).

[19] Y. Tokura, A. Urushibara, Y. Moritomo, T. Arima, A. Asamitsu, G. Kido, and N. Furukawa, J. Phys. Soc. Jpn. **63**, 3931 (1994).

[20] Y. Okimoto, T. Katsufuji, T. Ishikawa, A. Urushibara, T. Arima, and Y. Tokura, Phys. Rev. Lett. **75**, 109 (1995).

[21] Y. Konishi, Z. Fang, M. Izumi, T. Manako, M. Kasai, H. Kuwahara, M. Kawasaki, K. Terakura, and Y. Tokura, J. Phys. Soc. Jpn. **68**, 3790 (1999).

[22] Z. Fang, I. V. Solovyev, and K. Terakura, Phys. Rev. Lett. **84**, 3169 (2000).

[23] O. Chmaissem, B. Dabrowski, S. Kolesnik, J. Mais, J. D. Jorgensen, and S. Short, Phys. Rev. B **67**, 094431 (2003).

[24] I. V. Solovyev and K. Terakura, Phys. Rev. B **63**, 174425 (2001).

[25] H. Yoshizawa, H. Kawano, Y. Tomioka, and Y. Tokura, Phys. Rev. B **52**, R13145 (1995).

[26] A. Moreo, S. Yunoki, and E. Dagotto, Science **283**, 2034 (1999).

[27] E. L. Nagaev, Phys. Rev. B **58**, 2415 (1998).

[28] S. Yunoki, J. Hu, A. L. Malvezzi, A. Moreo, N. Furukawa, and E. Dagotto, Phys. Rev. Lett. **80**, 845 (1998).

[29] T. Koida, M. Lippmaa, T. Fukumura, K. Itaka, Y. Matsumoto, M. Kawasaki, and H. Koinuma, Phys. Rev. B **66**, 144418 (2002).

[30] H. Yamada, M. Kawasaki, T. Lottermoser, T. Arima, and Y. Tokura, Appl. Phys. Lett. **89**, 052506 (2006).

[31] M. Izumi, Y. Murakami, Y. Konishi, T. Manako, M. Kawasaki, and Y. Tokura, Phys. Rev. B **60**, 1211 (1999).

[32] U. von Barth and L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972).

[33] J. Kübler, Hock, J. Sticht, and A. R. Williams, J. Phys. F: Metal Phys. **18**, 469 (1988).

[34] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[35] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[36] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[37] T. Ozaki, Phys. Rev. B **67**, 155108 (2003).

[38] T. Ozaki and H. Kino, Phys. Rev. B **69**, 195113 (2004).

[39] P. Kurz, F. Förster, L. Nordström, G. Bihlmayer, and S. Blügel, Phys. Rev. B **69**, 024415 (2004).

[40] R. Gebauer and S. Baroni, Phys. Rev. B **61**, R6459 (2000).

[41] J. A. Alonso, M. J. Martinez-Lope, M. T. Casais, and M. T. Fernandez-Diaz, Inorg. Chem. **39**, 917 (2000).

[42] T. Negas and R. S. Roth, J. Phys. Chem. Solids **1**, 409 (1970), ISSN 0022-4596.

[43] C. Zener, Phys. Rev. **82**, 403 (1951).

A 6 Introduction to Multiferroics

Marjana Ležaić
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Ferroelectricity and ferromagnetism **4**

3 Ferromagnetism **5**

4 Ferroelectricity **7**
4.1 Displacive phase transitions 7
4.2 Landau theory: proper and improper phase transitions 10
4.3 Vibronic theory of ferroelectricity 12
4.4 Typical examples 13

5 Magnetic ferroelectrics **16**

1 Introduction

Multiferroic materials [1] are in broad sense defined as materials possessing, in a single phase, two or more *ferroic* orders:

- ***Ferromagnetic*** materials are characterized by a spontaneous magnetization, switchable hysteretically by an applied magnetic field. A lattice of a magnetic atoms can also be formed in a way that all moments are aligned, but neighboring moments point in opposite directions. If the resulting magnetization in such a lattice is zero, the material is ***antiferromagnetic***; in the case of a non-vanishing resulting magnetization, we are dealing with a ***ferrimagnetic*** material.
- ***Ferroelectric*** materials are characterized by a spontaneous electric polarization, switchable hysteretically by an applied electric field. If a material possesses ordered dipole moments which cancel each other completely within each crystallographic unit cell, it is said to be ***antiferroelectric***.
- ***Ferroelastic*** materials are characterized by a spontaneous deformation, switchable hysteretically by an applied stress.
- ***Ferrotoroidic*** materials are the newest addition to the family of ferroics. They are characterized by an order parameter which is taken to be the curl of a magnetization or polarization. It is anticipated that this order parameter, referred to as *toroidization* [2], is hysteretically switchable.

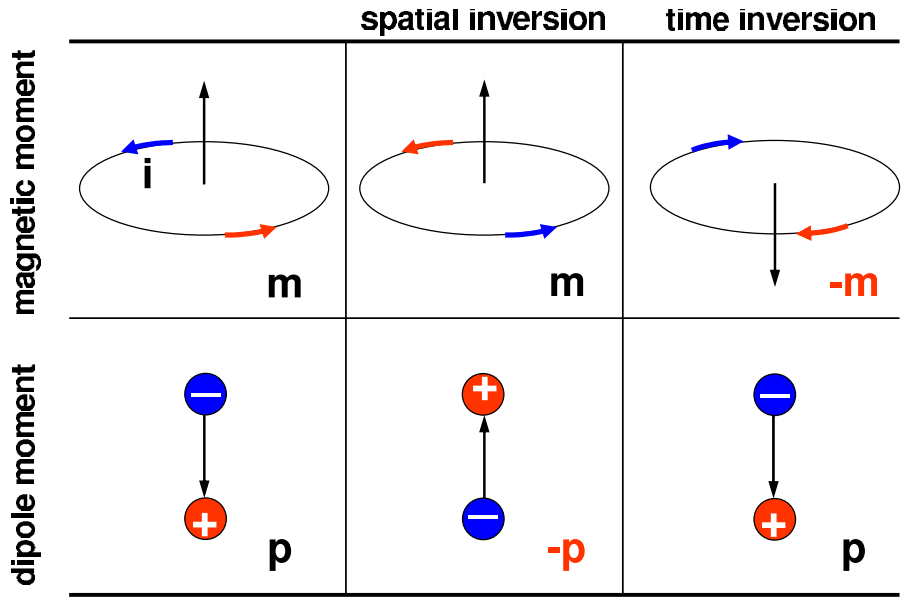


Fig. 1: *Symmetry of a dipole (***p***) and a magnetic (***m***) moment: spatial inversion reverses the direction of the dipole, but leaves the magnetic moment unchanged; on the contrary, time inversion switches the magnetic moment and leaves the dipole unchanged.*

Ferroic materials have certain symmetry properties which originate in the nature of the order parameters characterizing them. Let us observe how the spatial and time inversion act on the local magnetic and dipole moment (Fig. 1). Local magnetic moments originate in spins of

the electrons and in orbital angular momenta. Their symmetry properties, however, can be understood from a classical example of the magnetic moment of a current loop: spatial inversion will map the current vector **i** to its antipode on the current loop and will change its direction, leaving so the magnetic moment unchanged; time inversion will reverse the direction of the current and the magnetic moment will also be reversed. A dipole moment is produced by spatially separated positive and negative charge. Spatial inversion will obviously reverse its direction; since the charge separation is static in time, the dipole moment will be invariant under the time reversal. The symmetry of ferroic materials, summed up in Fig. 2, reflects these simple observations.

	spatial inversion	time inversion
ferromagnetic	✓	✗
ferroelectric	✗	✓
ferroelastic	✓	✓
ferrotoroidic	✗	✗

Fig. 2: *Symmetry of ferroic materials; note that a multiferroic which is at the same time ferromagnetic and ferroelectric cannot have either of the inversion symmetries.*

Order parameters which characterize ferroic materials can often exhibit weaker or stronger coupling to different external fields (Fig. 3):
Magnetoelectric coupling describes the change of magnetization (polarization) of a material when an external electric (magnetic) field is applied.
Piezoelectric coupling is what brings change to polarization (deformation) when an external stress (electric field) is applied.
Piezomagnetic coupling changes magnetization (deformation) upon application of stress (magnetic field).
Both piezomagnetism and piezoelectricity describe changes of order parameters which are linear functions of applied fields. Changes in deformation which are quadratic functions of applied magnetic (electric) fields are caused by the phenomenon of ***magnetostriction*** (***electrostriction***). It should be noted that strong couplings need not be a characteristic of a multiferroic material, neither are they restricted to multiferroics [3].
Although the broad definition of a multiferroic material involves coexistence of any two (or more) ferroic orders, the term *multiferroic* is in practice used when referring to a material which orders magnetically and ferroelectrically in a single phase (we will also restrict ourselves to this meaning from now on). Such materials offer new possibilities to information storage applications, such as the possibility of encoding information in the magnetization and the polarization state independently, in a single multiferroic element. Experimentally, four-state memory has already been demonstrated [4] (the current technology uses a 2-state, *i.e.* 2-bit, memory). Using materials where magnetization and polarization are coupled, it was suggested that information could be written by electric and read out by magnetic field. In this way, a couple of important problems one meets in the developing MRAM (magnetic random access memory) and FeRAM (ferroelectric random access memory) would be avoided. Namely, a large, spatially localized

	magnetic field	electric field	stress
magnetization	direct coupling	magnetoelectric coupling	piezomagnetism
electric polarization	magnetoelectric coupling	direct coupling	piezoelecticity
deformation	piezomagnetism	piezoelecticity	direct coupling

Fig. 3: *Coupling of order parameters to applied external fields.*

magnetic field would be needed for writing in a MRAM, while the readout action of a FeRAM destroys the read information. Due to the multifunctional character and easy manipulation of possible devices based on ferroelectric magnetic materials, it is exactly this class of multiferroics that has attracted most of the attention in the recent years.

2 Ferroelectricity and ferromagnetism

While ferromagnetic materials have been known for a long time, the study of the first material identified as a ferroelectric, the Rochelle salt [5, 6], started only a bit over a century ago. The observed spontaneous polarization which can be reversed when an external electric field is applied, led to the term *ferroelectricity*, by analogy with the known ferromagnetism.

Both classes of materials are characterized by hysteresis loops (Fig. 4a). In case of a ferro- or ferrimagnet the hysteresis connects the applied magnetic field H and the magnetization M (or magnetic induction B). In case of a ferroelectric there is a hysteresis between the applied electric field E and the polarization P (or the dielectric displacement D). The work done in reversing the magnetization (polarization) is in both cases given by the area enclosed by the hysteresis. A common feature of ferromagnetic and ferroelectric materials is the formation of domains, areas with unidirectional magnetization (polarization), which is why the as-prepared samples often lack a measurable macroscopic ferroic order; this state corresponds to the coordinate origin in Fig. 4a and to the middle panel of the Fig. 4b. When a sufficiently strong external magnetic (electric) field is applied, the sample reaches the *saturation* magnetization M_s (polarization P_s) (Fig. 4b, right). When the field is reduced to zero after saturation, the magnetization (polarization) decreases from M_s (P_s) to M_r (P_r), called the *residual* magnetization (polarization). The reversed field needed to reduce the magnetization (polarization) to zero is called the *coercive field* (H_c, E_c).

Another phenomenological similarity is that both ferromagnets and ferroelectrics have a characteristic Curie temperature T_C, beyond which the magnetic moments (electric dipoles) are no more aligned and the material is in a *paramagnetic* (*paraelectric*) phase. In both classes of

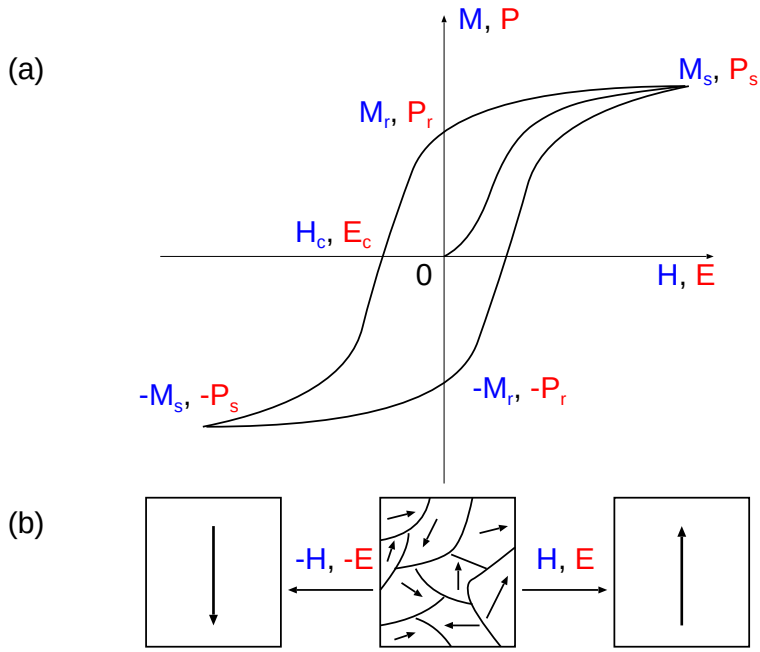


Fig. 4: *(a) Hysteresis loop in a ferromagnet or a ferroelectric. (b) Domains in a ferromagnet or in a ferroelectric; middle: disordered, without an external field; left/right: fully ordered, at the saturation points.*

materials, the ferroic order appears because it is energetically favorable, but this is where the analogy ends. The physical origins of the two phenomena are quite different. While magnetism originates in the interaction of electronic spins, ferroelectricity involves lattice distortions due to the interaction between electrons and nuclear displacements (vibronic interactions).

3 Ferromagnetism

Many properties of ferromagnets can be explained by two phenomenological theories, the Curie-Weiss localized moment theory and the Stoner band theory of ferromagnetism. Both theories show that localized electrons, such as transition metal d or rare earth f, are the basic requirement for magnetism to occur. This is an important point and we will come to it later again, when considering the conditions for simultaneous existence of magnetism and ferroelectricity in multiferroics.

Curie-Weiss theory developed on a postulate of Weiss [7] that there is an internal molecular field which acts on the magnetic moments in ferromagnetic materials and aligns them. Today this molecular field is understood to originate in the quantum mechanical exchange energy: due to the Pauli principle, electrons with parallel spins will stay away from each other, lowering in this way their energy due to the smaller Coulomb repulsion; exchange, however, does not limit the proximity of two electrons with antiparallel spins. Above the ordering temperature T_C, the alignment energy of the molecular field becomes smaller than the thermal energy kT and the directions of the magnetic moments become random, *i.e.* the material shows a paramagnetic behavior. The Curie-Weiss law,

Metal	Na	Al	Cr	Mn	Fe	Co	Ni	Cu	Pd	Pt
$n(E_{\mathrm{F}})$ [(eV)$^{-1}$]	0.23	0.21	0.35	0.77	1.54	1.72	2.02	0.14	1.14	0.79
I [eV]	1.82	1.22	0.76	0.82	0.93	0.99	1.01	0.73	0.68	0.63
$In(E_{\mathrm{F}})$ [eV]	0.41	0.25	0.27	0.63	1.43	1.70	2.04	0.11	0.78	0.50

Table 1: *Bulk density of states $n(E_F)$ at the Fermi energy (E_F) as calculated from nonmagnetic calculations, the Stoner parameter I, and the product of both, $In(E_F)$. All results are obtained with the density functional theory in the local density approximation [9, 10].*

$$\chi = \frac{C}{T - T_C}, \tag{1}$$

where χ is the magnetic susceptibility, C material-specific Curie constant and T temperature, successfully predicts the experimentally observed fact that at the ordering temperature, $T = T_C$, the magnetic susceptibility in many ferro- and antiferromagnetic materials diverges. However this theory also predicts magnetic moments on each atom to be formed by an integer number of electrons and this value should remain the same irrespective on whether the material is it a ferro- or antiferromagnetic state. Neither of the two is observed experimentally. The discrepancy is especially striking in metallic systems, where the Stoner band theory of ferromagnetism provides a better description.

Stoner theory [8] expresses a competition between the exchange interaction and the kinetic energy, where the former is described by the *exchange integral* I, and the latter by the density of electronic states at the Fermi energy $n(E_F)$, in the non-magnetic state. The exchange energy will again act as an aligning force on the electronic spins. However, the electrons which have to stay far away from each other will have an increased kinetic energy, since each of them will be more localized in space, much like an electron in a potential well the energy of which increases when the well width is reduced.
The exchange integral I is the measure of energy gained by aligning the electronic spins. The kinetic energy is higher for broad, dispersive bands, *i.e.* for a low density of states $n(E)$. The higher the value of $n(E_F)$, the lower is the kinetic energy of electrons at the Fermi level. Thus, what we need for magnetism to appear is a large I and a high density of states $n(E_F)$. This analysis is summed up in the simple *Stoner criterion* for ferromagnetism:

$$In(E_F) > 1. \tag{2}$$

Table 1 lists the exchange integral I, the local DOS at the Fermi energy $n(E_F)$ derived from nonmagnetic calculations and the product $In(E_F)$ for a number of elemental metals. It shows that the Stoner condition for ferromagnetism is only fulfilled for Fe, Co, and Ni, precisely those metals that show itinerant ferromagnetism.

Predicting the magnetic ground state of a magnetic system can be a highly nontrivial problem. In cases, for example, where competing exchange interactions between neighboring atoms exist, a multitude of possible spin-structures arises. A relatively simple model often used to describe

the magnetism of complex spin structures is the classical *Heisenberg Hamiltonian*,

$$H = -\sum_{\substack{i,j \\ i>j}} J_{ij}\, \mathbf{M}_i \cdot \mathbf{M}_j \ . \tag{3}$$

The magnetic moments localized on the lattice sites i, j are considered as classical vectors $\mathbf{M}$, with the assumption that their magnitudes M are constant. The exchange interaction between the magnetic moments is described by the pair interaction J_{ij}. From the first-principles calculations one can extract the pair interactions and then *e.g.* use a Monte Carlo method to obtain, from the Hamiltonian (3), the magnetization of the material as a function of temperature and determine the T_C. In localized spin systems the J_{ij} can often be restricted to the ferromagnetic ($J_1 > 0$) or antiferromagnetic ($J_1 < 0$) nearest-neighbor (n.n.) interaction, i.e. $J_{ij} = 0$ for all i, j, except for $J_{n.n.} = J_1$. Also in itinerant magnets J_1 often dominates over the rest of the further distant pairs. However, in this case the electrons that are responsible for the formation of the magnetic state participate in the formation of the Fermi-surface and hop across the lattice. Thus, the validity of the Heisenberg model here becomes questionable, although usually reasonable values of ordering temperatures T_C are obtained.

4 Ferroelectricity

In contrast to the structural phase transitions that involve significant diffusion of atoms, the ferroelectric transition belongs to the *displacive phase transitions*, which only require small collective displacements of individual atoms (amounting to fractions of the nearest neighbour interatomic distances). Displacive transitions occur spontaneously and reversibly at specific pressure and temperature conditions.

4.1 Displacive phase transitions

A well-studied structure which presents a series of different displacive phase transitions is that of the perovskite materials with a general chemical formula ABX_3 [11, 12, 13]. A and B are cations of different sizes, while X is an anion which bonds to both. This structure is adopted by many oxides ABO_3 (Fig. 5 left), where the Oxygen atoms form octahedra in the centers of which the B-site cation is positioned, while the A-site cation occupies the cube corner positions. For example, $SrTiO_3$ undergoes a displacive phase transition where the TiO_6 octahedra make a small rotation about the [001] axis. Similarly, in $CaTiO_3$ and $MgSiO_3$, the octahedra tilt by different amounts about all three axes. Another type of displacive phase transition is seen in $PbTiO_3$ [14, 15], where the Pb^{2+} and Ti^{4+} cations move off-center along [001], generating a ferroelectric phase transition (Fig. 5 right) . This results in a tetragonal unit cell with $c > a$, where c is the hight of the unit cell, while a is the lattice parameter in the plane perpendicular to the shift.

The ferroelectric phase transition in $BaTiO_3$ [16, 17], however, is peculiar: although it appears to be very similar to $PbTiO_3$, the behavior is qualitatively different: here the Ti^{4+} ions appear to occupy a central site in the high temperature cubic phase only on average, whereas in practice that site is always a potential-energy maximum. The potential-energy minima for the Ti^{4+} cations are located away from the central site along the eight $\langle 111 \rangle$ directions, so that in the high-temperature phase the Ti^{4+} cations are hopping among the eight different sites. The ferroelectric

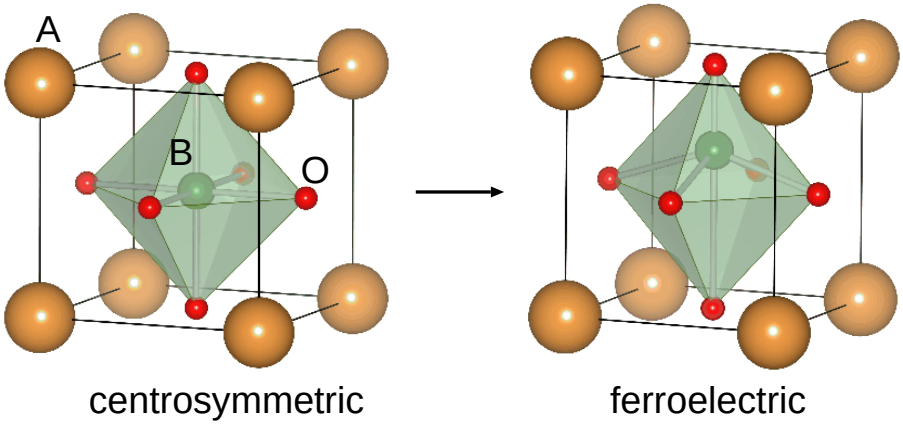


Fig. 5: *Perovskite structure ABO_3 in its centrosymmetric phase (left) and in one of possible ferroelectric phases (right).*

phase transition occurs when the Ti^{4+} cations begin to localize preferentially in the sites in the *e.g.* positive c direction. There are still four of these sites, so there are subsequent phase transitions when the material is cooled, till the Ti^{4+} cations all occupy the same one site in the unit cell.

Often, only the first type of phase transitions if tagged displacive, while the phase transitions like the one taking place in $BaTiO_3$ are referred to as the *order-disorder* transitions. It is not always easy to draw a line between the two. We will consider a simple model to investigate the factors which influence the nature of a transition [18]. The model consists of a 1-dimensional array of atoms, interacting with each other via harmonic forces, represented in Fig. 6 by the springs connecting the atoms. We will assume that each atom feels a double-well potential, without going into details of origin of such a potential shape (we will deal with this question later). For the moment, let us consider how our system is influenced by the two parameters which characterize it: one is the depth of the potential well V_0, and the other is the strength of the interaction between neighboring atoms (harmonic force constant) J. If the displacement of each atom i from the center of the potential well it is in is u_i, the local double-well potential can be expressed as

$$V_i \equiv V(u_i) = -\frac{1}{2}k_2 u_i^2 + \frac{1}{4}k_4 u_i^4, \tag{4}$$

where the parameters k_2 and k_4 are positive constants. The energy of each atom (of mass m)

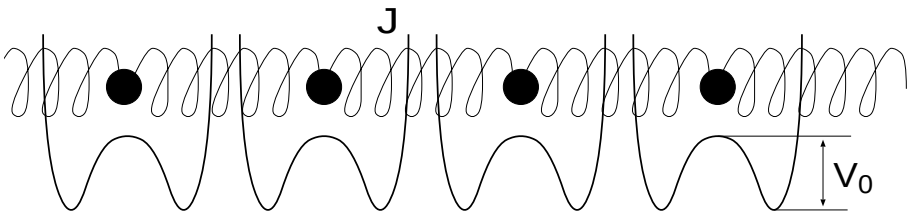


Fig. 6: *A one-dimensional model of a crystal which undergoes a displacive phase transition. Each atom is in a double-well potential of depth V_0 and interacts with its neighbors by harmonic forces represented by springs of stiffness J.*

also includes the kinetic term, $T_i \equiv T(\dot{u}_i) = m\dot{u}_i^2/2$ and the harmonic energy term due to the interactions of the atom i with its neighbor j, $W_{ij} = J(u_i - u_j)^2/2$, *i.e.* the Hamiltonian H of the system is

$$H = \sum_i T_i + \sum_i V_i + \sum_{i \neq j} W_{ij} = \sum_i T_i + \sum_i \left(-\frac{1}{2} k_2 u_i^2 + \frac{1}{4} k_4 u_i^4 \right) + \frac{1}{2} \sum_{i \neq j} J(u_i - u_j)^2. \quad (5)$$

Let us now see what the relative strengths of the two last terms in Eq. 5 can tell us about the system. The minima of the potential 4 are at $u_i = \pm u_0$, where $u_0^2 = k_2/k_4$, yielding the depth of the potential well,

$$V_0 = \frac{1}{4} k_2 u_0^2 \quad (6)$$

(we drop the subscript i since all the potential wells of the model are the same). The transition is determined by the strength of the interatomic interaction J. A stronger interaction means a higher transition temperature ($J \propto k_B T_C$). There are two limiting cases to the model:

$V_0 \gg J$ means that the atoms tend to remain near the minima of the potential wells, due to the very high potential barrier compared to the interaction between the neighboring atoms. Even at a temperature well above the transition T_C, the atoms will reside in one or the other minimum of the double potential well (because $V_0 \gg k_B T_C$), although the choice of the specific minimum will be random at first. This is the order-disorder case, an example of which is $BaTiO_3$. The ordering sequence for this case in presented in Fig. 7 on the left-hand side.

$V_0 \ll J$ is the displacive limit. Here the forces between atoms are much larger than the forces due to the local potential. Since now $V_0 \ll k_B T_C$, the atoms will not be forced to sit in one of the minima, but will vibrate about the origin instead. On cooling, the influence of the double potential well will become more important, and atoms start spending more time on one side of the origin. Because of the strong interatomic interactions, at the T_C these displacements will be in the same direction on all atoms (Fig. 7 right).

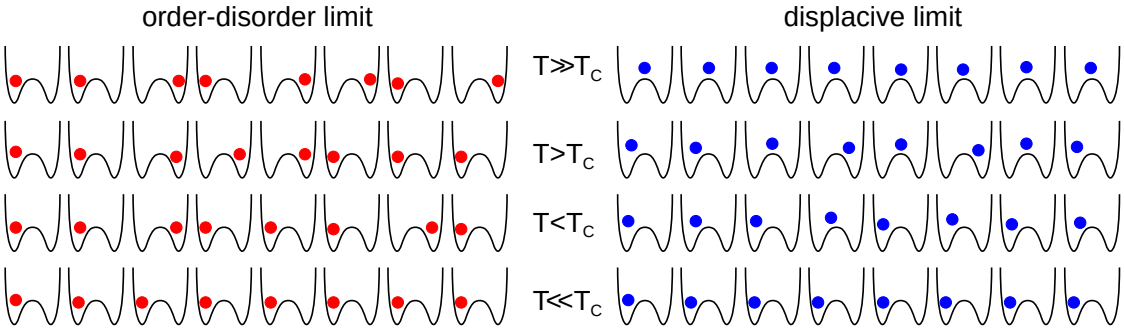


Fig. 7: *Ordering sequence in order-disorder (left) and in the displacive limit (right). The limits differ in strength of interatomic interactions relative to the depth of potential wells.*

One can also think of this case in the so-called *soft mode* picture [21]: the harmonic prefactor $-k_2$ in Eq. (4) is also equivalent to the square of the harmonic component of the oscillator frequency, ω^2. Since we assumed that k_2 is a positive constant, this means that the frequency

ω is imaginary. The oscillation mode the frequency of which becomes zero at T_C is called the soft mode and is said to *freeze in* at the transition temperature. At $T < T_C$ the frequency of the soft mode is imaginary. Phonon spectra of materials can be calculated from the first principles and the soft modes can be identified. This is a powerful tool in determining the ground-state structure. A phonon spectrum from Ref. [20] of $PbTiO_3$ is shown in Fig. 8, where the imaginary values identify the soft modes.

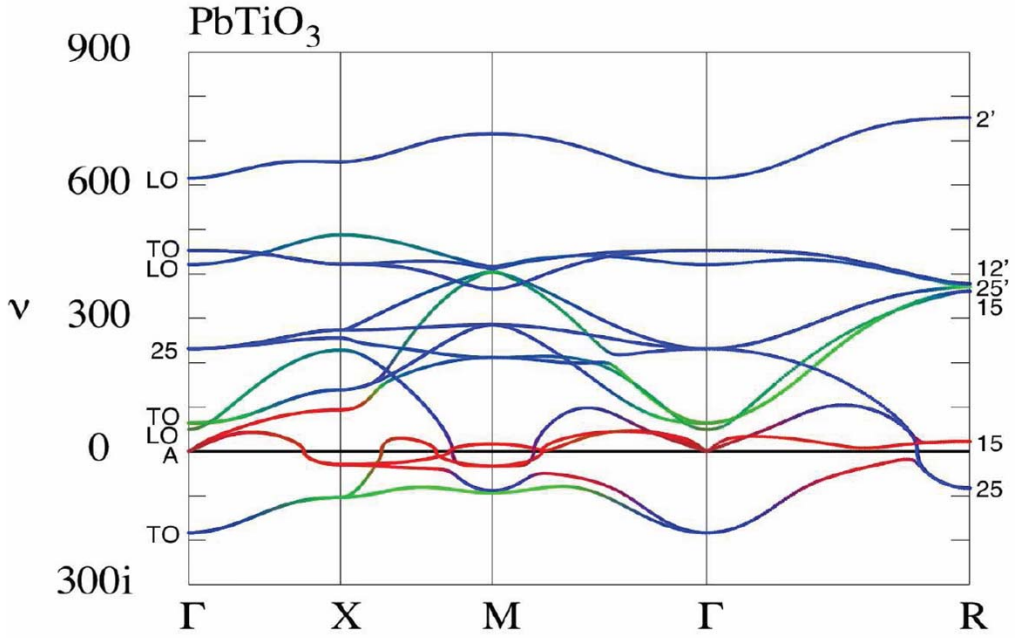


Fig. 8: *Phonon spectrum of $PbTiO_3$ (from Ref. [20]). Imaginary values indicate soft modes, the one with the lowest energy being at the center of the Brillouin zone (Γ-point phonon) and causing a ferroelectric distortion of the compound.*

4.2 Landau theory: proper and improper phase transitions

A theoretical treatment of phase transitions usually involves a construction of the associated (excess) free energy. The terms involved in the free-energy potential have to be chosen to fulfil certain restrictions imposed by symmetry. Such a phenomenological model uses thermodynamic variables to describe the macroscopic state of the crystal structure, irrespective of the actual microscopic properties (the actual topology of the crystal structure and the physical interactions between atoms). We have already met with this approach in Sec. 3.

A phase transition can be characterized by an order parameter, which quantifies the deviation from the high-symmetry phase (*paraphase*). Thus, the order parameter is zero by definition for the paraphase, while, provided appropriate normalization with respect to the ground-state structure is carried out, the order parameter is 1 at 0 K. In Landau theory, the free energy is expressed as a low-order Taylor expansion in terms of the order parameter, together with terms that couple the order parameter with other physical quantities (such as strain). From the free energy, with coefficients that are usually obtained by fitting experimental data, a lot of information about the system can be derived.

Assuming that the system under investigation can be described with only one order parameter

η, the excess Gibbs free energy of a phase transition (also called Landau free energy), $G_L(\eta) = G(\eta) - G(\eta = 0)$ can be expressed as

$$G_L(\eta) = \frac{1}{2}A\eta^2 + \frac{1}{3}b\eta^3 + \frac{1}{4}B\eta^4 + ... \tag{7}$$

In the first approximation, $A = a(T - T_C)$, where T is temperature (T_C is the ordering temperature), while all the other parameters are assumed to be constant. We will assume that G_L is symmetric with respect to η reversal and drop the odd terms in the expansion (7). In a simple case, the expansion can be truncated after the fourth-order term:

$$G_L(\eta) = \frac{1}{2}A\eta^2 + \frac{1}{4}B\eta^4. \tag{8}$$

From this expression, various thermodynamic quantities can be obtained, such as the (excess) entropy $S(\eta) = \partial G_L(\eta)/\partial T$, the specific heat $c_p = T\partial S/\partial T$, and the equilibrium order parameter $\eta(T) = \pm\sqrt{a(T_c - T)/B}$. The second derivative of $G_L(\eta)$ with respect to the order parameter yields the inverse susceptibility of η,

$$\frac{1}{\chi} = a(T - T_c) + 3B\eta^2, \tag{9}$$

corresponding to the curvature of the potential. For $T < T_c$ this can be simplified to $\chi^{-1} = 2a(T_c - T)$. When $T > Tc$, at $\eta = 0$, $\chi^{-1} = a(T - T_c)$ is obtained. We see that if χ is the magnetic susceptibility, this result reproduces the Curie-Weiss law (Eq. (1)). In the case of a ferroelectric material, the order parameter is the electric polarization $\mathbf{P}$. The susceptibility relates this polarization to an electric field $\mathbf{E}$ as $\mathbf{P} = \varepsilon_0\chi\mathbf{E}$ (ε_0 is the permittivity of vacuum) and is hence related to the static dielectric constant ε of the material: $\varepsilon = 1 + \chi$. Thus, at the ferroelectric phase transition the dielectric constant diverges.
A divergence of the susceptibility at the phase transition is what defines the *driving (intrinsic) order parameter*. If the system is characterized by other, coupled (or extrinsic) order parameters, their associated response functions (susceptibilities) usually do not show such a divergence [22].

Order parameter coupling

In systems described by two or more coupled order parameters, each one provides its own contribution to the Landau free energy and in addition coupling terms arise that may lower or raise G_L. We will see later (Sec. 5) that such a coupling of magnetization and polarization can be the driving force for the occurrence of multiferroicity. We denote the driving order parameter η_1 and the coupled order parameter η_2 and write:

$$G_L(\eta_1, \eta_2) = \frac{1}{2}A_1\eta_1^2 + \frac{1}{4}B_1\eta_1^4 + ... + \frac{1}{2}A_2\eta_2^2 + \frac{1}{4}B_2\eta_2^4 + ... + \lambda\eta_1^{n_1}\eta_2^{n_2} + ..., \tag{10}$$

where λ contains the energy associated with the coupling of the order parameters. We assume that only $A_1 = a(T - T_C)$ contains a temperature dependence and we exclude the odd-order terms. The exponents n_1 and n_2 have to be chosen such that they fulfill the symmetry requirements. The *biquadratic* coupling, $n_1 = n_2 = 2$, is always allowed since the square of the order parameter is invariant under the parameter's inversion, but it will rarely be the leading order term in coupling energy.

The *bilinear* coupling ($n_1 = n_2 = 1$) is allowed when both order parameters are associated with the same irreducible representation of the symmetry group of the paraphase. In this case, the transition observed in η_2 is termed *pseudo-proper*. The *linear-quadratic* coupling ($n_1 = 2$, $n_2 = 1$) occurs when the irreducible representations associated with η_1 and η_2 differ and the transition is labelled ***improper*** with respect to η_2. Finally, a ***proper*** transition is the transition in the driving order parameter η_1. For example, the phase transition in $PbTiO_3$ is a proper ferroelectric and improper ferroelastic at the same time. Further details of different types of transitions can be found in Ref. [22].

4.3 Vibronic theory of ferroelectricity

We will now slide down to the microscopic scale and look for the origin of the local double-well potential (Fig. 6) which is needed for a displacive phase transition to occur. The physical idea behind it can be traced back to the pseudo Jahn-Teller effects [23] and is expressed in the vibronic theory of ferroelectricity [24]. The basic statement of this theory is that, under certain conditions, the mixing of the electronic ground state with the near-lying excited ones caused by the dipole type nuclear displacements lead to an instability (or softening) of the high-symmetry nuclear configuration with regard to these displacements, resulting in a spontaneous polarization of the crystal.

We consider a system of electrons and nuclei oscillating about their equilibrium positions. The oscillation is described by a set of normal coordinates Q. We can write down the full Hamiltonian of the system as

$$H = H_r + H_Q + V(r, Q), \tag{11}$$

where r stands for the set of electronic variables, H_r is the electronic and H_Q the nuclear part, while $V(r, Q)$ describes the electron-nuclear interaction. We now expand the latter in a series with respect to the normal displacement of nuclei around the high symmetry configuration $Q = 0$:

$$V(r, Q) = V(r, 0) + \sum_{\alpha} \left(\frac{\partial V}{\partial Q_\alpha} \right)_0 Q_\alpha + \frac{1}{2} \sum_{\alpha\beta} \left(\frac{\partial^2 V}{\partial Q_\alpha \partial Q_\beta} \right)_0 Q_\alpha Q_\beta + ... \tag{12}$$

One can first solve the adiabatic part of the Hamiltonian (11), $H_A = H_r + V(r, 0)$. Let us assume its solutions are two energy levels, E_1 (the ground state) and E_2 (the excited state) with the corresponding wavefunctions ψ_1, ψ_2. The higher order terms in Eq. (12) we will take into account as perturbations, obtaining from the perturbation theory the secular equation

$$\begin{vmatrix} \Delta_{1,2} - \varepsilon & F_{1,2}Q \\ F^*_{1,2}Q & -\Delta_{1,2} - \varepsilon \end{vmatrix} = 0, \tag{13}$$

where $2\Delta_{1,2} = E_2 - E_1$. The *linear vibronic coupling constant*,

$$F_{1,2} = \left\langle \psi_1 \left| \left(\frac{\partial V}{\partial Q_\alpha} \right)_0 \right| \psi_2 \right\rangle \tag{14}$$

is non-zero only for the coordinate Q of the correct symmetry: the product of symmetry representations of ψ_1,ψ_2 and Q must contain the unit representation.

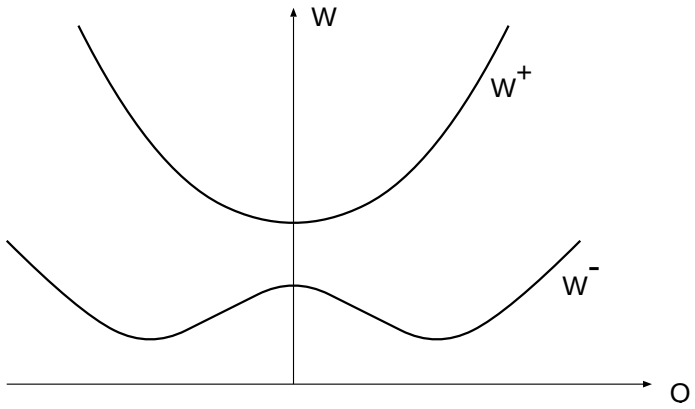


Fig. 9: *Adiabatic potential sheets (Eq. (15)) for a simple model-system with two energy levels*

The roots of Eq. (13) are $\varepsilon^{\pm}(Q) = \pm\sqrt{\Delta_{1,2}^2 + F_{1,2}^2 Q^2}$. Together with the ionic-core harmonic interaction term, $KQ^2/2$, where K is the force constant, these roots determine the *adiabatic potential sheets* (Fig. 9),

$$W^{\pm} = \frac{1}{2}KQ^2 \pm \sqrt{\Delta_{1,2}^2 + F_{1,2}^2 Q^2}. \tag{15}$$

We can now easily find the instability condition: the lower sheet $W^{-}(Q)$ will have a maximum at $Q_{max} = 0$ and two minima at $Q_{min} = \pm\sqrt{(F_{1,2}/K)^2 - (\Delta_{1,2}/F_{1,2})^2}$ if

$$K < \frac{F_{1,2}^2}{\Delta_{1,2}}. \tag{16}$$

In the case of many possible excited states, their effect on the instability is additive (this is the case of a *multilevel pseudo Jahn-Teller effect*). We label the excited states with $i > 1$, and write the instability condition under the influence of the normal displacements Q^{α} for this case as

$$K^{\alpha} < \sum_i \frac{F_{1,i}^2}{\Delta_{1,i}}. \tag{17}$$

The symmetry condition (selection rule) for a non-zero vibronic coupling constant mentioned earlier imposes restrictions on the transitions that can destabilize the system [25]. In addition to this, for higher excited states $\Delta_{1,i}$ is large thus reducing their contribution. Usually one or a few excited states contribute significantly to the instability of polyatomic systems.

4.4 Typical examples

Let us first consider the case of $BaTiO_3$. An analysis following the steps described on the simple example of Sec. 4.3 is also valid in this more complex case and yields eight minima of the adiabatic potential sheet along the $\langle 111 \rangle$ directions of $BaTiO_3$ lattice [25] (recall that Ti atoms in this compound off-center along the $\langle 111 \rangle$ directions in the low temperature phase).
Using first-principles calculations, we can investigate the change of the electronic structure of $BaTiO_3$ due to a displacement of Ti. Let us compare the density of states (DOS) of the centrosymmetric cubic structure (Fig. 10 up) and the one where the lattice is kept cubic, but the Ti atom is shifted towards one of the surrounding oxygens (in [001] direction) by 2% of the

lattice parameter (Fig. 10 down) [26]. In order to emphasize the distortion effect, this shift was here exaggerated with respect to the ones usually observed in ferroelectrics. First, note that the d states of Ti hybridize with the O $2p$ states. The electrons are in the ground state occupying the lowest energy levels, residing in this way mostly on O which thus becomes a negatively charged ion, O^{2-}; Ti, on the other hand, ends up as a Ti^{4+} ion, formally in a d^0 state. The lowest unoccupied energy levels have the character of Ti d-states, while the highest occupied ones have the O p-states character. Even in the cubic undistorted structure, the hybridization results in a partial occupation of the Ti e_g and t_{2g} states: the ionic configuration calculated in Ref. [26] is $Ba^{1.98+}Ti^{1.71+}O_3^{1.24-}$.

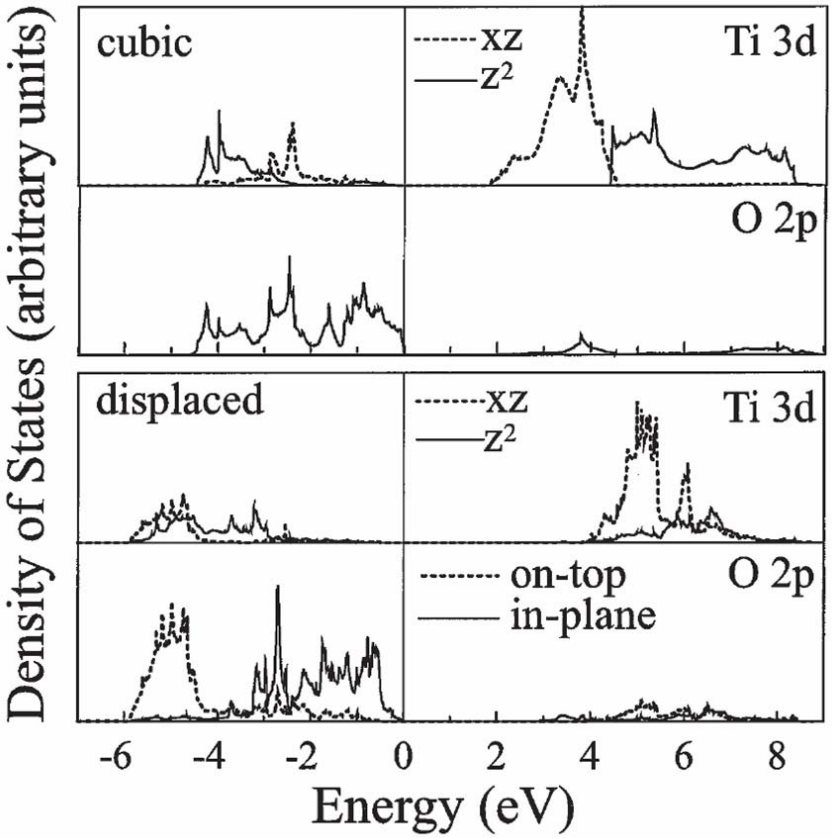


Fig. 10: *Density of states of O p and Ti d in $BaTiO_3$ (from Ref. [26]) in centrosymmetric cubic structure (up) and in a polar structure where the Ti atom was shifted along [001] direction by 2% of the lattice parameter (down). The zero of energy is at the Fermi level.*

The DOS changes significantly upon distortion (Fig. 10 down). The reduction of the cubic symmetry leads to a splitting of t_{2g} orbitals into one singlet (d_{xy}, lying in the plane perpendicular to the shift) and one doublet (d_{xz} and d_{yz}). Similarly, the e_g states are split into two singlets, d_{z^2} and $d_{x^2-y^2}$ which is, like the d_{xy}, perpendicular to the shift. The oxygen p states assume different energies depending on whether the O atom lies in the plane perpendicular to the shift or on the axis of the shift (on-top). In Fig. 10, the states which are affected by the shift are shown: the oxygen p, and the Ti d_{xz} (d_{yz} are the same) and d_{z^2}. The states perpendicular to the shift stay mostly unaffected. Analysis of the DOS reveals two kinds of hybridizations between O p and Ti d states: a pd_σ hybridization mixing p_z and d_{z^2} orbitals, and a pd_π hybridization involving p_x, p_y, d_{xz} and d_{yz} states [37].

It is very important to note that the shift causes a charge transfer from O to Ti and the occupancies of Ti d_{z^2}, d_{xz} and d_{yz} (the orbitals oriented in the direction of the displacement) increase (this event is also referred to as the *charge-transfer vibronic mixing* [27]). The fact that all the d states of Ti are (formally) empty (or, as it is often called, the d^0-*ness* of Ti) means that the charge transfer can occur for the shifts along every direction. This is consistent with the observation of several different phases of $BaTiO_3$ with Ti shifts along different directions. Figure 11 visualizes the formation of the new covalent bonds under the lattice distortion at the center of the Brillouin zone: in the high symmetry configuration the total overlap of the unoccupied Ti $3d_\pi$ and O $2p_\pi$ is zero. However, when the Ti atom is displaced the overlap is non-zero, resulting in the additional covalency and the charge transfer.

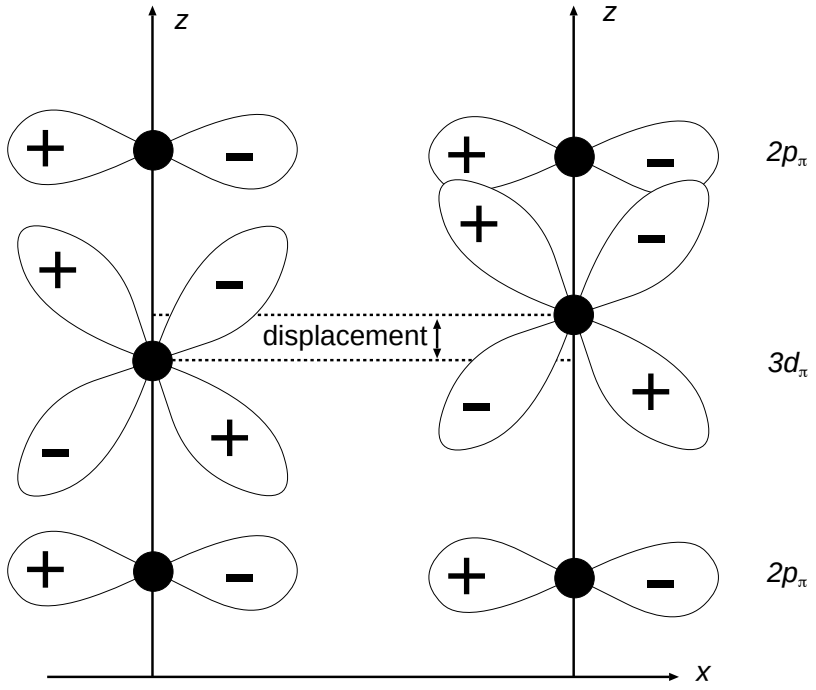


Fig. 11: *Formation of new covalent bonds under the lattice distortion: when the Ti atom shifts along the z-axis, overlap of Ti $3d_\pi$ and O $2p_\pi$ is non-zero and new bonds are formed.*

We will turn now to an other perovskite ferroelectric material mentioned earlier, $PbTiO_3$. Its ground-state structure is tetragonal, with a ferroelectric off-centering of Ti along [001] direction. The hybridization mechanism leading to the off-centering is the same as that in $BaTiO_3$, but in the case of $PbTiO_3$ the tetragonal structure is stable, while $BaTiO_3$ undergoes a series of transitions from the high-temperature cubic to tetragonal to orthorombic to rhombohedral with the low-temperature polarization along [111]. A comparison of the two sheds some light on the role of the A-site cation (here Ba^{2+} or Pb^{2+}).

Figure 12 compares the DOS of the two compounds [28], using the experimental ferroelectric $PbTiO_3$ structure for both, so that the differences are entirely due to the A-site cation properties. An obvious difference is that while the Ba $2p$ states do not hybridize with the valence band, the Pb $6s$ show a strong hybridization with the oxygen $2p$. This bonding, along with the smaller ionic radius of Pb^{2+} compared to that of Ba^{2+}, leads to the larger strain in $PbTiO_3$, stabilizing the tetragonal lattice. An indirect effect of the Pb-O bonding on the Ti-O hybridization is that the Ti $3d$ states are in $PbTiO_3$ lower than in $BaTiO_3$, even for the same Ti displacement.

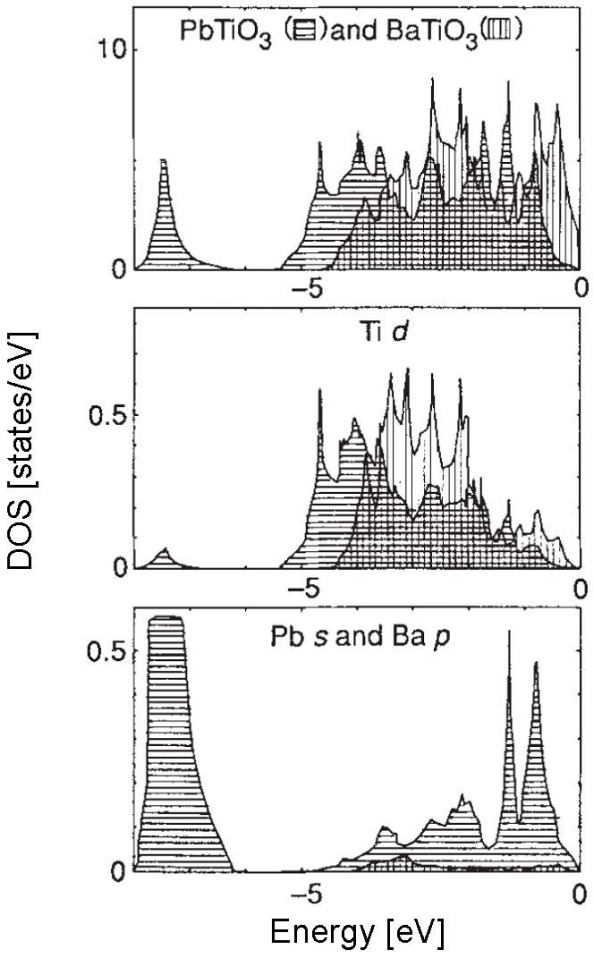


Fig. 12: *Comparison of densities of states in $BaTiO_3$ (vertical lines) and $PbTiO_3$ (horizontal lines). Up: total density of states; middle: density of states on the Ti atom; down: density of states of Pb s in comparison to the Ba p (from Ref. [28]). Both compounds were taken to have the same, experimental ferroelectric $PbTiO_3$ structure. The zero of energy is at the Fermi level.*

It should be noted that in many IV-VI semiconductors a structural distortion can actually be driven solely by the ns^2 valent electrons of the cation [29, 30, 31]. The ns^2, usually referred to as the *lone pair*, can lose inversion symmetry due to a mixing of the ns^2 ground state with the low-lying excited ns^1np^1 state. This mixing can only occur if the ionic site does not have inversion symmetry [32] and so when the energy gain due to the mixing is larger than the inter-ionic repulsion which opposes the ionic shift, the crystal distorts.

5 Magnetic ferroelectrics

We have now established that for magnetism the partially filled d-states are needed, while the ferroelectricity requires the "d^0-ness". This chemical incompatibility was pointed out by Hill [33] in 2000 as the reason for the rarity of both orders in a single material. This argument initiated a big search for alternative mechanisms which could lead to multiferroicity. Although several mechanisms have been identified [34, 35], a room temperature multiferroic is still a challenge: magnetic ordering temperatures are mostly lower than the ferroelectric ones, and

while the ferroelectricity might still hang on well above the room temperature, for the magnetic ordering to also occur the samples have to be cooled.
In this section we will see on examples how the mentioned chemical incompatibility is circumvented and ferroelectricity induced in several magnetic materials. We will start with the examples where the electronic pairing is the driving force for a structural distortion, similarly to the cases of $BaTiO_3$ and $PbTiO_3$. The ferroelectrics with this kind of distortion mechanism are called *proper*. However, ferroelectricity can also occur as an accidental by-product of a different kind of ordering. In this case, the resulting ferroelectric is tagged *improper*.

An example of a proper magnetic ferroelectric is $BiMnO_3$ in perovskite structure, where the $6s^2$ lone pair on the Bi ion plays a role in stabilizing the ferroelectric order [39, 40]. Bi cations move in the direction opposite to oxygen-manganese cage, the shift being induced by the Bi-O covalency. It should be noted that, surprisingly, this compound is also a ferromagnet, although an antiferromagnetic state would be expected due to the superexchange interaction. [43] One possible explanation [26] is that the electronegativity of Bi enhances Bi-O hybridization and in turn reduces the amount of Mn-O hybridization. The combination of structural distortion and reduced Mn $3d$-O $2p$ overlap reduces the strength of the antiferromagnetic superexchange interaction, making the observed ferromagnetic coupling more favorable. Although several experiments seem to confirm the ferroelectricity of $BiMnO_3$ [42], recent first-principles calculations [41] suggest that the bulk compound should actually be antiferroelectric. The experimentally observed ferroelectricity could be a consequence of strain or defects. This question is still open.
An other case of (proper) ferroelectricity driven by the $6s^2$ lone pair on Bi is $BiFeO_3$ [44, 45], with the polarization oriented along [111] direction. The magnetic moments on Fe atoms are coupled ferromagnetically in the (111) planes and antiferromagnetically between the adjacent planes. On the top of this antiferromagnetic magnetic structure, a long-wavelength spin-spiral is superimposed.
Because magnetism and ferroelectricity in these two compounds are due to different ions (responsible for ferroelectricity is Bi, while magnetism is introduced by the transition-metal ions), the coupling between the two order parameters is weak, limiting their possible technological applications. For this reason, the improper ferroelectrics are getting into focus of current research in multiferroics: it is understandable that if an ordering is a consequence of an other (the order parameter is extrinsic), it should be easily manipulated by the field corresponding to the driving order parameter.

There are several known mechanisms of inducing ferroelectricity as a secondary order [34]. One way is *charge ordering*: in many metals with strong electronic correlations charge carriers become localized at low temperatures, forming periodic superstructures, and the material undergoes a metal-insulator transition. If the charges order in a non-symmetric fashion, they induce electric polarization. An example of ferroelectricity induced by charge ordering is ferrimagnetic $LuFe_2O_4$, which has a bilayered structure with a triangular lattice of Fe ions in each layer. Below ~ 350 K, the charge ordering creates alternating layers with Fe^{2+}:Fe^{3+} ratios of 2:1 and 1:2 and a net polarization is induced (Fig.13) [34, 46, 47].
Another kind of improper ferroelectricity is *geometric ferroelectricity*, such as the one in hexagonal $YMnO_3$. This compound is experimentally established to be ferroelectric, with the hexagonal perovskite structure[36]. The magnetic ordering is A-type AFM, with noncollinear Mn spins oriented in a triangular arrangement. Due to the changes in the phonon spectrum at the

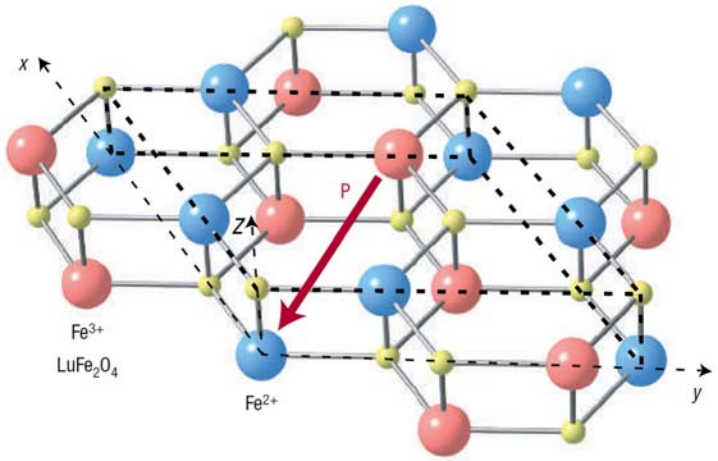


Fig. 13: *Ferroelectricity induced by charge ordering in $LuFe_2O_4$ (from Ref. [34]): charge is transferred from the upper to the lower layer resulting in electric polarization.*

antiferromagnetic transition [38], a coupling of ferroelectric and magnetic ordering occurs. Ferroelectricity in this case appears as a product of the nonlinear coupling to nonpolar lattice distortions, such as the buckling of Y-O planes and the tilts of Mn-O bipyramids [48].

Improper ferroelectricity which is most interesting from the technological point of view is the kind induced by magnetic ordering, enabling manipulation of electric polarization by an external magnetic field. This has been proven to be experimentally possible: the small electric polarization of perovskite $TbMnO_3$ was successfully rotated by 90° using a magnetic field [49]. The onset of ferroelectricity in $TbMnO_3$ correlates with the transition to a spiral spin-density wave. In order to understand the underlying mechanism, let us consider the case of cubic symmetry and the part of the thermodynamic potential due to magneto-electric coupling ($\mathbf{P}$ is polarization and $\mathbf{M}$ magnetization),

$$\Phi_{me}(\mathbf{P},\mathbf{M}) = \lambda \mathbf{P}\cdot[\mathbf{M}(\nabla\cdot\mathbf{M}) - (\mathbf{M}\cdot\nabla)\mathbf{M} + ...] \tag{18}$$

(the omitted terms do not contribute to the uniform polarization) [50]. We assume that there is no ferroelectric instability in the absence of magnetism and thus we keep only the quadratic part in the electric part of the thermodynamic potential, $\Phi_e = P^2/(2\chi)$, χ being the electric susceptibility in the absence of magnetic order. From the variation of $\Phi_e + \Phi_{me}$ with respect to $\mathbf{P}$,

$$\mathbf{P} = \lambda\chi[\mathbf{M}(\nabla\cdot\mathbf{M}) - (\mathbf{M}\cdot\nabla)\mathbf{M}]. \tag{19}$$

Let us assume now that a spin-density wave with a wave vector $\mathbf{q}$ is present:

$$\mathbf{M} = \mathbf{e}_1 M_1\cos(\mathbf{q}\cdot\mathbf{r}) + \mathbf{e}_2 M_2\sin(\mathbf{q}\cdot\mathbf{r}) + \mathbf{e}_3 M_3, \tag{20}$$

where $\mathbf{e}_{1,2,3}$ are orthogonal unit vectors. From Eq. (19) and (20), the space-average polarization is found to be independent of M_3 and orthogonal both to $\mathbf{e}_3$ and $\mathbf{q}$,

$$\overline{\mathbf{P}} = \lambda\chi M_1 M_2[\mathbf{e}_3\times\mathbf{q}]. \tag{21}$$

Note that, if only M_1 or only M_2 is non-zero, Eq. (20) defines a sinusoidal spin-density wave. For this case from Eq. (21) follows that no polarization is induced. The resulting polarization is non-zero only for a helicoidal spin-spiral, *i.e.* when M_1, $M_2 \neq 0$ (Fig. 14). Here one should pay attention to the terminology: many authors refer to a spin-spiral as helical only in the case when the wavevector q is perpendicular to the plane of spiraling (defined by the two vectors, $\mathbf{e}_1$ and $\mathbf{e}_2$), *i.e.* parallel to $\mathbf{e}_3$. From Eq. (21) one can see that this spiral induces no polarization. When the wavevector lies in the spiraling plane, a polarization is induced and the spin-spiral is labeled cycloidal.

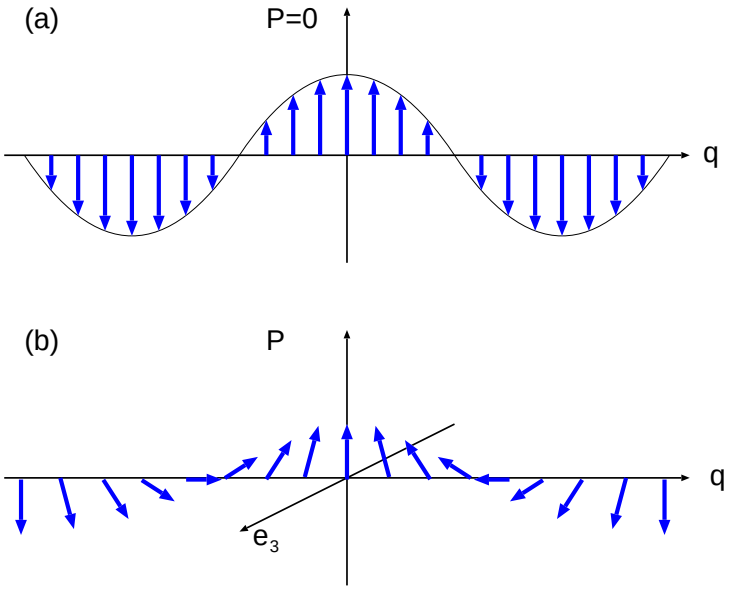


Fig. 14: *A sinusoidal (a) and a cycloidal (b) spiral defined by Eq. (20) According to Eq. (21), only the cycloidal spiral induces polarization [50].*

It has been experimentally observed [49] that in $TbMnO_3$ a transition to a phase with a sinusoidal spin-density wave at $T = 41$ K does not yield polarization. Ferroelectricity is induced only at the transition from the phase with sinusoidal to the one with helical spin-density wave at $T = 28$ K. This observation is consistent with the result above. Mostovoy [50] also showed that, due to the same mechanism, electric polarization can also be induced in domain walls and magnetic vortices.

Finally, it should also be noted that a non-collinear spin arrangement is not a necessary condition for magnetically induced improper ferroelectricity. Recently, Picozzi *et al.* have demonstrated that certain collinear antiferromagnetic configurations of spins in orthorombic $HoMnO_3$ and $TbMnO_3$ also induce electric polarization [51, 52].

Acknowledgements

I gratefully acknowledge the help I received from Nicola Spaldin in understanding the physics of ferroelectrics and multiferroics. For many useful discussions I owe a big Thank You to Phivos Mavropoulos, Yuriy Mokrousov and Stefan Blügel.

References

[1] H. Schmid, Ferroelectrics **162**, 317 (1994)

[2] K. M. Rabe, Nature **449**, 674 (2007)

[3] W. Eerenstein, N. D. Mathur, and J. F. Scott, Nature **442**, 759 (2006)

[4] M. Gajek, M. Bibes, S. Fusil, K. Bouzehouane, J. Fontcuberta, A. Barthélémy, and A. Fert, Nature Mat. **6**, 296 (2007)

[5] F. Pockels, Abh. Gött **39**, 1 (1894)

[6] J. Valasek, Phys. Rev. **17**, 475 (1921)

[7] P. Weiss, J. Phys. **6**, 661 (1907)

[8] E. C. Stoner, Phylos. Mag. **15**, 1080 (1933)

[9] O. Gunnarsson, J. Phys. F: Met. Phys. **6**, 587 (1976)

[10] J. F. Janak, Phys. Rev. B, **16**, 255 (1977)

[11] A. D. Bruce , and R. A. Cowley, *Structural Phase Transitions*, Taylor and Francis, London (1981)

[12] R. Blinc, and B. Zeks, *Soft Modes in Ferroelectrics and Antiferroelectrics*, North Holland, Amsterdam (1974)

[13] M. E. Lines, and A. M. Glass, *Principles and Applications of Ferroelectrics and Related Materials*, Clarendon Press, Oxford(1977)

[14] G. Shirane, J. D. Axe, J. Harada, and J.P. Remeika, Phys. Rev. B **2**, 155 (1970)

[15] G. Burns and B. A. Scott, Phys. Rev. Lett. **25**, 167 (1970)

[16] Y. Yamada, G. Shirane, and A. Linz, Phys. Rev. **177**, 848 (1969)

[17] J. Harada, J. D. Axe, and G. Shirane, Phys. Rev. B **4**, 155 (1971)

[18] M. T. Dove, *Introduction to lattice dynamics*, Cambridge University Press (1993)

[19] M. T. Dove, American Mineralogist **82** 213 (1997)

[20] P. Ghosez and J. Junquera, in *Handbook of theoretical and computational nanotechnology*, edited by M. Reith and W. Schommers, Volume 9, 623, American Scientific Publishers, Stevenson Ranch, California, USA (2006)

[21] W. Cochran, Phys. Rev. Lett. **3**, 412 (1959)

[22] T. Malcherek, EMU Notes in Mineralogy, **7**, 139 (2005)

[23] I. Perlin, M. Wagner, *The dynamical Jahn-Teller effect in localized systems*, North Holland, Amsterdam (1984)

[24] I. B. Bersuker, Phys. Lett. **20**, 5891 (1966)

[25] I. B. Bersuker, Ferroelectrics **164**, 75 (1995)

[26] N. A. Hill, Annu. Rev. Mater. Res. **32**, 1(2002)

[27] I. B. Bersuker, Ferroelectrics **153**, 1 (1994)

[28] R. E. Cohen, Nature **358**, 136 (1992)

[29] G. Trinquier, R. Hoffmann, J. Phys. Chem. **88**, 6696 (1984)

[30] G. W. Watson, S. C. Parker, J. Phys. Chem. B **103**, 1258 (1999)

[31] G. W. Watson, S. C. Parker, G. Kresse, Phys. Rev. B **59**, 8481 (1999)

[32] M. Atanasov, D Reinen, J. Phys. Chem. A **105**, 5450 (2001)

[33] N. A. Hill, J. Phys. Chem. B **104**, 6694 (2000)

[34] S.-W. Cheong and M. Mostovoy, Nature Mat. **6**, 13 (2007)

[35] N. A. Hill and A. Filippetti, J. Magn. Magn. Mater. **242**, 976 (2002).

[36] H. L. Yakel, W. C. Koehler, E. F. Bertaut, E. F. Forrat, Acta Crystallogr. **16**, 957 (1963)

[37] A. Filippetti and N. A. Hill, Phys. Rev. B, **65**, 195120 (2002)

[38] Z. J. Huang,Y. Cao,Y.Y. Sun,Y.Y. Xue,C. W. Chu, Phys. Rev. B **56**, 2623 (1997)

[39] N. A. Hill, K. M. Rabe, Phys. Rev. B **59**, 8759 (1999)

[40] R. Seshadri, N. A. Hill, Chem. Mater. **13**, 2892 (2001)

[41] P. Baettig, R. Seshadri, and N. A. Spaldin, J. Am. Chem. Soc., **129**, 9854 (2007)

[42] A. Moreira dos Santos, S. Parashar, A. R. Raju, Y. S. Zhao, A. K. Cheetham, and C. N. R. Rao, Solid State Commun. **122**, 49 (2002)

[43] J. B. Goodenough, Phys. Rev. **100**, 564 (1955)

[44] C. Ederer and N. Spaldin, Phys. Rev. B **71**, 060401(R) (2005)

[45] J. B. Neaton, C. Ederer, U. V. Waghmare, N. A. Spaldin, and K. M. Rabe, Phys. Rev. B **71**, 014113 (2005)

[46] N. Ikeda, H. Ohsumi, K. Ohwada, K. Ishii, T. Inami, K. Kakurai, Y. Murakami, K. Yoshii, S. Mori, Y. Horibe, and H. Kitô, Nature **436**, 1136 (2005)

[47] M. Naka, A. Nagano, and S. Ishihara, Phys. Rev. B. **77**, 224441 (2008)

[48] C. J. Fennie and K. M. Rabe, Phys. Rev. B. **72**, 100103(R) (2005)

[49] T. Kimura, T. Goto, H. Shintani, K. Ishizaka, T. Arima, and Y. Tokura, Nature **426**, 55 (2003)

[50] M. Mostovoy, Phys. Rev. Lett. **96**, 067601 (2006)

[51] S. Picozzi, K. Yamauchi, B. Sanyal, I. A. Sergienko, and E. Dagotto, Phys. Rev. Lett **99**, 227201 (2007)

[52] K. Yamauchi and S. Picozzi, J. Phys.:Condens. Matter **21**, 064203 (2009)

A 7 Berry phase in solid state physics

Ming-Che Chang
Dept. of Physics, National Taiwan Normal Univ.
Taipei, Taiwan

Contents

1 Anholonomy in geometry **2**
1.1 Parallel transport and anholonomy angle 2
1.2 Moving frame and curvature 2

2 Anholonomy in quantum mechanics **4**
2.1 Introducing the Berry phase 4
2.2 A rotating solenoid 6

3 Berry phase and spin systems **8**
3.1 Persistent spin current 8
3.2 Magnetic cluster 10

4 Berry phase and Bloch state **12**
4.1 Electric polarization 12
4.2 Quantum Hall effect 14
4.3 Anomalous Hall effect 17

5 Berry phase and wave-packet dynamics **19**
5.1 Wave-packet dynamics 19
5.2 Non-Abelian generalization 22

6 Concluding remarks **27**

1 Anholonomy in geometry

Before introducing the Berry phase, we review the elegant mathematical framework behind it. It helps explaining why the Berry phase is often also called the geometric phase.

1.1 Parallel transport and anholonomy angle

Consider a two-dimensional curved surface embedded in a three dimensional Euclidean space. At each point $x = (x_1, x_2)$ on the surface, there is a vector space T_x formed by the tangent vectors at that point. For an ant living on the surface, is it possible to judge if two vectors at different locations (1 and 2) of the surface are nearly parallel or far from it?

One possible way to calibrate the difference between two vectors at different locations is as follows: Starting from point 1, the ant can carry the vector around in such a way that it makes a fixed relative angle with the tangent vector along a path between 1 and 2 (see Fig. 1a). Such a vector is said to be *parallel transported*. One can then compare the vector already at point 2 with the parallel transported vector for difference.

Notice that, if we follow this rule, then "being parallel" is a path-dependent concept. That is, one cannot have a global definition of "being parallel" on the curved surface. The other way to say the same thing is that, if you parallel transport a vector along a closed loop on the surface, then the final vector $\mathbf{v}_f$ is generically different from the initial vector $\mathbf{v}_i$ (see Fig. 1b).

The angle between these two vectors is called the anholonomy angle (or defect angle). Such an angle is an indication of how curved the surface is. One can use it to define the intrinsic curvature of the surface. For example, for a sphere with radius R, the defect angle α for a vector transported around a spherical triangle is equal to the solid angle Ω subtended by this triangle,

$$\alpha = \Omega = \frac{A}{R^2}, \tag{1}$$

where A is the area enclosed by the triangle.

One can define the curvature at point x as the ratio between α and A for an infinitesimally closed loop around x. According to this definition, the sphere has a constant curvature $1/R^2$ everywhere on the surface.

You can apply the same definition to find out the intrinsic curvature of a cylinder. The result would be zero. That is, the cylinder has no intrinsic curvature. That is why we can cut it open and lay it down on top of a desk easily without stretching.

1.2 Moving frame and curvature

In practice, apart from a few simple curved surfaces, it is not easy to determine the curvature without using algebraic tools. At this point, it helps introducing the method of the moving frame. We follow a very nice article by M. Berry (see Berry's introductory article in Ref. [1]) and apply this method to calculate the curvature.

Instead of moving a vector, one now moves an orthonormal frame (a triad) along a path C between two points. At the starting point, the triad is $(\hat{r}, \hat{e}_1, \hat{e}_2)$, where $\hat{r}$ is the unit vector along the normal direction and $(\hat{e}_1, \hat{e}_2)$ is an orthonormal basis of the tangent vector space T_x.

As a rule of parallel transport, we require that, when moving along C, the triad should not twist around $\hat{r}$. That is, if $\boldsymbol{\omega}$ is the angular velocity of the triad, then

$$\boldsymbol{\omega} \cdot \hat{r} = 0. \tag{2}$$

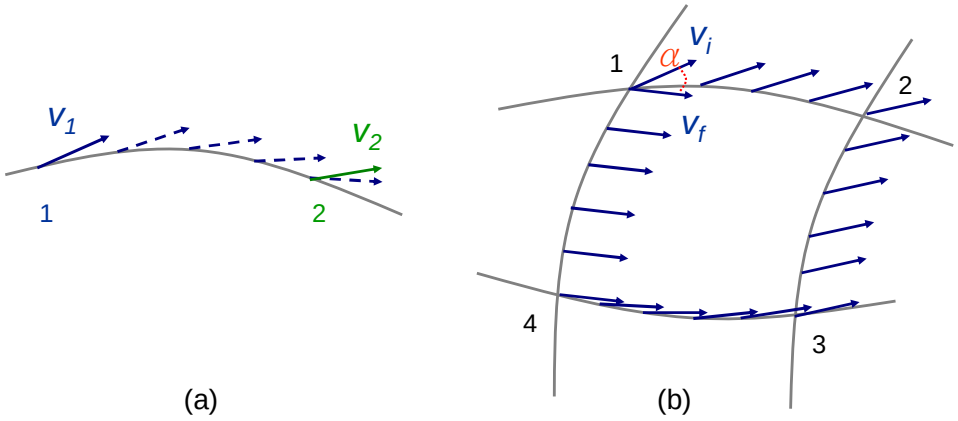


Fig. 1: *(a) Parallel transport of a vector from 1 to 2. It offers a way to compare* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *on a curved surface. (b) A vector is parallel transported around a closed path. When the surface is curved, the final vector would point to a different direction from the initial vector. The angle of difference* α *is called the anholonomy angle.*

Using the identity $\dot{\hat{e}}_1 = \boldsymbol{\omega} \times \hat{e}_1$ it follows from this requirement that $\dot{\hat{e}}_1 \cdot \hat{e}_2 = 0$:

$$\begin{aligned} \boldsymbol{\omega} \cdot \hat{r} &= \boldsymbol{\omega} \cdot \hat{e}_1 \times \hat{e}_2 \\ &= \boldsymbol{\omega} \times \hat{e}_1 \cdot \hat{e}_2 = \dot{\hat{e}}_1 \cdot \hat{e}_2 = 0. \end{aligned} \tag{3}$$

Likewise also the relation $\dot{\hat{e}}_2 \cdot \hat{e}_1 = 0$ is shown easily.
To make further analogy with the complex quantum phase in the next section, let us introduce the following complex vector,

$$\psi = \frac{1}{\sqrt{2}} \left(\hat{e}_1 + i\hat{e}_2 \right). \tag{4}$$

Then the parallel transport condition can be rephrased as,

$$\mathrm{Im} \left(\psi^* \cdot \dot{\psi} \right) = 0, \text{ or } i\psi^* \cdot \dot{\psi} = 0. \tag{5}$$

Notice that the real part of $\psi^* \cdot \dot{\psi}$ is always zero since $\hat{e}_1 \cdot \hat{e}_1$ and $\hat{e}_2 \cdot \hat{e}_2$ are time independent.
Instead of the moving triad, we could also erect a fixed triad, $(\hat{r}, \hat{u}, \hat{v})$, at each point of the surface and introduce

$$n = \frac{1}{\sqrt{2}} \left(\hat{u} + i\hat{v} \right). \tag{6}$$

Assuming these two triads differ by an angle $\alpha(x)$ (around the $\hat{r}$-axis), then $\psi(x) = n(x)e^{-i\alpha(x)}$. It follows that

$$\psi^* \cdot d\psi = n^* \cdot dn - id\alpha. \tag{7}$$

Because of the parallel transport condition in Eq. (5), one has $d\alpha = -in^* \cdot dn$. Finally, the twist angle accumulated by the moving triad after completing a closed loop C is,

$$\alpha(C) = -i \oint_C n^* \cdot \frac{dn}{dx} dx, \tag{8}$$

where we have changed the variable of integration to the coordinate on the surface. Therefore, the defect angle can be calculated conveniently using the fixed-triad basis.

With the help of the Stokes theorem, one can transform the line integral to a surface integral,

$$\alpha(C) = \int_S \frac{1}{i}\left(\frac{dn^*}{dx_1}\cdot\frac{dn}{dx_2} - \frac{dn^*}{dx_2}\cdot\frac{dn}{dx_1}\right)dx_1dx_2, \tag{9}$$

where S is the area enclosed by C. In the case of the sphere, one can choose (x_1, x_2) to be the spherical coordinates (θ, ϕ), and choose $\hat{u}$ and $\hat{v}$ to be the unit vectors $\hat{\theta}$ and $\hat{\phi}$ in spherical coordinates. That is, $\hat{u} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)$ and $\hat{v} = (-\sin\phi, \cos\phi, 0)$. It is not difficult to show that the integrand in Eq. (9) is $\sin\theta d\theta d\phi$. Therefore, $\alpha(C)$ is indeed the solid angle of the area S.

The integral in Eq. (9) over the whole sphere (the total curvature) is equal to its solid angle, 4π. In fact, any closed surface that has the same topology as a sphere would have the same total curvature $2\pi \times 2$. The value of 2 (Euler characteristic) can thus be regarded as a number characterizing the topology of sphere-like surfaces. In general, for a closed surface with g holes, the Euler characteristic is $2 - 2g$. For example, the total curvature of a donut ($g = 1$) is 0. This is the beautiful Gauss-Bonnet theorem in differential geometry.

2 Anholonomy in quantum mechanics

Similar to the parallel transported vector on a curved surface, the phase of a quantum state (not including the dynamical phase) may not return to its original value after a cyclic evolution in parameter space. This fact was first exposed clearly by Michael Berry [3] in his 1984 paper. In this section, we introduce the basic concept of the Berry phase, in later sections we will move on to examples of the Berry phase in condensed matter.

2.1 Introducing the Berry phase

Let us start from a time-*independent* system described by a Hamiltonian $H(\mathbf{r}, \mathbf{p})$. We denote the eigenstates by $|m\rangle$ and the eigenvalues by ϵ_m. For simplicity, the energy levels are assumed to be non-degenerate. An initial state $|\psi_0\rangle = \sum a_m|m\rangle$ evolves to a state $|\psi_t\rangle = \sum a_m e^{-i/\hbar\epsilon_m t}|m\rangle$ at time t. The probability of finding a particle in a particular level remains unchanged, even though each level acquires a different dynamical phase $e^{-i/\hbar\epsilon_m t}$. In particular, if one starts with an eigenstate of the Hamiltonian, $|\psi_0\rangle = |n\rangle$, with $a_m = \delta_{m,n}$, then the probability amplitude does not "leak" to other states.

Let us now consider a slightly more complicated system with two sets of dynamical variables $H(\mathbf{r}, \mathbf{p}; \mathbf{R}, \mathbf{P})$. The characteristic time scale of the upper-case set is assumed to be much longer than that of the lower-case set. For example, the system can be a diatomic molecule H_2^+. The electron and nuclei positions are represented by $\mathbf{r}$ and $\mathbf{R}$ respectively. Because of its larger mass, the nuclei move more slowly (roughly by a thousand times) compared to the electron. In the spirit of the Born-Oppenheimer approximation, one can first treat $\mathbf{R}$ as a time-dependent parameter, instead of a dynamical variable, and study the system at each "snapshot" of the evolution. The kinetic part of the slow variable is ignored for now.

Since the characteristic frequency of the nuclei is much smaller than the electron frequency, an electron initially in an electronic state $|n\rangle$ remains essentially in that state after time t,

$$|\psi_t\rangle = e^{i\gamma_n(\mathbf{R})}e^{-i/\hbar\int_0^t dt\epsilon_n(\mathbf{R}_t)}|n;\mathbf{R}\rangle. \tag{10}$$

Table 1: *Anholonomies in geometry and quantum state*

	geometry	quantum state
fixed basis	$n(x)$	$\|n;\mathbf{R}\rangle$
moving basis	$\psi(x)$	$\|\psi;\mathbf{R}\rangle$
parallel-transport condition	$i\psi^*\cdot\dot{\psi}=0$	$i\langle\psi\|\dot{\psi}\rangle=0$
anholonomy	anholonomy angle	Berry phase
classification of topology	Euler characteristic	Chern number

Apart from the dynamical phase, one is allowed to add an extra phase $e^{i\gamma_n(\mathbf{R})}$ for each snapshot state. Such a phase is usually removable by readjusting the phase of the basis $|n;\mathbf{R}\rangle$ [2]. In 1984, almost six decades after the birth of quantum mechanics, Berry [3] pointed out that this phase, like the vector in the previous section, may not return to its original value after a cyclic evolution. Therefore, it is not always removable.

To determine this phase, one substitutes Eq. (10) into the time-*dependent* Schrödinger equation. It is not difficult to get an equation for $\gamma_n(t)$,

$$\dot{\gamma}_n(t)=i\langle n|\dot{n}\rangle. \tag{11}$$

Therefore, after a cyclic evolution, one has

$$\gamma_n(C)=i\oint_C\langle n|\frac{\partial n}{\partial\mathbf{R}}\rangle\cdot d\mathbf{R}=\oint_C\mathbf{A}\cdot d\mathbf{R}, \tag{12}$$

where C is a closed path in the $\mathbf{R}$-space. The integrand $\mathbf{A}(\mathbf{R})\equiv i\langle n|\frac{\partial n}{\partial\mathbf{R}}\rangle$ is often called the *Berry connection*.

If the parameter space is two dimensional, then one can use Stokes' theorem to transform the line integral to a surface integral,

$$\gamma_n(C)=i\int_S\langle\frac{\partial n}{\partial\mathbf{R}}|\times|\frac{\partial n}{\partial\mathbf{R}}\rangle\cdot d^2\mathbf{R}=\int_S\mathbf{F}\cdot d^2\mathbf{R}. \tag{13}$$

The integrand $\mathbf{F}(\mathbf{R})\equiv\nabla_{\mathbf{R}}\times\mathbf{A}(\mathbf{R})$ is usually called the *Berry curvature*. For parameter spaces with higher dimensions, such a transformation can still be done using the language of the differential form.

By now, the analogy between Eqs. (8,9) and Eqs. (12,13) should be clear. Notice that $|n\rangle$ is a normalized basis with $\langle n|n\rangle=1$. Therefore, $\langle n|\dot{n}\rangle$ should be purely imaginary and $i\langle n|\dot{n}\rangle$ is a real number. The basis state $|n\rangle$ plays the role of the fixed triad n in the previous subsection. Both are single-valued. On the other hand, the parallel transported state $|\psi\rangle$ and the moving triad ψ are not single-valued.

A point-by-point re-assignment of the phase of the basis state, $|n;\mathbf{R}\rangle'=e^{ig(\mathbf{R})}|n;\mathbf{R}\rangle$, changes the Berry connection,

$$\mathbf{A}'=\mathbf{A}-\frac{\partial g}{\partial\mathbf{R}}. \tag{14}$$

However, the Berry curvature $\mathbf{F}$ and the Berry phase are not changed. This is similar to the gauge transformation in electromagnetism: one can choose different gauges for the potentials, but the fields are not changed. Such an analogy will be explored further in the next subsection.

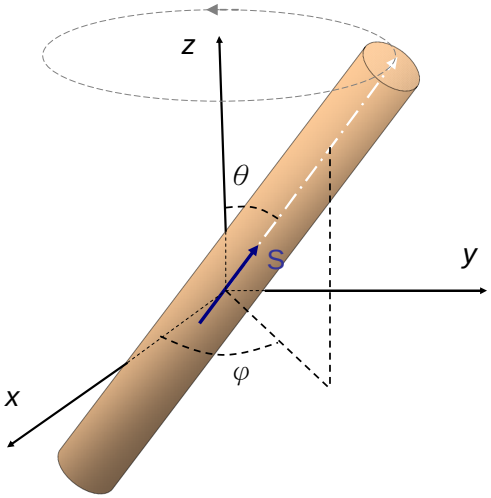


Fig. 2: *A long solenoid hinged at the origin is slowly rotating around the z-axis. At each instant, the spin at the origin aligns with the uniform magnetic field inside the solenoid.*

A short note: It is possible to rephrase the anholonomy of the quantum state using the mathematical theory of fiber bundles, which deals with geometrical spaces that can locally be decomposed into a product space (the "fiber" space times the "base" space), but globally show nontrivial topology. The Möbius band is the simplest example of such a geometric object: Locally it is a product of two one-dimensional spaces but globally it is not (because of the twisting). In our case, the fiber is the space of the quantum phase $\gamma(\mathbf{R})$ and the base is the space of $\mathbf{R}$. The concept of the parallel transport, the connection, and the curvature all can be rephrased rigorously in the language of fiber bundles [4]. Furthermore, there is also a topological number (similar to the Euler characteristic) for the fiber bundle, which is called the Chern number.
The analogy between geometric anholonomy and quantum anholonomy is summarized in Table 1.

2.2 A rotating solenoid

To illustrate the concept of the Berry phase, we study a simple system with both slow and fast degrees of freedom. Following M. Stone [5], we consider a rotating (long) solenoid with an electron spin at its center. The solenoid is tilted with a fixed angle θ and is slowly gyrating around the z-axis (see Fig. 2). Therefore, the electron spin feels a uniform magnetic field that changes direction gradually. This example is a slight generalization of the spin-in-magnetic-field example given by Berry in his 1984 paper. The Hamiltonian of this spin-in-solenoid system is,

$$H = \frac{L^2}{2I} + \mu_B \boldsymbol{\sigma} \cdot \mathbf{B}, \tag{15}$$

where L and I are the angular momentum and the moment of inertia of the solenoid, respectively, and the Bohr magneton is $\mu_B = e\hbar/2mc$.
The magnetic field $\mathbf{B}$ along the direction of the solenoid is our time-dependent parameter $\mathbf{R}$. In the quasi-static limit, the rotation energy of the solenoid is neglected. When the solenoid rotates to the angle (θ, ϕ), the spin eigenstates are

$$|+; \hat{B}\rangle = \begin{pmatrix} \cos\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} \end{pmatrix}, |-; \hat{B}\rangle = \begin{pmatrix} -e^{-i\phi}\sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \end{pmatrix}. \tag{16}$$

Table 2: *Analogy between electromagnetism and quantum anholonomy*

Electromagnetism	quantum anholonomy
vector potential $\mathbf{A}(\mathbf{r})$	Berry connection $\mathbf{A}(\mathbf{R})$
magnetic field $\mathbf{B}(\mathbf{r})$	Berry curvature $\mathbf{F}(\mathbf{R})$
magnetic monopole	point degeneracy
magnetic flux $\Phi(C)$	Berry phase $\gamma(C)$

These states can be obtained, for example, from the spin-up (-down) states $|\pm\rangle$ by a rotation $e^{-i\boldsymbol{\sigma}\cdot\hat{\theta}(\theta/2)}$, in which the rotation axis $\hat{\theta} = (-\sin\phi, \cos\phi, 0)$ is perpendicular to both $\hat{z}$ and $\hat{B}$. Using the definitions of the Berry connection and the Berry curvature in Eqs. (12) and (13), one obtains

$$\mathbf{A}_{\pm} = \mp\frac{1}{2}\frac{1-\cos\theta}{B\sin\theta}\hat{\phi} \tag{17}$$

$$\mathbf{F}_{\pm} = \mp\frac{1}{2}\frac{\hat{B}}{B^2}. \tag{18}$$

They have the same mathematical structure as the vector potential and the magnetic field of a magnetic monopole. The location of the "monopole" is at the origin of the parameter space, where a point degeneracy occurs. The strength of the monopole $(1/2)$ equals the value of the spin (this is true for larger spins also). That is why the Berry connection and the Berry curvature are sometimes called the Berry potential and the Berry field. In this picture, the Berry phase is equal to the flux of the Berry field passing through a loop C in parameter space. It is easy to see that,

$$\gamma_{\pm}(C) = \mp\frac{1}{2}\Omega(C), \tag{19}$$

where $\Omega(C)$ is the solid angle subtended by loop C with respect to the origin. The similarity between the theory of Berry phase and electromagnetism is summarized in Table 2.

The Berry phase of the fast motion is only half of the story. When the quantum state of the fast variable acquires a Berry phase, there will be an interesting "back action" to the slow motion. For example, for the rotating solenoid, the wave function of the whole system can be expanded as

$$|\Psi\rangle = \sum_{n=\pm}\psi_n(\mathbf{R})|n;\mathbf{R}\rangle, \tag{20}$$

in which $\psi_n(\mathbf{R})$ describes the slow quantum state. From the Schrödinger equation, $H|\Psi\rangle = E|\Psi\rangle$, one can show that,

$$\left[\frac{\hbar^2}{2I\sin^2\theta}\left(\frac{1}{i}\frac{d}{d\phi} - \mathbf{A}_n\right)^2 + \epsilon_n\right]\psi_n = E\psi_n, \tag{21}$$

where ϵ_n is the eigen-energy for the fast degree of freedom, and $\mathbf{A}_n \equiv i\langle n;\mathbf{R}|\frac{d}{d\phi}|n;\mathbf{R}\rangle$. The off-diagonal coupling between $|+\rangle$ and $|-\rangle$ has been ignored. Therefore, the effective Hamiltonian for the slow variable acquires a Berry potential $\mathbf{A}_n(\mathbf{R})$. Such a potential could shift the spectrum and results in a force (proportional to the Berry curvature) upon the slow motion, much like the effect of vector potential $\mathbf{A}(\mathbf{r})$ and magnetic field on a charged particle.

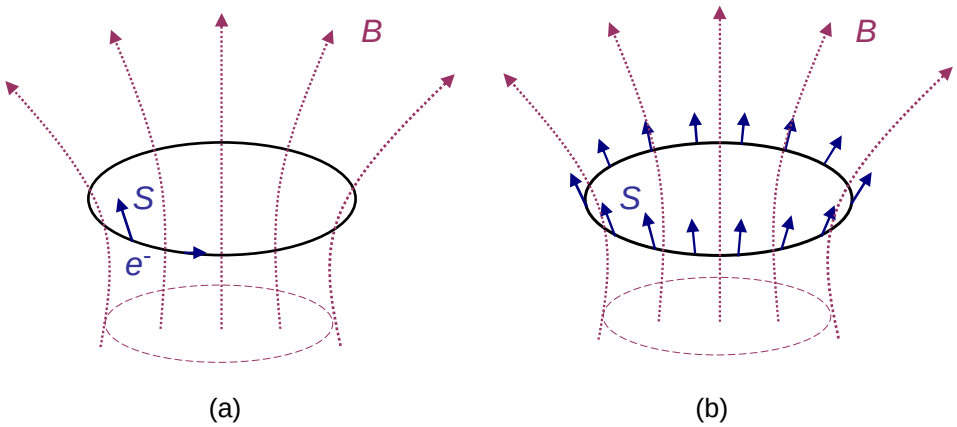


Fig. 3: *(a) A metal ring in a non-uniform magnetic field. The spin of the electron that is circling the ring would align with the magnetic field and trace out a solid angle in its own reference frame. (b) A ferromagnetic ring in a non-uniform magnetic field. The spins on the ring are bent outward because of the magnetic field.*

3 Berry phase and spin systems

A natural place to find the Berry phase is in spin systems. Numerous researches related to this subject can be found in the literature [6]. Here we only mention two examples, one is related to the persistent spin current in a mesoscopic ring, the other relates to quantum tunneling in a magnetic cluster.

3.1 Persistent spin current

We know that an electron moving in a periodic system feels no resistance. The electric resistance is a result of incoherent scatterings from impurities and phonons. If one fabricates a clean one-dimensional wire, wraps it around to form a ring, and lowers the temperature to reduce the phonon scattering, then the electron inside feels like living in a periodic lattice without electric resistance.

For such a design to work, two ingredients are essential: First, the electron has to remain phase coherent (at least partially) after one revolution. Therefore, a mesoscopic ring at very low temperature is usually required. Second, to have a traveling wave, there has to be a phase advance (or lag) after one revolution. This can be achieved by threading a magnetic flux ϕ through the ring, so that the electron acquires an Aharonov-Bohm (AB) phase $(e/\hbar)\phi = 2\pi(\phi/\phi_0)$ after one cycle, where ϕ_0 is the flux quantum h/e. When this does happen, it is possible to observe the resulting *persistent charge current* in the mesoscopic ring.

Soon after this fascinating phenomenon was observed [7], it was proposed that, in addition to the AB phase, a spinful electron can (with proper design) acquire a Berry phase after one cycle, and this can result in a persistent *spin* current [8]. The design is as follows: Instead of a uniform magnetic field, a textured magnetic field is used, so that during one revolution, the electron spin follows the direction of the field and traces out a non-zero solid angle Ω (see Fig. 3a). According to Eq. (19), this gives rise to a spin-dependent Berry phase $\gamma_\sigma(C) = -(\sigma/2)\Omega$, where $\sigma = \pm$. After combining this with the (spin-independent) AB phase, spin-up and spin-down electrons have different phase shifts, generating different amounts of persistent particle current I_+, I_-. Therefore, a spin current defined as $I_s = (\hbar/2)(I_+ - I_-)$ is not zero.

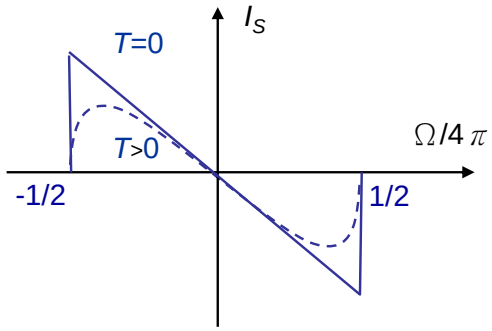


Fig. 4: *Persistent spin current as a function of the solid angle. At non-zero temperature, the sharp edges of the sawtooth become smooth.*

To illustrate the physics just mentioned, consider a ring that allows only angular motion. Before applying the magnetic flux, the electron with wave vector k picks up a phase kL from circling the ring, where $L = 2\pi R$ and R is the radius of the ring. Because of the periodic boundary condition, one has $kL = 2\pi n$ ($n \in Z$). After adding the AB phase and the Berry phase, it becomes $kL = 2\pi n + 2\pi(\phi/\phi_0) - \sigma(\Omega/2)$. Therefore, the energy of an electron in the n-th mode is

$$\epsilon_{n\sigma} = \frac{\hbar^2 k^2}{2m} + \mu_B B\sigma = \frac{\hbar^2}{2mR^2}\left(n + \frac{\phi}{\phi_0} - \sigma\frac{\phi_\Omega}{\phi_0}\right)^2 + \mu_B B\sigma, \tag{22}$$

where $\phi_\Omega/\phi_0 \equiv \Omega/4\pi$.
The spin current can be calculated from

$$I_s = \frac{1}{L}\sum_{n,\sigma}\left(\frac{\hbar}{2}\sigma\right)\frac{\partial\epsilon_{n\sigma}}{\hbar\partial k}P_{n\sigma}, \tag{23}$$

where $P_{n\sigma} = \exp(-\epsilon_{n\sigma}/k_BT)/Z$ is the probability of the electron in the (n,σ)-state, and $Z = \sum_{n,\sigma} e^{-\epsilon_{n\sigma}/k_BT}$. For a particular k and ϕ, the current can also be written as

$$I_s = -\sum_{n,\sigma}\frac{\partial\epsilon_{n\sigma}}{\partial\Omega}P_{n\sigma}. \tag{24}$$

To get a rough understanding, we consider the simplest case, where the $n = 1$ mode is populated with equal numbers of spin-up and -down electrons (if the Zeeman splitting is negligible). The higher modes are all empty at low enough temperature. In this case, the spin current $I_s = -(\hbar^2/4\pi mR^2)(\Omega/4\pi)$ is proportional to the solid angle of the textured magnetic field (see Fig. 4). At higher temperature, the sawtooth curve will become smooth.
The mesoscopic ring considered above is a metal ring with moving electrons that carry the spins with them. A different type of spin current has also been proposed in a ferromagnetic ring with no moving charges [9]. Again the ring is subject to a textured magnetic field, such that when one moves round the ring, one sees a changing spin vector that traces out a solid angle Ω (see Fig. 3b). As a result, the spin wave picks up a Berry phase when traveling around the ring, resulting in a persistent spin current. So far neither type of persistent spin current has been observed experimentally.

3.2 Magnetic cluster

Berry phase plays a dramatic role in the quantum tunneling of nano-sized magnetic clusters. The tunneling between two degenerate spin states of the cluster depends on whether the total spin of the particle is an integer or a half-integer. In the latter case, the tunneling is completely suppressed because different tunneling paths interfere destructively as a result of the Berry phase [10].

Consider a single-domain ferromagnetic particle without itinerant spin. Its total spin J can be of order ten or larger, as long as tunneling is still possible. Assume that the particle lives in an anisotropic environment with the Hamiltonian,

$$H = -k_1 \frac{J_z^2}{J^2} + k_2 \left(\frac{J_x^2}{J^2} - \frac{J_y^2}{J^2} \right), \ (k_1 > k_2). \tag{25}$$

That is, the easy axis is along the z-axis and the easy plane is the yz-plane. The cluster is in the ground state when the spin points to the north pole or to the south pole of the Bloch sphere. Even though these two degenerate states are separated by a barrier, the particle can switch its direction of spin via quantum tunneling.

To study the Berry phase effect on the tunneling probability, the best tool is the method of path integrals. In the following, we give a brief sketch of its formulation.

The fully polarized spin state $|\hat{n}, J\rangle$ along a direction $\hat{n}$ with spherical angles (θ, ϕ) can be written as,

$$\begin{aligned} |\hat{n}, J\rangle &= |\hat{n}, +\rangle \otimes |\hat{n}, +\rangle \cdots \otimes |\hat{n}, +\rangle \\ &= \prod_{l=1}^{2J} e^{-i\frac{\theta}{2}\boldsymbol{\sigma}_l \cdot \hat{\theta}} |\hat{z}, +\rangle_l, \end{aligned} \tag{26}$$

where $|\hat{n}, +\rangle$ is the spin-1/2 "up" state along the $\hat{n}$-axis and $\hat{\theta}$ is a unit vector along the $\hat{z} \times \hat{n}$ direction. Such a so-called *spin coherent state* can be used to "resolve" the identity operator [11],

$$I = \frac{2J+1}{4\pi} \int d\Omega |\hat{n}\rangle\langle\hat{n}|, \tag{27}$$

where $|\hat{n}\rangle$ is an abbreviation of $|\hat{n}, J\rangle$.

In order to calculate the transition probability amplitude $\langle \hat{n}_f | \exp[-(i/\hbar)HT] | \hat{n}_i \rangle$, one first divides the time evolution into steps, $\exp(-i/\hbar HT) = [\exp(-i/\hbar H dt)]^N$, $dt = T/N$, then insert the resolution of identity in Eq. (27) between neighboring steps. The transition amplitude then becomes a product of factors with the following form,

$$\begin{aligned} \langle \hat{n}(t+dt) | e^{-\frac{i}{\hbar}Hdt} | \hat{n}(t) \rangle &\simeq \langle \hat{n}(t+dt) | \hat{n}(t) \rangle - \frac{i}{\hbar} \langle \hat{n}(t+dt) | H(\mathbf{J}) | \hat{n}(t) \rangle dt \\ &\simeq 1 - \langle \hat{n} | \dot{\hat{n}} \rangle dt - \frac{i}{\hbar} H(J\hat{n}) dt. \end{aligned} \tag{28}$$

In the final step, we have replaced the quantum Hamiltonian by a classical Hamiltonian. That is, $\langle H(\mathbf{J}) \rangle = H(\langle \mathbf{J} \rangle)$. This holds exactly if the Hamiltonian is linear in $\mathbf{J}$, but is only an approximation in general. The correction due to the non-commutativity of the spin operator is roughly of the fraction $1/J$ and can be ignored for large spins.

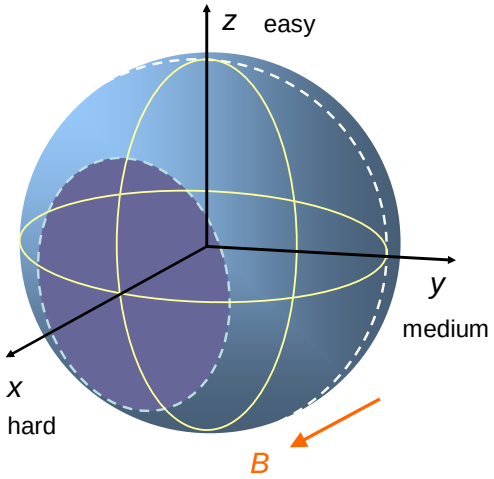


Fig. 5: *According to the Hamiltonian in Eq. (25), the z-axis and the x-axis are the easy axis and the hard axis, respectively. There are two (degenerate) ground states at the north pole and the south pole of the Bloch sphere. Tunneling from one ground state to the other follows the dashed line on the $y-z$ plane. Applying a magnetic field along the x-direction moves the locations of the ground states and shrinks the tunneling path to a smaller loop.*

Finally, by summing over paths in the $\hat{n}$-space, one has

$$\langle \hat{n}_f | e^{-\frac{i}{\hbar}HT} | \hat{n}_i \rangle = \int [\mathcal{D}\hat{n}] \exp\left\{ \frac{i}{\hbar} \int_{t_i}^{t_f} \left[i\hbar \langle \hat{n} | \dot{\hat{n}} \rangle - H(J\hat{n}) \right] dt \right\}. \tag{29}$$

Notice that the first integral in the exponent generates a Berry phase for a path (see Eq. (12)). In the semiclassical regime, the functional integral in Eq. (29) is dominated by the classical path $\hat{n}_c$ with least action, which is determined from the dynamical equation of $\hat{n}$ (see below). During tunneling, the paths under the barrier are classically inaccessible and $\hat{n}$ becomes an imaginary vector. It is customary to sacrifice the reality of time t to keep $\hat{n}$ real. The good news is that the final result does not depend on which imaginary world you choose to live in.
Define $\tau = it$, then the transition amplitude dominated by the classical action is,

$$\langle \hat{n}_f | e^{-\frac{i}{\hbar}HT} | \hat{n}_i \rangle \propto e^{i\int_i^f \mathbf{A}\cdot d\hat{n}_c} e^{-1/\hbar \int_i^f H(J\hat{n}_c) d\tau}, \tag{30}$$

where $\mathbf{A} = i\langle \hat{n} | \nabla \hat{n} \rangle$ is the Berry potential. The integral of the Berry potential is gauge dependent if the path is open. It is well defined for a closed loop, such as the classical path on the yz-plane in Fig. 5. The Berry phase for such a loop is $2\pi J$ since it encloses an area with solid angle 2π (Cf. Eq. (19)). This is also the phase difference between the two classical paths from the north pole to the south pole. Therefore,

$$\langle -\hat{z} | e^{-\frac{i}{\hbar}HT} | \hat{z} \rangle \propto \cos(\pi J) e^{-1/\hbar \int_i^f H(J\hat{n}_c) d\tau}. \tag{31}$$

When J is a half integer, the transition process is completely suppressed because of the Berry phase. The conclusion remains valid if one considers classical paths with higher winding numbers [10].
As a reference, we also write down the equation of motion for $\hat{n}_c$ that is determined from the classical action in Eq. (30),

$$J\frac{d\hat{n}}{dt} = \hat{n} \times \frac{\partial H(J\hat{n})}{\partial \hat{n}}. \tag{32}$$

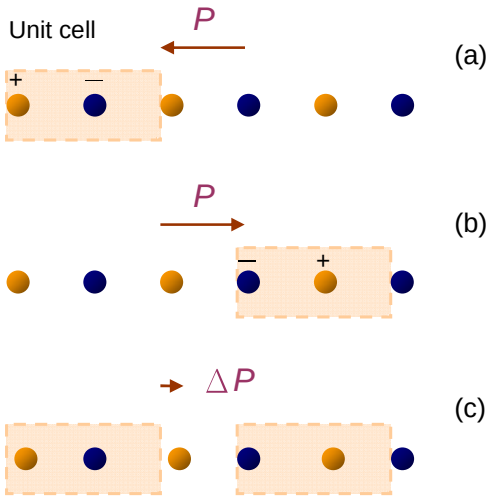


Fig. 6: *An one-dimensional solid with infinite length. Different choices of the unit cell give different electric polarization vectors ((a), (b)). On the other hand, the change of polarization does not depend on the choice of the unit cell (c).*

This is the Bloch equation for spin precession, in which $\partial H/\partial \hat{n}$ plays the role of an effective magnetic field.
One comment is in order: One can apply a magnetic field along the x-axis that shifts the energy minima along that direction and shrinks the classical loop (see Fig. 5). In an increasingly stronger field, the size of the loop C eventually would shrink to zero. That is, the Berry phase γ_C would decrease from the maximum value of $2\pi J$ to zero. During the process, one expects to encounter the no-tunneling situation several times whenever $\gamma_C/2\pi$ hits a half-integer. Such a dramatic Berry phase effect has been observed [12].

4 Berry phase and Bloch state

In the second half of this article, we focus on the Berry phase in *periodic* solids. It has been playing an ever more important role in recent years due to several discoveries and "re-discoveries", in which the Berry phase either plays a crucial role or offers a fresh perspective.

4.1 Electric polarization

It may come as a surprise to some people that the electric polarization $\mathbf{P}$ of an infinite periodic solid (or a solid with periodic boundary conditions) is generically not well defined. The reason is that, in a periodic solid, the electric polarization depends on your choice of the unit cell (see Fig. 6a,b). The theory of electric polarization in conventional textbooks applies only to solids consisting of well localized charges, such as ionic or molecular solids (Clausius-Mossotti theory). It fails, for example, in a covalent solid with bond charges such that no natural unit cell can be defined.
A crucial observation made by R. Resta [13] is that, even though the value of $\mathbf{P}$ may be ambiguous, its change is well defined (see Fig. 6c). It was later pointed out by King-Smith and Vanderbilt [14] that $\Delta\mathbf{P}$ has a deep connection with the Berry phase of the electronic states. The outline of their theory below is based on one-particle states. However, the same scheme

applies to real solids with electronic interactions, as long as one replaces the one-particle states by the Kohn-Sham orbitals in the density functional theory.
We will use λ to label the degree of ion displacement. It varies from 0 to 1 as the ions shift adiabatically from an initial state to a final state. The difference of polarizations between these two states is given by $\int_0^1 d\lambda d\mathbf{P}/d\lambda$, where

$$\mathbf{P}(\lambda) = \frac{q}{V}\sum_i \langle\phi_i|\mathbf{r}|\phi_i\rangle. \tag{33}$$

The summation runs over filled Bloch states ϕ_i (with λ-dependence) and V is the volume of the material. For an infinite crystal, the expectation value of $\mathbf{r}$ is ill-defined. Therefore, we consider a finite system at first, and let $V \to \infty$ when the mathematical expression becomes well-defined.
The Bloch states are solutions of the Schrödinger equation,

$$H_\lambda|\phi_i\rangle = \left(\frac{p^2}{2m} + V_\lambda\right)|\phi_i\rangle = \epsilon_i|\phi_i\rangle, \tag{34}$$

where V_λ is the crystal potential. From Eq. (34), it is not difficult to show that, for $j \neq i$, one has

$$(\epsilon_i - \epsilon_j)\,\langle\phi_j|\frac{\partial\phi_i}{\partial\lambda}\rangle = \langle\phi_j|\frac{\partial V_\lambda}{\partial\lambda}|\phi_i\rangle. \tag{35}$$

Therefore,

$$\frac{d\mathbf{P}}{d\lambda} = \frac{q}{V}\sum_i\sum_{j\neq i}\left[\langle\phi_i|\mathbf{r}|\phi_j\rangle\frac{\langle\phi_j|V'_\lambda|\phi_i\rangle}{\epsilon_i - \epsilon_j} + H.c.\right]. \tag{36}$$

There is a standard procedure to convert the matrix elements of $\mathbf{r}$ to those of $\mathbf{p}$: Start with the commutation relation, $[\mathbf{r}, H_\lambda] = i\hbar\mathbf{p}/m$, and sandwich it between the i-state and the j-state (again $j \neq i$), we can get an useful identity,

$$\langle\phi_i|\mathbf{r}|\phi_j\rangle = \frac{i\hbar}{m}\frac{\langle\phi_i|\mathbf{p}|\phi_j\rangle}{\epsilon_j - \epsilon_i}. \tag{37}$$

With the help of this identity, Eq. (36) becomes the following expression derived by Resta [13],

$$\frac{d\mathbf{P}}{d\lambda} = \frac{q\hbar}{imV}\sum_i\sum_{j\neq i}\left[\frac{\langle\phi_i|\mathbf{p}|\phi_j\rangle\langle\phi_j|V'_\lambda|\phi_i\rangle}{(\epsilon_i - \epsilon_j)^2} - H.c.\right]. \tag{38}$$

Now all of the matrix elements are well-defined and the volume V can be made infinite. After integrating with respect to λ, the resulting $\Delta\mathbf{P}$ is free of ambiguity, even for an infinite covalent solid.
For Bloch states, the subscripts are $i = (m, \mathbf{k})$ and $j = (n, \mathbf{k})$, where m, n are the band indices and $\mathbf{k}$ is the Bloch momentum defined in the first Brillouin zone. Eq. (38) can be transformed to a very elegant form, revealing its connection with the Berry curvature [14]. One first defines a $\mathbf{k}$-dependent Hamiltonian, $\tilde{H} = e^{-i\mathbf{k}\cdot\mathbf{r}}He^{i\mathbf{k}\cdot\mathbf{r}}$. It is the Hamiltonian of the cell-periodic function $u_{n\mathbf{k}}$. That is, $\tilde{H}|u_{n\mathbf{k}}\rangle = \epsilon_{n\mathbf{k}}|u_{n\mathbf{k}}\rangle$, where $\phi_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}$. It is then straightforward to show that,

$$\langle\phi_{m\mathbf{k}}|\mathbf{p}|\phi_{n\mathbf{k}}\rangle = \frac{m}{\hbar}\langle u_{m\mathbf{k}}|\left[\frac{\partial}{\partial\mathbf{k}}, \tilde{H}\right]|u_{n\mathbf{k}}\rangle = \frac{m}{\hbar}(\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}})\langle u_{m\mathbf{k}}|\frac{\partial u_{n\mathbf{k}}}{\partial\mathbf{k}}\rangle. \tag{39}$$

With the help of this equation and another one very similar to Eq. (35) (just replace the ϕ_i's by the u_i's), we finally get ($\alpha = x, y, z$)

$$\begin{aligned} \frac{dP_\alpha}{d\lambda} &= -\frac{iq}{V}\sum_{n\mathbf{k}}\left(\left\langle\frac{\partial u_{n\mathbf{k}}}{\partial k_\alpha}|\frac{\partial u_{n\mathbf{k}}}{\partial\lambda}\right\rangle - \left\langle\frac{\partial u_{n\mathbf{k}}}{\partial\lambda}|\frac{\partial u_{n\mathbf{k}}}{\partial k_\alpha}\right\rangle\right) \\ &= -\frac{q}{V}\sum_{n\mathbf{k}}\Omega^n_{k_\alpha\lambda}(\mathbf{k}), \end{aligned} \tag{40}$$

where $\Omega^n_{k_\alpha\lambda} \equiv i\left(\langle\frac{\partial u}{\partial k_\alpha}|\frac{\partial u}{\partial\lambda}\rangle - c.c.\right)$ is the Berry curvature for the n-th band in the parameter space of k_α and λ (Cf. Eq. (13)).

Let us take a one-dimensional system as an example. Assuming the lattice constant is a. Then the difference of polarization is ($q = -e$),

$$\Delta P = \frac{e}{2\pi}\sum_n\int_0^{2\pi/a}dk\int_0^1 d\lambda\Omega^n_{k\lambda}. \tag{41}$$

The area of integration is a rectangle with lengths 1 and $2\pi/a$ on each side. The area integral can be converted to a line integral around the boundary of the rectangle, which gives the Berry phase γ_n of such a loop. Therefore,

$$\Delta P = e\sum_n\frac{\gamma_n}{2\pi}. \tag{42}$$

In the special case where the final state of the deformation V_1 is the same as the initial state V_0, the Berry phase γ_n can only be integer multiples of 2π [14]. Therefore, the polarization P for a crystal state is uncertain by an integer charge Q.

One the other hand, this integer charge Q does carry a physical meaning when it is the difference ΔP between two controlled states. For example, when the lattice potential is shifted by one lattice constant to the right, this Q is equivalent to the total charge being transported. Based on such a principle, it is possible to design a quantum charge pump using a time-dependent potential [15].

4.2 Quantum Hall effect

The quantum Hall effect (QHE) has been discovered by K. von Klitzing *et al.* [16] in a two-dimensional electron gas (2DEG) at low temperature and strong magnetic field. Under such conditions, the Hall conductivity σ_H develops plateaus in the $\sigma_H(B)$ plot. For the integer QHE, these plateaus always locate at integer multiples of e^2/h to great precision, irrespective of the samples being used. Such a behavior is reminiscent of macroscopic quantum phenomena, such as the flux quantization in a superconductor ring.

To explain the integer QHE, Laughlin wraps the sheet of the 2DEG to a cylinder to simulate the superconductor ring, and studies the response of the current with respect to a (fictitious) magnetic flux through the cylinder (see Fig. 7). He found that, as the flux increases by one flux quantum h/e, integer charges $Q = ne$ are transported from one edge of the cylinder to the other [17]. This charge transport in the transverse direction gives the Hall current, and the integer n can be identified with the integer of the Hall conductance ne^2/h [18].

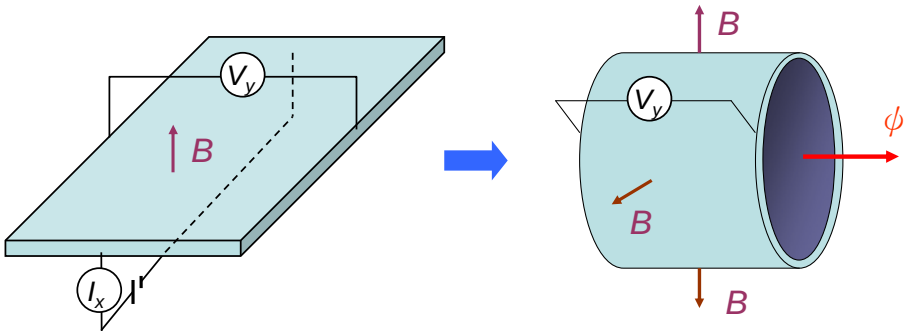


Fig. 7: *In Laughlin's argument, the 2DEG is on the surface of a cylinder. The real magnetic field B now points radially outward. In addition, there is a fictitious flux threading through the cylinder. When the fictitious flux changes by one flux quantum, integer number of electrons are be transported from one edge of the cylinder to the other.*

Soon afterwards, Thouless *et al.* (TKNdN) [19] found that the Hall conductivity is closely related to the Berry curvature (not yet discovered by Berry at that time) of the Bloch state. We now briefly review the TKNdN theory.

Consider a 2DEG subject to a perpendicular magnetic field and a weak in-plane electric field. In order not to break the periodicity of the scalar potential, we choose a time-dependent gauge for the electric field. That is, $\mathbf{E} = -\partial \mathbf{A}_E/\partial t$, $\mathbf{A}_E = -\mathbf{E}t$. The Hamiltonian is,

$$H = \frac{(\boldsymbol{\pi} - e\mathbf{E}t)^2}{2m} + V_L(\mathbf{r}), \tag{43}$$

where $\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}_0$ has included the vector potential of the magnetic field, and V_L is the lattice potential. Similar to the formulation the in previous subsection, it is convenient to use the k-dependent Hamiltonian $\tilde{H}$ and the cell-periodic function $u_{n\mathbf{k}}$ in our discussion. They are related by $\tilde{H}|u_{n\mathbf{k}}\rangle = E_{n\mathbf{k}}|u_{n\mathbf{k}}\rangle$.

We will assume that the system can be solved with known eigenvalues and eigenstates, $\tilde{H}_0|u^{(0)}_{n\mathbf{k}}\rangle = E^{(0)}_{n\mathbf{k}}|u^{(0)}_{n\mathbf{k}}\rangle$ in the absence of an external electric field [20]. The electric field is then treated as a perturbation. To the first-order perturbation, one has

$$|u_{n\mathbf{k}(t)}\rangle = |n\rangle - i\hbar \sum_{n' \neq n} \frac{|n'\rangle\langle n'|\frac{\partial}{\partial t}|n\rangle}{\epsilon_n - \epsilon_{n'}}, \tag{44}$$

where $\mathbf{k}(t) = \mathbf{k}_0 - e\mathbf{E}t/\hbar$, and $|n\rangle$ and ϵ_n are abbreviations of $|u^{(0)}_{n\mathbf{k}(t)}\rangle$ and $E^{(0)}_{n\mathbf{k}(t)}$.

The velocity of a particle in the n-th band is given by $\mathbf{v}_n(\mathbf{k}) = \langle u_{n\mathbf{k}}|\partial\tilde{H}/\hbar\partial\mathbf{k}|u_{n\mathbf{k}}\rangle$. After substituting the states in Eq. (44), we find

$$\mathbf{v}_n(\mathbf{k}) = \frac{\partial \epsilon_n}{\hbar \partial \mathbf{k}} - i \sum_{n' \neq n} \left(\frac{\langle n|\frac{\partial \tilde{H}}{\partial \mathbf{k}}|n'\rangle\langle n'|\frac{\partial n}{\partial t}\rangle}{\epsilon_n - \epsilon_{n'}} - c.c. \right). \tag{45}$$

The first term is the group velocity in the absence of the electric perturbation. With the help of an equation similar to Eq. (39),

$$\langle n|\frac{\partial \tilde{H}}{\partial \mathbf{k}}|n'\rangle = (\epsilon_n - \epsilon_{n'})\,\langle \frac{\partial n}{\partial \mathbf{k}}|n'\rangle, \tag{46}$$

one finally gets a neat expression,

$$\mathbf{v}_n(\mathbf{k}) = \frac{\partial \epsilon_n}{\hbar \partial \mathbf{k}} - i\left(\left\langle \frac{\partial n}{\partial \mathbf{k}} | \frac{\partial n}{\partial t} \right\rangle - \left\langle \frac{\partial n}{\partial t} | \frac{\partial n}{\partial \mathbf{k}} \right\rangle\right). \tag{47}$$

By a change of variable, the second term becomes $\mathbf{\Omega}_n \times \dot{\mathbf{k}} = -(e/\hbar)\mathbf{\Omega}_n \times \mathbf{E}$, where $\Omega_{n\alpha} = i\epsilon_{\alpha\beta\gamma}\langle \frac{\partial n}{\partial k_\beta} | \frac{\partial n}{\partial k_\gamma} \rangle$ is the Berry curvature in momentum space.
For a 2DEG, $\mathbf{\Omega}_n = \Omega_n \hat{z}$. All states below the Fermi energy contribute to the current density,

$$\mathbf{j} = \frac{1}{V}\sum_{n\mathbf{k}} -e\mathbf{v}_n(\mathbf{k}) = \frac{e^2}{\hbar}\sum_n \int \frac{d^2\mathbf{k}}{(2\pi)^2}\mathbf{\Omega}_n(\mathbf{k}) \times \mathbf{E}. \tag{48}$$

Notice that the first term in Eq. (47) does not contribute to the current. From Eq. (48), it is clear that the Hall conductivity is given by,

$$\sigma_{yx} = \frac{e^2}{h}\sum_n \frac{1}{2\pi}\int d^2\mathbf{k}\Omega_n(\mathbf{k}). \tag{49}$$

Thouless *et al.* have shown that the integral of the Berry curvature over the whole BZ divided by 2π must be an integer c_n. Such an integer (the Chern number mentioned in Sec. 2.2) characterizes the topological property of the fiber bundle space, in which the base space is the two-dimensional BZ, and the fiber is the phase of the Bloch state (see the discussion near the end of Sec. 2.1). Therefore, the Hall conductivity of a filled band is always an integer multiple of e^2/h. Such a topological property is the reason why the QHE is so robust against disorders and sample varieties. Even though the discussion here is based on single-particle Bloch states, the conclusion remains valid for many-body states [21].
Some comments are in order. First, the formulas behind the change of electric polarization $\Delta\mathbf{P}$ in Sec. 4.1 and those of the quantum Hall conductivity here look very similar. Both are based on the linear response theory. In fact, the analogy can be carried further if $\Delta\mathbf{P}$ is considered as the time integral of a polarization current $\mathbf{j}_P = \partial\mathbf{P}/\partial t$. The latter, similar to the quantum Hall current in Eq. (48), can be related to the Berry curvature directly.
Second, if a solid is invariant under space inversion, then the cell-periodic state has the symmetry,

$$u_{n-\mathbf{k}}(-\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}). \tag{50}$$

On the other hand, if the system has time-reversal symmetry, then

$$u^*_{n-\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}). \tag{51}$$

As a result, if both symmetries exist, then one can show that the Berry potential $\mathbf{A}_n = i\langle n|\frac{\partial n}{\partial \mathbf{k}}\rangle$ (and therefore the Berry curvature) is zero for all $\mathbf{k}$. The conclusion, however, does not hold if there is band crossing or spin-orbit interaction (not considered so far).
That is, the Berry potential (or curvature) can be non-zero if (i) the lattice does not have space inversion symmetry. This applies to the polarization discussed in the previous subsection. (ii) Time-reversal symmetry is broken, e.g., by a magnetic field. This applies to the quantum Hall system in this subsection. In the next subsection, we consider a system with spin-orbit interaction, in which the Berry curvature plays an important role.

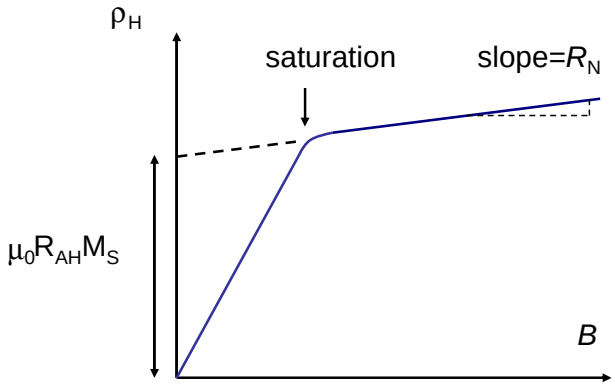


Fig. 8: *When one increases the magnetic field, the Hall resistivity of a ferromagnetic material rises quickly. It levels off after the sample is fully magnetized.*

4.3 Anomalous Hall effect

Soon after Edwin Hall discovered the effect that bears his name in 1879 (at that time he was a graduate student at Johns Hopkins university), he made a similar measurement on iron foil and found a much larger Hall effect. Such a Hall effect in ferromagnetic materials is called the anomalous Hall effect (AHE).

The Hall resistivity of the AHE can be divided into two terms with very different physics (proposed by Smith and Sears in 1929) [22],

$$\rho_H = \rho_N + \rho_{AH} = R_N(T)B + R_{AH}(T)\mu_0 M(T,H), \tag{52}$$

where $B = \mu_0(H + M)$. The first (normal) term is proportional to the magnetic field in the sample. The second (anomalous) term grows roughly linearly with the magnetization M and the coefficient R_{AH} is larger than R_N by one order of magnitude or more. If the applied field is so strong that the material is fully magnetized, then there is no more enhancement from the anomalous term and the Hall coefficient suddenly drops by orders of magnitude (see Fig. 8). Since the normal term is usually much smaller than the anomalous term, we will neglect it in the following discussion.

Unlike the ordinary Hall effect, the Hall *resistivity* in the AHE increases rapidly with temperature. However, the Hall *conductivity*,

$$\sigma_H = \frac{\rho_H}{\rho_L^2 + \rho_H^2} \simeq \frac{\rho_H}{\rho_L^2} \quad (\text{if } \rho_L \gg \rho_H), \tag{53}$$

shows less temperature dependence, where ρ_L is the longitudinal resistivity. The reason will become clear later.

Since the AHE is observed in ferromagnetic materials, the magnetization (or the majority spin) must play a role here. Also, one needs the spin-orbit (SO) interaction to convert the direction of the magnetization to a preferred direction of the transverse electron motion.

Among many attempts to explain the AHE, there are two popular explanations [23], both involve the SO interaction,

$$H_{SO} = -\frac{\hbar}{4m^2c^2}\boldsymbol{\sigma} \cdot (\mathbf{p} \times \nabla V). \tag{54}$$

The first theory was proposed by Karplus and Luttinger (KL) in 1954 [24]. It requires no impurity (the intrinsic scenario) and the V in Eq. (54) is the lattice potential. The Hall resistivity

ρ_{AH} is found to be proportional to ρ_L^2. The other explanation is proposed by Smit in 1958 [25]. It requires (non-magnetic) impurities (the extrinsic scenario) and V is the impurity potential. It predicts $\rho_{AH} \propto \rho_L$. When both mechanisms exist, one has

$$\rho_{AH} = a(M)\rho_L + b(M)\rho_L^2. \tag{55}$$

The Smit term is a result of the skewness of the electron-impurity scattering due to the SO interaction. That is, the spin-up electrons prefer scattering to one side, and the spin-down electrons to the opposite side. Because of the majority spins of the ferromagnetic state, such skew-scatterings produce a net transverse current. Smit's proposal started as an opposition to KL's theory and gained popularity in the early years. As a result, the KL scenario seems to have been ignored for decades.

At the turn of this century, however, several theorists picked up the KL theory and put it under the new light of the Berry curvature [26]. Subsequently, increasing experimental evidences indicate that, in several ferromagnetic materials, the KL mechanism does play a much more important role than the skew-scattering. These works published in renowned journals have attracted much attention, partly because of the beauty of the Berry curvature scenario.

KL's theory, in essence, is very similar to the ones in the previous two subsections. One can first regard the Hamiltonian with the SO interaction as solvable, then treat the electric field as a perturbation. To the first order of the perturbation, one can get the electron velocity with exactly the same form as the one in Eq. (47). The difference is that the state $|n\rangle$ now is modified by the SO interaction and the solid is three dimensional. That is, one simply needs to consider a periodic solid without impurities and apply the Kubo formula, which (in these cases) can be written in Berry curvatures,

$$\sigma_{AH} = \frac{e^2}{\hbar}\frac{1}{V}\sum_{n,\mathbf{k}}\Omega_n(\mathbf{k}). \tag{56}$$

However, not every solid with the SO interaction has the AHE. The transverse velocities (also called the anomalous velocity) in general have opposite signs for opposite spins in the spin-degenerate bands. Therefore, these two Hall currents will get canceled. Again the ferromagnetic state (which spontaneously breaks the time reversal symmetry) is crucial for a net transverse current.

From Eq. (53), one has $\rho_H \simeq \rho_{AH} = \sigma_H \rho_L^2$. Also, the anomalous current generated from the Berry curvature is independent of the relaxation time τ. This explains why the Hall conductivity in the KL theory is proportional to ρ_L^2.

In dilute magnetic semiconductors, one can show that $\mathbf{A}(\mathbf{k}) = \xi\mathbf{S} \times \mathbf{k}$ for the conduction band of the host semiconductor, where ξ is the strength of the SO coupling (more details in Sec. 5.2). Therefore, $\mathbf{\Omega} = \nabla \times \mathbf{A} = 2\xi\mathbf{S}$. In this case, the coefficient $b(M)$ in Eq. (55) is proportional to M. In ferromagnetic materials with a more complex band structure, the Berry curvature shows non-monotonic behavior in magnetization. For one thing, in density-functional-theory calculations, the Berry curvature can be dramatically enhanced when the Fermi energy is near a small energy gap [27]. However, spin fluctuations may smear out the erratic behavior and lead to a smooth variation (see Fig. 9) [28].

The Berry curvature is an intrinsic property of the electronic states. It appears not only at the quantum level, but also in the semiclassical theory of electron dynamics. In the next section, we will see that the QHE, the AHE, and the spin Hall effect can all be unified in the same semiclassical theory.

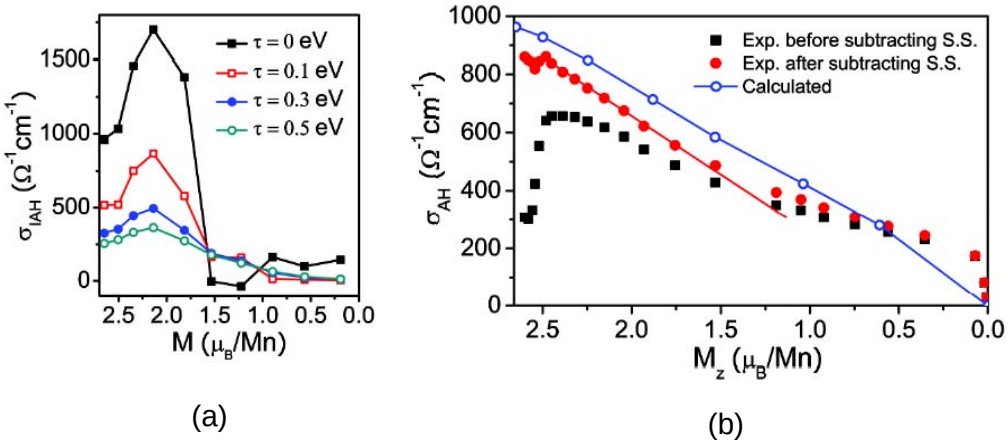


Fig. 9: *(a) Calculated anomalous Hall conductivity (the intrinsic part) versus magnetization for* $\mathrm{Mn_5Ge_3}$ *using different relaxation times. (b) After averaging over long-wavelength spin fluctuations, the calculated anomalous Hall conductivity becomes roughly linear in* M*. The initials S.S. refers to skew scattering. The figures are from Ref. [28].*

5 Berry phase and wave-packet dynamics

When talking about electron transport in solids, people use two different languages: It is either particle scattering, mean free path, cyclotron orbit ..., or localized state, mobility edge, Landau level ... etc. In this section, we use the first language and treat the electrons as particles with trajectories. Besides being intuitive, this approach has the following advantage: The electromagnetic potentials in the Schrödinger equation are often linear in $\mathbf{r}$ and diverge with system size. Such a divergence can be avoided if the wave function of the electron is localized.

5.1 Wave-packet dynamics

Consider an energy band that is isolated from the other bands by finite gaps. Also, the energy band is not degenerate with respect to spin or quasi-spin. The energy band with internal (e.g., spin) degrees of freedom is the subject of the next subsection. When inter-band tunneling can be neglected, the electron dynamics in this energy band can be described very well using a wave-packet formalism.

The wave packet can be built by a superposition of Bloch states $\psi_{n\mathbf{q}}$ in band n (one band approximation),

$$|W\rangle = \int_{BZ} d^3q a(\mathbf{q},t)|\psi_{n\mathbf{q}}\rangle. \tag{57}$$

It is not only localized in position space, but also in momentum space,

$$\langle W|\mathbf{r}|W\rangle = \mathbf{r}_c; \quad \int_{BZ} d^3q\mathbf{q}|a(\mathbf{q})|^2 = \mathbf{q}_c, \tag{58}$$

where $\mathbf{r}_c$ and $\mathbf{q}_c$ are the centers of mass. The shape of the wave packet is not crucial, as long as the electromagnetic field applied is nearly uniform throughout the wave packet.

Instead of solving the Schrödinger equation, we use the time-dependent variational principle to study the dynamics of the wave packet. Recall that in the usual (time-independent) variational principle, one first proposes a sensible wave function with unknown parameters, then minimizes

its energy to determine these parameters. Here, the wave packet is parametrized by its center of mass $(\mathbf{r}_c(t), \mathbf{q}_c(t))$. Therefore, instead of minimizing the energy, one needs to extremize the action $S[C] = \int_C dtL$, which is a *functional* of the trajectory C in phase space.
One starts from the following effective Langrangian,

$$L(\mathbf{r}_c, \mathbf{q}_c; \dot{\mathbf{r}}_c, \dot{\mathbf{q}}_c) = i\hbar\langle W|\frac{d}{dt}|W\rangle - \langle W|H|W\rangle. \tag{59}$$

Notice the resemblance between this $S[C]$ and the the action in the coherent-state path integral (Eq. (29)). The Hamiltonian for a Bloch electron in an electromagnetic field is

$$H = \frac{1}{2m}(\mathbf{p} + e\mathbf{A})^2 + V_L(\mathbf{r}) - e\phi(\mathbf{r}) \simeq H_0 - e\phi + \frac{e}{2m}\mathbf{r} \times \mathbf{p} \cdot \mathbf{B}, \tag{60}$$

in which $H_0 = p^2/2m + V_L$ and ϕ and $\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}$ are treated as perturbations. The fields are allowed to change slowly in space and time, as long as it is approximately uniform and quasi-static (adiabatic) from the wave packet's perspective.
To evaluate the Lagrangian approximately, one can Taylor-expand the potentials with respect to the center of the wave packet and keep only the linear terms. Using this gradient approximation, the wave-packet energy $\langle W|H|W\rangle$ is evaluated as [29],

$$E = E_0(\mathbf{q}_c) - e\phi(\mathbf{r}_c) + \frac{e}{2m}\mathbf{L}(\mathbf{q}_c) \cdot \mathbf{B}, \tag{61}$$

where E_0 is the unperturbed Bloch energy of the band under consideration, and $\mathbf{L}(\mathbf{k}_c) = \langle W|(\mathbf{r} - \mathbf{r}_c) \times \mathbf{p}|W\rangle$ is the self-rotating angular momentum of the wave packet.
On the other hand, the first term in Eq. (59) can be written as

$$i\hbar\langle W|\frac{d}{dt}|W\rangle = \hbar\langle u|i\frac{du}{dt}\rangle + \hbar\mathbf{q}_c \cdot \dot{\mathbf{r}}_c, \tag{62}$$

in which $|u\rangle$ is the unperturbed cell-periodic function. Therefore, the effective Lagrangian is

$$L = \hbar\dot{\mathbf{k}}_c \cdot \mathbf{R}_c + (\hbar\mathbf{k}_c - e\mathbf{A}_c) \cdot \dot{\mathbf{r}}_c - E(\mathbf{r}_c, \mathbf{k}_c), \tag{63}$$

where $\hbar\mathbf{k}_c = \hbar\mathbf{q}_c + e\mathbf{A}_c$ is the gauge-invariant quasi-momentum, $\mathbf{R}_c = i\langle n|\frac{\partial n}{\partial \mathbf{k}_c}\rangle$ is the Berry potential, and $\mathbf{A}_c = \mathbf{A}(\mathbf{r}_c)$.
Treating both $\mathbf{r}_c$ and $\mathbf{k}_c$ as generalized coordinates and using the Euler-Lagrange equation, it is not very difficult to get the following (coupled) equations of motion (EOM) for the wave packet [29],

$$\hbar\dot{\mathbf{k}}_c = -e\mathbf{E} - e\dot{\mathbf{r}}_c \times \mathbf{B}, \tag{64}$$

$$\hbar\dot{\mathbf{r}}_c = \frac{\partial E}{\partial \mathbf{k}_c} - \hbar\dot{\mathbf{k}}_c \times \mathbf{\Omega}_c, \tag{65}$$

where $\mathbf{\Omega}_c = \nabla_{\mathbf{k}_c} \times \mathbf{R}_c$ is the Berry curvature of the band under consideration.
Compared to the usual semiclassical EOM in textbooks, there are two new quantities in Eqs. (64,65), and both lead to important consequences. The first is the Berry curvature $\mathbf{\Omega}$. It generates the so-called anomalous velocity. In the presence of a perturbing electric field, the anomalous velocity is $e\mathbf{E} \times \mathbf{\Omega}$, which is perpendicular to the driving electric field and gives rise to, e.g., the AHE.

The second is the spinning angular momentum $\mathbf{L}$ in Eq. (61). It is closely related to the orbital magnetization of a solid [30]. For a spinful wave packet (Sec. 5.2), this $\mathbf{L}$ modifies the electron spin and is the origin of the anomalous g-factor in solids. In fact, starting from Dirac's relativistic electron theory (which has no explicit spin in the Hamiltonian), we have shown that, the wave packet in the positive-energy branch of the Dirac spectrum has an intrinsic spinning angular momentum [31]. That is, it explains why an electron has spin.
In the semiclassical theory of electron transport, the current density is given by

$$\mathbf{j} = -\frac{e}{V}\sum_{n\mathbf{k}} f\dot{\mathbf{r}}, \tag{66}$$

where $f = f_0 + \delta f$ is the distribution function away from equilibrium. The distribution function f is determined from the Boltzmann equation,

$$\dot{\mathbf{r}}\cdot\frac{\partial f}{\partial \mathbf{r}} + \dot{\mathbf{k}}\cdot\frac{\partial f}{\partial \mathbf{k}} = -\frac{\delta f}{\tau}, \tag{67}$$

where τ is the relaxation time. For a homogeneous system in an electric field, $\delta f \simeq \tau\frac{e}{\hbar}\mathbf{E}\cdot\frac{\partial f_0}{\partial \mathbf{k}}$, and

$$\mathbf{j} \simeq -\frac{e}{V}\sum_{n\mathbf{k}}\left(\delta f\frac{\partial E_n}{\hbar\partial \mathbf{k}} + f_0\frac{e}{\hbar}\mathbf{E}\times\mathbf{\Omega}_n\right). \tag{68}$$

The usual current (the first term) depends on carrier relaxation time τ through the change of the distribution function δf. On the other hand, the second term gives the Hall current. Clearly, this $\mathbf{\Omega}$ is also the one in the Kubo formula of QHE and AHE. (The latter involves spin-degenerate band and belongs more properly to the next subsection.)
We emphasize that, just like the Bloch energy $E_0(\mathbf{k})$, both $\mathbf{\Omega}(\mathbf{k})$ and $\mathbf{L}(\mathbf{k})$ are intrinsic to the energy band (not induced by the applied field). They are the three main pillars of band theory. Unlike the Bloch energy that has been around for a very long time, the other two quantities are relatively new players, but their importance should increase over time.
If there is only a magnetic field, then combining Eq. (64) and Eq. (65) gives

$$\hbar\dot{\mathbf{k}}_c = \frac{-\frac{e}{\hbar}\frac{\partial E}{\partial \mathbf{k}_c}\times\mathbf{B}}{1+\frac{e}{\hbar}\mathbf{B}\cdot\mathbf{\Omega}}. \tag{69}$$

It describes a cyclotron orbit moving on a plane perpendicular to the magnetic field. The orbit is an energy contour on the Fermi surface. Its size can change continuously, depending on the electron's initial condition.
One can apply a Bohr-Sommerfeld quantization rule to get quantized orbits, which have discrete energies (the Landau levels). The EOM in momentum space, Eq. (69), follows from the effective Lagrangian,

$$L(\mathbf{k}_c;\dot{\mathbf{k}}_c) = \frac{\hbar^2}{2eB}\mathbf{k}_c\times\dot{\mathbf{k}}_c\cdot\hat{B} + \hbar\dot{\mathbf{k}}_c\cdot\mathbf{R}_c - E(\mathbf{k}_c). \tag{70}$$

This gives the generalized momentum,

$$\boldsymbol{\pi} = \frac{\partial L}{\partial\dot{\mathbf{k}}_c} = -\frac{\hbar^2}{2eB}\mathbf{k}_c\times\hat{B} + \hbar\mathbf{R}_c. \tag{71}$$

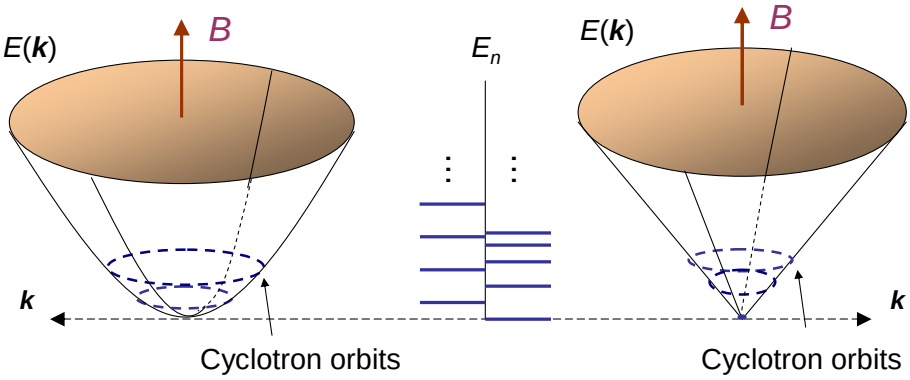


Fig. 10: *The quantized cyclotron orbits on two different energy surfaces. The one on the left is a paraboloid near its band edge; the one on the right is a conical surface. Without Berry phase correction, the Landau-level energies are* $E_n = (n + 1/2)\hbar\omega_c$ *and* $E_n = v_F\sqrt{2eB\hbar(n + 1/2)}$ *respectively. In graphene, an orbit circling the Dirac point acquires a Berry phase of* π*, which cancels the 1/2 in the square root.*

The quantization condition is given by $\oint \boldsymbol{\pi} \cdot d\mathbf{k}_c = (m+\gamma)h$, where m is a non-negative integer and $\gamma = 1/2$ for the cyclotron motion. Therefore, we have

$$\frac{\hat{B}}{2} \cdot \oint_{C_m} (\mathbf{k}_c \times d\mathbf{k}_c) = 2\pi \left(m + \frac{1}{2} - \frac{\Gamma(C_m)}{2\pi} \right) \frac{eB}{\hbar}, \tag{72}$$

where $\Gamma(C_m) = \oint_{C_m} \mathbf{R}_c \cdot d\mathbf{k}_c$ is the Berry phase for orbit C_m.
This equation determines the allowed size (and therefore energy) of the cyclotron orbit. The Berry phase correction slightly shifts the Landau-level energies. For example, the orbit around the Dirac point of graphene picks up a Berry phase of π due to the monopole at the origin. This cancels the other $1/2$ in Eq. (72) and results in a zero-energy level at the Dirac point (see Fig. 10). This agrees nicely with experimental measurements [32].

5.2 Non-Abelian generalization

In the one-band theory without internal degrees of freedom, the Bloch state has only one component and the gauge structure of the Berry phase is Abelian. When the band has internal degrees of freedom (henceforth simply called the spin), the Bloch state has several components and the gauge structure becomes non-Abelian. This happens, for example, in energy bands with Kramer's degeneracy. By extending the semiclassical dynamics to such cases, one is able to investigate problems involving spin dynamics and spin transport.
The scheme for building such a theory is the same as the one in the previous subsection. Therefore, we only give a very brief outline below. One first constructs a wave packet from the Bloch states $\psi_{n\mathbf{q}}$,

$$|W\rangle = \sum_{n=1}^{D} \int_{BZ} d^3q a(\mathbf{q}, t)\eta_n(\mathbf{q}, t)|\psi_{n\mathbf{q}}\rangle. \tag{73}$$

Here n is a spinor index for an isolated band with D-fold degeneracy, $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_D)^T$ is a normalized spinor at each $\mathbf{q}$, and $a(\mathbf{q}, t)$ is again a narrow distribution centered at $\mathbf{q}_c(t)$.
Similar to the non-degenerate case, there are three basic quantities in such a formalism, the Bloch energy $\mathcal{H}_0(\mathbf{q})$, the Berry connection $\boldsymbol{\mathcal{R}}(\mathbf{q})$ (and related curvature, now written as $\boldsymbol{\mathcal{F}}(\mathbf{q})$),

and the spinning angular momentum $\boldsymbol{\mathcal{L}}(\mathbf{q})$ [33]. They all become matrix-valued functions and are denoted by calligraphic fonts. The Bloch energy is simply an identity matrix multiplied by $E_0(\mathbf{q})$ since all spinor states have the same energy.
The matrix elements of the Berry connection are,

$$\mathbf{R}_{mn}(\mathbf{q}) = i\left\langle u_{m\mathbf{q}}|\frac{\partial u_{n\mathbf{q}}}{\partial \mathbf{q}}\right\rangle. \tag{74}$$

The Berry curvature is given by,

$$\boldsymbol{\mathcal{F}}(\mathbf{q}) = \nabla_{\mathbf{q}} \times \boldsymbol{\mathcal{R}} - i\boldsymbol{\mathcal{R}} \times \boldsymbol{\mathcal{R}}. \tag{75}$$

Recall that the Berry connection and Berry curvature in the Abelian case are analogous to the vector potential and the magnetic field in electromagnetism (see Sec. 2.1). Here, $\boldsymbol{\mathcal{R}}$ and $\boldsymbol{\mathcal{F}}$ also are analogous to the gauge potential and gauge field in the non-Abelian $SU(2)$ gauge field theory [34].
The expectation value of the third basic quantity, the spinning angular momentum, is again given by $\mathbf{L}(\mathbf{q}_c) = \langle W|(\mathbf{r}-\mathbf{r}_c)\times\mathbf{p}|W\rangle$. However, it is often written in an alternative (Rammal-Wilkinson) form easier for evaluation,

$$\mathbf{L}(\mathbf{q}) = i\frac{m}{\hbar}\left\langle \frac{\partial u}{\partial \mathbf{q}}\right| \times \left[\tilde{H}_0 - E_0(\mathbf{q})\right] \left|\frac{\partial u}{\partial \mathbf{q}}\right\rangle, \tag{76}$$

where the cell-periodic function without a subscript is defined as $|u\rangle = \sum_{n=1}^{D}\eta_n|u_n\rangle$ and $\tilde{H}_0$ is the Hamiltonian for $|u\rangle$. The corresponding matrix-valued function $\boldsymbol{\mathcal{L}}$ therefore has the matrix elements,

$$\mathbf{L}_{nl}(\mathbf{q}) = i\frac{m}{\hbar}\left\langle \frac{\partial u_n}{\partial \mathbf{q}}\right| \times \left[\tilde{H}_0 - E_0(\mathbf{q})\right] \left|\frac{\partial u_l}{\partial \mathbf{q}}\right\rangle. \tag{77}$$

Obviously, after taking the spinor average, one has the angular momentum in Eq. (76), $\mathbf{L} = \langle\boldsymbol{\mathcal{L}}\rangle \equiv \boldsymbol{\eta}^\dagger\boldsymbol{\mathcal{L}}\boldsymbol{\eta} = \sum_{nl}\eta_n^*\mathbf{L}_{nl}\eta_l$.

Equations of motion
So far we have laid out the necessary ingredients in the non-Abelian wave packet theory. Similar to Sec. 5.2, we can use Eq. (59) to get the effective Lagrangian for the center of mass, $(\mathbf{r}_c, \mathbf{k}_c)$, and the spinor $\boldsymbol{\eta}$. Afterwards, the Euler-Lagrange equation for this effective Lagrangian leads to the following EOM [33],

$$\hbar\dot{\mathbf{k}}_c = -e\mathbf{E} - e\dot{\mathbf{r}}_c\times\mathbf{B}, \tag{78}$$

$$\hbar\dot{\mathbf{r}}_c = \left\langle\left[\frac{\mathcal{D}}{\mathcal{D}\mathbf{k}_c}, \mathcal{H}\right]\right\rangle - \hbar\dot{\mathbf{k}}_c\times\mathbf{F}, \tag{79}$$

$$i\hbar\dot{\boldsymbol{\eta}} = \left(\frac{e}{2m}\boldsymbol{\mathcal{L}}\cdot\mathbf{B} - \hbar\dot{\mathbf{k}}_c\cdot\boldsymbol{\mathcal{R}}\right)\boldsymbol{\eta}, \tag{80}$$

where $\mathbf{F} = \langle\boldsymbol{\mathcal{F}}\rangle$, and the covariant derivative $\mathcal{D}/\mathcal{D}\mathbf{k}_c \equiv \partial/\partial\mathbf{k}_c - i\boldsymbol{\mathcal{R}}$. The semiclassical Hamiltonian inside the commutator in Eq. (79) is

$$\mathcal{H}(\mathbf{r}_c, \mathbf{k}_c) = \mathcal{H}_0(\mathbf{k}_c) - e\phi(\mathbf{r}_c) + \frac{e}{2m}\boldsymbol{\mathcal{L}}(\mathbf{k_c})\cdot\mathbf{B}, \tag{81}$$

where $\mathbf{k}_c = \mathbf{q}_c + (e/\hbar)\mathbf{A}(\mathbf{r}_c)$.

Even though these equations look a little complicated, the physics is very similar to that of the simpler Abelian case in Sec. 5.1. There are two differences, however. First, the anomalous velocity in Eq. (79) is now spin-dependent in general. In some interesting cases (see below), $\mathbf{F}$ is proportional to the spin vector $\mathbf{S} = \langle \boldsymbol{\mathcal{S}} \rangle$, where $\boldsymbol{\mathcal{S}}$ is the spin matrix. Therefore, if one applies an electric field to such a system, the spin-up and spin-down electrons will move to opposite transverse directions. This is the cause of the AHE and the spin Hall effect.
Second, we now have an additional equation (Eq. (80)) governing the spinor dynamics. From Eq. (80) we can derive the equation for $\mathbf{S}$,

$$i\hbar\dot{\mathbf{S}} = \left\langle \left[\boldsymbol{\mathcal{S}}, \mathcal{H} - \hbar\dot{\mathbf{k}}_c \cdot \boldsymbol{\mathcal{R}} \right] \right\rangle . \tag{82}$$

The spin dynamics in Eq. (82) is influenced by the Zeeman energy in $\mathcal{H}$, as it should be. We will demonstrate below that the term with the Berry connection is in fact the spin-orbit energy. Such an energy is not explicit in $\mathcal{H}$, but only reveals itself after $\mathcal{H}$ is being re-quantized.

Re-quantization
As we have shown in Sec. 5.1, re-quantization of the semiclassical theory is necessary when one is interested in, for example, the quantized cyclotron orbits that correspond to the Landau levels. Here we introduce the method of canonical quantization, which is more appropriate for the non-Abelian case compared to the Bohr-Sommerfeld method.
In this approach, one needs to find variables with canonical Poisson brackets,

$$\begin{aligned} \{r_\alpha, r_\beta\} &= 0, \\ \{p_\alpha, p_\beta\} &= 0, \\ \{r_\alpha, p_\beta\} &= \delta_{\alpha\beta}, \end{aligned} \tag{83}$$

then promote these brackets to quantum commutators. As a result, the variables become non-commutating operators and the classical theory is quantized.
An easier way to judge if the variables are canonical is by checking if they satisfy the canonical EOM,

$$\dot{\mathbf{r}} = \frac{\partial E}{\partial \mathbf{p}}; \dot{\mathbf{p}} = -\frac{\partial E}{\partial \mathbf{r}}. \tag{84}$$

The variables $\mathbf{r}_c$ and $\mathbf{k}_c$ that depict the trajectory of the wave packet are not canonical variables because their EOM are not of this form. This is due to the vector potential and the Berry connection, $\mathbf{A}(\mathbf{r}_c)$ and $\mathbf{R}(\mathbf{k}_c)$, in the Lagrangian (see Eq. (63)).
In fact, if one can remove these two gauge potentials from the Lagrangian by a change of variables,

$$L = \mathbf{p} \cdot \dot{\mathbf{r}} - E(\mathbf{r}, \mathbf{p}), \tag{85}$$

then these new variables will automatically be canonical. Such a transformation is in general non-linear and cannot be implemented easily. However, if one only requires an accuracy to linear order of the electromagnetic fields (consistent with the limit of our semiclassical theory), then the new variables can indeed be found.
The canonical variables $\mathbf{r}$ and $\mathbf{p}$ accurate to linear order in the fields are related to the center-of-mass variables as follows [35],

$$\begin{aligned} \mathbf{r}_c &= \mathbf{r} + \boldsymbol{\mathcal{R}}(\boldsymbol{\pi}) + \boldsymbol{\mathcal{G}}(\boldsymbol{\pi}), \\ \hbar\mathbf{k}_c &= \mathbf{p} + e\mathbf{A}(\mathbf{r}) + e\mathbf{B} \times \boldsymbol{\mathcal{R}}(\boldsymbol{\pi}), \end{aligned} \tag{86}$$

where $\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}(\mathbf{r})$, and $\mathcal{G}_\alpha(\boldsymbol{\pi}) \equiv (e/\hbar)(\boldsymbol{\mathcal{R}} \times \mathbf{B}) \cdot \partial\boldsymbol{\mathcal{R}}/\partial\pi_\alpha$. The last terms in both equations can be neglected in some cases. For example, they will not change the force and the velocity in Eqs. (78) and (79). These relations constitute a generalization of the Peierls substitution.
When expressed in the new variables, the semiclassical Hamiltonian in Eq. (81) becomes,

$$\begin{aligned} \mathcal{H}(\mathbf{r}, \mathbf{p}) &= \mathcal{H}_0(\boldsymbol{\pi}) - e\phi(\mathbf{r}) + e\mathbf{E} \cdot \boldsymbol{\mathcal{R}}(\boldsymbol{\pi}) \\ &+ \mathbf{B} \cdot \left[\frac{e}{2m} \boldsymbol{\mathcal{L}}(\boldsymbol{\pi}) + e\boldsymbol{\mathcal{R}} \times \frac{\partial \mathcal{H}_0}{\partial \boldsymbol{\pi}} \right], \end{aligned} \tag{87}$$

where we have used the Taylor expansion and neglected terms nonlinear in the fields. Finally, one promotes the canonical variables to quantum conjugate variables to convert $\mathcal{H}$ to an effective quantum Hamiltonian.
The dipole-energy term $e\mathbf{E} \cdot \boldsymbol{\mathcal{R}}$ is originates from the shift between the charge center $\mathbf{r}_c$ and the canonical variable $\mathbf{r}$. We will show below that for a semiconductor electron, the dipole term is in fact the spin-orbit coupling.
The correction to the Zeeman energy is also related to the Berry connection. Near a band edge, where the effective mass approximation is applicable and $E_0 = \pi^2/2m^*$, this term can be written as $e\boldsymbol{\mathcal{R}} \cdot \mathbf{v} \times \mathbf{B}$, where $\mathbf{v} = \boldsymbol{\pi}/m^*$. We know that an electron moving in a static magnetic field feels an effective electric field $\mathbf{E}_{eff} = \mathbf{v} \times \mathbf{B}$. Therefore, this term arises as a result of the electric dipole energy in electron's own reference frame.

Semiconductor electron
A necessary requirement for the non-Abelian property is that the Bloch electron has to have internal degrees of freedom. In a semiconductor with both space-inversion and time-reversal symmetries, every Bloch state is two-fold degenerate due to Kramer's degeneracy. But where do we expect to see the non-Abelian Berry connection and curvature?
Instead of the full band structure, one can start from a simpler band structure using the $\mathbf{k} \cdot \mathbf{p}$ expansion. Assuming the fundamental gap is located at $\mathbf{k} = 0$, then for small k, one has an effective Hamiltonian with 4 bands, 6 bands, 8 bands, or more, depending on the truncation. In the following discussion, we use a 8-band Kane Hamiltonian that includes the conduction band, the HH-LH bands, and the spin-orbit (SO) split-off band, each with 2-fold degeneracy (see Fig. 11). The explicit Kane Hamiltonian can be found in Ref. [36].
We focus only on the wave packet in the conduction band. Without going into details, we first show the Berry connection that is essential to the wave packet formulation. The result correct to order k^1 and up to a gauge rotation is [35],

$$\boldsymbol{\mathcal{R}} = \frac{V^2}{3} \left[\frac{1}{E_g^2} - \frac{1}{(E_g + \Delta)^2} \right] \boldsymbol{\sigma} \times \mathbf{k}, \tag{88}$$

where $V = \frac{\hbar}{m}\langle S|\hat{p}_x|X\rangle$, E_g is the energy gap, and Δ is the SO gap. Therefore, the dipole term $e\mathbf{E} \cdot \boldsymbol{\mathcal{R}}$ in Eq. (87) becomes,

$$H_{so} = e\mathbf{E} \cdot \boldsymbol{\mathcal{R}} = \alpha \mathbf{E} \cdot \boldsymbol{\sigma} \times \mathbf{k}, \tag{89}$$

where $\alpha \equiv (eV^2/3)[1/E_g^2 - 1/(E_g + \Delta)^2]$. It coincides precisely with the spin-orbit coupling of a conduction electron. This shows that the SO coupling has a very interesting connection with the Berry connection. This is also the case for the SO coupling in Dirac's relativistic electron theory [37].

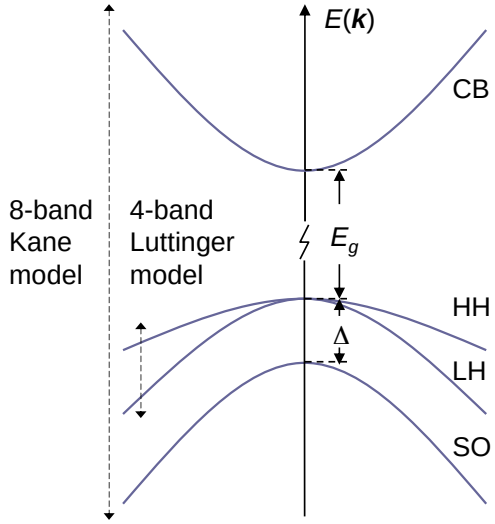


Fig. 11: *One can use the 4-band Luttinger model or the 8-band Kane model to approximate the energy bands near the fundamental gap.*

The Berry curvature calculated from Eq. (75) gives (to the lowest order) $\boldsymbol{\mathcal{F}} = \alpha/e\boldsymbol{\sigma}$, which is proportional to spin. Therefore, the anomalous velocity $e\mathbf{E} \times \mathbf{F}$ in Eq. (79) is $\alpha\mathbf{E} \times \langle\boldsymbol{\sigma}\rangle$. That is, spin-up and spin-down electrons acquire opposite transverse velocities. In non-magnetic materials, these two species have the same population and we do not expect to see a net transverse current. However, "if" one defines a spin current as the *difference* of these two transverse currents, then there will be a net *spin* current, giving rise to the spin Hall effect [38].
One can also calculate the spinning angular momentum of the conduction electron from Eq. (77). The result is,

$$\boldsymbol{\mathcal{L}} = -\frac{2mV^2}{3\hbar}\left(\frac{1}{E_g} - \frac{1}{E_g + \Delta}\right)\boldsymbol{\sigma}. \tag{90}$$

Through the Zeeman energy in Eq. (87), the orbital magnetic moment generated from Eq. (90) contributes an extra g-factor,

$$\delta g = -\frac{4}{3}\frac{mV^2}{\hbar^2}\left(\frac{1}{E_g} - \frac{1}{E_g + \Delta}\right). \tag{91}$$

This is the anomalous g-factor of the conduction electron [39]. Therefore, the anomalous g-factor in solid is indeed a result of the self-rotating motion of the electron wave packet.
Finally, the effective quantum Hamiltonian in Eq. (87) for the conduction band has the following form,

$$H(\mathbf{r}, \mathbf{p}) = E_0(\boldsymbol{\pi}) - e\phi(\mathbf{r}) + \alpha\mathbf{E} \cdot \boldsymbol{\sigma} \times \boldsymbol{\pi} + \frac{\delta g}{2}\mu_B \mathbf{B} \cdot \boldsymbol{\sigma}, \tag{92}$$

where E_0 includes the Zeeman energy from the bare spin, α is given below Eq. (89), δg is given in Eq. (91), and the correction to the Zeeman energy has been neglected. This Hamiltonian agrees with the one obtained from block diagonalization [36]. The wave packet approach is not only simpler, but also reveals the deep connections between various effective couplings and the Berry potential.
Some comments are in order: First, we emphasize again that it is necessary to include the Berry curvature and orbital moment in order to account for physical effects to first order in

external fields. Furthermore, from the discussions above, we can see that these quantities are also sufficient for building a correct quantum theory.
Second, starting from a quantum theory, one can construct a semiclassical theory in a specific subspace. This theory can later be re-quantized. The re-quantized effective theory applies to a smaller Hilbert space compared to the original quantum theory. Nevertheless, it can still have its own semiclassical theory, which in turn can again be re-quantized. As a result, a hierarchy of effective theories and gauge structures can be produced, all within the wave packet approach (see Ref. [35] for more discussions).

6 Concluding remarks

In this review, selected topics related to Berry phase in solid state physics are reported. Many of these topics have been fully developed over the years. The exposition here only serves as an introduction, without going into details and more recent development. Readers interested in certain topics can consult some of the following books or review articles: [1] and [40] on Berry phase in general, [41] and [42] on electric polarization, [43] on quantum Hall effect, [44] and [45] on anomalous Hall effect, [46] and [47] on dynamics of Bloch electrons, and [35] on non-Abelian wave packet dynamics.
In optics, the Berry curvature is related to a transverse shift (side jump) of a light beam reflected off an interface.[48] The shift is roughly the order of the wavelength. Its direction depends on the circular polarization of the incident beam. This is called the *optical Hall effect*, or the Imbert-Federov effect,[49] which is not covered here. The side jump of a light beam is similar to the analogous "jump" of an electron scattering off an impurity in the anomalous Hall effect [22]. A more detailed study of the optical transport involving spin can be found in Ref. [50].
Several topics not covered here can be found in an upcoming review on Berry phase in solid state physics [51]. These topics include the orbital magnetization of a solid, dipole moment of the wave packet, anomalous thermoelectric transport, and inhomogeneous electric polarization. It is amazing that the Berry phase plays such a versatile role in so many solid-state phenomena. On the other hand, several challenging subjects still remain largely unexplored. For example, the effect of the Berry phase in systems in which non-adiabatic processes or many-body interaction is crucial. Therefore, one can expect to see more of the intriguing Berry phase effects in solid state systems.

References

[1] A. Shapere and F. Wilczek, (ed) *Geometric Phases in Physics* (Singapore: World Scientific, 1989)

[2] For example, see p. 290 in L. I. Schiff, *Quantum Mechanics* (McGraw Hill, 3rd ed 1968)

[3] M. V. Berry, Proc. R. Soc. A **392**, 45 (1984)

[4] B. F. Schutz, *Geometrical methods of mathematical physics* (Cambridge University Press, 1980)

[5] M. Stone, Phys. Rev. D **33**, 1191 (1986)

[6] For example, see K. Ohgushi, S. Murakami, N. Nagaosa, Phys. Rev. B **62**, R6065 (2000); Y. Taguchi, Y. Oohara, H. Yoshizawa, *et al.* Science **291**, 2573 (2001); R. Shindou, N. Nagaosa, Phys. Rev. Lett. **87**, 116801 (2001)

[7] L. P. Levy *et al.* , Phys. Rev. Lett. **64**, 2074 (1990); V. Chandrasekhar *et al.* , Phys. Rev. Lett. **67**, 3578 (1991)

[8] D. Loss, P. Golbart, and A. V. Balatsky, Phys. Rev. Lett. **65**, 1655 (1990)

[9] F. Schütz, M. Kollar, and P. Kopietz, Phys. Rev. Lett. **91**, 017205 (2003)

[10] D. Loss, D. P. DiVincenzo, and G. Grinstein, Phys. Rev. Lett. **69**, 3232 (1992); J. von Delft, and C. L. Henley, Phys. Rev. Lett. **69**, 3236 (1992)

[11] See Chap 21 in R. Shankar, *Quantum Mechanics* (Springer, 2nd ed. 1994)

[12] Y. Zhang *et al.* , Nature **438**, 201 (2005).

[13] R. Resta, Ferroelectrics **136**, 51 (1992)

[14] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993)

[15] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983); Q. Niu, Phys. Rev. Lett. **64**, 1812 (1990)

[16] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980)

[17] R. B. Laughlin, Phys. Rev. B. **23**, 5632 (1981)

[18] These mobile charges are carried through an extended state in the middle of every Landau level. Such an extended state is flanked by localized states on both sides of the energy. When the chemical potential falls within the localized states for a range of magnetic field, no charges can be transported and the Hall conductance exhibits a plateau over this range of magnetic field. Without disorder, there will be no localized states, thus no plateaus. Therefore, disorder also plays a crucial role in the quantum Hall effect.

[19] J. D. Thouless, M. Kohmoto, P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982); Also see M. Kohmoto Ann. Phys. (N.Y.) **160**, 355 (1985)

[20] Because the vector potential $\mathbf{A}_0$ in the unperturbed Hamiltonian H_0 breaks the lattice symmetry, one needs to solve the Schrödinger equation based on the so-called magnetic translation symmetry. Therefore, to be precise, the Bloch state, the Bloch energy, and the Brillouin zone mentioned here for the QHE should actually be the magnetic Bloch state, the magnetic Bloch energy, and the magnetic Brillouin zone, respectively. This distinction is not emphasized in the text.

[21] Q. Niu, D. J. Thouless, and Y. S. Wu, Phys. Rev. B **31**, 3372 (1985)

[22] L. Berger and G. Bergmann, in *The Hall Effect and Its Applications* edited by C. L. Chien, and C. R. Westgate (Plenum, New York, 1979) p. 55.

[23] Another often mentioned explanation is the side-jump mechanism proposed by Berger in 1970. It involves the impurity scattering but the underlying formalism can be rephrased in a way related to the Berry connection. It also predicts $\rho_H \propto \rho_L^2$. One can consult Ref. [22] for more details.

[24] R. Karplus and J. M. Luttinger, Phys. Rev. **95**, 1154 (1954)

[25] J. Smit, Physica **21**, 877 (1955)

[26] Y. Taguchi, Y. Oohara, H. Yoshizawa, N. Nagaosa, and Y. Tokura, Science **291**, 2573 (2001); T. Jungwirth, Q. Niu, and A. H. MacDonald, Phys. Rev. Lett. **88**, 207208 (2002)

[27] Z. Fang *et al.* , Science **302**, 92 (2003)

[28] C. Zeng, Y. Yao, Q. Niu, and H. Weitering, Phys. Rev. Lett. **96**, 37204 (2006)

[29] M. C. Chang, and Q. Niu, Phys. Rev. B **53**, 7010 (1996); G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999)

[30] D. Xiao, Y. Yao, Z. Fang, and Q. Niu, Phys. Rev. Lett. **97**, 026603 (2006).

[31] P. C. Chuu, M. Chang, and Q. Niu to be published

[32] S. E. Novoselov *et al.* , Nature **438**, 197 (2005); Y. Zhang *et al.* , Nature **438**, 201 (2005)

[33] D. Culcer, Y. Yao, and Q. Niu, Phys. Rev. B **72**, 85110 (2005)

[34] F. Wilczek, and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984)

[35] M. C. Chang, and Q. Niu, J. Phys.: Condens. Matter **20**, 193202 (2008)

[36] R. Winkler *Spin-orbit coupling effects in two-dimensional electron and hole systems* (Springer, 2003)

[37] H. Mathur, Phys. Rev. Lett. **67**, 3325 (1991); R. Shankar, and Mathur, Phys. Rev. Lett. **73**, 1565 (1994)

[38] S. Murakami, N. Nagaosa, and S. C. Zhang, Science **301**, 1348 (2003)

[39] See Problem 9.16 in P. Y. Yu and M. Cardona *Fundamentals of Semiconductors* the third ed. (Springer, 2003)

[40] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, *The Geometric Phase in Quantum Systems* (Springer-Verlag, Heidelberg, 2003)

[41] R. Resta, Rev. Mod. Phys. **66**, 899 (1994)

[42] R. Resta, J. Phys.: Condens. Matter **12**, R107 (2000)

[43] R. E. Prange, and S. M. Girvin, *The quantum Hall effect* (Springer, 2nd ed., 1990)

[44] N. Nagaosa, J. Phys. Soc. Jpn. **75**, 042001 (2006)

[45] N. A. Sinitsyn, J. Phys: Condens. Matter **20**, 023201 (2008)

[46] E. I. Blount, *Solid State Physics*, edited by Seitz F and Turnbull D (Academic Press Inc., New York, 1962), Vol. 13

[47] G. Nenciu, Rev. Mod. Phys. **63**, 91 (1991)

[48] M. Onoda, S. Murakami, and N. Nagaosa, Phys. Rev. Lett. **93**, 83901 (2004); K. Sawada, and N. Nagaosa, Phys. Rev. Lett. **95**, 237402 (2005); M. Onoda, S. Murakami, and N. Nagaosa, Phys. Rev. E **74**, 66610 (2006)

[49] F. I. Fedorov, Dokl. Akad. Nauk SSSR **105**, 465 (1955); C. Imbert, Phys. Rev. D **5**, 787 (1972)

[50] K. Y. Bliokh *et al.*, Phys. Rev. Lett. **96**, 073903 (2006); K. Bliokh, Phys. Rev. Lett. **97** 043901 (2006); C. Duval et al., Phys. Rev. D **74**, 021701 (R) (2006)

[51] D. Xiao, M. C. Chang, and Q. Niu, to be published

B 1 Spin polarization: From the EPR paradox to the GMR effect

P. Grünberg
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Stern-Gerlach experiments 2
1.1 History 2
1.2 Detection of the quantization of spin-polarization by Stern-Gerlach experiments 2
1.3 The EPR paradox in the Bohm version 2

2 Optical experiments 3
2.1 Some important features of polarized light 3
2.2 Distinction between local realism and quantum theory: Bell's inequality 4

3 Spin polarization in 3d-metals and its relation to band structure 6

4 Interlayer Exchange Coupling 7
4.1 Discovery 7
4.2 Phenomenological description 7
4.3 Microscopic origin of oscillatory coupling across metallic interlayers 9
4.4 Simple model to explain coupling across insulators and semiconductors 11

5 Giant magnetoresistance (GMR) 11
5.1 First observations 11
5.2 Microscopic origin 12

6 Conclusion 14

References 14

1 Stern-Gerlach experiments

1.1 History

The discovery of the electron as a charged particle has been attributed to J.J. Thomson, who received the Nobel Prize in 1906 for this achievement. The electron spin has been discovered in 1925 by Goudsmit, Uhlenbeck, and Pauli while trying to understand the optical spectra of alkali halides. Already in 1922 Stern and Gerlach had performed an experiment with a clear signature of the spin but had misinterpreted it by the assumption that they had verified experimentally the quantization of orbital angular momentum. The Nobel Prize was never awarded to the discovery of the spin *per se*.

Nevertheless the Stern-Gerlach (SG) experiment has become one of the key experiments for the detection of spin polarization and its spatial quantization. It is often used to introduce quantum mechanics and to discuss the Einstein-Podolsky-Rosen (EPR) paradox in the Bohm version. This is relevant in the context of quantum teleportation and quantum computing, which are topics of this school. We therefore discuss the SG-experiment and the EPR paradox in some more detail. However as will be seen, the answer to the EPR paradox comes from optical experiments using polarized light and not from SG experiments.

1.2 Detection of the quantization of spin-polarization by Stern-Gerlach experiments

Figure 1 shows a schematic of the general SG setup. Ag atoms carry a magnetic moment, which is due to an unpaired s-electron. In an inhomogeneous magnetic field a beam of Ag atoms therefore experiences a deflection, which according to $S_z = \pm 1/2$ (as displayed on the right-hand side) is in the direction of the field gradient or opposite to it thus demonstrating the spatial quantization.

1.3 The EPR paradox in the Bohm version

For the following discussion we represent an SG device by the magnets with their peculiar shape. The letter *z* in Fig. 2 indicates that the field gradient and separation of the two spins is in the *z*-direction, likewise for *y* that it is in the *y*-direction. The ball indicated in pink is meant to be a source of entangled electrons, which are the result of a decay of some other particle

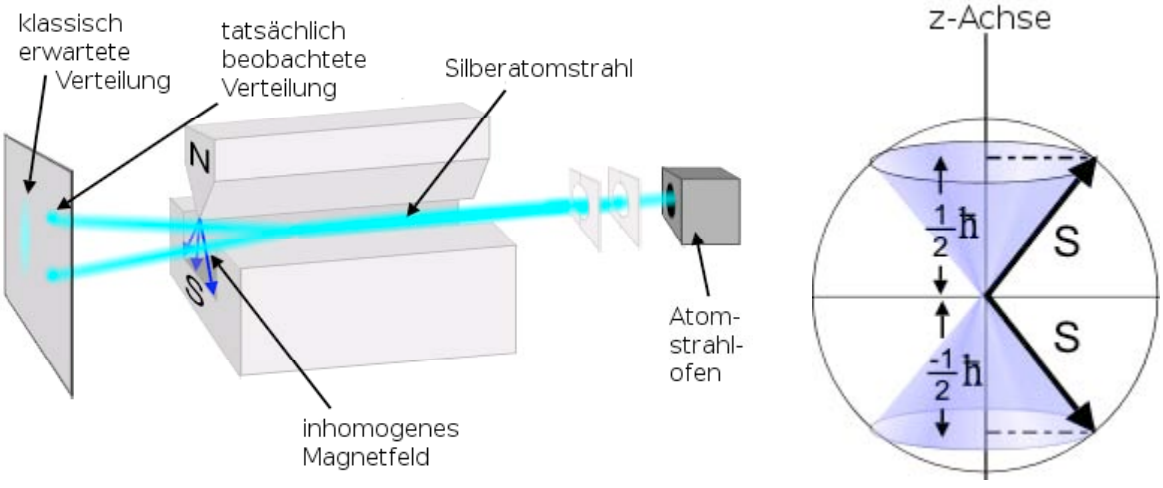


Fig. 1: *The Stern-Gerlach experiment.*

with zero angular momentum. Under these conditions a certain spin component of an electron going to the right (S_r) must be the same but with reversed sign of an electron going to the left S_l, *i.e.* $S_r = -S_l$. That means if we know S_y for the electron going to the right we know it also for the electron going to the left. With the other SG device we can measure in the same way S_z, so we know it also for both electrons. With this result we can reach the conclusion that it is possible to measure simultaneously both S_y and S_z, here even for both electrons.
However, simultaneous knowledge of non-commuting variables violates the uncertainty principle of quantum mechanics. The reason why the thought experiment by EPR was invented in the first place was to disprove this principle, *i.e.* the statistical nature of quantum mechanics. The idea to consider spin components came from Bohm. The inventors were aware that an obvious way out could be that measurement of *e.g.* S_y could destroy information about S_z gained just an instant before. But since the two SG setups of Fig. 2 could basically be arbitrarily far away, it could mean that information under certain conditions would have to be transmitted faster then with the speed of light. Einstein called it "spooky action at a distance" and never believed it. As we will see, Einstein was wrong in that case, but the „spooky action" does in fact not violate the mentioned law on the maximum speed of signals.

2 Optical experiments

2.1 Some important features of polarized light

(1) Light energy is quantized and comes in the form of multiples of an elemental quantum $h\nu$ called photon.
(2) Suppose a polarizer P1 is set at angle α and the light then passes an analyzer P2 set at β. The probability normalized to 1 that a photon is transmitted is then given by

$$p_{trans}(\alpha-\beta) = \cos^2(\alpha-\beta) \qquad (1a)$$

and the probability that it is blocked by

$$p_{block}(\alpha-\beta) = 1 - \cos^2(\alpha-\beta) = \sin^2(\alpha-\beta) \qquad (1b)$$

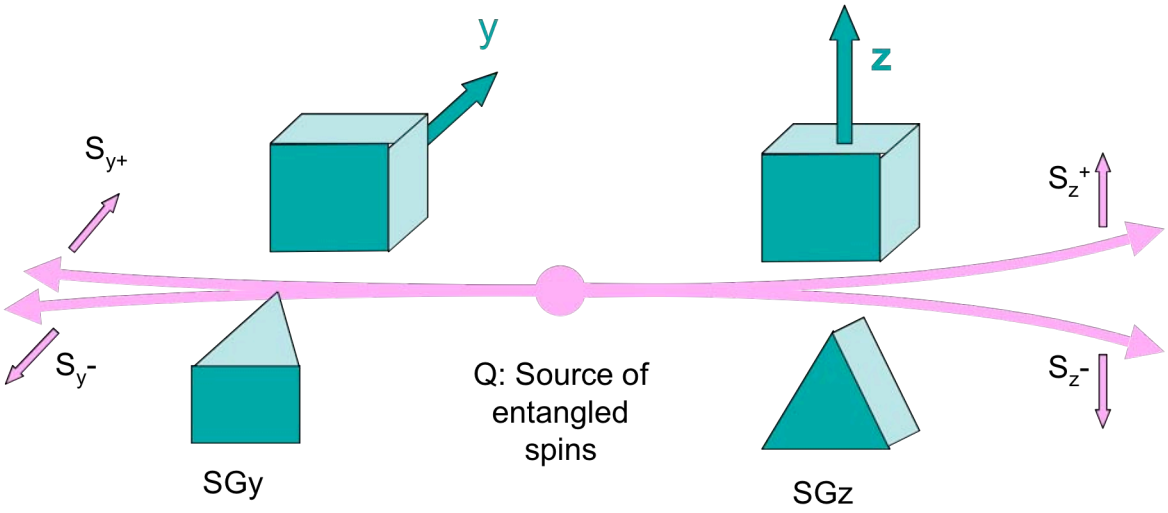


Fig. 2: *The EPR thought experiment.*

(3) The statement given under (2) is true for the statistical average. For the individual event a photon due to its quantum nature is either completely absorbed or completely transmitted.
(4) In the case of transmission through a polarizer the photon acquires the direction of *E* given by the polarizer.

2.2 Distinction between local realism and quantum theory: Bell's inequality

We follow a lecture presented in the internet [1] and discuss now the experimental situation displayed in Fig. 3. Suppose there is a source of entangled photons Q characterized further by the assumption that the total angular momentum is zero. Then at the source the two outgoing photons must have opposite linear momenta and the same polarization. However due to the rotational symmetry of the experimental setup around the optical axis the polarization is undetermined. On the other hand due to statement (4) photons (ph1) after passage of P1 will have polarization α and photons ph2 after transmission through P2 will have polarization β. Two scenarios have been discussed for what happens to the photons between the source and the polarizers P1 and P2.

In the classical picture also called „local realism" at source Q both photons have some unknown, but well-defined or real polarization. It is assumed that –provided all parameters of influence were known– passage of a photon at the analyzer would be predictable and reproducible. As will be seen below, this situation could be characterized by an inequality relation named after John Bell.
In the quantum mechanical picture the unknown angle α is not assumed to exist in the traditional sense. It is not part of reality. Then what kind of prediction can we make? If we assume that ph1 arrives at P1 earlier than ph2 at P2, then ph1 after passage has polarization α. Since ph1 and ph2 are entangled, ph1 immediately also acquires α but after passage of P2 acquires β. This is the situation assumed in Fig. 3, hence Eqs. (1a) and (1b) should be valid.
A further mathematical criterion to distinguish between the two cases was suggested by Bell. He found that „local realism" leads to an inequality relation, which can be –but doesn't have to be– violated by the quantum picture. However, in case of violation "local realism" is led *ad absurdum* and the quantum picture is true.

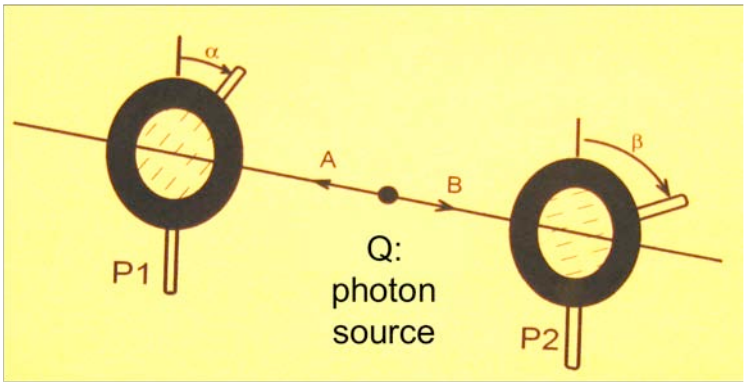


Fig. 3: *The EPR thought experiment with photons.*

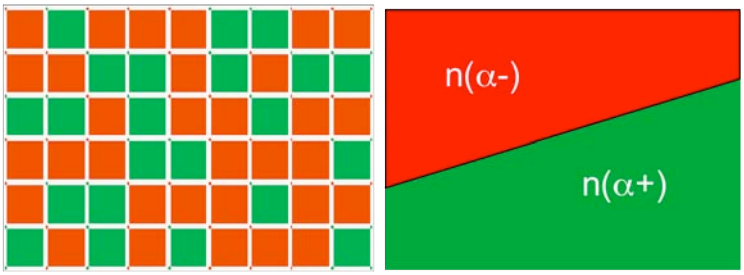


Fig. 4: *Local realism: Depending on the value of the hidden parameter (here the color) a photon is transmitted (green) or not (red).*

For a plausibility argumentation we consider first only a single analyzer set at a fixed angle α and photons impinging on it. If for the transmitted or reflected photons a statistical distribution is found, then Einstein would ascribe this to a hidden parameter attaining different values. The relation between parameter and transmission is still believed to be causal. Suppose we knew the parameter we could then arrange its values in the form of a chart as in Fig. 4, where the color of each little square stands for a certain value of the hidden parameter. Without loss in generality we can arrange the squares such that a continuous area for each color is obtained. After laying two such charts upon each other we obtain intersections as in the lower part of Fig. 5. and we consider what they represent.
For example the red area $n(\alpha+,\beta+)$ represents the number of events that in the case shown in the upper part of Fig. 5 one photon passes at P1 set at α and another at P2 set at β. This would be a coincidence because they are generated by the source at the same time. Likewise $n(\alpha+,\gamma+)$ counts the events that one photon passes P1 set at α and P2 set at γ. Finally, the red area $n(\alpha+,\gamma-)$ corresponds to transmission of a photon at one of the polarizers set at α and blockage at the other set at γ. Comparison of the red areas demonstrates Bell's inequality relation

$$n(\alpha+,\beta+) \leq n(\alpha+,\gamma+) + n(\beta+,\gamma-) \qquad (2)$$

Quantum mechanics has led us to Eqs. (1). Lets see if they fulfill Bell's relation (2). From

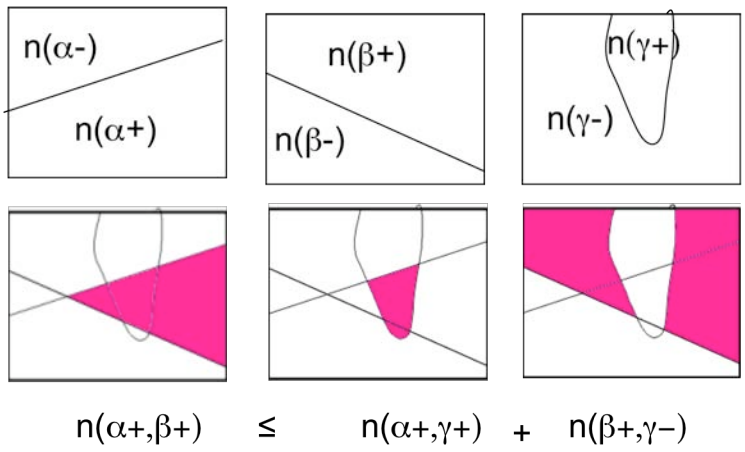


Fig. 5: *Presentation of Bell's inequality by means of areas.*

Eqs. (1) follows

$$n(\alpha+,\beta+) = \cos^2(\alpha-\beta)$$
$$n(\alpha+,\gamma+) = \cos^2(\alpha-\gamma)$$
$$n(\beta+,\gamma-) = \sin^2(\alpha-\gamma)$$

hence

$$\cos^2(\alpha-\beta) \leq \cos^2(\alpha-\gamma) + \sin^2(\beta-\gamma). \qquad (3)$$

For most values of α,β, and γ Eq. (3) is fulfilled but not for α=0°, β=30°, and γ=60° because 3/4 ≤1/4+1/4 is not true. Hence, the quantum-mechanical prediction here is at variance with local realism. This statement is supported by most experiments.

3 Spin polarization in 3d-metals and its relation to band structure

In magnetic solids the individual spins add up to the total magnetization, where we have to include generally also orbital contributions. In 3d-metals the latter are known to be quenched, so we consider only the spins. In an itinerant picture we obtain the magnetization simply by adding up the moments of the occupied states, *i.e.* those below the Fermi level. These are the states of the 3d bands, with some small contribution from 5s bands. This is illustrated in Fig. 6, right-hand side.

In the rigid band model it is assumed that the displayed band structure is essentially the same for all 3d metals and the up-shift of the spin-down band relative to spin-up band is also constant. Magnetization M then is proportional to the hatched area and it is easy to see that due to the up-shift of E_F for increasing number of valence electrons, M should first increase and then after a maximum again decrease. Experimental data as collected by Slater and Pauling and displayed in Fig. 6 indeed show this dependence.

The effect of the spatial quantization on the magnetization M is twofold. First in adding up the spin moments to the total magnetization we consider only up and down spins, which corresponds to the two split beams in Fig. 1. Second the total moment is also subject to a

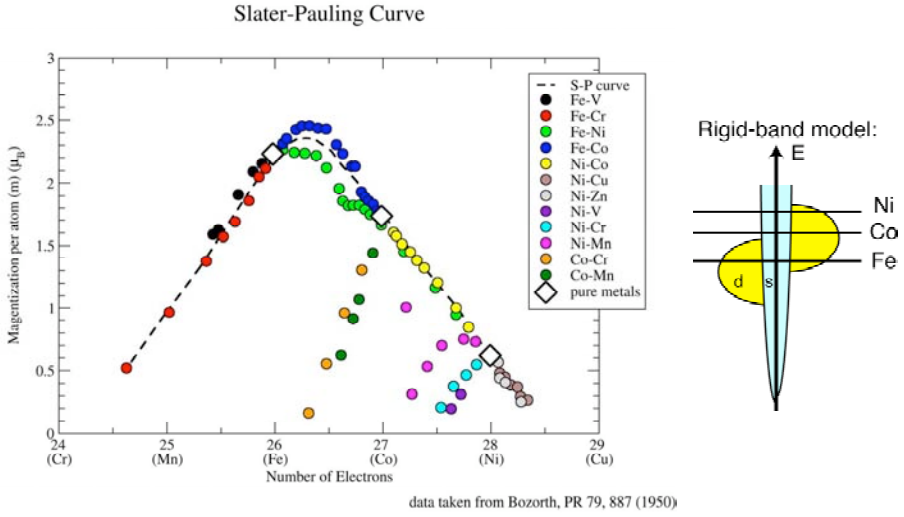


Fig. 6: *The Slater-Pauling curve (left) is explained by the rigid-band model (right).*

quantization effect. It can only change corresponding to an "umklapp" of one or more spins. However that doesn't have to be at the same site but can be distributed over a whole chain of moments in the form of an excitation. In this picture the "umklapp" of one spin corresponds to the excitation of one magnon.
For transport phenomena only the electrons with energies close to E_F are important. Again, their spins can be up or down with respect to M. This is quantified by means of the effective spin polarization

$$P = (n\uparrow - n\downarrow) / (n\uparrow + n\downarrow), \qquad (4)$$

where $n\uparrow$ and $n\downarrow$ are the number of electrons at E_F with spin up and spin down, respectively. Depending on whether $(n\uparrow > n\downarrow)$ or $(n\uparrow < n\downarrow)$ the polarization P will be positive or negative. From the above consideration within the rigid band model it follows that $P(E_F)$ is positive for the ascending sections of the Slater-Pauling curve and negative for the descending.

4 Interlayer Exchange Coupling

4.1 Discovery

The exchange coupling of magnetic films across metallic interlayers was first observed in 1986 for dysprosium and gadolinium films separated by yttrium interlayers, and for iron films separated by chromium interlayers (Salamon *et al.* 1986, Majkrzak *et al.* 1986, Grünberg *et al.* 1986). For references and a review on interlayer coupling see [2]. Due to the work of various groups it could later be shown that the coupling is not restricted to metallic interlayers but can also be observed across MgO and Si.
Three types of coupling, namely ferromagnetic, antiferromagnetic, and 90°-type have been observed. In Fig. 7 this is demonstrated by means of magnetic domains in Fe/Cr/Fe samples with "wedge-type" interlayers (see bottom part for Fig. 7). The upper part shows magnetic domains as observed by a Kerr microscope. In the middle part we see the corresponding alignments of the magnetizations in the upper and lower Fe film, which are evaluated from the contrasts of the experimental image in the top part. Since 90°-type coupling is believed to be extrinsic, but due to interface roughness or magnetic impurities, we discuss here only ferromagnetic or antiferromagnetic coupling.

4.2 Phenomenological description

A phenomenological description of the coupling in order to link experimental observations by different methods is given by the interlayer coupling energy per unit area, E_{coupl}, as

$$E_{coupl} = -J_1\cos(\theta) - J_2\cos^2(\theta), \qquad (5)$$

where θ is the angle between the magnetizations of the films on both sides of the spacer layer. The parameters describe the type and the strength of the coupling. If the term with J_1 dominates, then from the minima of Eq. (5) the coupling is ferromagnetic (antiferromagnetic) for positive (negative) J_1. If the term with J_2 dominates and is negative, we obtain 90°-type coupling. The first term of Eq. (5) is often called bilinear coupling, and the second biquadratic coupling. There are various methods to measure the parameters J_1 and J_2 and, thus, the coupling. We mention only measurement of spin wave frequencies by microwave absorption

or inelastic light scattering and remagnetization curves by means of the magneto-optic Kerr effect or SQUID.

For metallic interlayers the coupling generally oscillates between ferromagnetic and antiferromagnetic as a function of the interlayer thickness. This is demonstrated by the black and white stripes in Fig. 8. Here again a sample with a wedge-type interlayer has been used. The stripes indicate magnetic domains with opposite magnetization direction for black and

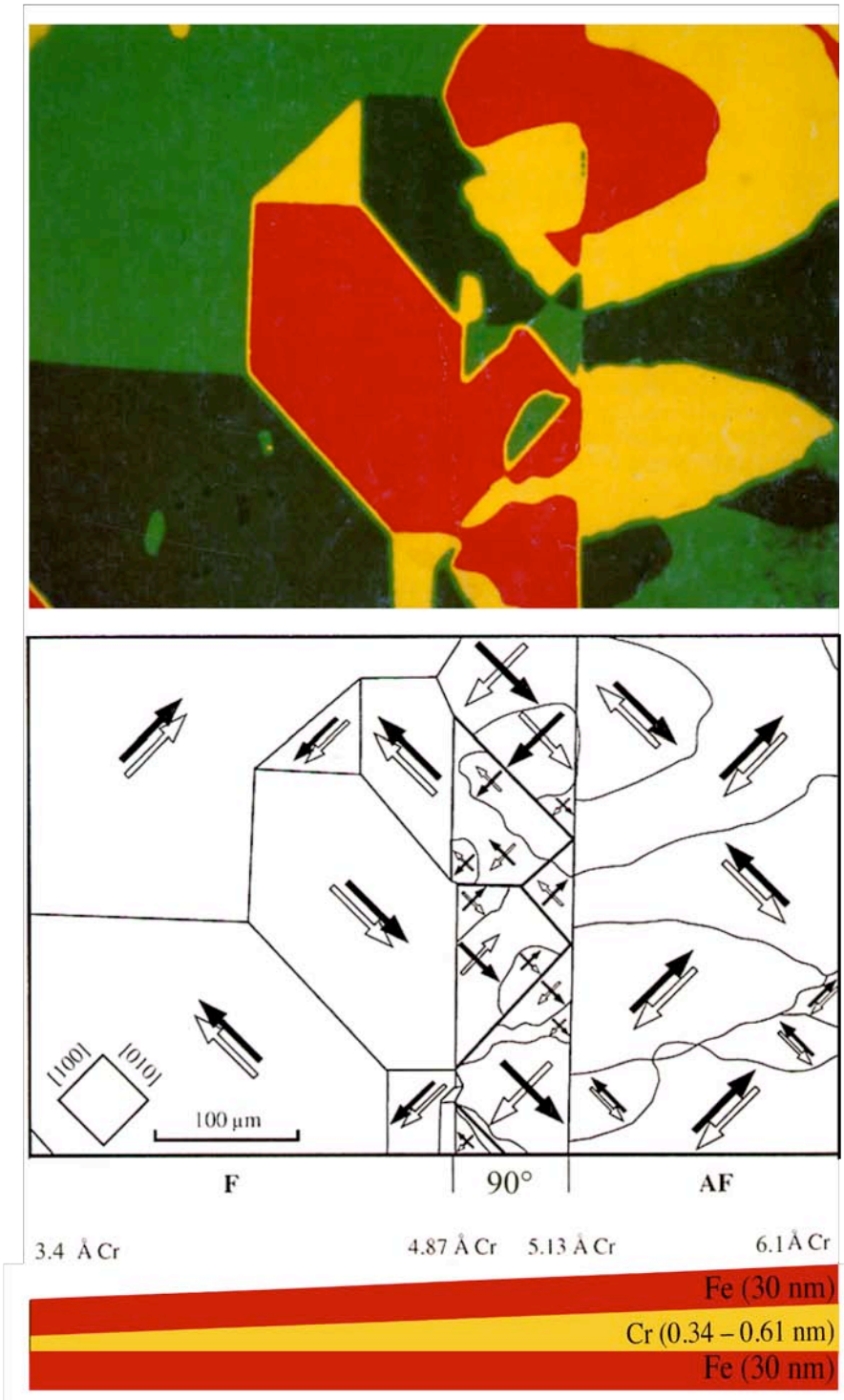


Fig. 7: Magnetic domains of a Fe/Cr/Fe trilayer as observed by Kerr microscopy and their dependence on interlayer coupling, which is varying along the Cr wedge.

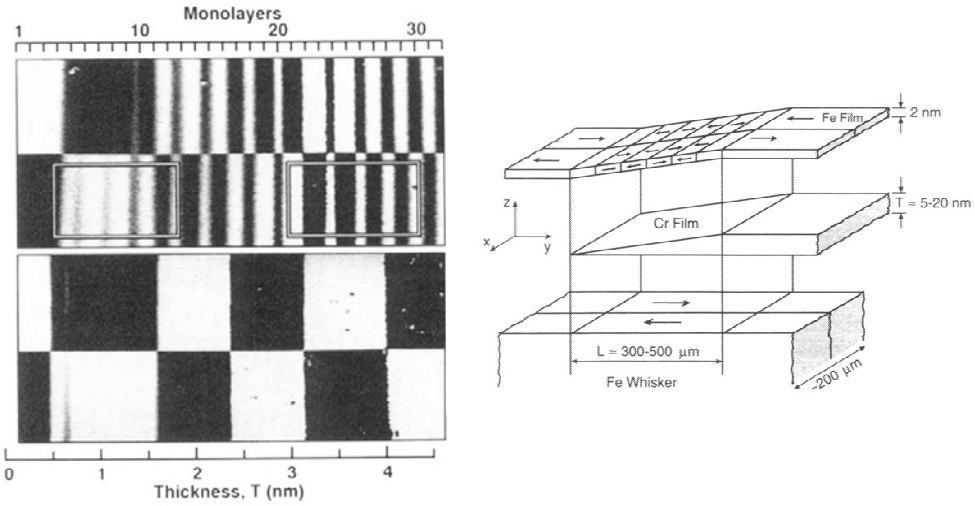


Fig. 8: *Oscillatory coupling as observed by scanning electron microscopy with spin analysis (SEMPA) in Fe/C-wedge/Fe structures. The diagram on the right-hand side schematically indicates the domain configuration.*

white, which come as a result of the coupling, thus demonstrating the oscillatory behavior. Each pattern is seen twice but with reversed contrast. This corresponds to two domains in the Fe substrate beneath. Note that in the upper part the period is much shorter (2 monolayers) than in the lower. This is due to the better quality of the sample in the upper part.
Whereas there are numerous examples for oscillatory coupling across metallic interlayers, examples for antiferromagnetic coupling across insulators or semiconductors are still rare. In fact up to now only Si and MgO interlayers are established examples (for a review see [3]). On the other hand antiferromagnetic coupling across Si turns out to be surprisingly strong. In terms of the parameters defined in Eq. (5) we have for example J_1 = -1.6 mJ/m^2 for Cr, which is already considered to be strong, but J_1 = -6.3 mJ/m^2 for Si interlayers.

4.3 Microscopic origin of oscillatory coupling across metallic interlayers

The basic assumption to explain oscillatory coupling within the quantum well approach is spin-dependent reflectivity of electrons at the non-magnetic/magnetic interfaces. In Fig. 9 left-hand side strong reflectivity at the interfaces is assumed for those electrons, which have their spins opposite to the local magnetization, and weak reflectivity for the others. The reason for such behavior is indicated in the top right part of Fig. 9 using schematic band structures for the magnetic 3d transition metals and noble metals as examples. For reasons that will become clearer below, we can restrict ourselves to electrons at the Fermi level. Then, for the chosen example, for spin-up electrons we have s-states at the Fermi energy for both the 3d transition metals and the noble metals, as is indicated by the similar densities of states on the right-hand side. This leads to a good transmission. For spin down electrons on the other hand, due to the splitting of the energies in the magnetic films we have mixed d- and s-states at the Fermi level for the transition metal, hence the transmission of the electrons from the noble metal is reduced. Based on the spin-dependent reflectivity there is a strong (weak) confinement in the interlayer for spin down (up) for parallel magnetization alignment (left-hand side of Fig. 9), whereas for the antiparallel alignment the reflectivities are as shown on the right-hand side and the confinement is lost.

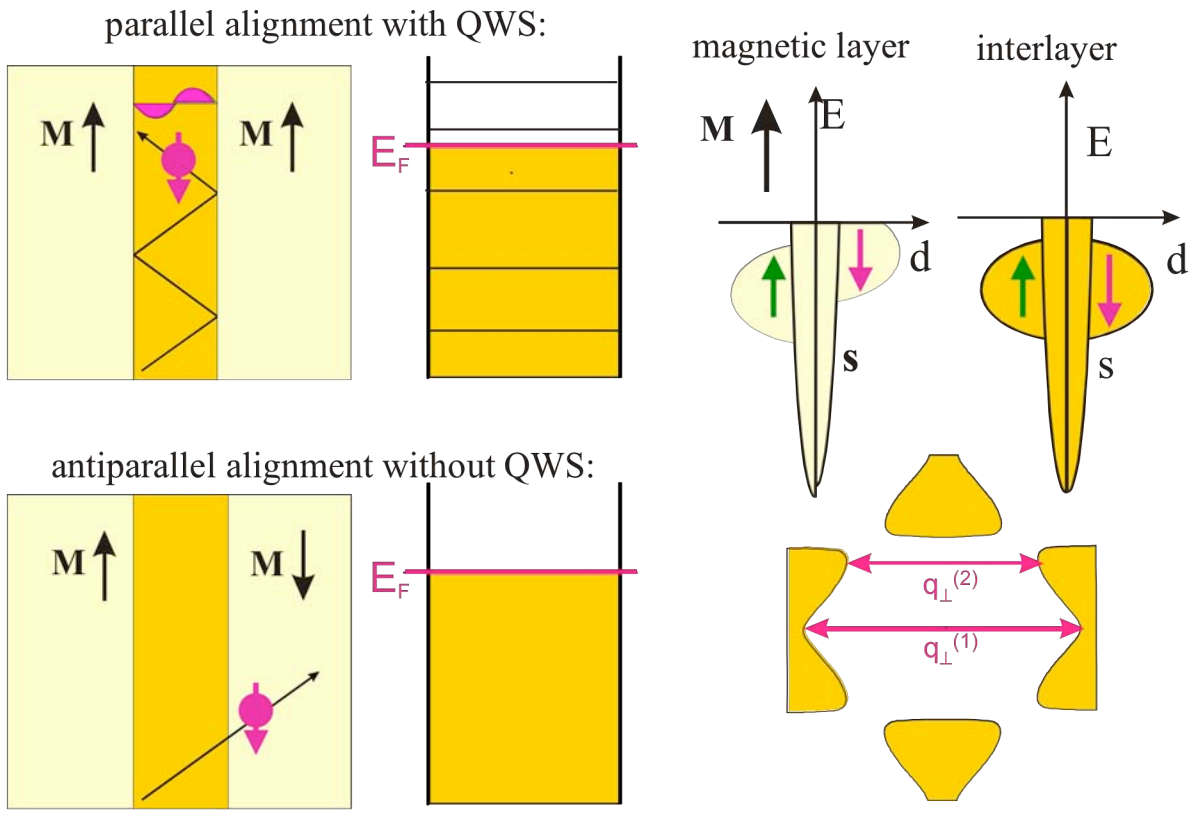


Fig. 9: *Spin-dependent reflectivity leads to the formation of quantum-well states (QWS) for parallel alignment but not for antiparallel alignment and is the main mechanism leading to oscillatory interlayer exchange coupling.*

Hence, parallel magnetization alignment is characterized by a confinement of part of the electrons and there are characteristic energy changes associated with this, which makes this situation for certain interlayer thickness more and for others less favorable than antiparallel alignment, for which case the confinement is lost. Due to the confinement the motion of the electrons perpendicular to the interfaces becomes quantized and we obtain a spectrum of discrete energy levels corresponding to the formation of standing electron waves. Such a standing wave is indicated in Fig. 9 in the top left part and is the result of the superposition of two propagating waves with wave-vector components $\pm q_\perp$. To form a standing wave we must have

$$|2q_\perp| = n\, 2\pi/D \qquad n = 1, 2, 3, \ldots$$

where D is the interlayer thickness. When the interlayer thickness is increased the discrete levels shift downwards and are populated upon crossing the Fermi energy E_F. Hence it is plausible that there are oscillations of the electronic energy due to the fact that discrete energy levels become populated. It turns out that these oscillations favor parallel alignment for certain thicknesses and antiparallel alignment for others. Hence the interlayer coupling oscillates as a function of the interlayer thickness D due to oscillations in the electronic energy. The oscillation period λ_D is given by the difference in D, where two subsequent discrete energy levels cross the Fermi energy, hence $\lambda_D = 2\pi/|2q_\perp|$, where $q_\perp$ is a caliper of the Fermi surface.

Hence oscillatory coupling can be traced back to changes in the densities of states, which come as a result of confinement. The stronger the confinement and the higher the changes in the density of states the larger will be the associated amplitudes. At this point it is important to consider that not all possible wave-vectors $q_\perp$ from the Fermi surface contribute in the same way. Some are associated with higher densities of states than others and therefore contribute more in the above consideration. Let us discuss this situation for our example of noble metal interlayers and choose the [100] orientation for the normal to the interface. The cross-section through the Fermi surface in Fig. 9 (bottom right) reveals two wave-vectors $q_\perp^{(1,2)}$ along the [100] direction. They are associated with particularly high densities of states, because their lengths change only very little upon a slight shift up or down. They are called stationary vectors and generate two superimposed oscillations of the coupling as a function of the interlayer thickness. In this way for given growth direction the oscillation periods can be predicted from the crystallographic structure and the Brillouin zone. In a similar way a dependence on the thickness of the layer can also be considered and has indeed been observed.

4.4 Simple model to explain coupling across insulators and semiconductors

For interlayers of this kind transmission is suppressed for larger interlayer thickness D and, therefore, also the oscillations discussed in the previous paragraph are replaced by a simple attenuation. For small D, however, electrons can be transmitted and again we consider their spin polarization. As a result of spin-dependent reflectivity, scattering, transmission, *etc*. the net flow of electrons can be stronger for spin-up (spin-down) electrons and it is easy to see that this favors ferromagnetic (antiferromagnetic) coupling.

5 Giant magnetoresistance (GMR)

5.1 First observations

Figure 10 shows the first observations of GMR in multilayers (a) and double layers (b). References to the original publications and review articles can be found in [4,5]. The current flows in both cases in the plane of the layers. Due to antiferromagnetic interlayer coupling the magnetization alignment of the neighboring Fe films in small fields is antiparallel and the resistance is high. Increasing the field aligns the magnetizations parallel and the resistance drops. In multilayers (a) the effect is much stronger than in double layers (b), which is an indication that the number of available interfaces plays an important role. The term *giant* magnetoresistance referred originally to the large size of the effect in multilayers, but is now generally used for magnetoresistance due to non-parallel magnetization alignment. The largest effect occurs when an antiparallel alignment by an applied field is changed into a parallel alignment. The antiparallel alignment can be provided by antiferromagnetic interlayer exchange as in Fig. 10 or by other means (see below). Oscillatory coupling also gives rise to oscillations of the GMR effect as a function of the spacer thickness.

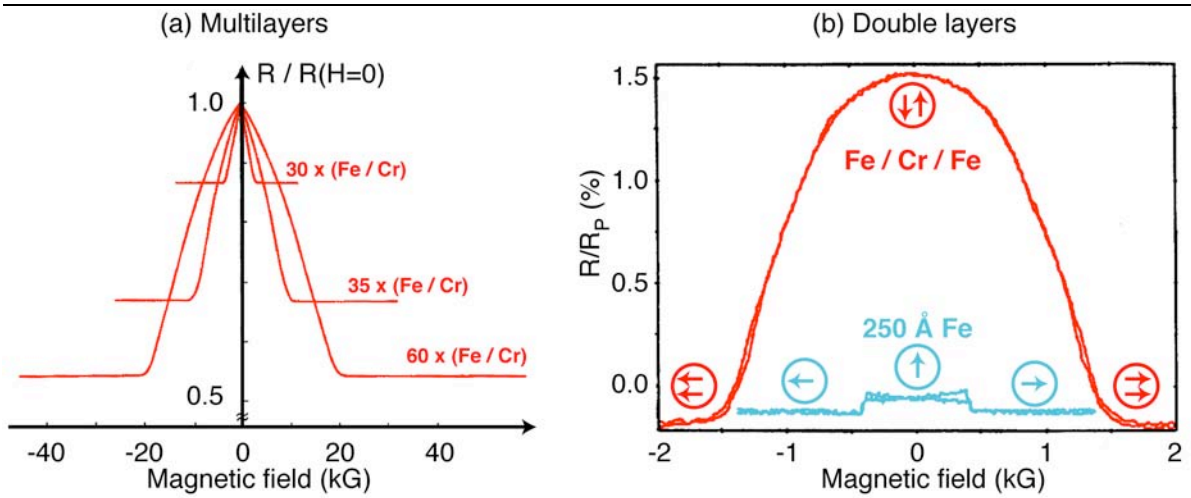


Fig. 10: First observations of *GMR in multilayers (a) and double layers (b). In both cases the antiparallel alignment at zero field is due to the antiferromagnetic coupling of the Fe/Cr system.*

Antiparallel arrangement can also be obtained by other means, for example by different coercivities of successive magnetic layers or by pinning the magnetization using an antiferromagnetic material in direct contact with one of the ferromagnetic layers, known as "exchange biasing". If GMR is obtained *via* one of these methods and not *via* antiferromagnetic interlayer coupling, usually the term "spin valve system" is used in the literature although there is no difference concerning the mechanism of the magnetoresistive effect.

The GMR effect has been investigated in two different geometries, namely the "CIP" (current in plane) and the "CPP" (current perpendicular plane) geometry. The relative effect is stronger in the CPP geometry, but due to the fact that the contact diameter is some orders of magnitude larger than the film thickness the voltage drop perpendicular to the layers, in the CPP geometry, is very difficult to detect, unless the contact diameter is lithographically reduced to sub-μ dimensions.

Apart from the normal GMR effect, where the resistivity is largest for the antiparallel magnetic state, there is also an inverse effect, where it is largest for the parallel state. The latter only occurs for asymmetric systems when the two involved ferromagnetic materials sit on different slopes of the Slater-Pauling curve shown in Fig. 6 (see below).

5.2 Microscopic origin

In Section 4.3 it was seen that interlayer coupling can be explained on the basis of a spin-dependent interface reflectivity. Similarly GMR can be explained to be due to spin-dependent scattering.

To see the origin of the normal and the inverse GMR effect we discuss Fig. 11, where for both cases double layers in parallel and antiparallel alignment are displayed. For both kinds of spins paths between two reflections at outer surfaces are shown, with scattering events in between, which is assumed to be representative. In order not to confuse the picture the changes in direction due to the scattering events are suppressed. The scattering processes are the cause of electrical resistivity. Usually it is believed that conduction electrons are s electrons, but their scattering rates are given by the density of d states with the same energy, *i.e.* E_F, and the same spin. Hence some idea on $n\uparrow$ and $n\downarrow$ can be obtained from the right-hand side of Fig. 6: $n\uparrow > n\downarrow$ for the ascending parts of the Slater-Pauling curves and $n\uparrow < n\downarrow$ for the

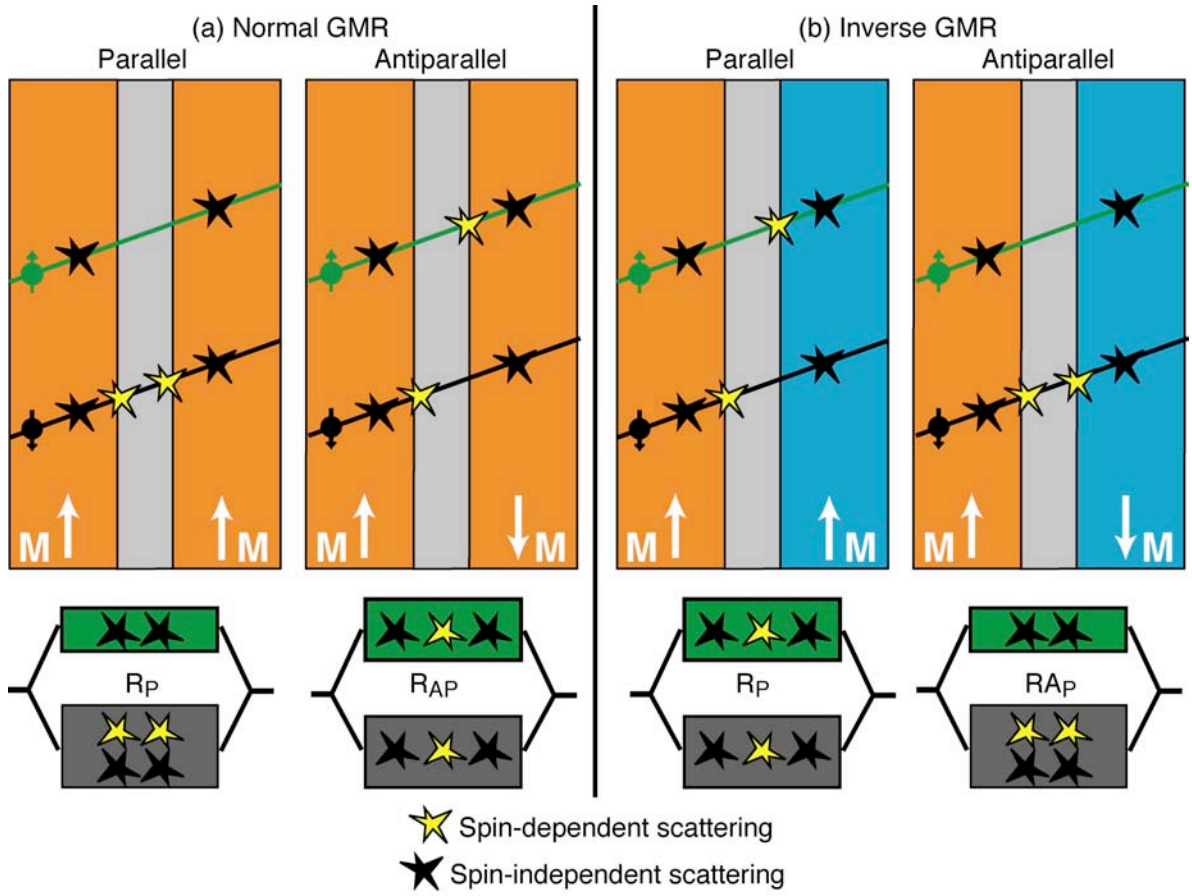


Fig. 11: *Phenomenological picture for normal (a) and inverse (b) GMR.*

descending parts. Therefore, the spin polarization at the Fermi level P according to Eq. (4) is positive for the ascending and negative for the descending part. We treat interfaces by taking the values for the corresponding alloys.

In order to explain the normal and the inverse case a magnetic double layer in parallel and antiparallel alignment is displayed in Fig. 11. For the normal case same materials are used on both sides of the interlayer for the inverse effect they are different. But how?

We use the simple consideration displayed in Fig. 11. Here we neglect scattering inside the interlayer and there are equal rates of spin-dependent and spin-independent scattering in the ferromagnetic films. The spin-dependent scattering has been assumed to take place at the interfaces, where for the normal GMR effect displayed in Fig. 11(a) only electrons with spin antiparallel to the local magnetization are assumed to be scattered. According to the two-current model invented by Mott the total current can be divided into two currents flowing in parallel. One with spin-up (I_+) and one with spin-down (I_-) electrons. If we assume that one scattering event contributes to the total resistance by an amount r then on the left-hand side I_+ is associated with resistance $2r$ and I_- with resistance $4r$. Hence for parallel alignment the total current $I = I_+ + I_-$ has resistance $R_p = 2r \times 4r/(2r + 4r) = 8r/6$. In the same way we obtain for antiparallel alignment for the total current a resistance $R_{ap} = 3r \times 3r/(3r + 3r) = 9r/6$. Hence, there is an increase of the resistance due to antiparallel magnetization alignment and the GMR ratio is $(R_{ap} - R_p)/R_p$ in the present case would have a value of 12.5%. For double layer systems the strongest measured GMR effects are around 17%, hence a distribution of the scattering rates as in Fig. 11 seems realistic.

Generally an inverse GMR effect [Fig. 11(b)] occurs when the materials on both sides of the interlayer have different signs of the spin polarization *P* [see Eq. (4) and Fig. 6]. Normal GMR occurs when they have the same sign, no matter whether negative or positive.

6 Conclusion

In this lecture various aspects of spin polarization have been discussed. We started with the experiment by which spin polarization and its spatial quantization has for the first time been observed. It is clear that for any physical quantity, where the spin is important, its quantization is important, too. We discussed as examples the magnetization of ferromagnetic 3d metals and alloys as well as two important effects in layered magnetic structures, namely interlayer exchange coupling and giant magnetoresistance. The role of polarization (of photons or electrons) for the development of quantum mechanics was also addressed. Implications thereof will possibly lead to applications in quantum information, which is also subject of this school.

References

[1] www.ikg.rt.bw.schule.de
[2] D.E. Bürgler, P. Grünberg, S.O. Demokritov, and M.T. Johnson, *Handbook of Magnetic Materials*, Vol. 13, ed. by K.H.J. Buschow, Chapter “Interlayer Exchange Coupling in Layered Magnetic Structures” (Elsevier Amsterdam, 2001).
[3] P. Grünberg, D.E. Bürgler, H. Dassow, A.D. Rata, and C.M. Schneider, Acta Materialia **55**, 1171 (2007).
[4] A. Barthélémy, A. Fert, and F. Petroff, *Handbook of Magnetic Materials*, Vol. 12, ed. by K.H.J. Buschow, Chapter “Giant Magnetoresistance in Magnetic Multilayers” (Elsevier Amsterdam, 1999).
[5] P. Grünberg, Acta Materialia **48**, 239 (2000).

B 2 Theory of tunneling magneto resistance

Daniel Wortmann
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Theory of TMR **3**
2.1 Tunneling through a one-dimensional rectangular barrier 3
2.2 The Julliere model of TMR 5
2.3 Landauer transport 6
2.4 Interpretation of the Landauer formula 8
2.5 The Bardeen approach to tunneling 10

3 TMR in crystalline MTJs: Fe/MgO/Fe **13**
3.1 Basic electronic properties of the Fe/MgO interface 14
3.2 MgO as a tunneling barrier 15
3.3 Electronic tunneling at normal incidence 16
3.4 TMR of the Fe/MgO/Fe MTJ 17
3.5 TMR in more complex oxides 19

4 Summary **20**

1 Introduction

Spin-polarized electric conductance is the very heart of the whole field of spintronics. Nearly all physical phenomena and proposed and realized devices rely on the transport of spin-polarized electrons and hence the spin-polarized electron transport is one of the most basic phenomena to study. The most well-known effect directly linked to spin-polarized transport is the giant magnetoresistance (GMR) effect discovered by Grünberg [1] and Fert[2]. The magnetoresistance(MR) is the actual key figure of merit in the GMR as well as in other similar effects. It describes the ratio between the resistance of two different magnetic states of the system, mostly between an parallel(P) and antiparallel(AP) alignment of magnetic leads as

$$MR = \frac{R_P - R_{AP}}{\min(R_P, R_{AP})}. \tag{1}$$

This is the so called optimist definition of magnetoresistance. An alternative definition in which one divides by the sum $R_{AP} + R_P$ can also be found, leading to a maximum MR of 100% while the optimistic definition is not bounded. Besides the GMR, many further incarnations of magnetoresistance effects have been discovered and utilized:

- The Anisotropic MR (AMR) is actually known for a long time and describes the change of resistance induced by a change of the direction of the magnetization. The spin-orbit interaction, coupling the spin degrees of freedom of the conducting electrons to the lattice orientation, leads to differences in the resistance depending on the orientation of the magnetization with respect to the underlying crystal lattice. This effect is usually considered to be a bulk effect.

- The Colossal MR (CMR) effect in which a very large resistance change occurs after application of a magnetic field in some oxides which are close to a phase transition.

- The Tunneling MR (TMR) effect which will be the main topic of this lecture.

- MR effects which combine features of these most basic effects in nanostructures such as the ballistic MR (BMR) or the ballistic anisotropic (tunneling) MR.

This lecture considers the TMR effect in more detail. The basic setup to consider for this effect is a nanosize magnetic tunneljunction (MTJ) in which a thin barrier layer of an insulator separates two magnetic metal leads. If the insulator is sufficiently thin, electrons can tunnel through the insulator enabling the flow of a current between the two leads. As both leads are magnetic this current depends on the relative orientation of the magnetization of the leads and a MR can be measured.
The first realization of such a TMR setup was actually obtained by Julliere [3] in 1975 at very low temperatures. TMR at room temperature was achieved much later in the 90th by Moodera [4] and by Miyazaki [5]. These experiments used MTJ with amorphous barrier materials like Al_2O_3 and relatively low TMR ratios below 100% were obtained. In 2004 two groups in Japan (S. Yuasa et al. [6]) and in the US (S. Parkin et al. [7]) managed to fabricate epitaxially grown Fe/MgO/Fe MTJs with much larger TMR values. The development of such MTJs with MgO barrier still continues today with TMR values exceeding 1000%.
Different from GMR which has proven to by applicable mostly to build sensors, TMR based devices seem to have an even larger field of potential application. In particular, the high MR values possible and the low currents in typical tunneljunctions make the TMR effect the key

ingredients in many proposed new devices of magnetoelectronic. The most prominent of those is probably the magnetic random access memory (MRAM). In MRAM-cells the information is stored in the relative orientation of the magnetization of the two ferromagnetic metallic sides of the junctions and it can be accessed by using the tunnel-magneto resistance (TMR) effect, i.e. the strong dependence of the resistance of the junction on this relative orientation.

2 Theory of TMR

Calculating the electronic current flowing in a system due to some bias voltage is one of the most difficult problems in theoretical solid state physics. Different levels of approximations have been applied to the problem and theoretical models for many different aspects of the problem have been discussed. Among these, one of the most well known method to treat electronic transport in solids is based on the Boltzmann formalism, which describes the time-dependent change in the electronic distribution function due to the applied field $\vec{E}$. It is a classical approximation as it considers the electrons to be moving freely between individual scattering events. Interference effects due to the quantum nature of the electrons are neglected. On the other hand, the very fact that the electrons move in a solid without scattering at every atomic site, i.e. the possibility to describe the electrons as (quasi) particles moving freely around is of course a quantum mechanical phenomenon.
The approximations used in the Boltzmann approach restrict its applicability to the case in which the sample dimensions are much larger than the mean-free path and at the same time the scattering events can be viewed as independent from each other, i.e. quantum interference effects can be neglected. A more basic approach to the problem of electric conductance uses the idea of treating the current as a response of the quantum system to the applied electric field. In its most rigorous formulation this idea can be used in the linear response limit in which one assumes that the current depends linearly on the field. Using the corresponding time-dependent quantum mechanical formulation one obtains the so called Kubo formalism which can be used to obtain the conductance of the system. While this approach is very general and would allow to include all kinds of scattering, it is not easy to apply to realistic systems.
In the following, we will concentrate on calculating the electronic transport in tunnel junctions and hence adopt a rather simplified quantum mechanical point of view. In particular, we will consider systems with typical dimensions much smaller than the mean-free path due to scattering at structural impurities, by interface roughness, by phonons, magnons or other temperature dependent excitations present in real experiments. In the nanosize setups we consider, the resistance is due to the scattering of the electrons on the potential of the insulating barrier.
This lecture will cover two main subjects. First, we will discuss the basic models to describe electron transport in a single particle picture. After a brief reminder of the basic phenomenon of quantum mechanical tunneling, and the introduction of the phenomenological description of Julliere, this part will cover the famous Landauer approach to ballistic transport (Sec. 2.3) as well as a more specialized approach to the tunneling problem – Bardeen's approach based on perturbation theory (Sec. 2.5).

2.1 Tunneling through a one-dimensional rectangular barrier

Of course the quantum mechanical tunneling effect is a very basic phenomenon discussed in every introductory course of quantum mechanics. In brief it describes the fact that in contrast

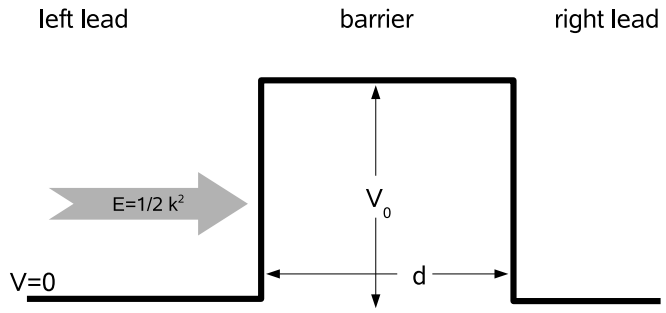


Figure 1: *Simple one-dimensional model for quantum mechanical tunneling. A electron of energy $E = \frac{\hbar^2}{2m}k^2$ is incident to a rectangular barrier potential.*

to the classical mechanics which prohibits a particle to enter any area in which the potential level surpasses the particles energy, quantum mechanics assigns a finite non-zero probability find the particle in such areas. Mathematically, this is reflected by the fact that the wavefunction is non-zero in such areas of a repulsive potential. To elucidate this effect a little more, let us consider the probably simplest model for electronic tunneling.
Fig. 1 shows the setup chosen for this simple model, a rectangular barrier of height V_0 and width d between leads in which the electrons are described by free electron wavefunctions. In this system it is an trivial problem to construct the wavefunction as

$$\psi(x) = \begin{cases} \exp(ikx) + r\,\exp(-ikx) & x < 0 \text{ in left region} \\ a\,\exp(-\kappa x) + b\,\exp(\kappa x) & 0 < x < d \text{ in the barrier} \\ t\,\exp(ikx) & x > d \text{ in right region.} \end{cases} \tag{2}$$

The decay constant is given by $\kappa = \sqrt{\frac{2m}{\hbar^2}V_0 - k^2}$, the coefficients a, b and r, t can be determined by wavefunction matching, i.e. by the requirement that the wavefunction and its derivative are continuous at $x = 0$ and $x = d$. Simple algebra reveals the well know formula

$$t = \frac{4ik\kappa\, e^{-ikd}}{(ik+\kappa)^2 e^{-\kappa d} + (k+i\kappa)^2 e^{\kappa d}}. \tag{3}$$

For the case of a sufficiently thick and/or high barrier, i.e. large d and/or large κ this expression for t can be simplified by neglecting higher order terms in $e^{-\kappa d}$ to

$$t \sim \frac{4ik\kappa\, e^{-ikd}}{(k+i\kappa)^2} e^{-\kappa d}.$$

The wavefunction now actually leads to an electric current flowing across the barrier. Applying the quantum mechanical current operator one obtains for the current density (which can be most simple evaluated in the right electrode region but is of course conserved in all space)

$$j(x) \propto \frac{1}{2i}\left(\psi^*(x)\partial_x\psi(x) - \partial_x\psi^*(x)\psi(x)\right) \propto t^2. \tag{4}$$

Hence we find that there is a finite electric current flowing through the barrier which is proportional to the square of the so called transmission amplitude t, a quantity that can be interpreted as a transmission probability. We should note that we only considered the proportionality here as the actual value of the current will of course depend on the normalization of the wavefunction, an issue that will be re-occur in Sec. 2.3.

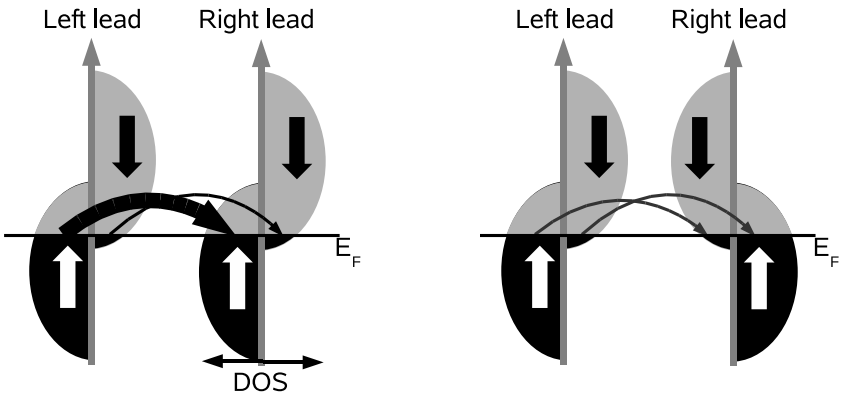


Figure 2: *Julliere model of spin-polarized tunneling. The spin-polarized density of states(DOS) of the parallel(left) and antiparallel(right) MTJ is sketched. The tunneling current is indicated by curved arrows at the Fermi energy E_F. In the parallel situation the large DOS at both sides of the junction in the $\uparrow$-spin leads to a large current while all other currents are small due to the small number of available states.*

2.2 The Julliere model of TMR

In a magnetic tunneljunction the additional spin-degree of freedom of the electrons has to be included in the description of the transport process. In a ferromagnetic metal electrons of different spin exhibit a different electronic bandstructure, their wavefunctions and transport properties differ. As the electronic density of states for the two spins differ, usually also the number of states relevant for transport, i.e. those states at or close to the Fermi level differ for the two spins. In the problem we are interested in here, the electronic tunneling through an insulating barrier, one usually assumes that the spin of the electrons is not altered. Hence, we assume that the electron spin is a conserved quantum number, no scattering processes coupling electrons of different spins are included in our theory. While this approximation is justified in most cases one should be aware of its limitations. As the tunneling process itself is a coupling phenomena with a very low transition rate, the neglected spin-flip scattering can actually become a major effect as it was demonstrated e.g. in the case of surface states in half-metallic MTJs.
If the spin is not changed during the transport process across the MTJ, one can decompose the total current into a spin-up and a spin-down component

$$I = I_\uparrow + I_\downarrow \tag{5}$$

in which the up-spin electrons form $I_\uparrow$ and the down-spin electrons $I_\downarrow$. This two-current model, which can be equivalently expressed as a "two resistor" model in which the junctions is considered as two parallel resistors with $\frac{1}{R} = \frac{1}{R_\uparrow} + \frac{1}{R_\downarrow}$, is an extremely popular and successful concept in spintronics. Starting from this ansatz Julliere [3] constructed a very basic model explaining the TMR effect. His basic assumption was that the current across the junction is proportional to the product of the density of states(DOS) of the two sides (see Fig. 2), i.e. including the two current model one obtains the following expression for the total current

$$I_P \propto n_{L\uparrow} n_{R\uparrow} + n_{L\downarrow} n_{R\downarrow}. \tag{6}$$

This we will assume to be the expression for the parallel current, if we now switch to an antiparallel alignment to of the electrodes, we will have to flip the spins of one side with respect

to the other. If we assign this switching to the right electrode we would obtain the following expression for the current in the antiparallel case

$$I_{AP} \propto n_{L\uparrow} n_{R\downarrow} + n_{L\downarrow} n_{R\uparrow} \tag{7}$$

and consequently we get for the TMR value

$$TMR = \frac{I_P - I_{AP}}{I_P + I_{AP}} = \frac{(n_{L\uparrow} - n_{L\downarrow})(n_{R\uparrow} - n_{R\downarrow})}{(n_{L\uparrow} + n_{L\downarrow})(n_{R\uparrow} + n_{R\downarrow})} = P_L P_R, \tag{8}$$

where P_L and P_R are the spin-polarizations of the density of states

$$P_{L/R} = \frac{(n_{L/R\uparrow} - n_{L/R\downarrow})}{(n_{L/R\uparrow} + n_{L/R\downarrow})} \tag{9}$$

for the left and right electrodes, respectively. This very simple expression, known as Julliere's formula can already explain many basic features of TMR:

- TMR only occurs if both electrodes are magnetic, i.e. if both have a non-vanishing spin-polarization.
- The maximal TMR of 100% is expected if both electrodes are 100% spin-polarized, i.e. if both electrodes behave like half-metals.
- If both electrodes are equivalent, i.e. if $P_L = P_R$, the TMR effect is always positive.

At the same time one can easily spot several shortcomings of this theory of which the most significant is the use of the rather ill-defined densities of states $n_{L/R}$. Obviously, these quantities have to be somehow related to the electronic density of states at or close to the Fermi level as these electrons will carry the electric current. Furthermore, it must be somehow related to the local density of states at the metal/insulator interface since this is the region of space from which tunneling "takes place". However, the exact definition of these quantities is unclear and in consequence the predictive and explanatory power of Julliere's formula is strongly limited.

2.3 Landauer transport

Landauer [8] proposed a theory of the transport process which is well adapted to describe the tunneling transport. A very intuitive and simple derivation will be presented here. The Landauer equation can also be derived more rigorously starting from linear response theory. In the Landauer approach to transport one considers the region Ω – in which the electrons travel ballistically – to be attached to two reservoirs L and R.
The conductance Γ of the region Ω is defined by the current I_{LR} divided by the potential difference between the two reservoirs. The current I_{LR} on the other hand is given by the current due to all electrons traveling from L to R minus the current due to the electrons traveling vice versa

$$I_{LR} = I_{L\to R} - I_{R\to L}. \tag{10}$$

To arrive at an equation for these currents, one can start with a simple one-dimensional model. The current from the left to the right is determined by all electrons leaving the left reservoir, entering the scattering region Ω, and leave this scattering region by passing into the right reservoir. If one now assumes a very simple picture of the region Ω in which its electronic structure

is described by single band in which states with $k > 0$ propagate from the left to the right the current is given by an integral over all states with $k > 0$ up to the Fermi wave-vector k_F

$$I_{L\to R} = \int_0^{k_F} ev(k)dk, \tag{11}$$

where v denotes the group velocity of the state. Since

$$v = \frac{1}{\hbar}\frac{\partial E}{\partial k} \tag{12}$$

and converting the integral over k into an energy integration using the density of states $n(E)$,

$$\begin{aligned} I_{L\to R} &= \int_0^{\mu_L} \frac{e}{\hbar}\frac{\partial E}{\partial k} n(E)\, dE \\ &= \int_0^{\mu_L} \frac{e}{\hbar}\frac{\partial E}{\partial k}\, \frac{1}{\partial E/\partial k}\, \frac{1}{2\pi}\, dE \\ &= \int_0^{\mu_L} \frac{e}{h}\, dE = \frac{e}{h}\,\mu_L\,, \end{aligned} \tag{13}$$

where the energy integration has to be performed over all energies up to the Fermi energy (the chemical potential) of the left reservoir. This can be understood from the requirement that the electrons were assumed to be incoming from the left and therefore must be occupied in the reservoir.
Using the same derivation for the states incoming from the right reservoir one obtains

$$I_{LR} = \frac{e}{h}(\mu_L - \mu_R). \tag{14}$$

Identifying the difference in the chemical potentials μ_L and μ_R with the applied voltage $eV = (\mu_L - \mu_R)$ one obtains the following interesting equation for the conductance

$$\Gamma = \frac{I_{LR}}{V} = \frac{e^2}{h}. \tag{15}$$

This equation is truly remarkable since it states that each conducting band contributes the same to the conductance. Irrespectively of the density of states or the group velocity of the conducting states the conductance is always given by the fundamental quantum of conductance $\frac{e^2}{h}$. Indeed, as Eq. (12) shows, states with a low velocity and therefore a low current $j = ev$ are compensated by their higher density of states such that the conductance remains constant.
In the case of multiple bands, the derivation has to be modified by the inclusion of an extra sum over the different bands. Therefore in the general case of N conducting bands one obtains

$$\Gamma = \frac{e^2}{h}N. \tag{16}$$

The different „bands“ in this discussions are usually called „channels“. The argumentation presented so far did not care about the proper definition of these channels. These were simple assumed to form some kind of „band“ within Ω described by the usual formalism of a wave-vector k and a dispersion relation $E(k)$. Strictly speaking, since the system is not periodic, one cannot speak of Bloch states with some wave-vector having a component k in the direction of the current.

Since a key point in the discussion was the preparation of a state traveling from within the reservoirs through the region Ω one should clarify this idea. For such a state traveling to the right, one might assume the typical scattering problem. Within the left reservoir one considers a wavefunction being a Bloch state propagating towards the region Ω. „Propagate towards“ in this context should be understood as a state having a current flowing towards Ω. Within the reservoirs the resulting scattering state can be written in terms of reflected ψ_r and transmitted ψ_t states which are all solutions of the bulk Schrödinger equation in the reservoirs with the same energy as the incoming state ψ_{in} . The k values of these transmitted and reflected states have to be chosen such that the states „propagate away“ from Ω.

$$\psi(\vec{r}) = \begin{cases} \psi_{\text{in}}(\vec{r}) + \sum_n r_{\text{in},n}\, \psi_r^n(\vec{r}) & \vec{r} \text{ in left reservoir} \\ \sum_{n'} t_{\text{in},n'}\, \psi_t^{n'}(\vec{r}) & \vec{r} \text{ in right reservoir} \end{cases} . \tag{17}$$

Where, the summations can be considered to be performed over all reflected Bloch states or all transmitted Bloch states. In principle, also states decaying away from the interfaces into the reservoirs must be included in this expansion. However, since these do not carry any current and by shifting the interface far enough into the reservoirs one can eliminate these decaying states.
Looking back to the derivation of the Landauer formula, an important change has to be made. While in Eq. (11) and Eq. (12) the summation over the incoming states and the evaluation of the current from their group velocities were all performed within the same single band picture, now one has to distinguish more carefully. The k integration in Eq. (11) has to be performed over the „in“ label of the expansion in Eq. (17). The sum over the velocities on the other hand is best performed in the right electrode. This is possible since current is conserved and can be very easily be done if all transmitted states and the incoming state are normalized to carry unit current. Using the orthogonality of the Bloch states one can perform the same steps as in Eq. (10) to (14) again to derive the more general Landauer equation for ballistic transport in the presence of some scattering of the incoming electrons,

$$\Gamma = \frac{e^2}{h} \sum |t_{ij}|^2, \tag{18}$$

where i, j label the Bloch states in the reservoirs traveling from the left to the right.
Eq. (18) allows a simple interpretation of the transport in terms of the underlying quantum mechanical property of the transmission probability $P_{ij} = |t_{ij}|^2$ of an electron from the incoming Bloch state i into the transmitted Bloch state j. This interpretation makes the requirement of normalizing the incoming and transmitted Bloch states to unit current very clear, since in this normalization the direct interpretation of this probability is reasonably well defined and Eq. 18 can be seen as a simple generalization of Eq. 16.

2.4 Interpretation of the Landauer formula

The Landauer formula Eq. (18) was the source of some confusion for quite some time after its first formulation [9, 10]. The most striking feature of the equation might be its limit for a perfectly transmitting region, i.e. for a region with $P_{ij} = |t_{ij}|^2 = 1$ for some set of i, j. For example if one would consider a perfect bulk crystal sandwiched between reservoirs of the same bulk material the expansion of Eq. (17) would collapse to

$$\psi(\vec{r}) = \begin{cases} \psi_{\text{in}}(\vec{r}) + \sum 0\psi_r & \vec{r} \text{ in left reservoir} \\ 1\,\psi_t = \psi_{\text{in}}(\vec{r}) & \vec{r} \text{ in right reservoir} \end{cases} , \tag{19}$$

and one would rediscover Eq. (16) with N denoting the number of incoming Bloch states. At first glance, this means that the Landauer equation predicts a limited conductance of a system without any de-coherent scattering, i.e. of a perfect bulk crystal. In the same way the Landauer equation would also give a finite conductivity of a free electron gas. In this case a question which can always be asked only becomes more obvious to ask: How can a region with ballistic transport, i.e. without any dissipative processes, have a finite conductance? Since there is a voltage drop over the region and a current is flowing, some energy must dissipate. The key to the answer to this question lies in the definition of the reservoirs which were assumed to be in thermal equilibrium with some chemical potential μ attached to them. This is only possible, if there are actually dissipative processes in the reservoirs leading to the „thermalization“ of the „hot“ electrons being transfered across the region of ballistic transport.

The surprising result of a finite conductance in the case of a perfect crystal can now been interpreted in different ways. Either the setup described was not correct, since the reservoirs could not remain in thermal equilibrium and being perfect crystals like the region of ballistic transport at the same time or, which is actually very much the same, no finite voltage can be applied across such a system. The finite conductance of such a system with perfect ballistic transmission can now be interpreted as due the finite resistance at the interface between the reservoir and the ballistic region. This is also called the Sharvin-resistance of the system.

Another point to mention in the discussion of the physical significance of the Landauer equation is its formulation in terms of a two-terminal device. Both the current and the voltage drop are defined between the same two reservoirs. In many experiments, especially in mesoscopic physics a four point measurement is performed in which the current is driven between electrodes different than those between the voltage drop is measured. Büttiker [9] presented a generalization of the Landauer equation to these multi-terminal case. While this approach is very appropriate for mesoscopic physics, on the atomic scale multi-terminal arrangements are not the typical experimental arrangement and thus Eq. (18) will be sufficient. Additional resistances present in the current circuit are frequently eliminated in a four-point measurement, in which two additional potential probes are attached close to the scattering volume. However, for scattering volumes on the atomic scale, these geometries are not appropriate and thus we will restrict ourself to simple two point geometries.

While the Landauer equation is valid in many cases reaching from systems with high conductivity to systems in the tunneling regime, one has to be careful in its application in some cases. Only states in which the incoming and transmitted waves can be described by Bloch states contribute to the tunneling current. This excludes any state which is localized within the region of ballistic transport to contribute. This corresponds to the fact that these states do not carry any current within the simple one electron picture of transport chosen. In reality, there exist processes beyond this picture which lead to some coupling of these localized states to the otherwise orthogonal Bloch states in the reservoirs. For example the many-body electron-electron interaction, electron-phonon scattering, or structural defects not included in the description can provide such a coupling. Thus, while the Landauer approach will be correct for cases of high transmission through Bloch states, one could imagine that in the limit of a very low transmission probability another processes of transport across the ballistic region becomes important. In the one electron picture these processes could be thought of as the transition of an electron from the reservoirs into some localized state of the reservoir, the transition of the electron from one side of the reservoir to the other and than the transition of the electron into a state of the other reservoir. The validity of the Landauer model is now limited by the transmission probability between the reservoir states and the localized state. If this probability becomes comparable to

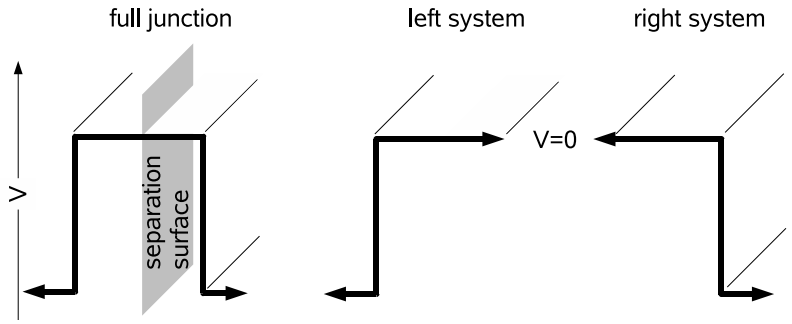


Figure 3: *Setup considered in Bardeen's approach to tunneling. The full junction is split at the separation surface. The potential at this surface (assumed to be $V = 0$ for simplicity) is extended to infinity to create two insulated systems with semi-infinite leads and semi-infinite barriers.*

the probabilities $P_{ij} = |t_{ij}|^2$ the Landauer equation breaks down.
On the other hand, one can of course treat the other limit in which the transition probability between the two sides of the reservoir becomes very small and the details of the scattering processes needed to couple the states can be neglected. This limit can be successfully described by theories for the quantum mechanical tunneling process.

2.5 The Bardeen approach to tunneling

The following description of the tunneling process is based on Bardeen's approach [11] to tunneling which essentially applies time dependent perturbation theory to the problem. Fig. 3 shows the tunneling setup used in this approach. Two semi-infinite crystals are separated by a barrier region, which will be assumed to be a vacuum barrier for simplicity. If this vacuum barrier is sufficiently high and wide one can think the total setup to consist of of two independent systems: one at the left (L) and one at the right(R) side.
This total separation of the systems leads to two independent Schrödinger equations for the two sides

$$(T + V_L)\psi_L = \epsilon_L \psi_L$$

$$(T + V_R)\psi_R = \epsilon_R \psi_R \tag{20}$$

where T denotes the operator of the kinetic energy of a single electron and V_L and V_R are the potentials of the left and right system respectively. The single particle wavefunction $\psi(t)$ of the entire setup is determined by the total Hamiltonian $H = T + V_L + V_R$.
Now one can apply time dependent perturbation theory to describe the tunneling of an electron across the vacuum barrier. Tunneling from the left to the right is assumed, the case of an electron tunneling vice versa may be treated completely analogously. The initial state of the tunneling process is localized in the left system. Therefore, there exists an eigenstate ψ_L^μ with $|\psi(t \to -\infty)\rangle = |\psi_L^\mu\rangle$. The time dependence of the state $|\Psi(t)\rangle$ is governed by the Hamiltonian of the whole system.

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle \,. \tag{21}$$

The tunneling probability is given by the overlap of this time-dependent wavefunction with a

wavefunction $|\psi_R^\nu>$ of the right system. Multiplying Eq. (21) from the left with $\langle\psi_R^\nu|$leads to

$$\langle\psi_R^\nu| \left(i\hbar\frac{\partial}{\partial t}\right) |\psi(t)\rangle = \langle\psi_R^\nu| H |\psi(t)\rangle . \tag{22}$$

Using the Schrödinger equation for the left state one obtains

$$\begin{aligned} i\hbar\tfrac{\partial}{\partial t} \langle\psi_R^\nu|\, \psi\rangle &= \langle\psi_R^\nu| H |\psi\rangle - \langle\psi_R^\nu| H_R |\psi\rangle \\ &= \langle\psi_R^\nu| V_L |\psi\rangle \end{aligned} \tag{23}$$

Substituting $|\psi(t \to -\infty)\rangle = |\psi_L^\mu\rangle$ for $|\psi\rangle$ at the right hand side of Eq. (23) leads to first order perturbation theory

$$i\hbar\frac{\partial}{\partial t} \langle\psi_R^\nu|\, \psi\rangle = \langle\psi_R^\nu| V_L |\psi_L^\mu\rangle . \tag{24}$$

Even though this equation looks familiar one has to emphasize that this is not a result obtained by standard time-dependent perturbation theory. The states $|\psi_L\rangle$ and $|\psi_R\rangle$ are eigenstates of the Hamiltonians H_L and H_R respectively. Therefore, they do not form a complete orthogonal basis of the eigenspace of the total Hamiltonian $H = T + V_L + V_R$ and the matrix elements at the left side of Eq. (24) are not sufficient to determine the total time dependence of $|\psi >$. This is a basic weakness of Bardeen's approach. However many applications of this formalism have shown that Bardeen's approximation produces reliable results for systems which are well separated, i.e. systems where the overlap of the two wavefunctions ψ_R and ψ_L is small.
Since the potential V_L is not small in the left region, the question arises whether one is allowed to use perturbation theory at all. However, it can be seen from Eq. (24) that the quantity which in fact determines the strength of the perturbation of the initial state is $\langle\psi_R^\nu| V_L |\psi_L^\mu\rangle$. Since the final wavefunction $|\psi_R\rangle$ is localized in the right region in which the left potential V_L is very weak this perturbation might still be regarded as a small perturbation and thus time depended perturbation will lead to reasonable results.
By separating the time-dependence of the states $|\psi_L^\mu\rangle = e^{i\epsilon_\mu t} |\Psi_L^\mu\rangle$ and $|\psi_R^\nu\rangle = e^{i\epsilon_\nu t} |\Psi_R^\nu\rangle$, integrating Eq. (24) and performing the limit $t \to \infty$, one obtains an expression for the tunneling-probability per time interval

$$P_{\mu\nu}^{LR} = \lim_{t\to\infty} \frac{1}{t}\frac{1}{\hbar^2} \int_0^t |\langle\psi_R^\nu| V_L |\psi_L^\mu\rangle|^2 \tag{25}$$

$$= \lim_{t\to\infty} \frac{4}{\hbar} \frac{\sin^2\left(\frac{\epsilon_\nu - \epsilon_\mu}{2\hbar}t\right)}{(\epsilon_\nu - \epsilon_\mu)^2 t} \left|M_{\mu\nu}^{RL}\right|^2 , \tag{26}$$

where the matrix element $M_{\mu\nu}^{LR}$ is given by the stationary-state matrix element of the potential

$$M_{\mu\nu}^{LR} = \langle\psi_R^\nu| V_L |\psi_L^\mu\rangle . \tag{27}$$

Assuming a continuous range of energy levels ϵ_μ (or ϵ_ν) the limit of Eq. (26) can be evaluated directly. One obtains

$$P_{\mu\nu}^{LR} = \frac{2\pi}{\hbar}\delta(\epsilon_\nu - \epsilon_\mu)|M_{\mu\nu}^{LR}|^2. \tag{28}$$

This result is similar to the well known 'Golden Rule' Fermi obtained for standard time-dependent perturbation theory. It describes elastic tunneling with energy $\epsilon_\nu = \epsilon_\mu$ only. Formally this condition is taken care of by the δ-function in Eq. (28).

To evaluate this matrix element one can introduce an additional approximation. He assumed the potential V_L to be zero in the right region of space. Similar the right potential should be zero in the left region. More formal one assumes a separation surface S which separates the regions in which the two potentials differ from zero. This can be written down by the condition $V_L V_R = 0$ for any point in space. Figure 3 shows the setup as used in this additional approximation. Of course, this approximation will become better if the potentials V_L and V_R are reasonably small at and beyond the separation surface. This will be the case if the separation surface is located far out in the vacuum.
Using the Schrödinger equation for the left wavefunction and having in mind that the potential V_L is zero in the right space one can now rewrite the matrix element as an integral over the left region only

$$M_{\mu\nu}^{LR} = \int_L \Psi_R^\nu(\vec{r})^* (\epsilon_\mu + \frac{\hbar^2}{2m}\vec{\nabla}^2)\Psi_L^\mu(\vec{r}) dV \tag{29}$$

which can be written in a more symmetric form

$$\begin{aligned} M_{\mu\nu}^{RL} &= \int_L \left\{ \Psi_R^\nu(\vec{r})^* \epsilon_\nu \Psi_L^\mu(\vec{r}) + \Psi_R^\nu(\vec{r})^* \tfrac{\hbar^2}{2m}\vec{\nabla}^2 \Psi_L^\mu(\vec{r}) \right\} dV \\ &= \int_L \left\{ \Psi_R^\nu(\vec{r})^* \overleftarrow{(T+V_R)} \Psi_L^\mu(\vec{r}) + \Psi_R^\nu(\vec{r})^* \tfrac{\hbar^2}{2m}\vec{\nabla}^2 \Psi_L^\mu(\vec{r}) \right\} dV \\ &= -\tfrac{\hbar^2}{2m} \int_L \left\{ \Psi_L^\mu(\vec{r}) \vec{\nabla}^2 \Psi_R^\nu(\vec{r})^* - \Psi_R^\nu(\vec{r})^* \vec{\nabla}^2 \Psi_L^\mu(\vec{r}) \right\} dV \end{aligned} \tag{30}$$

In these transformations in the first step the eigenvalue ϵ_μ was substituted by ϵ_ν because energy conservation requires the calculation of matrix elements with $\epsilon_\mu = \epsilon_\nu$ only. In the second step the Schrödinger equation for the right state was used (the arrow indicates the wavefunction the operators acts on). The integration area is the left region. Since the potential V_R is assumed to be zero in this region, it was dropped in the last step. Using Greens theorem and the boundary condition that the right wavefunction is zero at infinite distance from the separation surface this integral can be transformed into an integral over the separation surface

$$M_{\mu\nu}^{LR} = -\frac{\hbar^2}{2m} \int_S \left(\Psi_L^\mu(\vec{r}) \vec{\nabla} \Psi_R^\nu(\vec{r})^* - \Psi_R^\nu(\vec{r})^* \vec{\nabla} \Psi_L^\mu(\vec{r}) \right) dS. \tag{31}$$

So far only an expression for the probability of the transition of an electron from a left state into a right state was obtained.
Slightly modifying Eq. (28) this probability can be written as

$$P_{\mu\nu}^{LR} = \frac{2\pi}{\hbar} \delta(\epsilon_\mu^L - \epsilon_\nu^R - eV) |M_{\mu\nu}^{LR}|^2, \tag{32}$$

where the additional term eV is introduced to account for the bias voltage V applied between the two sides. To calculate the tunneling current one has to sum over all different possible left and right states and one has to keep in mind that the electrons might tunnel from the left to the right as well as vice versa. The total current therefore is given by

$$\begin{aligned} I &= I^{L\to R} - I^{R\to L} \\ &= e \sum_{\mu\nu} f(\epsilon_\mu)(1 - f(\epsilon_\nu + eV)) P_{\mu\nu}^{LR} - e \sum_{\mu\nu} (1 - f(\epsilon_\mu)) f(\epsilon_\nu + eV) P_{\nu\mu}^{RL} \\ &= e \sum_{\mu\nu} \left(f(\epsilon_\mu) - f(\epsilon_\nu + eV) \right) P_{\mu\nu}^{LR} \end{aligned} \tag{33}$$

where $f(\epsilon)$ denotes the Fermi-distribution function which is introduced to ensure that only tunneling from occupied to unoccupied states can occur. In Eq. (33) the symmetry of the tunneling

probability $P_{\mu\nu}^{LR} = P_{\nu\mu}^{RL}$ which can easily be deduced from Eq. (31) was used. The sum in Eq. (33) has to be performed over all right states labeled by ν and all left states labeled by μ. No further assumption is made on the nature of these left and right states, i.e. both Bloch states and surface states decaying into the bulk contribute to the current and therefor this formula differs significantly from the Landauer formula [12]. At the same time Eq. (32) and Eq. (33) show quite some similarity to the Julliere model. However, the additional matrix element in Eq. (32) which contains many details of the wavefunction.

3 TMR in crystalline MTJs: Fe/MgO/Fe

In contrast to the widely used but amorphous barrier material Al_2O_3, MgO allows to grow well-defined tunneljunctions. The Fe/MgO/Fe(100) system represents an ideal candidate for a comparison with a theoretical description. Because of the small lattice mismatch between MgO and Fe, it is possible to grow MgO epitaxially on the Fe substrate. Thus one might expect the MgO/Fe(100) interfaces obtained experimentally to be very close to the ideal theoretical structure. As discussed already, the Fe/MgO/Fe MTJs recently enabled the production of most successful TMR devices with high TMR ratios at room temperature.

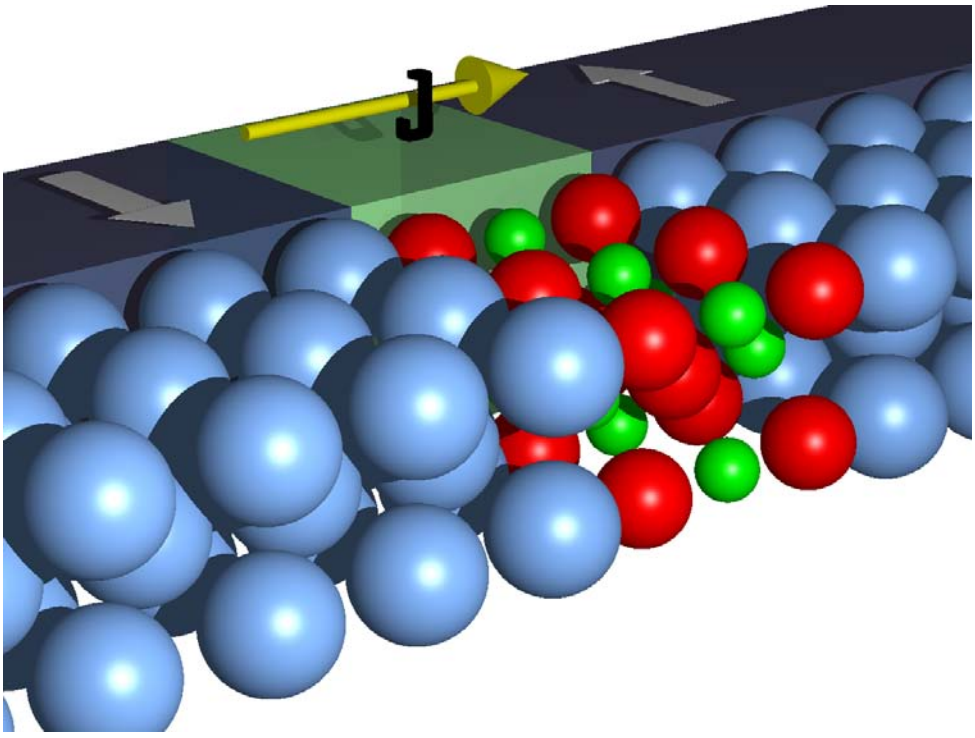


Figure 4: *Sketch of an Fe/MgO/Fe MTJ with 3 monolayer (ML) of MgO sandwiched between antiparallel magnetized Fe leads.*

The basis of the current understanding of the properties of such crystallineFe/MgO/Fe MTJs is its electronic structure. Actually, calculations [13, 14] establishing the arguments we are going to discuss have been predicting a high TMR in this junction even before the first sucessful experiments could be performed. Hence, we will first discuss the basic electronic properties of the interface with special focus on the states relevant for electronic transport. We will then apply the Landauer formalism by calculating the transmission probabilities of the MTJ. The

results have been obtained by using the density functional theory code FLEUR [15] developed in Jülich and its extension to calculate semi-infinite systems and transport [16].

3.1 Basic electronic properties of the Fe/MgO interface

Fig. 5 shows the majority (positive) and minority (negative) contribution to the local density of states (LDOS) at different positions for the MgO/Fe interface. The energy range $e_F - 3eV < \epsilon < e_F + 3eV$ in the plots was chosen to include only states near the Fermi energy e_F which are relevant for transport. The uppermost panel displays the Fe-LDOS. To the left the LDOS of bulk Fe layer is plotted, to the right the LDOS of the interface Fe is shown. In both plots arrows at the Fermi level ϵ_F are used to indicate the value of the LDOS in both spins. One can see that for the bulk Fe the LDOS at Fermi level is slightly larger for the majority spin contribution, hence the spin polarization at Fermi level is positive in bulk Fe. The spin polarization at the Fermi level becomes negative at the interface Fe layer. This result might already point out some problem of applying the Julliere model as it is not clear which polarization to use in this model. The two central panels of Fig. 5 display the LDOS in the MgO interface layer, the lower two panels the LDOS in next MgO layer. At the energy range shown here, bulk MgO has a bandgap and therefore the LDOS in these plots are purely due to the decaying LDOS induced by the Fe. This can be seen very clearly in the right panel showing the O LDOS. The LDOS in these plots mirrors that of the Fe at the interface. These induced states decay exponentional into the MgO, hence the LDOS in the interface MgO layer is about a factor of 10 smaller than that of the Fe layer, and it decays by another factor of $\sim$10 in the next MgO layer.

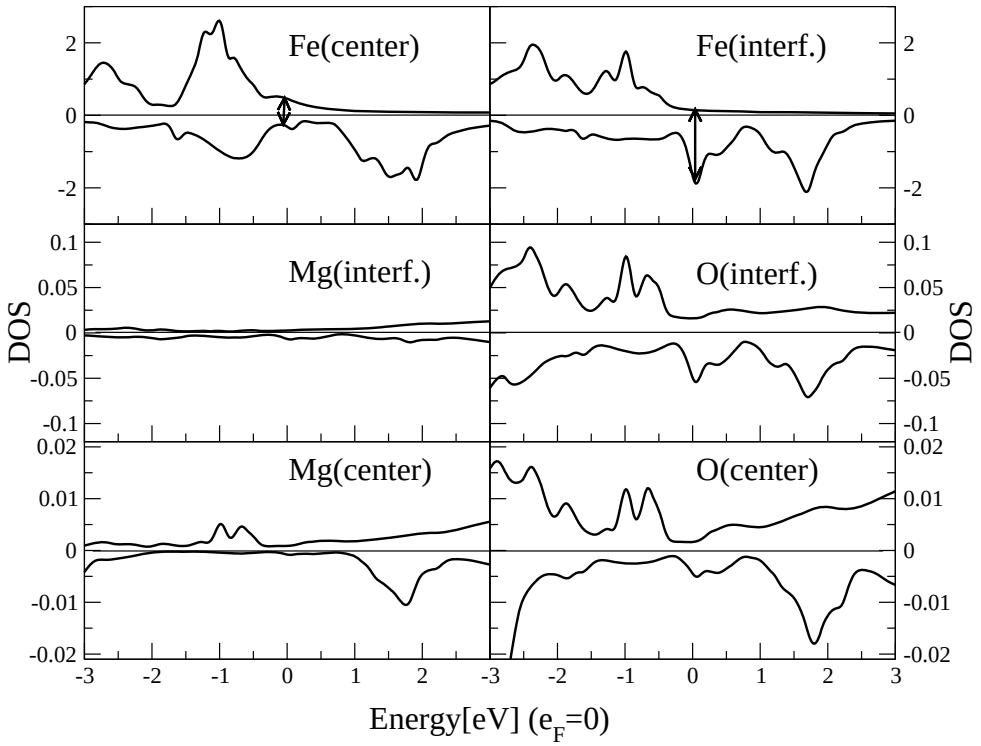


Figure 5: *Local density of states at different atomic positions of the Fe/MgO/Fe junction*

3.2 MgO as a tunneling barrier

The induced LDOS as discussed so-far is already an indication of the decay of the wavefunctions originating in the Fe into the MgO. As we have seen in our simple discussion of Sec. 2.1 this decay is one of the fundamental ingredients of the tunneling process. In particular, the rate κ of this decay, will determine the transmission of electrons through a tunnel barrier. In our simple model this decay rate was simply given in terms of the height of the barrier as

$$\kappa = \sqrt{\frac{2m}{\hbar^2}V_0 - \frac{2m}{\hbar^2}\epsilon}, \tag{34}$$

where V_0 is the potential in the barrier and ϵ the energy of the tunneling electron. If we now consider a realistic barrier with its crystal structure, we will have to switch from a simple free electron model to the full bandstructure of the material. However, in contrast to the usual discussion of the bandstructure in terms of Bloch states delocalized in all space, we must focus our interest at bandgaps in which no such Bloch states exists. Hence, we will have to consider the so-called complex bandstructure (CBS) in which not only the usual Bloch states are plotted but also states which decay exponentially. Such states exist also at energy at which no ordinary Bloch states can be found, i.e. in the bandgaps of the usual bandstructure. They can be understood as generalizations of Bloch states in which the $\vec{k}$-vector is not purely real but has an imaginary part. In fact, for a state with the complex vector $\vec{k} = \vec{q} + i\vec{\kappa}$ one finds

$$\psi(\vec{r}) = e^{i\vec{k}\vec{r}}u(\vec{r}) = e^{-\vec{\kappa}\vec{r}}e^{i\vec{q}\vec{r}}u(\vec{r}), \tag{35}$$

i.e. an exponentially decaying state with decay constant $\vec{\kappa}$. While such states are not normalizable in an infinite bulk and thus are no valid solutions for bulk wavefunctions, at interfaces and in particular in thin tunnelbarriers these evanescent states form the generalization of the simple exponentially decaying solution used in Sec. 2.1. More details of the properties of the CBS can be found in the literature, e.g. in [17].

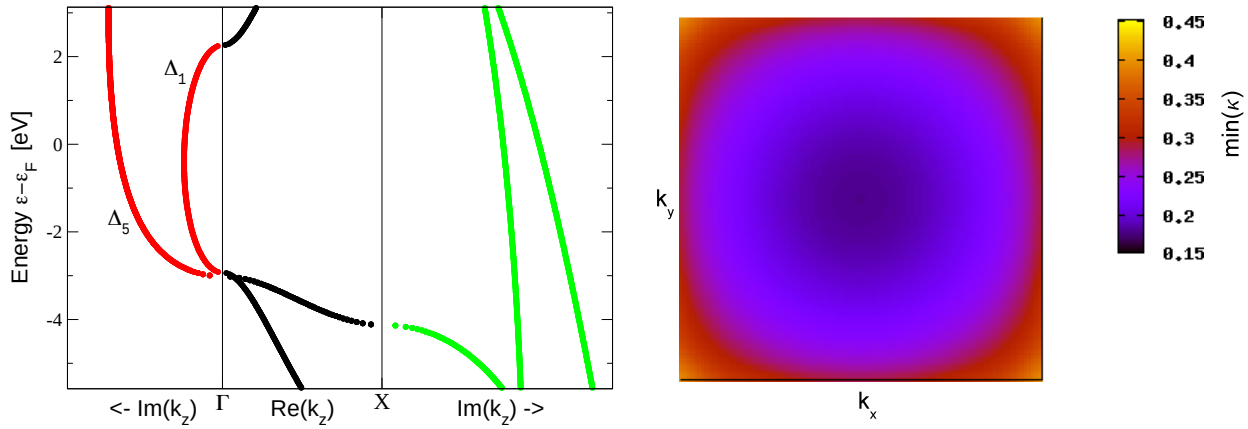


Figure 6: *a) Complex bandstructure of MgO at* $\vec{k}_{\|} = 0$. *b) Smallest decay constant* $\kappa = \min(\Im k_z(\vec{k}_{\|}))$ *for all* $\vec{k}_{\|}$*-values of the two dimensional Brillouin zone.*

We now focus on the CBS of MgO as shown in Fig. 6a). The bandgap has a value of about 4.5 eV as typical for LDA calculations and reaches from about -2.5 eV up to 2 eV. The Fermi

energy is chosen according to its position in a Fe/MgO junction. At $\vec{k}_{\|} = 0$ the highest occupied band of the valence band is connected by a complex band forming a loop with purely imaginary k_z to the lowest unoccupied band. This complex band is the slowest decaying evanescent state within the band gap. At Fermi energy the red line shows that this slowest decaying band has a k_z value of $k_z = 0 + i\,0.34\,\text{Å}^{-1}$. According to Eq. (4) this will lead to a tunneling transmission proportional to $\exp(-0.68\,\text{Å}^{-1}\,d)$ where d is the thickness of the barrier. As all other bands decay more rapidly, this is the expected asymptotic decay constant if no interface effects are present, i.e. it is the decay expected in the limit of very thick barriers.
To confirm that this state is actually the slowest decaying evanescent state one has to investigate not only the electronic states of normal incidence at $\vec{k}_{\|} = 0$ but all possible $\vec{k}_{\|}$-values from the two-dimensional Brillouin zone(2D-BZ). In Fig 6b) for each of these $\vec{k}_{\|}$-values the smallest imaginary κ value is plotted yielding the slowest decaying wavefunction. On can see that this minimal κ is smallest around the center of the 2D-BZ and grows quickly for $\vec{k}$-vectors with non-vanishing $|\vec{k}_{\|}|$.
Summarizing these results, one might state that MgO as a barrier material shows properties similar to a simple flat-potential model. In particular:

- Electrons of normal incidence have the slowest decay into the barrier.
- The decay constant κ raises with increasing $|\vec{k}_{\|}|$, little anisotropy is found, i.e. all $\vec{k}_{\|}$-directions show the same increase in the decay constant.
- The state with slowest decay at $\vec{k}_{\|} = 0$ is the highest symmetric, so called Δ_1 state.

The last of these points will be of special interest later, as the symmetry of this slowest decaying state determines the tunneling characteristics significantly.

3.3 Electronic tunneling at normal incidence

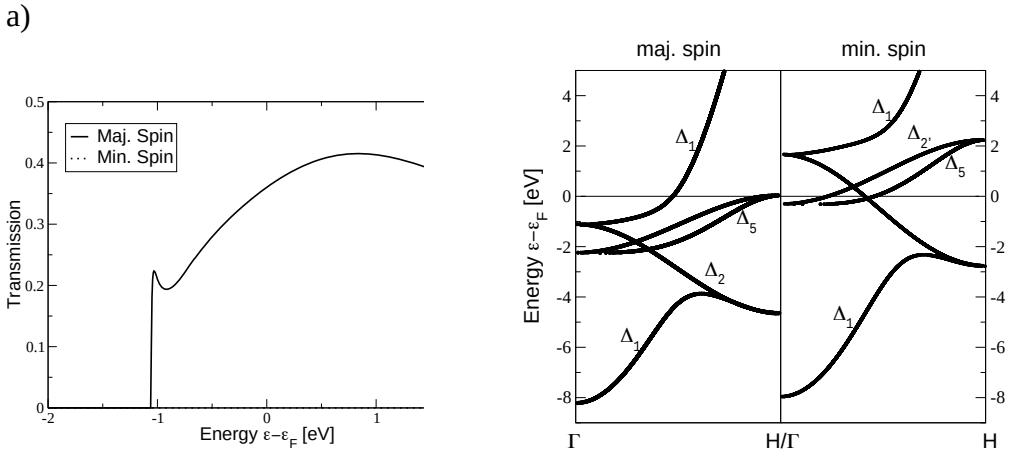


Figure 7: *a) Transmission through a 3ML MgO barrier at $\vec{k}_{\|} = 0$ as a function of energy in a parallel magnetized MTJ. b) Bandstructure of incident electrons in Fe for the same k-point and energy range.*

We start our discussion of the electronic tunneling by investigating the transmission for electrons at $\vec{k}_{\|} = 0$. Fig. 7a) shows the corresponding energy dependent transmission probability

$T = \sum |t|^2$ for energies close to the Fermi energy for a parallel alignment of the lead magnetizations. The two most striking features of this plot are: the very small transmission in the minority spin channel around the Fermi energy, which actually can not be distinguished from zero transmission and the strong transmission starting at around ~ -1 eV in the majority spin channel. Obviously, minority spin electrons are strongly reflected at the MgO barrier while some majority spin electrons – those in a band starting at ~ -1 eV – have a significant transmission probability. Actually, the same strong increase of the transmission can also be observed in the minority spin channel at a higher energy of around ~ 1.8 eV.
By comparison with the bandstructure of Fe, shown in Fig 7b), one can easily recognize the origin of this effect. In the majority spin channel of Fe a Δ_1 band starts at ~ -1 eV which can couple to the slowly decaying evanescent band in MgO and, in the case of parallel alignment of the magnetization, to the same Fe band on the other side of the tunneling barrier. All other bands in Fe have a different symmetry and hence can not be matched to the slowest decaying state in MgO. Consequently, these other states will decay much stronger into the barrier and will have a very low transmission probability. On therefore could call the MgO barrier a symmetry filter as only those Fe states with the correct Δ_1 symmetry can lead to a significant transmission. The low transmission in the minority spin close to the Fermi energy is therefore a result of the exchange splitted bandstructure in which the Δ_1 is pushed well above the Fermi level. In the minority spin channel the transmission also abruptly raises as soon as the Δ_1 can contribute at slightly above 2 eV.

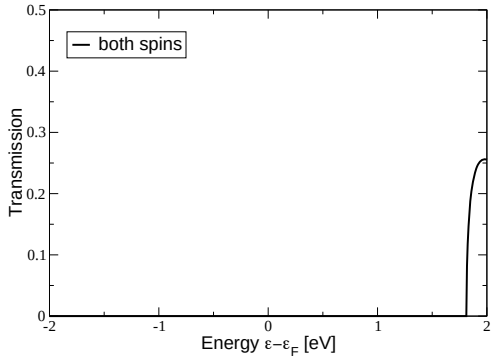


Figure 8: *Transmission through a 3ML MgO barrier at $\vec{k}_{\|} = 0$ as a function of energy in a antiparallel magnetized MTJ.*

This interpretation in terms of symmetry also explains the low tunneling transmission in the case of an antiparallel alignment of the magnetizations (Fig. 8). In this case the Δ_1 states present on one side of the barrier do not find corresponding "partners" at the other side of the MTJ as the spins are flipped. Consequently, the transmission in the anti-parallel configuration of the junction looks qualitatively similar to that of the minority-spin transmission of the parallel junction. Due to symmetry, in this case both spins of course give the same transmission.

3.4 TMR of the Fe/MgO/Fe MTJ

While the results presented so far give an indication of the origin of high TMR in all-epitaxial Fe/MgO/Fe MTJs, one important problem still has to be studied: the influence of electrons with $\vec{k}_{\|} \neq 0$. In this case two competing effects could be expected. On the one hand electrons with non-normal incidence, i.e. with non-vanishing $\vec{k}_{\|}$, decay faster into the MgO layer as we have

discussed in Sec. 3.2. This effect is expected to be stronger for thicker barriers and thus in the limit of thick MgO only $\vec{k}_{\|} = 0$ should contribute to the tunneling current and consequently a huge TMR should be expected in this limit.

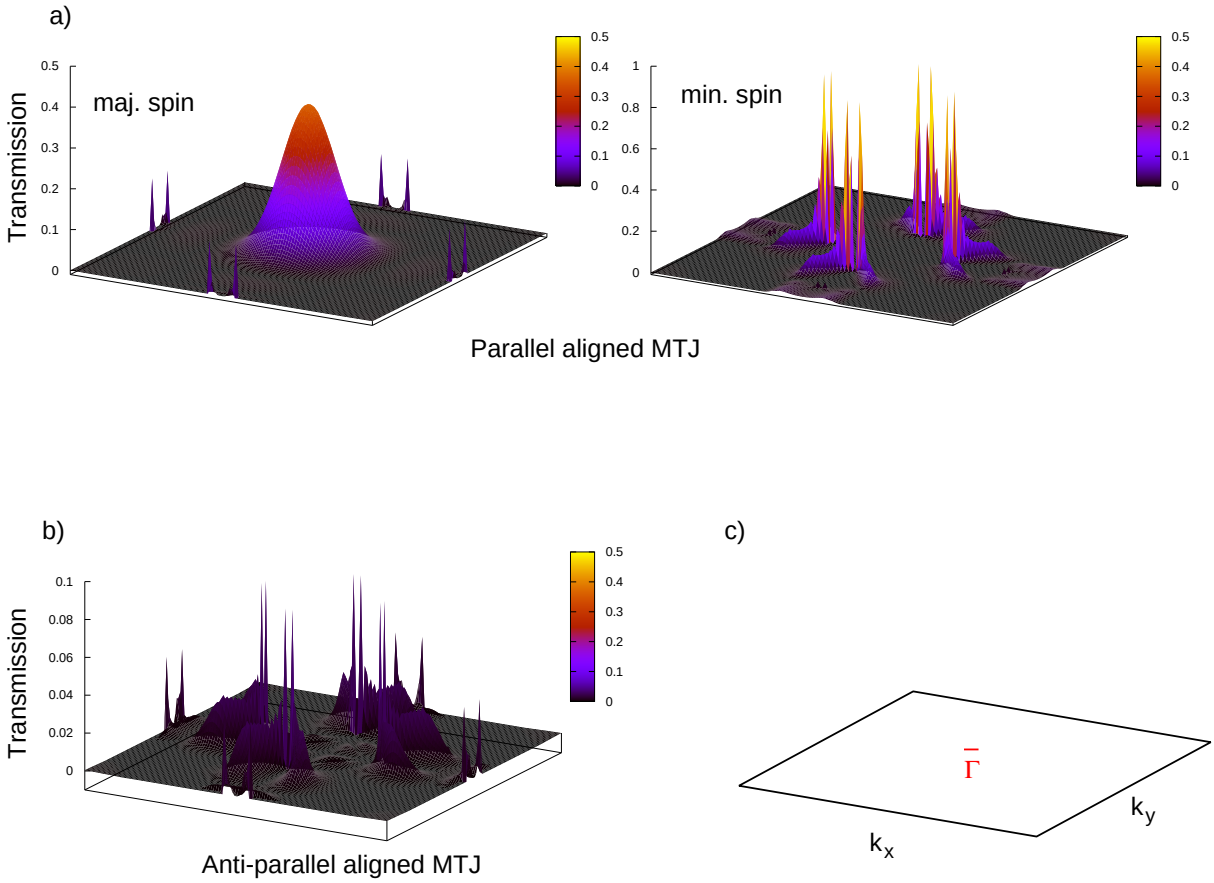


Figure 9: *a) Transmission through a parallel magnetized 3ML MgO barrier at the Fermi level for all $\vec{k}_{\|}$ and both spins. b) Transmission through an antiparallel magnetized 3ML MgO barrier at the Fermi level for all $\vec{k}_{\|}$. c) Sketch of the 2D-BZ.* ***Note:*** *the plots have different scalings while the colorscale is the same in all plots.*

On the other hand, the symmetry analysis we presented and the strong selection of tunneling states in MgO is only valid for the $\overline{\Gamma}$-point and thus one could expect that the suppression of tunneling minority spin states of Fe and the huge polarization of the transmission is only present at $\vec{k}_{\|} = 0$. To investigate this effect we also have to consider the transmission in all of the 2D-BZ (Fig 9c)).

Fig. 9a) shows the transmission as a function of $\vec{k}_{\|}$ for the majority and minority spin of a parallel aligned magnetization through 3ML MgO. This transmission was evaluated for states at the Fermi level so that the integral over the 2D-BZ of this transmission is the zero-bias Landauer conductance of the MTJ. Concentrating first at the majority-spin transmission one can identify a peak of high transmission at the center of the 2D-BZ. At the very center of this peak, the value of highest transmission is the normal incidence transmission we discussed in the last section. This peak of transmission due to the Fe Δ_1-band clearly dominates the majority-spin conductance in the parallel alignment. As expected, the minority spin channel shows no transmission directly at the $\vec{\Gamma}$-point, and also in its vicinity around the center of the BZ no

transmission is found. However, several sharp spikes of very high (up to perfect transmission of 1) can be seen at some $\vec{k}_{\|}$-points. A detailed study of these spikes identifies them as resonance effects in which surface resonance states at the interface lead to high transmission. These effects have been studied in great detail and are expected to decay rapidly with interface roughness[18] present in a realistic junction. The transmission of the antiparallel MTJ as shown in Fig. 9b) again shows some similarities to the minority spin conductance in the parallel MTJ. Only at a few special $\vec{k}_{\|}$-points resonances lead to a high transmission.
From the discussion of the various transmission values we can now draw conclusions on the expected TMR in a Fe/MgO/Fe MTJ. From the zero-bias conductance of Fig. 9 one obtains a TMR value of approximately 480%. This value already shows the high TMR also found experimentally but should not taken too serious as the very sharp peaks in the conductance are numerically hard to integrate and physically not so relevant. Additionally, this junction is at or beyond the minimal barrier thickness which could be expected in any experimental setup and a larger thickness will lead to a relative suppression of all $\vec{k}_{\|} \neq 0$ transmission and therefore to an increase in TMR. Calculations indeed have obtained TMR values well above several thousand percent.

3.5 TMR in more complex oxides

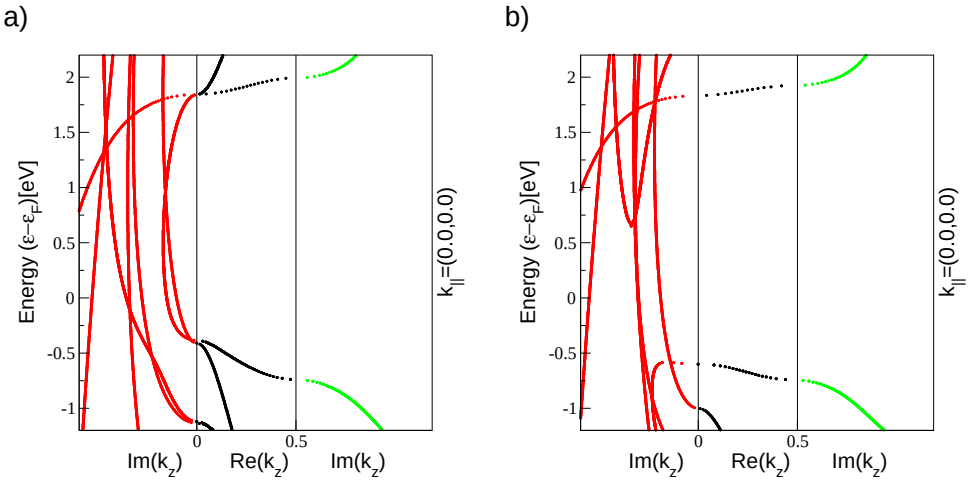


Figure 10: *Complex bandstructure of $SrTiO_3$ (a) and $BaTiO_3$ (b) for electrons at $\vec{k}_{\|} = 0$.*

Advancing beyond MTJs with MgO as a barrier material the investigation of more complex material combinations for the leads and the insulating barrier is aiming at producing MTJs with complex behavior, including possibly new physical phenomena and improved functionality. A special focus at the moment is at complex oxide materials as this class of mostly transition metal oxides show an extremely rich phasespace of new effects which might be utilized in MTJ based devices. In particular the idea of including ferroelectric or multiferroic materials into an tunneljunction seems very promising as this enables to use the strong magneto-electric coupling phenomena in these materials to easily switch the resistance state of the device or to achieve multiple resistance states depending on the relative orientation of magnetic and ferroelectric order[19].
The most promising class of materials in this context are the transition metal oxides in the perovskite structure as many of them have similar composition and lattice constant and structure

so that one can grow fully epitaxial multilayers. Much experimental progress has be reported recently on the growth of such heterostructures and many new effects at these interfaces have be discovered. As strong electronic correlation effects are important in many of these materials the theoretical description of the electronic properties and of the electronic transport processes can be quite difficult. However, some materials like the popular insulator and substrate material $SrTiO_3$ or the simple ferroelectric oxide $BaTiO_3$ are simple band insulators and their properties can be described in analogy to the discussion of MgO presented so far.
Comparing the electronic structure of $SrTiO_3$ with that of MgO, in particular looking at the character of the slowly decaying evanescent states in the barrier, the complex bandstructure as seen in Fig. 10a) shows some remarkable differences. Even though the material is not a direct-gap insulator, a rather slowly decaying state can be found at the $\overline{\Gamma}$-point. However, several bands of different effective mass are present both in the valence and the conduction band. Furthermore, states with very little dispersion in some directions can be identified by flat bands with a huge effective mass. As shown in Fig. 11 this leads to a rather anisotropic distribution of the decay constant with a "cross-like" structure in the center of the Brillouin zone.
Even more unusual tunneling characteristics are expected from a ferroelectric $BaTiO_3$ barrier (Fig. 10b)). In this case structural distortions connected with the ferroelectric state move the flat bands into the bandgap, i.e. the top of the valence and bottom of the conductance band are now formed by such bands. However, due to the high effective mass of these bands, they lead to strongly decaying states in the bandgap and the states which will dominate the tunneling are no longer derived from the states at the edges of the valence and the conduction band. This leads to a rather uniform decay rate for all $\vec{k}_{\parallel}$-points, i.e. in contrast to MgO and to $SrTiO_3$ states close to $\overline{\Gamma}$ are no longer expected to dominate the tunneling current.
As we will see more examples of novel effects and proposed functionality in these complex oxide materials in other contributions to this spring school we will not discuss those in more detail here.

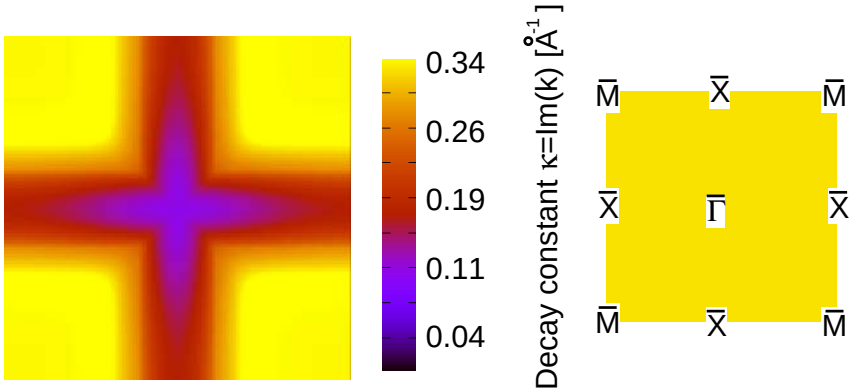


Figure 11: *Smallest decay constant as a function of $\vec{k}_{\parallel}$ for $SrTiO_3$ at an energy 0.1 eV below the conduction band minimum. The 2D-BZ with its high symmetry points is indicated as well.*

4 Summary

We discussed several different theoretical approaches to TMR. While the Julliere model is most simple and gives an very intuitive interpretation of the TMR value in terms of the polarization

its predictive power and consequently also its ability to interpret results is limited. Most popular is the Landauer picture of tunneling transport in which the quantum mechanical tunneling probability is the key quantity to determine the conductance. This is the de-facto standard at the moment in ab initio calculations of tunneling transport. As this approach includes all the details of the electronic wavefunction and its scattering on the tunneling barrier, the Landauer conductance is able to describe realistic MTJs and predict TMR values. However, the requirement of full coherence in the transport process might be a too strong assumption in some cases. One therefore could expect that Bardeen's approach being more realistic in this respect while still capturing all the physics included in the Landauer picture is an ansatz that will be used more often in the future.
We also applied the Landauer theory of TMR to the most simple and most promising MTJ Fe/MgO/Fe. Here we discussed the basic electronic properties of Fe and MgO and introduced the symmetry argument that explains the very high TMR ratios found in such MTJs. However, one should be aware of the fact that this was only a very basic summary of the results as we did not discuss or investigated many details such as e.g. the effects the electronic, atomistic, magnetic or chemical interface properties or modifications. Neither did we study the bias dependence of TMR, details of the barrier thickness dependence of TMR, inelastic transport processes or the temperature dependence of TMR. Some of those effects will be subject of other contributions to this spring school while some other are still unsolved.

References

[1] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, Phys. Rev. B **39**, 4828 (1989).

[2] M. Baibich, J. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, Phys. Rev. Lett. **61**, 2472 (1988).

[3] M. Julliere., Phys. Lett. A **54A**, 225 (1975).

[4] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, Phys. Rev. Lett. **74**(16), 3273 (1995).

[5] T. Yaoi, S. Ishio, and T. Miyazaki, J. of Magnetism and Magnetic Materials **126**(1), 430 (1993).

[6] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, Nature Materials **3**, 868 (2004).

[7] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, Nature Materials **3**, 862 (2004).

[8] R. Landauer, IBM Journal Res. Dev. **1**, 223 (1957).

[9] M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[10] A. Stone and A. Szafer, IBM Journal Res. Dev. **32**, 384 (1988).

[11] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1960).

[12] D. Wortmann, H. Ishida, and S. Blügel, Phys. Rev. B **72**(23), 235113 (2005).

[13] W. Butler, X.-G. Zhang, T. Schulthess, and J. MacLaren, Phys. Rev. B **63**, 054416/1 (2001).

[14] J. Mathon and A. Umerski, Phys. Rev. B **63**, 220403/1 (2001).

[15] http://www.flapw.de.

[16] D. Wortmann, H. Ishida, and S. Blügel, Phys. Rev. B **66**(7), 075113 (2002).

[17] V. Heine, Surface Science **2**, 1 (1964).

[18] K. Xia, P. Kelly, G. Bauer, I. Turek, J. Kudrnovský, and V. Drchal, Phys. Rev. B **63**, 64407 (2001).

[19] E. Y. Tsymbal and H. Kohlstedt, Science **313**(5784), 181 (2006).

B 3 Spin injection into semiconductors

B. Beschoten
II. Physikalisches Institut
RWTH Aachen University

Contents

1 Introduction 2

2 Band structure of direct band gap semiconductors 6

3 Optical selection rules 7

4 Electrical spin injection 8
4.1 Spin LEDs 8
4.2 Spin injection into bulk GaAs 11

5 Spin precession: from single spin to spin ensemble 14
5.1 Single spin precession 14
5.2 Spin precession of a spin ensemble 15
5.3 Continuous spin injection: Hanlé effect 16

6 Time-resolved electrical spin injection 18

References 19

1 Introduction

Modern information technology, i.e. data processing and storage, is based on semiconductors, such as silicon, and ferromagnetic materials, such as iron. Information processing and computing takes place in semiconductor transistors and integrated circuits, while information is magnetically stored on high density hard discs. The evolving field of semiconductor 'spintronics' is aimed at combining ferromagnets with semiconductors to develop electronic devices, which integrate information processing with information storage [1,2]. Up to now, information processing technology is relying on moving electron charges, ignoring the spin (which is closely connected to magnetism) attached to each electron. In ordinary electric circuits the spins are oriented randomly and have no effect on the current flow. On the other hand, spintronic devices create spin-polarized currents in which electrons are in a spin-aligned state, either spin 'up' or spin 'down', and use the spin to control current flow.

Spin polarized currents (and therefore electron spins) have nevertheless been of importance for information storage in read-out heads for computer hard drives during the last decade [3]. The read-out heads exploit an effect called giant magnetoresistance (GMR) [4, 5], which occurs in multilayer heterostructures consisting of alternating thin films of a ferromagnetic metal, for example cobalt, and a non-magnetic metal such as copper. The electrical resistance of such structures can be switched by a magnetic field. When the ferromagnetic material is magnetized, all magnetic moments of the individual atoms or of the conduction electrons align in one direction (see Fig. 1, for the sake of simplicity, here, the current is flowing perpendicular to the magnetic layers like in spin-valve structures [3], contrary to GMR devices in which the current flows parallel to the layers). Unpolarized electrons from the circuit line acquire the same magnetic moment direction when passing through the first

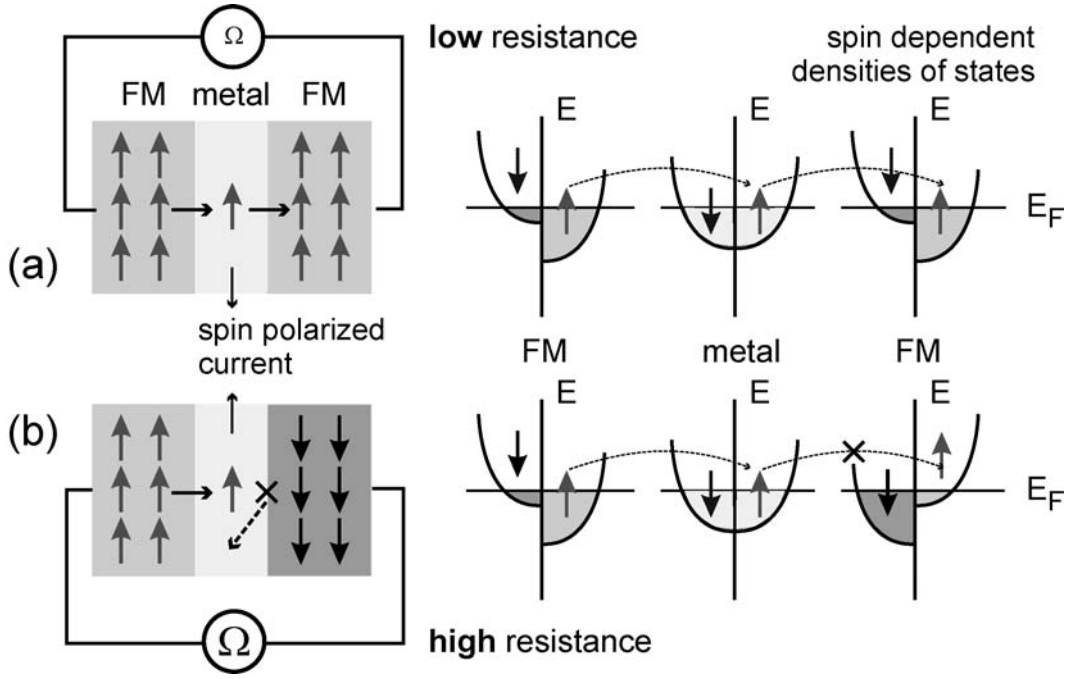


Fig. 1 Schematic illustration of a magnetic spin-valve structure consisting of a non-magnetic metal sandwiched by two ferromagnetic metallic layers of, e.g., different thickness or coercive fields. (a) If the magnetization of the two ferromagnetic layers is aligned parallel, spin polarized electron currents generated by the first layer can easily enter the second layer as they scatter into a high density of free electronic states of the same spin orientation leading to a low resistance state. (b) For oppositely oriented ferromagnetic layers only a small number of free electron states of the same spin orientation is available leading to a high resistance state [3].

ferromagnetic layer thus creating a spin polarized current. If this spin current passes through the second magnetic layer depends on whether the magnetic moments of its electrons are aligned parallel or antiparallel to the magnetic moments of the ferromagnetic layer. Only for parallel alignment the spin current can pass freely. With no external magnetic field applied, the magnetization of both ferromagnetic layers can alternate depending on the thickness of the intervening non-magnetic spacer metal. Then the flow of a spin polarized current is hindered when passing the second layer. However, if an external magnetic field is applied, which aligns the magnetic moments of all ferromagnetic layers in one direction, the barrier to spin polarized transport is reduced and the resistance drops.

After switching, the magnetic layer in a spin-valve keeps its direction of magnetization until it is switched again. This means that the device can act as a memory element and it even retains its spin configuration when the external power is turned off. Metal-based spintronics therefore lead to the more sophisticated storage technology of 'non-volatile' magnetic random-access memories (MRAMs). For practical purposes MRAM devices use a magnetoresistance effect that depends on spin dependent electron tunneling. A tunneling magnetoresistance (TMR) device has a similar sandwich structure like the spin valves with the nonmagnetic metallic layer being replaced by a tunneling barrier [6]. Spin dependent tunneling is a quantum-mechanical property of electrons that allows them to travel across an insulating barrier between two ferromagnetic layers (a few mono-layer thick aluminum-oxide (Al_2O_3) is often chosen as the tunneling barrier) even when, according to classical physics, they do not have sufficient energy to do so. Like in the spin-valve structure the magnetoresistance in TMR elements switches from a high to a low resistance for antiparallel and parallel alignment of the adjacent ferromagnetic layers.

Today's magnetic read-out heads and MRAMs are made of ferromagnetic metallic alloys. However, micro-electronics companies are solely oriented to semiconductors and not to metals. An important goal is therefore to design and to build all semiconductor spintronic devices using semiconductors, which are compatible with existing chip technology. In addition, semiconductor spintronics may even offer more interesting possibilities for information processing since semiconductors have the ability to amplify both optical and electrical signals, which is not possible in metallic devices.

The spin field-effect transistor (spin FET) has been the model device for many years in the field of semiconductor spintronics. It was proposed by S. Datta and B. Das in 1990 [7]. A schematic illustration is depicted in Fig. 2(b). In a conventional FET, a narrow semiconductor channel is placed between two opposing electrodes, the 'source' and the 'drain'. A third 'gate' electrode is located above the channel. An electrical field caused by a negative voltage applied to the latter gate drives electrons out of the channel, which turns the channel insulating. The spin FET has a ferromagnetic source and drain, so that the current injected into the semiconductor channel is spin polarized. If this spin current is not affected when traversing the channel, it will pass the ferromagnetic drain if the magnetizations of both the source and the drain are aligned in the same direction (Fig. 2(a)) (similar mechanism as for the spin valves in Fig. 1).

In contrast to spin transport through metals, however, in semiconductors the electron spin can easily precess around magnetic fields at the Larmor frequency. This precession leads to a continuous rotation of the spin orientation of the electrons when travelling through the semiconductor channel. Possible magnetic fields may exist due to intrinsic material properties (spin-orbit-coupling) or may be caused by additional electrical fields (in their rest system the electrons also see a magnetic field component). These electrical fields can either be built-in

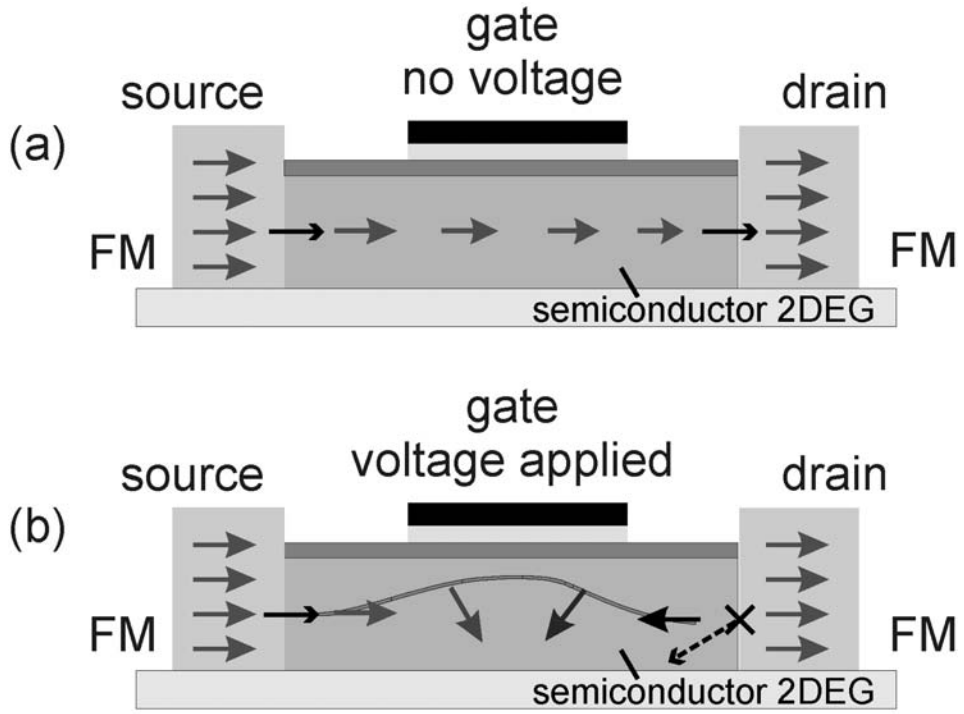


Fig. 2 (a) Schematic illustration of a ferromagnet-semiconductor-ferromagnet spintronic device. (b) In the Datta-Das spin FET [7], the electrical field generated by a gate voltage can control the phase (orientation) of the electron spins by changing their Larmor frequency within the semiconductor channel. Unlike in the all-metal spin-valves (Fig. 1), the high and low resistance states can be switched by the gate voltage alone leaving the magnetization of the two ferromagnetic layers unchanged.

fields (caused by confining potentials in semiconductor heterostructures) or external fields generated by a gate voltage. By carefully adjusting the gate voltage, the orientation (phase) of the electron spins can therefore be controlled within the semiconductor. When these fields are adjusted that the electron spins are aligned antiparallel to the drain magnetization, the electrons get rejected at the drain interface (high resistance state) (Fig. 2(b)). In contrast to the all-metal spin-valves, the resistance of the spin-FET can be switched from high to low by means of electrical fields only and no magnetization reversal of the ferromagnetic electrodes is required.

Up to now, no working spin FET prototype has been built. A major obstacle for its realization is the difficulty in effectively generating electron spin currents from a ferromagnetic metal into a semiconductor. In fact, it has been shown by Schmidt *et al.* [8] that electrical spin injection (in the diffusive limit) across an ohmic metal/semiconductor contact is only efficient if the ferromagnetic metal is nearly 100 % spin polarized (all magnetic moments of the ferromagnet then point in one direction), which is not observed in conventional ferromagnetic metals such as iron. This obstacle for spin injection can be bypassed by placing a tunneling or a Schottky barrier between the ferromagnetic metal and the semiconductor [9-16].

Recent optical experiments at various laboratories around the world show that efficient electrical spin injection into semiconductors can be achieved without the need of spin polarized tunneling using novel materials, called diluted magnetic semiconductors as a source for spin polarized carriers [17-19].

In addition to electrical spin injection, researchers must address and answer several important questions before potential semiconductor spintronic devices such as a spin FET can be utilized in a new technology :

- Can diluted magnetic semiconductors or ferromagnetic metals be used in integrated circuits?
- Can we design semiconductors, which exhibit ferromagnetism with a high spin polarization far above room temperature?
- How efficient can spin polarized currents be injected into semiconductors?
- What is the role of the ferromagnet-semiconductor interfaces for spin injection?
- On which length scales can spin polarized carriers be transported in semiconductors?
- How can we control the state of (individual) electron spins by external magnetic or electric fields inside the semiconductor?
- What are the fundamental advantages of a spin-based vs. a (conventional) charge-based electronic with respect to speed, power consumption, functionality, scalability, and profitability?

Besides electrical spin injection, the successful realization of semiconductor spintronic devices necessitates the ability to preserve spin information over practical length and time scales inside the semiconductor. Recent observations revealing extremely long spin coherence times (100 ns) for optically injected spins in non-magnetic semiconductors (GaAs) [20], as well as spin transport over macroscopic distances (100 μm in *n*-GaAs) [21, 22], and through semiconductor heterointerfaces [23, 24], has additionally raised the possibility that these spin coherent properties may eventually enable quantum computational operations in solid state systems [21] (see also lecture by Th. Schäpers). Spin quantum computation is a good example for a broader goal in spintronics, which is the development of new functionality that does not exist separately in a ferromagnet or in a semiconductor.

The lecture is devoted to address some of the above issues on semiconductor spintronic. In particular, it will be focused on the electrical spin injection into III-V semiconductors. The next section describes some basic semiconductor properties. In section 3 optical selection rules are presented. Spin injection into a spin LED and into bulk GaAs are the topics in section 4. Spin precession of single spins and of spin ensembles will be addressed in section 5. The last section scopes with a time-resolved experiment on electrical spin injection, which allows temporal phase triggering of the ensemble phase of an electrically injected spin packet.

2 Band structure of direct band gap semiconductors

Before discussing electrical spin injection into semiconductors and its optical probes, a brief summary of several basic band structure features of direct band gap semiconductors is given [26]. As depicted in Fig. 3, direct band gap semiconductors have a band gap at the centre of the Brillouin zone (Γ–point). The conduction band (*s*-like) is energetically separated from the valence band (*p*-like) by the gap energy E_g (E_g = 1.51 eV in GaAs at T = 4 K).

In atomic physics, orbital atomic wave functions are classified as *s*, *p*, *d*, etc., according to their orbital angular momentum $\boldsymbol{l}$. The *p* states (l = 1) are six-fold degenerate while the *s* states are two-fold degenerate (l = 0). These states can be chosen to be eigenstates of l_z, the *z* component of $\boldsymbol{l}$. The eigenvalues of l_z are known as the magnetic quantum numbers m_l with m_l = 1, 0, –1 for the *p* states. On the other hand, the spin angular momentum $\boldsymbol{s}$ has the eigenstates s_z with eigenvalues m_s = + 1/2, – 1/2. The spin-orbit interaction couples the orbital angular momentum $\boldsymbol{l}$ to the spin momentum $\boldsymbol{s}$ by

$$H_{\mathrm{SO}} = \lambda \boldsymbol{l} \cdot \boldsymbol{s} \,, \tag{1}$$

where λ is the spin-orbit coupling constant. The eigenfunctions of Eq. (1) are eigenstates of the total angular momentum $\boldsymbol{j} = \boldsymbol{l} + \boldsymbol{s}$ and its *z* component j_z. For *p* states with l = 1 and s = 1/2 the eigenvalues of j are: $j = l + s$ = 3/2 and $j = l – s$ = 1/2. The eigenvalues of j_z (denoted by m_j) can take the $2j + 1$ values j, $j – 1$, … , $– j + 1$, $– j$. The spin-orbit interaction splits the j = 1/2 state from the j = 3/2 state in the valence band of the semiconductor (Fig. 3, right panel). The splitting is known as the spin-orbit splitting Δ_o of the valence band at the Γ-point, which is typically hundreds of meV (0.35 eV in GaAs). States with j = 3/2 are four-fold degenerate and are called heavy and light hole states with $j_z = m_j = \pm 3/2$ and $j_z = m_j = \pm 1/2$, respectively. These hole states are energetically closest to the conduction band, and transitions between these states and the conduction band states dominate the majority of optical measurements.

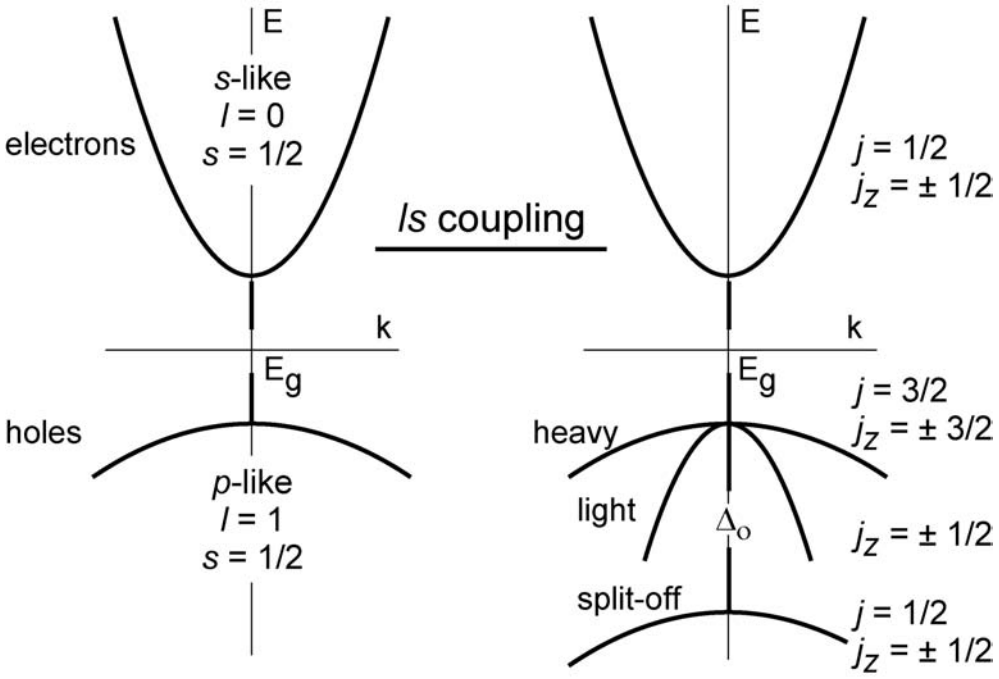


Fig. 3 Band structure of a direct semiconductor (e.g. GaAs): *s*-like conduction band and *p*-like valence band (left), band structure including spin-orbit coupling (right).

3 Optical selection rules

Photon absorption and emission (luminescence) are powerful tools for characterizing the band structure of semiconductors. The hole to electron transition obeys selection rules which preserve the angular momentum of the incoming or outgoing photon. If the photons are right or left circularly polarized they carry an angular momentum of + 1 or – 1 (denoted by σ^+ and σ^-), respectively.

The optical selection rules for absorption or emission of circularly polarized photons near the band edge between $j = 3/2$ hole states of the valence band and $j = 1/2$ electron states of the conduction band are illustrated in Fig. 4.

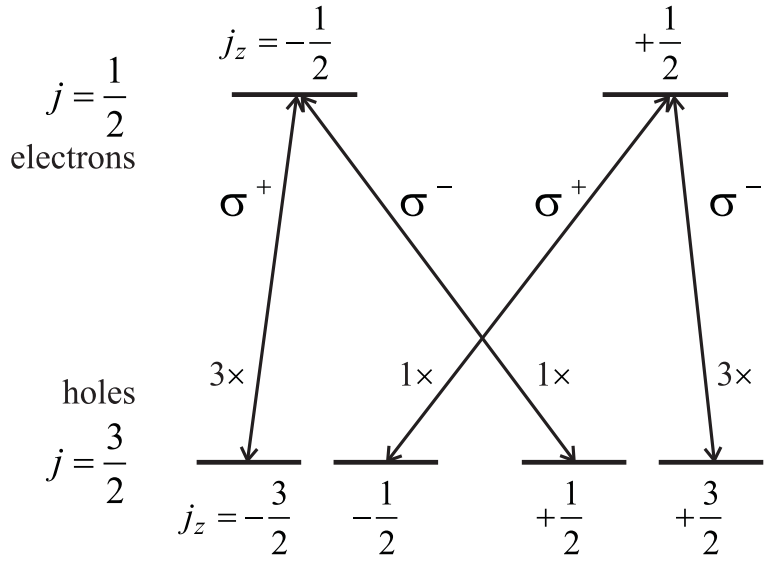


Fig. 4 Optical selection rules between $j = 3/2$ holes (valence band) and $j = 1/2$ electrons (conduction band). The probability for heavy hole transitions ($j_z = 3/2$) is three times as large as for the light hole transitions ($j_z = 1/2$).

Four different absorption and emission processes are allowed. Two of them involve heavy hole states, the others involve light hole states. It is important to note that an electron in the valence band with spin $j_z = –3/2$ leaves a hole of opposite spin after it is excited into the conduction band. The absorption process can thus be viewed in two equivalent ways: (I) a σ^+ photon transfers its angular momentum to an electron, promoting it from $j_z = – 3/2$ to $j_z = – 1/2$, or (II) a σ^+ photon creates an electron-hole pair with a total z angular momentum ($j_{z,\,heavy\,hole} + j_{z,\,electron} = 3/2 – 1/2 = + 1$) equal to that of the photon (+ 1).

The absorption of σ^+ photons will create spin – 1/2 electrons for heavy hole transitions and at the same spin + 1/2 electrons for light hole transitions. As the absorption probabilities for both transitions differ by a factor of 3, the absorption of σ^+ photons results in a large net electron spin polarisation of $-\,50\,\%$, where the sign reflects the relation between photon and spin polarization.

The inverse process, which is the recombination of electrons with holes, obeys the identical selection rules. A spin – 1/2 electron can recombine with a – 3/2 heavy hole state and emit a σ^+ photon. The luminescence emitted by spin polarized electrons recombining with unpolarized holes will thus be circularly polarized. These processes allow the conversion of spin polarization into an optical polarization. In particular the luminescence process can be used as a spin-detector for electrical spin injection [10][11][18][19].

4 Electrical spin injection

Early attempts to electrically inject spins into semiconductors using ohmic contacts only yielded small signals that could be traced back to parasitic effects [27]. After it became clear that efficient electrical spin injection is prohibited by the conductivity mismatch between the metal and the semiconductor [8](see also lecture by J. Fabian), a large variety of successful spin injection contacts have been developed. Many of these include spin-polarized tunneling that matches the resistance of the ferromagnetic contact to the semiconductor channel through an oxide tunnel barrier, through a highly doped Schottky tunnel barrier or through a Zener tunnel junction. Half-metallic contacts are a further theoretical possibility to obtain electrical spin injection. Both diluted magnetic semiconductors (DMS) and ferromagnetic semiconductors can be considered being halfmetallic at low temperatures. The materials resulted in the highest reported spin polarization, which exceeds 90% [19, 28]. “Traditional” half-metallic injectors such as Heusler alloys, CrO_2, Fe_3O_4 or $La_{2/3}Sr_{1/3}MnO_3$ have also been used. However, the few experiments testing Heusler alloys as injectors all used tunneling barriers and yielded smaller spin injection efficiencies than obtained for regular ferromagnets. The largest room temperature spin polarization (70 %) has been obtained for epitaxial Fe/MgO injectors on GaAs [15].

Most experiments on electrical spin injection have been performed on spin LEDs (light emitting diodes) as they straightforwardly allow a quantitative analysis of both the spin polarization in the semiconductor after injection (see section 4.1) and the spin injection efficiency. In contrast, for studying spin injection and spin transport in more complex devices either electrical (magneto-transport) or magneto-optical methods are applied for spin detection. The latter method will be discussed in section 4.2.

4.1 Spin LEDs

A spin LED consists of a rather simple semiconductor *p*-*i*-*n* like heterostructure (Fig. 5) in which a hole doped *p*-layer is separated from an electron doped *n*-layer by an undoped quantum well. The quantum well has the lowest energy gap (see Figure 5(b)). When the spin LED is reverse biased, hole and electron currents are injected into the quantum well from the *p* and the *n* regions, respectively. As both electrons and holes are spatially overlapping in the quantum well layer (here GaAs), they can recombine and emit photons of energies equal to their energy difference. Such a process is called electroluminescence (EL).

In conventional LEDs, both electron and hole currents are unpolarized. This results in quantum well emission of unpolarized light, because all carrier spin states are equally populated, and all dipole-allowed radiative transition occur with equal probability. In a spin LED, spin polarized carriers are injected from one magnetic contact into the quantum well, where they radiatively recombine. As discussed in section 3, radiative recombination of spin-polarized carriers results in emission of right and left circularly polarized light as determined by the selection rules. The light polarization is usually analyzed in the emission direction along the surface normal of the quantum well.

If electrons are injected from a ferromagnetic metal such as Fe (Fig. 5(b)), they become spin polarized in the *n*-layer with an electrical spin polarization of

$$P_e = \frac{n^{\uparrow} - n^{\downarrow}}{n^{\uparrow} + n^{\downarrow}}, \tag{2}$$

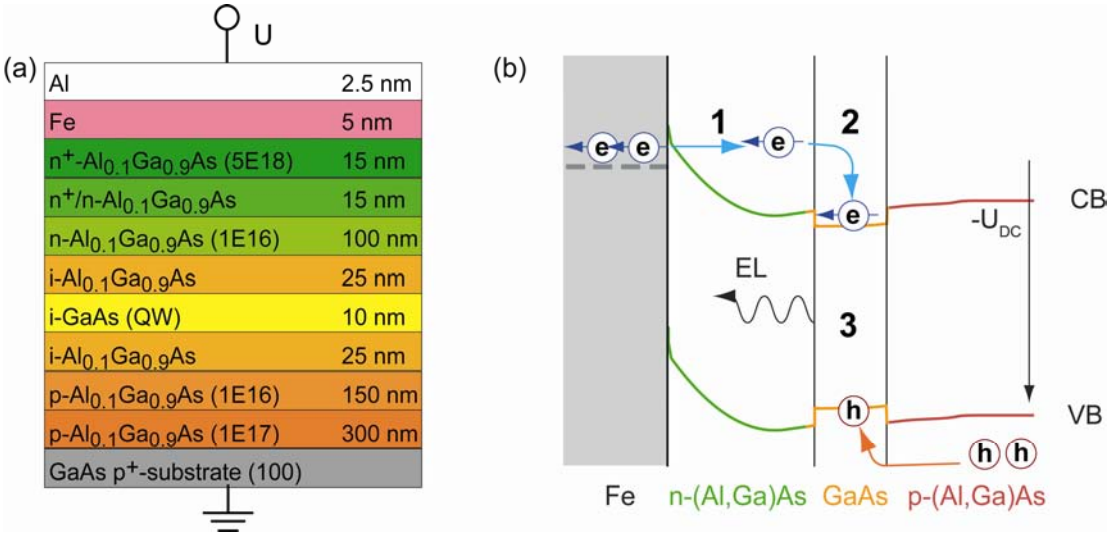


Fig. 5 (a) Layer structure of spin LED. A Schottky barrier is placed between the ferromagnetic Fe injector and a *p-i-n* diode with an embedded undoped GaAs quantum well. (b) Corresponding band diagram. If the spin LED is reverse biased ($U_{dc} < 0$), spin polarized electrons tunnel through the Schottky barrier (1) and lose excess energy by momentum relaxation (2). The polarized electrons recombine with unpolarized holes from the p-doped layer in the GaAs quantum well. The emitted electroluminescence (EL) will be detected through the Fe layer under normal emission.

where $n^{\uparrow}$ and $n^{\downarrow}$ is the spin density of spin-up and spin-down electrons of the ferromagnet. If the j = 3/2 hole states are degenerate (see Fig. 4), then the recombination with unpolarized holes yield the following optical polarization of the EL intensity

$$P_{opt} = \frac{I^+ - I^-}{I^+ + I^-} = \frac{1}{2}\frac{n^{\uparrow} - n^{\downarrow}}{n^{\uparrow} + n^{\downarrow}} = \frac{1}{2}P_e \,. \qquad (3)$$

Here, I^+ and I^- are the energy dependent intensities of the right (σ^+) and left (σ^-) circularly polarized components of the EL intensity, respectively.

In order to overcome the conductivity mismatch between the Fe layer and the *p-i-n* diode a Schottky barrier may be placed at the interface (Fig. 5(a)). It consist of a 15 nm thick transition layer with an silicon doping gradient from n(Si) = 5×10^{18}cm^{-3} to n(Si) = 1 - 5×10^{16}cm^{-3} followed by a 15 nm thick highly doped n(Si) = 5×10^{18}cm^{-3} layer. This approach was first described in Ref. 12. It minimizes the width of the interfacial depletion region, which results in a large tunnel current through the Schottky barrier under reverse bias. The epitaxial Fe layer is 5 nm thick and is capped with Al. Its magnetic easy axis is oriented in the layer plane. The moderate thicknesses of the metal layers guarantee optical transmissivity, which is important for EL detection.

The EL spectra are recorded by the spectrometer with a liquid nitrogen cooled CCD detector. Their helicities are decomposed by a circular polarization analyzer which consists of an achromatic quarter wave plate and a calcite polarizer. Fig. 6(a) shows the EL intensities I^+ and I^- which were taken under reverse bias (U_{DC} = -2.2 V) at T = 25 K and a magnetic field of B = 0T and -6 T. The magnetic field was applied perpendicular to the quantum well plane, i.e. perpendicular to the in-plane magnetic easy axis.

At B = 0 T, both spectra are identical, while a strong intensity difference is observed for B = - 6 T, which unambiguously demonstrates electrical spin injection. The optical polarization is extracted from Eq. 3 by integrating the full spectra for each light helicity. The

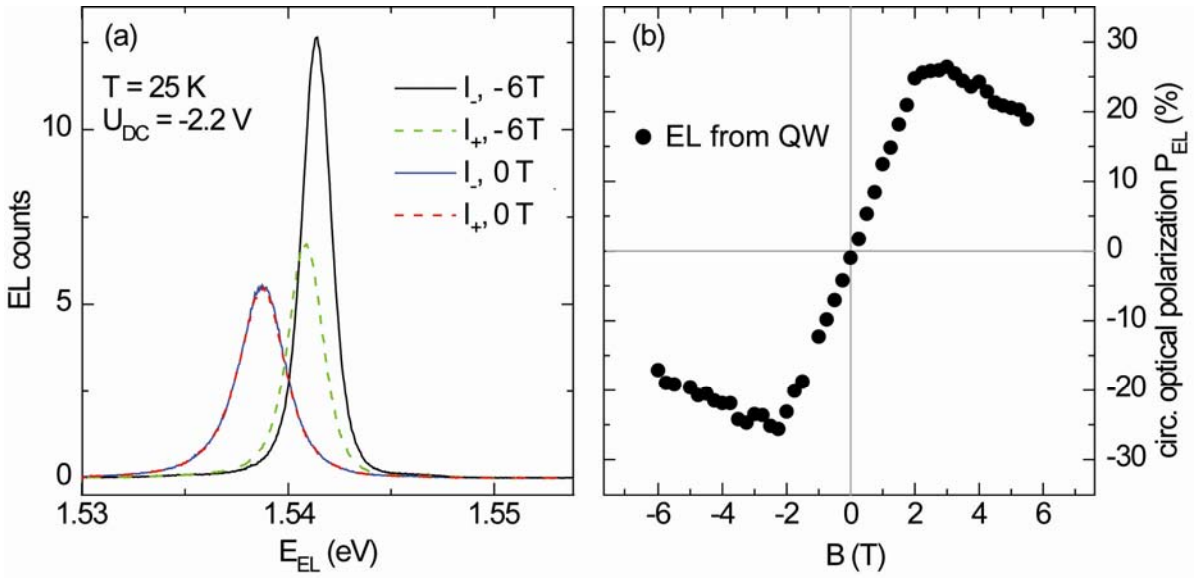


Fig.6 (a) EL spectra of GaAs emission from spin LED. The spectra are recorded under normal emission through the Fe layer (see Fig 5(B)). The magnetic field is applied in Faraday geometry, i.e parallel to the detection direction. (b) Magnetic field dependent optical polarization as determined by Eq. 3. Note that the optical polarization reaches values of 27 %.

resulting $P_{opt}(B)$ is depicted in Fig 6(b). The optical polarization equals zero at $B = 0$ T. This is expected since the magnetization vector is oriented in the layer plane for $B = 0$ T. Electron spins, which are injected from the Fe layer into the quantum well are therefore also aligned in-plane. According to optical selection rules, both I^+ and I^- intensities are not selective to the in-plane spin orientation under perpendicular light propagation direction, but only to out-of-plane spin components. The increase of P_{opt} for positive fields is therefore proportional to the hard axis out-of-plane magnetization loop (compare to Fig. 7) and saturates at B ~ 2.2 T. The decrease of P_{opt} at even larger fields is not related to the magnetization of the Fe layer but rather linked to the Zeeman polarization of the quantum well. Note that the optical polarization reaches values of 27 %.

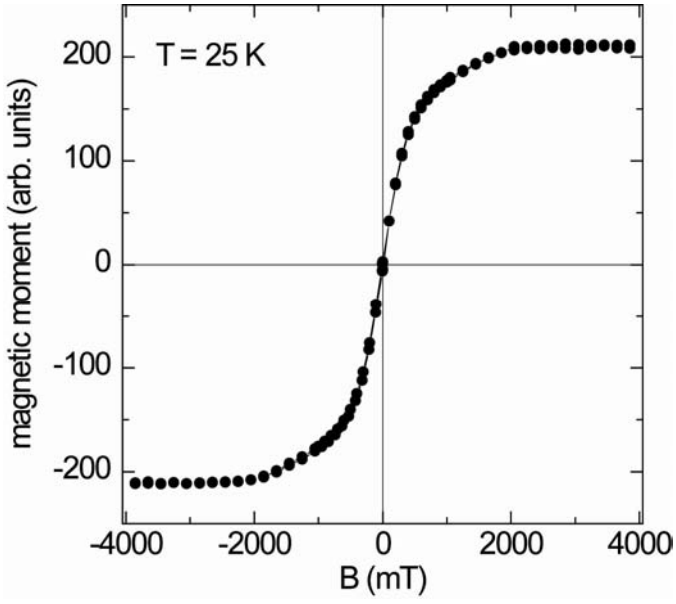


Fig. 7 Out-of-plane magnetization loop of the Fe-layer from the spin LED as a function of magnetic field. The magnetization saturates at |B| ~ 2.1 T.

4.2 Spin injection into bulk GaAs

In the previous section we have seen that spin LEDs are ideal devices to determine electrical injected spin polarizations. However, the injected spins get trapped and annihilated in the quantum well detection layer. In more complex devices (see Fig. 2), electron spins need to be transported over macroscopic distances after they have been injected into the semiconductor channel. Moreover, spin manipulation and subsequent electrical read-out might be important. In the next paragraph, we will discuss the magneto-optical Kerr effect, which is an ideal tool to probe and to image electron spins and spin currents in semiconductors.

Magneto-optical Kerr effect

In order to probe the net magnetization of electron spins, magneto-optical techniques can be used which either measure the Kerr or the Faraday rotation. Simply put, the polarization of the incident linearly polarized light is being rotated after passing through the sample (Faraday rotation) or after being reflected at the sample surface (Kerr rotation) (see Fig. 8). In both cases the resulting polarization is no longer strictly linear but rather slightly elliptic. Both effects can theoretically be described by Maxwell's equations through off-diagonal elements of the dielectric tensor. The strength of the polarization rotation is strongly enhanced in the spectral vicinity of optically-allowed band-to-band transitions. The origin of the large resonant Faraday rotation in semiconductors is the Zeeman effect. In (100) GaAs this corresponds to the well-known (spin-up and spin-down) splitting of the s-state conduction band electrons, and the fourfold splitting of the p-type valence band into spin-up and spin-down light- and heavy-holes. Thus the optical transitions associated with left and right circularly polarized light are split in energy. Consider two such Zeeman-split states whose absorption resonances appear as shown in Fig. 9(a) assuming a Lorentzian lineshape for simplicity. Their associated indices of refraction which are also split in energy (Fig. 9(b)) give rise to a large resonant Faraday rotation shown in Fig. 9(c) through the relation $\Theta_F(\omega) \propto \eta_-(\omega) - \eta_+(\omega)$. In practice, the spectral shape of the Faraday resonance strongly depends on the exact lineshape of the absorption edges and is rarely so symmetric.

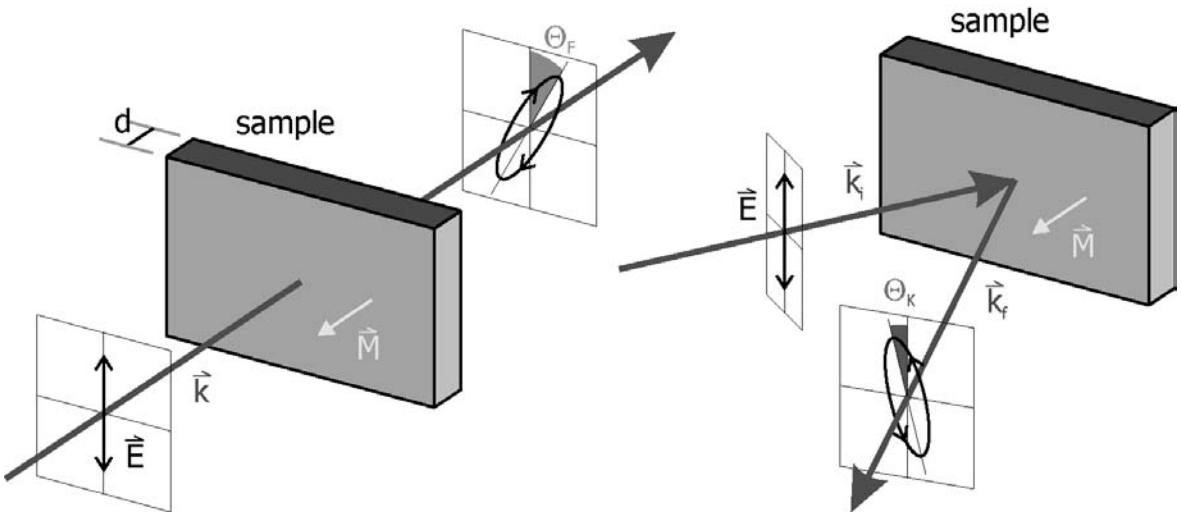


Fig. 8 (a) Geometries of the Faraday effect (left) and the magneto-optical Kerr effect (right). For both effects the sample's magnetization causes a rotation of the plane of polarization and a slight ellipticity.

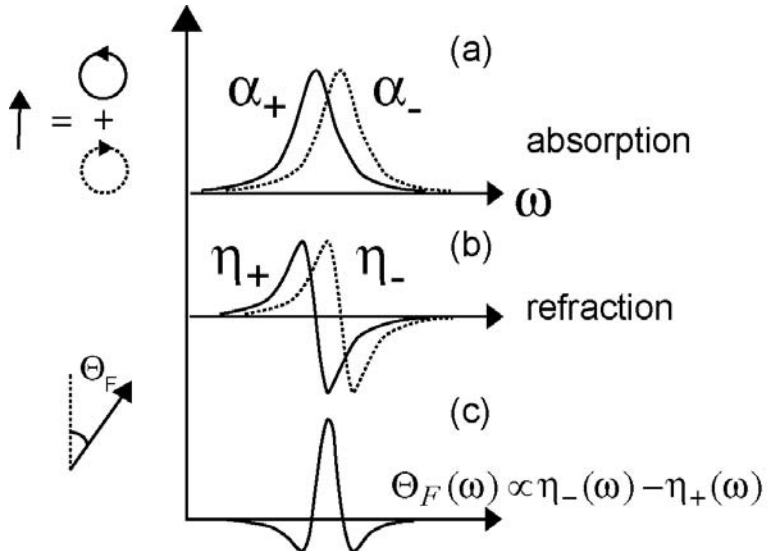


Fig. 9 (a) Idealized Zeeman-split absorption resonances (α_+, α_-) corresponding to right (σ^+) and left (σ^-) circularly polarized light. (b) Associated indices of refraction (η_+, η_-). (c) Resonant Faraday rotation.

It is important to emphasize that the Kerr effect can be used as a probe of any spin imbalance near the chemical potential of the semiconductor. For this technique, no light emission from electron hole recombination is needed.

In spin transport devices as depicted in Fig 2, electron spins are either injected into a 2-dimensional electron gas or into a bulk material. A heterostructure for the latter case is depicted in Fig. 10(a). The Schottky barrier has a similar doping profile as in the spin LED. This guarantees a high spin polarization after injection. Spins are injected into a 5µm thick *n*-doped GaAs layer with a Si doping concentration of 2×10^{16} cm^{-3}. This is the optimum doping for spin transport studies over large distances as the spin coherence length may exceed 100 µm at low temperatures [21]. After tunneling through the Schottky barrier (step 1 in Fig. 10(b)), electron spins lose their kinetic excess energy by momentum relaxation (step 2 in Fig.

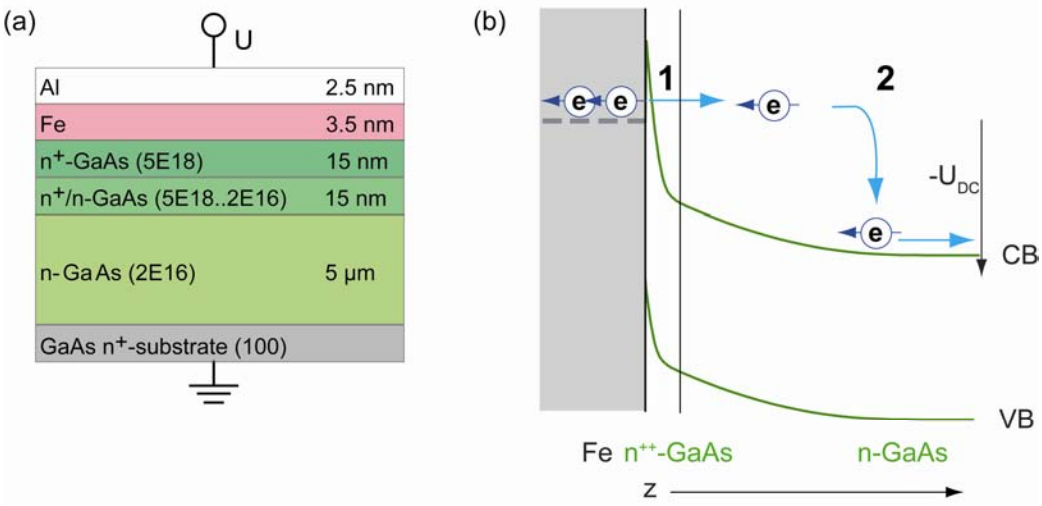


Fig. 10 (a) Layer structure for spin injection device. A Schottky barrier is placed between the ferromagnetic Fe injector and the bulk *n*-GaAs detection layer. (b) Corresponding band diagram under reverse bias. Spin polarized electrons tunnel trough the Schottky barrier (1) and loses excess energy by momentum relaxation (2).

10(b)), while they drift into the *n*-GaAs layer.

Fig. 11(a) illustrates a lateral spin injection device as used by S. Crooker et al. [13]. It consists of two Fe-GaAs Schottky contacts separated by an optimally doped *n*-GaAs bulk channel. When applying a DC bias, a spin current is generated from one of the Schottky contacts. The injected spins are oriented parallel to the magnetization direction of the Fe injector, i.e. they initially point in their propagation direction (x-direction). The spins shall be imaged by magneto-optical probes. A microscope objective is used with laser light propagation in the perpendicular to plane direction. Note that the injected spins cannot be seen as magneto-optical probes are only sensitive to an out-of-plane spin component in this polar configuration (similar optical selection rules as for the spin LED).

The spin direction can, however, be rotated towards the observation direction by Larmor precession around a small perpendicular magnetic field applied along the *y*-direction. The polarization of the injected spins is probed by means of the polar Kerr rotation angle $\Theta_K(x, y)$ of a linearly polarized reflected laser beam, which is scanned over the sample. The resulting map of the Kerr rotation angle is shown in Fig. 11(b) for $B_y = 3.6$ Gauss at a temperature of $T = 4$ K. Note that Θ_K is largest (see color code white) near the injector contact and gets diminished (color code changes to light blue) along the electron flow direction. As Θ_K is directly proportional to the spin polarization in the GaAs, it is a direct measure of the decay length of the injected spin polarization (~50 µm), which is much less than the 300 µm channel length of the GaAs. Therefore, the injected electrons lose their polarization long before they reach the opposite drain contact. Surprisingly, there is also a spin signal visible near the right-hand side of Fig. 11(b) in the GaAs channel just in front of the drain contact, which results from spin filtering and spin accumulation.

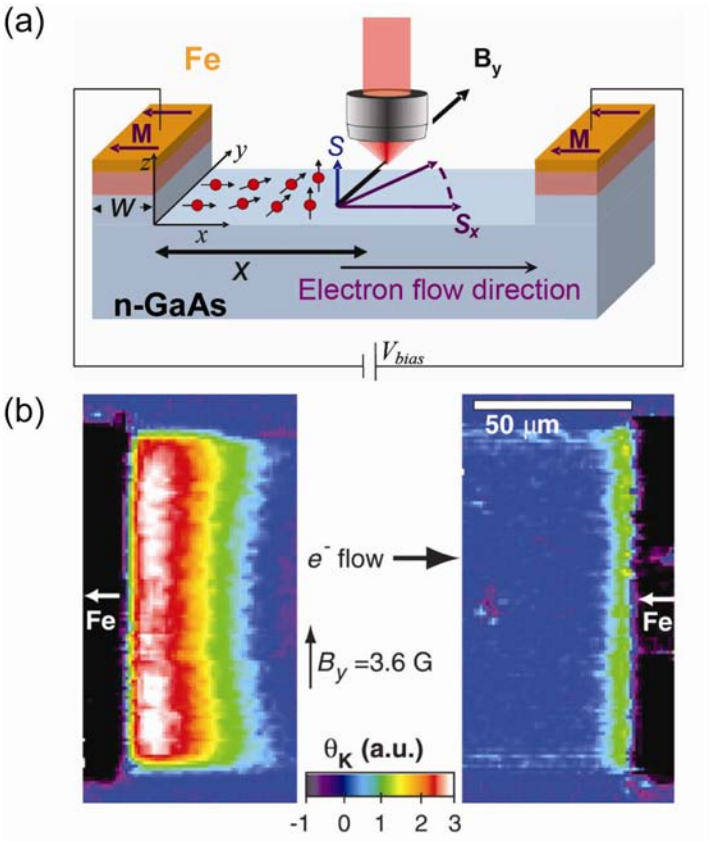


Fig. 11 Spin current flowing between Fe injectors contacts in an *n*-GaAs channel measured with polar Kerr rotation microscopy. (a) The spins are continuously injected collinear to the magnetization vector of the Fe layer. A small magnetic field B_y along the *y*-direction rotates the spins in GaAs towards the out-of-plane direction. (b) False color plot of Kerr rotation angle Θ_K as measured along the GaAs channel (adopted from [13]).

5 Spin precession: from single spin to spin ensemble

In the previous section it has been shown that spins can be generated into bulk GaAs by spin injection from a ferromagnetic source. Their initial spin orientation near the ferromagnet/semiconductor interface is defined by the magnetization direction of the ferromagnet. We have seen that the net spin orientation of the spin polarization in the semiconductor channel can be controlled by a transverse magnetic field, i.e., it rotates into the out-of-plane direction at B_y = 3.6 Gauss. The precise control of the spin orientation (its quantum mechanical phase) is of utmost importance for many device applications (see Fig. 2). In the above DC experiments individual spins start to precess at all times because spins are injected continuously in the time-domain. This, together with a velocity distribution of the electrons (diffusive transport regime) yields to a rapid depolarization of the steady-state spin polarization (the Hanlè effect). As a result, full Larmor precessions are usually not observed in injection experiments.

In order to understand the Hanlé depolarization in more details, single spin precssion and spin precession of ensembles are discussed in the next two subsection, while the Hanlé effect will be addressed in subsection 5.3.

5.1 Single spin precession

In the above experiment (see Fig. 11), electron spins are injected into GaAs with a spin orientation along the *x*-direction, while a perpendicular magnetic field is applied along the *y*-direction. In this "Voigt geometry" the individual spins can be viewed as a coherent superposition of the eigenstates for spin-up and spin-down defined by the magnetic field direction. The corresponding energy eigenvalues are

$$E_{\uparrow\downarrow} = \pm\frac{e\hbar B_y}{2m_e} \tag{4}$$

The resulting spin splitting can be related to the Larmor frequency ω_L by

$$\omega_L = \frac{g\mu_B}{\hbar} B_y \tag{5}$$

As the spin states along the *y*-axis (quantization axis) are energy eigenstates, the application of the time-evolution operator

$$\mathrm{A}(t,0) = \exp\left(-i\frac{\omega_L S_y t}{\hbar}\right) \tag{6}$$

on the eigenstates $|s_y;\uparrow\rangle$ and $|s_y;\downarrow\rangle$ yields the states themselves independent of time. Along the field axis (longitudinal direction) these quantum-mechanical spin states are thus stationary. Any loss of coherence is induced by coupling of those states with the environment. Applying the time-evolution operator on the spin states perpendicular to the magnetic field direction

$$\begin{aligned} |s_x;\uparrow\rangle &= \frac{1}{\sqrt{2}}\left[|s_y;\uparrow\rangle + |s_y;\downarrow\rangle\right] \\ |s_x;\downarrow\rangle &= \frac{1}{\sqrt{2}}\left[|s_y;\uparrow\rangle - |s_y;\downarrow\rangle\right] \end{aligned} \tag{7}$$

and

$$\begin{aligned} \left|s_z;\uparrow\right\rangle &= \frac{1}{\sqrt{2}}\left[\left|s_y;\uparrow\right\rangle + i\left|s_y;\downarrow\right\rangle\right] \\ \left|s_z;\downarrow\right\rangle &= \frac{1}{\sqrt{2}}\left[\left|s_y;\uparrow\right\rangle - i\left|s_y;\downarrow\right\rangle\right] \end{aligned} \tag{8}$$

yields the following expectation values [30]:

$$\left\langle s_x\right\rangle = \frac{1}{2}\hbar\cos(\omega_L t), \quad \left\langle s_z\right\rangle = \frac{1}{2}\hbar\sin(\omega_L t), \quad \text{and} \quad \left\langle s_y\right\rangle = 0. \tag{9}$$

The time-evolution of the spin vector may be viewed semi-classically as the Larmor precession of the classical spin vector in a plane perpendicular to the applied field. This results in an oscillatory cosine projection (Eq. (9)) of the electron spin polarization along its injection direction. Quantum-mechanically, such spin precession can be viewed as arising from the excitation of a coherent superposition of the spin states energy split by the magnetic field along the *y*-direction, which results in quantum beatings between the spin-up and the spin-down eigenstates. The coherent state gets destroyed once the phase relation between the two eigenstates is lost. The corresponding decoherence time is called T_2.

5.2 Spin precession of a spin ensemble

In most experiments there is not a single spin state but rather a spin ensemble with an ensemble magnetization $\boldsymbol{M_s}$. We again assume a constant magnetic field along the *y*-direction. The system then consists of a multiple of the single spin quantum system given in the previous section. We introduce T_2^* as the transverse spin dephasing time of the spin ensemble. It differs from the coherence time T_2, if the single spin systems are not accurately copied, but exhibit small deviations for example due to inhomogeneous effects such as internal inhomogeneous magnetic fields.

We furthermore assume that all the spins are impulsively generated at time t = 0 along the *x*-direction. Taking spin dephasing into account, the temporal evolution of the ensemble magnetization can be described by

$$M_S(t, B) = A\exp\left(-\frac{t}{T_2^*}\right)\cdot\exp\left(i\omega_L t\right). \tag{10}$$

with

$$\begin{aligned} M_{S,x}(t, B) &= \Re e\left(M_s\right) \\ M_{S,z}(t, B) &= \Im m\left(M_s\right). \end{aligned} \tag{11}$$

Fig. 12 (a) and (b) shows simulations of $M_{S,x}(t, B)$ and $M_{S,y}(z, B)$, respectively. The material parameters T_2^* and the effective *g*-factor are set to typical values for bulk *n*-GaAs. As expected from Eq. 5, the spin precession frequency increases with B_y. The exponential decay is a result of spin dephasing. Note that there is no spin precession at $B = 0$ mT. The spins therefore remain along the *x*-direction and cannot be observed in the *y*-direction ($\Im m\left(M_s, B = 0\text{mT}\right) = 0$).

It is important to emphasize that these time-domain information require an impulsive excitation of the spin ensemble, which triggers the ensemble phase. Such a phase triggering can be realized by ultrafast laser pulses [20-24]. In fact, time-resolved optical pump-probe

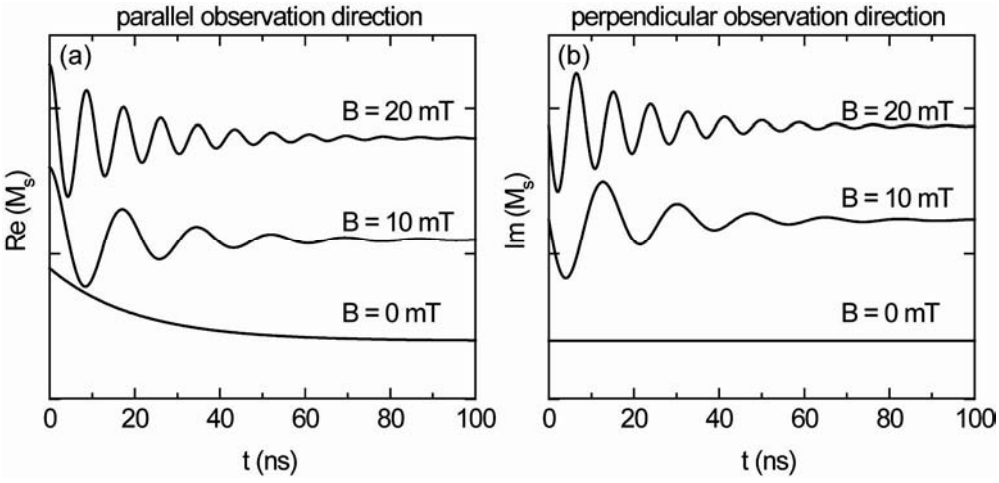


Fig. 12 Larmor precession of spin ensemble after impulsive excitation as a function of time *t*. The projection of the ensemble magnetization M_s parallel (a) and perpendicular (b) to both the excitation direction and the magnetic field is plotted for the same injection rate r_s. Typical bulk GaAs values of the spin dephasing time T_2^* = 20 ns and the electron *g*-factor (g = -0.41) are used.

experiments are nowadays standard tools for studying spin precession in the time-domain. Note that there is no such phase triggering DC electrical spin injection experiments as spins are injected at all times. In the next section, we will discuss the crossover from impulsive to continuous spin injection and its implications on the Kerr effect.

5.3 Continuous spin injection: Hanlé effect

For continuous spin injection we assume a constant spin injection rate along the *x*-direction. The resulting net magnetization $M_{cw}(B)$ is calculated by integrating the single spin systems over time *t*:

$$\begin{aligned} M_{cw} &= A' \int_0^\infty r_s \exp\left(-\frac{t}{T_2^*}\right) \exp\left(i\omega_L t\right) \\ &= A' r_s T_2^* \frac{1 + i\omega_L T_2^*}{1 + \left(\omega_L T_2^*\right)^2} \end{aligned} \tag{12}$$

$$\begin{aligned} M_{cw,\,x}(B) &= \Re e\left(M_{cw}(B)\right) = A' r_s T_2^* \frac{1}{1 + \left(\omega_L T_2^*\right)^2} \\ M_{cw,\,z}(B) &= \Im m\left(M_{cw}(B)\right) = A' r_s T_2^* \frac{\omega_L T_2^*}{1 + \left(\omega_L T_2^*\right)^2} \end{aligned} \tag{13}$$

The simulated transverse to field magnetization components $M_{cw,\,x}(B)$ and $M_{cw,\,y}(B)$ are plotted in Fig. 13. (a) and (b), respectively, for equal r_s and A'. In contrast to the impulsive excitation (see section 5.2), now the net magnetization is in dynamic equilibrium and does not depend on B and T_2^*.

For large |*B*|, the net magnetization vanishes for both observation directions as the spins increasingly occupy all precession angles within their dephasing time T_2^* and start to average out each other. This process is called Hanlé depolarization. Note that this is not a quantum

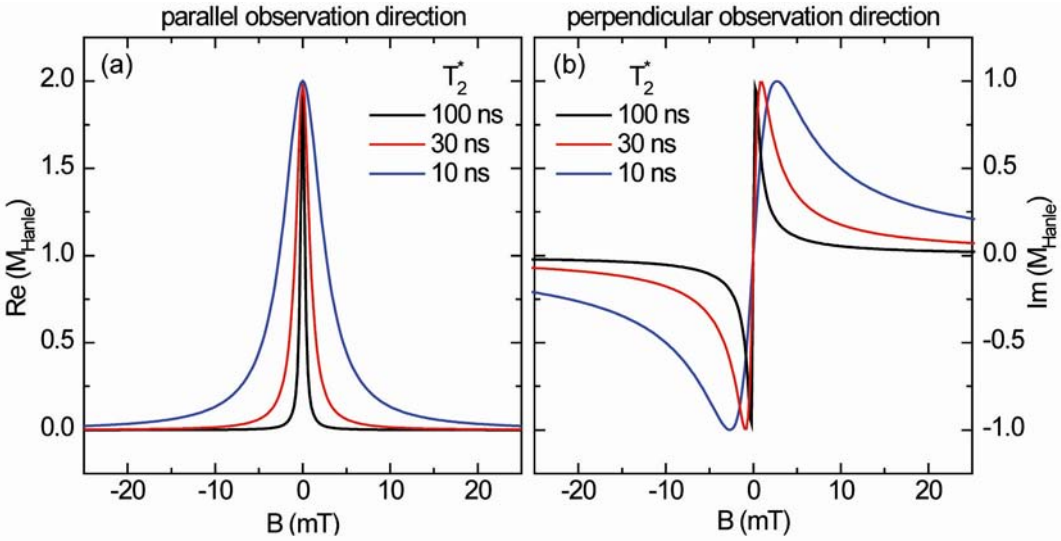


Fig. 13 Hanlé depolarization for continuous spin injection of electron spins with different spin dephasing times T_2* (*g*-factor: g = -0.41). The external magnetic field $\boldsymbol{B}$ is perpendicular to the spin injection direction. The projection if the total magnetization M_{cw} parallel to the injection direction (a) and perpendicular to both the injection direction and the magnetic field direction (b) is plotted for the same injection rate.

mechanical decoherence. While the individual spins may still be coherently precessing, the ensemble macrospin dephases, i.e. the macrospin magnetization vanishes according to phase spreading. The characteristic width of the Hanlé depolarization curve is $\Delta B = \omega_L T_2^*$, which can be used to determine T_2^* close to B = 0, if ω_L is known.

The Hanlé curves can easily be measured by the Kerr effect. Fig. 14 shows a series of Hanlé curves, which were measured on the lateral spin injection device (see Fig. 11) at various distances measured from the left ferromagnetic injector. All Hanlé curves are asymmetric, which is consistent with the observation direction along the *z*-axis, which is perpendicular to both the external magnetic field direction (*y*-axis) and the spin direction of the injected carriers (*x*-axis). Close to the injector contact (8 µm), the Hanlé curve has the largest amplitude and exhibit the typical shape as shown in Fig. 13(b). The amplitude decreases with

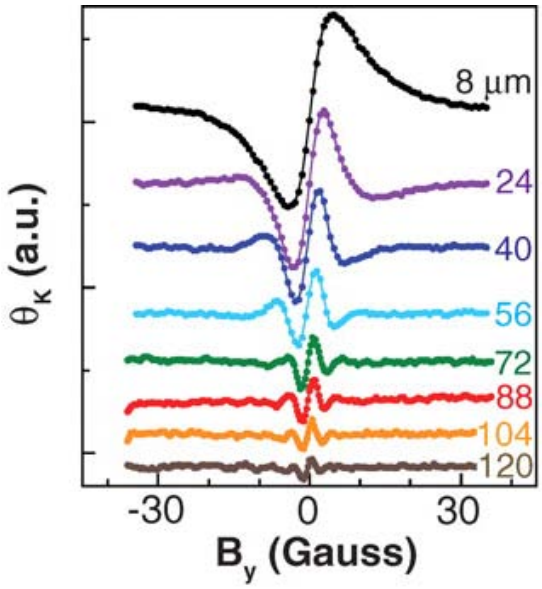


Fig. 14 Kerr rotation as a function of transverse magnetic field B_y measured as different laser spot positions (compare to Fig. 11). The distance between the laser spot and the Fe injector along the *x*-direction is given (adopted from [13])

increasing distance to the contact, which results from spin dephasing during spin transport.

Surprisingly, additional oscillations become visible at larger distances. The oscillations indicate that the Hanlé depolarization gets partly be compensated during spin transport. By this compensation the macrospin of the ensemble (probed within the laser spot) is not completely dephased during the first presession cycle. In diffusive transport, the finite spin dephasing time acts as a velocity filter of the electron spins. This in combination with the propagation distance (diffusion time) make the observation of the macrospin precession feasible.

6 Time-resolved electrical spin injection

In the previous chapter we have seen that the initial spin orientation near the ferromagnet/semiconductor interface is defined by the magnetization direction. Thus all injected spins initially have the same phase. However, the ensemble phase gets rapidly lost in a transverse magnetic field. In the following it will be shown that the Hanlé depolarization can be strongly suppressed by electrical phase triggering of the ensemble phase. This phase triggering is realized by fast current pulses, which electrically generates spin packets during injection from the ferromagnetic source into the n-GaAs. Spin precession can be optically probed in n-GaAs by time-resolved Faraday rotation.

The measurement setup and sample geometry are depicted in Figure 15. In the electrical pump / optical probe experiment, spins packets are injected by current pulses with a width of 2 ns at a repetition time of 125 ns. An injected spin packet is illustrated in Fig. 15 in the n-GaAs as a spin polarized sheet layer. Note that all spins are now spatially and temporally phase triggered. Linearly polarized ps laser pulses at normal incidence to the sample plane and phase-locked to the electrical pulses monitor the $\pm z$ component of the injected spins by detecting the Faraday rotation angle Θ_F. The time-delay Δt between the current pump pulses and the optical probe pulses can be adjusted by an electronic phase shifter.

At zero magnetic field, the spins will not precess, but rather remain in the in-plane orientation. However, when a finite magnetic field is applied, the spins start to precess into the

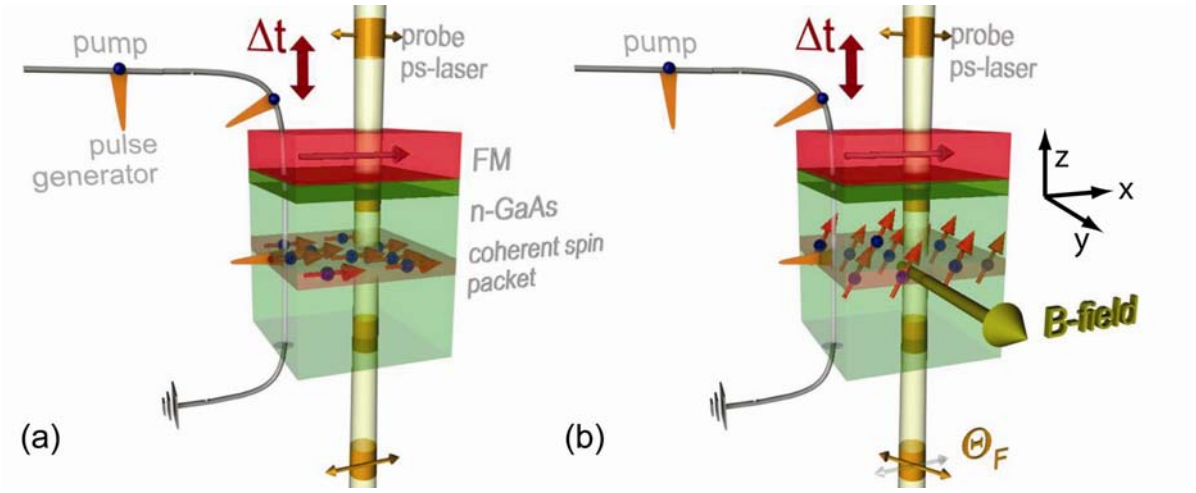


Fig. 15 Electrical pump and optical probe setup for time-resolved spin injection experiments. (a) A phase triggered spin packet in electrically injected from the ferromagnetic Fe injector into n-GaAs by a fast current pulse. (b) Spin precession in a transverse magnetic field by a time-delayed ps laser pulse, which measures the Faraday rotation in polar geometry.

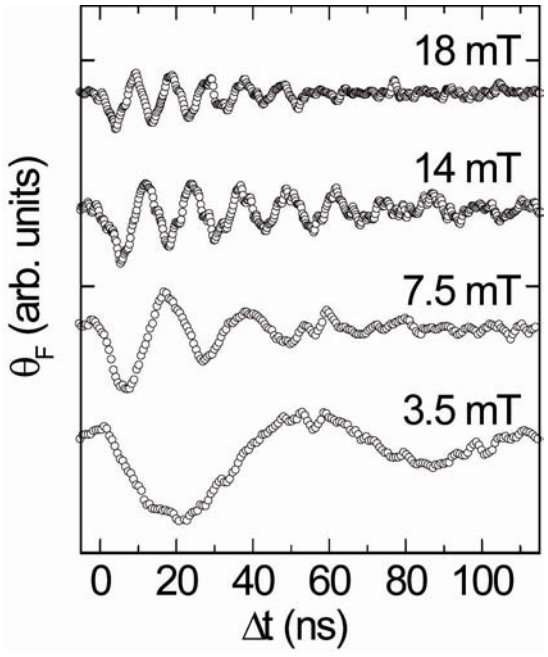


Fig. 16 Time evolution of Faraday rotation after impulsive electrical spin injection across an Fe/GaAs Schottky barrier as a function of pump probe delay Δt with vertical offsets for clarity.

observation direction (see Fig. 15(b)). If the ensemble phase was indeed temporally triggered multiple Larmor precessions of the ensemble should be observable (compare to Fig. 12). Several time-resolved Faraday rotation data are shown in Fig. 16 for various magnetic fields. Most strikingly, the expected Larmor precessions of the injected spin packets are clearly observed, demonstrating that the current pulses indeed trigger the macro-phase of the spin packet.

References

An excellent review on the various aspects of semiconductor spintronics is:
Semiconductor Spintronics and Quantum Computation, edited by D. D. Awschalom and N. Samarth (Vol. Eds.), Springer Series of Nanoscience and Technology, (Springer, Berlin Heidelberg New York, 2002).

[1] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnár, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, Science **294**, 1488 (2001).

[2] D. D. Awschalom and J. M. Kikkawa, Phys. Today **52**, 33 (1999).

[3] G. Prinz, Science **282**, 1660 (1998).

[4] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff , P. Eitenne, G. Creuzet, A. Friederich, and J. Chazelas, Phys. Rev. Lett **61**, 2472 (1988).

[5] J. Barnas, A. Fuss, R. Camley, P. Grünberg, and W. Zinn, Phys. Rev. B **42**, 8110 (1990).

[6] J. S. Moodera, L. R. Kinder, T. M. Wong, R. Meservey, Phys. Rev. Lett. **74**, 3273 (1995).

[7] S. Datta and B. Das, Appl. Phys. Lett. **56**, 665 (1990).

[8] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip and B. J. van Wees, Phys. Rev. B **62**, 4790 (2000).

[9] E. I. Rashba, Phys. Rev. B **62**, R16267 (2000).

[10] H. J. Zhu, M. Ramsteiner, H. Kostial, M. Wassermeier, H.-P. Schönherr, and K. H. Ploog, Phys. Rev. Lett. 87, 016601 (2001).

[11] A. T. Hanbicki, B. T. Jonker, G. Itskos, G. Kioseoglou, and A. Petrou, Appl. Phys. Lett. **80**, 1240 (2002).

[12] A. T. Hanbicki, O.M.J. van't Erve, R. Magno, G. Kioseoglou, and C.H. Li, Appl. Phys. Lett. **82**, 4092 (2003).

[13] S. A. Crooker, M. Furis, X. Lou, C. Adelmann, D. L. Smith, C. J. Palmstrøm, and P. A. Crowell, Science **309**, 2191 (2005).

[14] X. Jiang, R. Wang, R. M. Shelby, R. M. Macfarlane, S. R. Bank, J. S. Harris, and S. S. P. Parkin, Phys. Rev. Lett. **94**, 056601 (2005).

[15] Xiaohua Lou, Christoph Adelmann, Scott A. Crooker, Eric S. Garlid, Jianjie Zhang, K. S. Madhukar Reddy, Soren D. Flexner, Chris J. Palmstrom, and Paul A. Crowell, Nature Phys. **3**, 197 (2007).

[16] Ian Appelbaum, Biqin Huang, Douwe J. Monsma, Nature **447**, 295 (2007).

[17] For reviews on diluted magnetic semiconductors, see: J. K. Furdyna, J. Appl. Phys. **64**, R29 (1988), J. K. Furdyna, J. Kossut: Semiconductors and Semimetals, Vol. 25, (Academic, Boston, MA 1988), Vol. 25.

[18] Y. Ohno, D. K. Young, B. Beschoten, F. Matsukura, H. Ohno, and D. D. Awschalom, Nature (London) **402**, 790 (1999).

[19] R. Fiederling, M. Keim, G. Reuscher, W. Ossau, G. Schmidt, A. Waag, and L.W. Molenkamo, Nature (London) **402**, 787 (1999).

[20] J. M. Kikkawa and D. D. Awschalom, Phys. Rev. Lett. **80**, 4313 (1998).

[21] J. M. Kikkawa and D. D. Awschalom, Nature (London) **397**, 139 (1999).

[22] D. Hägele, M. Oestreich, W. W. Rühle, N. Nestle, and K. Eberl, Appl. Phys. Lett. **73**, 1580 (1998).

[23] I. Malajovich, J. M. Kikkawa, D. D. Awschalom, J. J. Berry, and N. Samarth, Phys. Rev. Lett. **84**, 1015 (2000).

[24] I. Malajovich, J. J. Berry, N. Samarth, and D. D. Awschalom, Nature (London) **411**, 770 (2001).

[25] D. P. DiVincenzo, Science **270**, 255 (1995).

[26] P. Y. Yu and M. Cardona: Fundamentals of Semiconductors (Springer-Verlag, Berlin Heidelberg New York, 1996).

[27] A. T. Filip, B. H. Hoving, F. J. Jedema, B. J. van Wees, B. Dutta and S. Borghs, Phys. Rev. B **62**, 9996 (2000).

[28] P. Van Dorpe, Z. Liu, W. Van Roy, V. F. Motsnyi, M. Sawicki, G. Borghs and J. De Boeck, Appl. Phys. Lett. **84**, 3495 (2004).

B 4 Spin Filter

Martina Müller
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Spin filter tunneling **3**
2.1 Phenomenological Description 3
2.2 Routes to spin polarized tunneling 4
2.3 Properties of spin filter materials 6
2.4 Quantifying the spin filter effect 9

3 Spintronic devices with spin filters **10**
3.1 Quasimagnetic tunnel junctions 10
3.2 Double spin filter tunnel junctions 13
3.3 Spin filtering in quantum wells 14
3.4 Spin injection through spin filters 16

4 Concluding remarks **17**

1 Introduction

Spin electronics relies on the fundamental principle that electrons carry not only charge, the property exploited in traditional semiconductor devices, but also *spin*. In semiconductor devices, an electric field is used to control the transport of conduction electrons, and hence, electrical currents. In magnetic materials, where spin-up and spin-down electron populations are unequal, an electrical current should naturally be spin-polarized. In this case, a *magnetic* field can be used to manipulate spin polarized electrical currents providing an additional channel of information and an additional degree of freedom for designing novel "spintronic" devices.
One major class of spintronic devices is based on spin polarized electron tunneling. Electron tunneling is a purely quantum mechanical phenomenon in which electrons can pass from one conducting electrode, through a thin insulating layer, into a second conducting electrode. Spin polarized tunneling is in principle no more complex than it sounds, it simply refers to a tunneling device in which the tunneling current has more electrons of one spin orientation than of the other. In practice, there are several ways to accomplish this. The lecture at hand primarily focuses on the use of "spin filter" tunnel barriers, barriers which have a different degree of opacity for each spin state. Thus, even if the source of electrons (i.e. the metallic electrodes) have no spin polarization, the resulting tunneling current will.
An interesting aspect of spin polarized tunneling arises when two magnetic materials are used in a trilayer tunneling device, typically referred as a magnetic tunnel junction (MTJ). In ferromagnetic materials, the spin orientation of the majority of electrons can be defined by the orientation of the magnetization. An electrical current flowing through a magnetic tunnel junction depends on the relative direction of magnetization in the magnetic layers: If both magnetic layers have parallel magnetization, a large tunnel current can flow, resulting in a low electrical resistance of the tunnel device. By switching from a parallel to an antiparallel configuration, a strong increase in electrical resistance is encountered. Since the relative magnetization (i.e., parallel or antiparallel) can be set by an external magnetic field, this means that the current through the tunnel junction can be modulated by the external magnetic field. Such devices display *magnetoresistance*, and comprise one class of spintronic devices in that they rely on the generation and manipulation of spin-polarized currents.
A natural question induced from the use of the word "device" is how can magnetoresistance be utilized? In principle, the resistance state of a magnetic tunnel junction, once set, can be maintained in the absence of a magnetic field or any power input – it "remembers" what resistance state it is in. One magnetization state may be used to signify a logical "0" and the other a logical "1". Consider an array of a large number of tunnel devices, where each element can be addressed individually. With appropriate addressing and recording schemes, in principle such an array provides the functions of a random access memory (RAM), but one which requires no power to maintain information, in contrast to currently used memory. Such a magnetic RAM (MRAM) can replace not only hard disk drives, but conventional RAM as well.
This lecture is intended to provide a review on the "spin filter" effect in magnetic tunnel barriers and its application in (magnetoresistive) spintronic devices. A description of the spin filter phenomenon and the fundamental physical mechanism will be given in Chapter 2. Spin filter tunneling occurs in (ferro-)magnetic insulators, a material class whose electronic and magnetic properties will be introduced. Examples of how to quantify the spin filter effect will given and the basic operating principle of spin-filter magnetic tunnel junctions explained. Chapter 3 serves as an overview on how to apply spin filters in spintronic devices. We will exemplarily discuss device architectures whose operating principles go beyond standard magnetic tunnel junctions.

2 Spin filter tunneling

2.1 Phenomenological Description

The tunnel effect is a well-known example which reveals the quantum mechanical nature of electrons. It describes the phenomenon, that an electron's wavefunction can penetrate a potential barrier which is higher than the electron's total energy, whereas classically their energy would be insufficient to pass that region. We will introduce the tunnel effect in solid state structures, in which an electrical current can flow from one electrode, through an insulating barrier, into another electrode. The most straightforward realization of this structure is a metal-insulator-metal (M/I/M) trilayer, commonly called a tunnel junction. The tunnel barrier thereby is modeled as a potential step as shown in Figure 1(a). Using the Wenzel-Kramers-Brillouin (WKB) approximation, the electron's transmission probability $T(E)$ through the tunnel barrier of height Φ_0 and thickness d decays exponentially with increasing thickness as

$$T(E) \propto \exp\left(-2d\sqrt{\frac{2m}{\hbar^2}(\Phi_0 - E)}\right), \tag{1}$$

with m being the electron mass and $\hbar$ the Planck constant. In most cases, electron tunneling in M/I/M structures is studied by observing the current (or its derivative) as a function of applied voltage across the junction. Without a voltage applied, the Fermi levels of the two metals will be at the same energies for the two electrodes. When a bias voltage is applied across the junction, one Fermi level will shift by eV with respect to the other (Figure 1(b)). The number of electrons tunneling from one to the other metallic lead is given by the product of density of states $N(E)$ of the left and right electrodes, weight by the transmission probability $T(E)$. Moreover, one has to take into account the probabilities that the electronic states in the left lead are occupied, $f(E)$, and the states in the right electrode are empty, $1 - f(E + eV)$, where $f(E)$ is the Fermi-Dirac function. The total current I flowing from the left (l) to the right (r) electrode is thus given by

$$I(V) = \int_{-\infty}^{+\infty} N_l(E)N_r(E + eV)|T|^2 f(E)[1 - f(E + eV)]dE. \tag{2}$$

In conventional M/I/M tunnel junctions, two nonmagnetic metallic electrodes are separated by an insulator typically provided by a nonmagnetic metal oxide (e.g., Al_2O_3). In that case, the tunnel probability $T(E)$ decays with barrier thickness in the same way for electrons with spin-up and spin-down orientation. If, however, a (ferro-)magnetic insulator is inserted as the tunnel barrier, the tunnel probability becomes spin-dependent. In a magnetic barrier, electrons are selectively transmitted due to its spin orientation, meaning that either spin-up or spin-down spins tunnels preferentially. This mechanism is referred as "spin-filter tunneling" and gives rise to a spin-polarized tunnel current. In this context, the spin polarization P of the tunnel current is defined as

$$P = \frac{I_\uparrow - I_\downarrow}{I_\uparrow + I_\downarrow}. \tag{3}$$

Spin filter tunneling occurs in magnetic insulators, i.e. materials that have a wide bandgap around the Fermi level and, in addition, show spontaneous magnetic ordering when cooled below a critical temperature T_C. As a consequence, the conduction band experiences a magnetic exchange splitting ΔE_{xc}, if the temperature falls below the Curie temperature T_C. Thus, the conduction band is split into a spin-up ($\uparrow$) and a spin-down ($\downarrow$) sublevel, as illustrated in Fig.

1(c). The effective barrier height of an insulating tunnel barrier corresponds to the energy difference between the Fermi level and the bottom of the conduction band. In a magnetic insulator, the spin-split conduction bands therefore effectively represent two different barrier heights, one for spin-up electrons ($\Phi_\uparrow$) and one for spin-down electrons ($\Phi_\downarrow$). Compared to the nonmagnetic case, where Φ_0 denotes the average barrier height above T_C, the spin-dependent barrier heights $\Phi_{\uparrow,\downarrow}$ are given by

$$\Phi_{\uparrow,\downarrow} = \Phi_0 \pm \frac{1}{2}\Delta E_{xc}. \tag{4}$$

In many magnetic materials, one can treat the electronic transport of spin-up ($\uparrow$) and spin-down ($\downarrow$) orientation separately. In addition, one generally assumes that the spin orientation is conserved during a tunnel process. Thus, two different transmission probabilities for spin-up and spin down electrons $T_{\uparrow,\downarrow}(E)$ must be considered. In a magnetic barrier, the tunneling probability $T(E)$ is different for the two spin directions $T_\uparrow(E) \neq T_\downarrow(E)$ because they exponentially depend on the corresponding barrier height $\Phi_{\uparrow(\downarrow)}$, as evident from Eqn.(1). This exponential dependence of $T(E)$ on the spin-dependent barrier heights finally leads to two very different currents $I_\uparrow \neq I_\downarrow$ for spin-up and spin-down electrons (recall Eqn.(2)), even with modest difference in the barrier heights $|\Delta\Phi| = |\Phi_\uparrow - \Phi_\downarrow| = \Delta E_{xc}$. Therefore, a very efficient spin filter process is at work, that in principle can result in a fully spin-polarized tunnel current ($P = 100\%$). It is obvious from this simple model, that the magnitude of exchange splitting ΔE_{xc} represents a substantial property of spin filter materials, since it is directly related to the spin filter efficiency. In the thickness regime of tunnel barriers of few nanometers, one has to take into consideration that ΔE_{xc} may be modified compared to bulk material.
Up to now, we have focused on a spin-dependent barrier height as the source of tunneling spin polarization. Apart from spin filter tunneling, there are also different routes to realize spin-polarized tunnel currents. One way is to use (ferro-)magnetic materials, which naturally provide a spin polarized density of states. One may anticipate that with magnetic electrodes in a M/I/M junction the tunnel current naturally is spin-polarized. Using magnetic electrodes is indeed the most conventional approach to realize spin polarized tunneling, and is applied in standard tunnel magnetoresistance (TMR) devices. In the following section, we will briefly review alternative routes to produce spin polarized tunnel currents.

2.2 Routes to spin polarized tunneling

The conventional approach realizing spin-polarized tunneling is by tunneling from a ferromagnetic metal through a nonmagnetic insulating barrier. Typically, $3d$ transition metals are utilized as the electrode material, because they provide a spin-split density of states N at the Fermi level $N_\uparrow(E_F) \neq N_\downarrow(E_F)$. As the barrier material in most cases polycrystalline or amorphous Al_2O_3 is used. Contrary to a magnetic barrier, which actively filters spins, the nonmagnetic barrier represents the "passive" element in a conventional tunnel device that simply selects the electron's momentum. Since ferromagnetic metals like Fe or Co provide only modest spin polarizations around $P = 40\%$ at the Fermi level, currents with just limited spin polarization can be realized. A way to optimize the efficiency of the electrode as the source of spin-polarized carriers is to use so-called half-metals. These materials possess a finite density of states at the Fermi level for one spin direction and a gap for the other spin orientation, resulting theoretically in a 100% spin polarization. Among the materials that have been predicted as half-metals are $3d$ transition metal oxides (e.g., CrO_2 and Fe_3O_4), manganites (e.g., $La_{1-x}Sr_xMnO_3$) and the material class of Heusler alloys. In practice, however, many of these materials are difficult to stabilize as or-

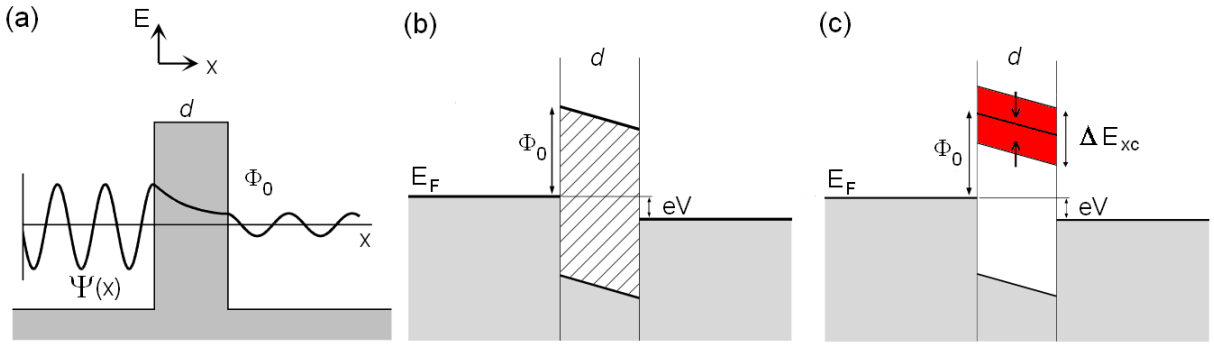


Fig. 1: *Tunneling in metal-insulator-metal structures. (a) Electron wave functions decay exponentially in the barrier of height Φ_0 and thickness d, and for thin barriers, some intensity remains at the right side. (b) Potential diagram for a M/I/M structure with applied bias voltage eV. The bias voltage defines an energy window of electronic states contributing to the tunnel current. (c) Principle of spin-filter tunneling with magnetic barriers. Below the Curie temperature T_C of the spin filter, the conduction band is exchange split by an amount ΔE_{xc} resulting in two spin dependent barrier heights $\Phi_{\uparrow,\downarrow} = \Phi_0 \pm 1/2\Delta E_{xc}$.*

dered crystalline structures and even in perfect crystals, a complete spin polarization can only be expected at $T = 0\,\mathrm{K}$.

A totally different mechanism of spin polarized tunneling is observed in epitaxial or highly oriented MgO(001) barriers combined with $3d$ transition metal electrodes (see also the lecture by D. Wortmann). Although an MgO tunnel barrier is nonmagnetic and therefore cannot directly select tunneling electrons according to their spin orientation, it filters electronic wave functions according to their symmetry. This transport process is termed coherent tunneling, and requires a matching of Bloch state symmetries at the electrode–barrier interface. At the Fe/MgO interface, for example, only totally symmetric wave functions with normal incidence can be connected to the electronic structure of the MgO tunnel barrier. It turns out, that the Fe Δ_1 wavefunctions have the required symmetry. Whereas the Δ_1 majority (spin-up) band has states at the Fermi level, no spin-down states with Δ_1 symmetry are present. Therefore, only Δ_1 spin-up electrons pass the MgO barrier efficiently and the tunneling current is highly polarized. It becomes clear, that even tunnel junctions based on $3d$ transition metals can generate nearly completely spin-polarized currents because of the symmetry filtering process. Although the tunnel process is spin-dependent, the sorting mechanism actually is a symmetry selection and therefore shall be termed “symmetry filtering”.

In addition to momentum and symmetry filter effects, a magnetic tunnel barrier itself can act as the spin-selective element of a tunnel junction, as discussed in Section 2.1. Therefore, a spin-polarized current can be created with *nonmagnetic* electrodes. This approach is in contrast to those mentioned above, which rely on ferromagnetic metals as (part of) the source of spin-polarized charge carriers.

Until now, various materials have been successfully used as spin filter tunnel barriers. Among those are binary rare earth chalcogenides, transition metal oxides and manganese-based perovskites. In the following, we will review electronic properties of selected spin filter materials that are relevant for the understanding of spin filter tunneling.

2.3 Properties of spin filter materials

Materials that filter spins by tunneling have to fulfill two criteria: On the one hand, they have to possess a spontaneous magnetic moment below a critical temperature and, on the other hand, they must have a bandgap around the Fermi level. In terms of electronic structure, such materials are either wide-gap semiconductors or insulators. Only a few compounds, however, simultaneously show magnetic and insulating properties. Until today, three material classes have been experimentally confirmed as spin filter tunnel barriers. A list of these known spin filters with some of their structural, magnetic and electronic properties is given in Table 1.

Europium chalcogenides The material class of Europium chalcogenides (EuO, EuS, EuSe and EuTe) was identified already in the early 1960s as magnetic insulators. For more than twenty years, these compounds have been subject of tremendous interest, but it has not been until recently, that the Eu chalcogenides are revisited for spintronics studies. Whereas EuO and EuS are ferromagnetic oxides, EuSe and EuTe show antiferromagnetic ordering. EuO and EuS have Curie temperatures of $T_C = 69.9\,\mathrm{K}$ and 16.6 K, respectively, and thus filter spins only at liquid helium temperatures ($T_{\mathrm{LHe}} = 4.2\,\mathrm{K}$).
Pure Europium (Eu) is a rare earth metal with the outermost electronic configuration $4f^7\ 5d^0\ 6s^2$. Due to the strong electron affinity of the chalcogens, the two $6s$ electrons of the Eu atoms are transferred to saturate the p orbitals of X in the EuX compound. The Eu-X bonding is therefore mainly ionic in nature, which explains why these materials crystallize in a compact rocksalt structure. If the europium compound is fully stoichiometric, the material is insulating with a valence band built of the p states of the anion (X) and the conduction band built up with $6s$ and $5d$ states of the Eu cation. The Eu $4f$ states are located within this energy gap, and the energy difference between the $4f$ states and the $5d$ conduction band is referred as the optical bandgap E_g.
The magnetic moment of the EuX compounds originates from the seven unpaired spins in the $4f$ levels of the Eu^{2+} ion. One can derive from Hund's rule, that the ground state spin configuration of the Eu $4f^7$ electrons is maximal $S = 7/2$, whereas the orbital momentum vanishes $L = 0$. The Eu chalcogenides are considered as typical substances where the Heisenberg model applies. The reason is that the Eu $4f$ wave functions are very localized, i.e. their overlap with other orbitals is small and the atomic character is mainly conserved in a crystal. Magnetic exchange is then described by the Hamiltonian

$$H = -\sum_{i,j} J_{ij} \mathbf{S_i} \mathbf{S_j}, \tag{5}$$

where J_{ij} is the exchange constant (positive for ferromagnets and negative for antiferromagnets) and $\mathbf{S_{i,j}}$ are the neighboring spins. In order to quantify the magnetic ordering in EuX compounds, one has to take into account the exchange integrals J_1 and J_2 between Eu^{2+} nearest neighbors and next-nearest neighbors, respectively. It turns out, that magnetic exchange originates from virtual transitions between the occupied $4f$ states and $5d$ conduction band states. Whereas the exchange J_1 between europium nearest neighbors is an indirect mechanism generated by intra-atomic $d - f$ exchange ($J_1 \simeq J_{df}$), the exchange integral J_2 is the result of several contributions of so-called superexchange. In EuO and EuS, the ferromagnetic indirect exchange J_1 is dominant and responsible for the ferromagnetic ordering at $T_C = 69.9\,\mathrm{K}$ and $T_C = 16.6\,\mathrm{K}$, respectively.
Up to this point, we considered perfectly stoichiometric europium chalcogenides, which are

Material	magnetic behavior	T_C (K)	moment (μ_B)	structure a (nm)	E_g (eV)	ΔE_{xc} (eV)	P (%)	References
EuO	FM	69.9	7.0	fcc 0.514	1.12	0.54	29	[2]
EuS	FM	16.6	7.0	fcc 0.596	1.65	0.36	86	[4]
EuSe	AFM	4.6	7.0	fcc 0.619	1.80		100	[2]
$NiFe_2O_4$	FerriM	850	2.0	spinel	1.20		22	[7]
$CoFe_2O_4$	FerriM	793	3.0	spinel	0.57	1.28		[9]
$CoCrO_4$	FerriM	95		spinel				[10]
$BiMnO_3$	FM	105	3.6	perovskite			22	[8]

Table 1: *Overview on known spin filter materials and selected magnetic, electronic and structural properties.*

insulating in nature. A small departure from stoichiometry, e.g. due to electron doping, generates free carriers in the EuX matrix which are responsible for a modification of the magnetic exchange interaction. In that case, the indirect exchange between the localized $4f$ spins S of the Eu^{2+} ions and the spins σ of the $5d$ conduction electrons is mediated by a free electron gas. The effect of the $d-f$ exchange constant J_{df} is to split the conduction band in spin-up ($\uparrow$) and spin-down ($\downarrow$) subbands by an amount

$$\Delta E_{xc} = J_{df} S\sigma, \tag{6}$$

with S being the spin carried by the Eu^{2+} ions and σ being the localized spin polarization. The exchange splitting ΔE_{xc} of the conduction bands results in a large variation of the bandgap E_g as a function of temperature, as shown in Figure 2(a). This thermal shift of E_g affects the optical, magnetic and electronic transport properties of EuX. An optical absorption experiment, which directly probes the electronic structure of a material, can visualize this effect (Figure 2(b)): A large shift ΔE of the optical absorption edge towards lower energies can be observed when the temperature is lowered and EuX becomes ferromagnetic. This shift is a result of the increasing J_{df} exchange interaction between the localized Eu $4f$ spins and the $5d$ conduction electrons, which pushes the spin-down conduction band to higher energies and the spin-up band to lower energies, thus reducing the effective bandgap E_g. Apart from optical measurements, various other experiments have confirmed the thermal shift of E_g, for example photoemission, field emission and tunneling experiments [1].

Whereas many studies have been presented for bulk europium chalcogenides, much less is known about the electronic and magnetic properties of thin films. In the thickness regime of spin filter tunnel barriers of few nanometers, it is important to achieve sizable magnetic ordering and exchange splitting. Ab initio calculations of single crystalline EuX(100) predict the Curie temperature to be strongly thickness dependent, i.e. to be significantly reduced for films of just few monolayers [3]. Qualitatively, this behavior can be explained by the lower coordination number of the surface atoms (four for an fcc structure) and thus a reduced exchange interaction compared to that of the atoms in the center of the film (twelve for bulk). One may anticipate that if T_C is reduced for thin EuX films, also a reduced exchange splitting ΔE_{xc} can be at play,

which recently has been confirmed by tunnel experiments. However, ΔE_{xc} is still large enough such that EuX spin filter barriers typically produce a relatively high spin polarization, as listed in Table 1. Since growing ultrathin films of Eu chalcogenides is relatively easy, this material class is commonly utilized for fundamental studies on spin filter tunneling [2, 4].

Spinel Ferrites Candidates for spin filter barriers are the transition metal oxides $NiFe_2O_4$, $CoFe_2O_4$ and $MnFe_2O_4$. These compounds with the general formula AB_2O_4 belong to the material class of spinel ferrites. Whereas the AB_2O_4 systems are magnetic and insulating, their parent compound Fe_3O_4 (magnetite) shows half-metallic character. Ferrites have magnetic ordering temperatures of about 800 K and thus could potentially filter spins at room temperature. The spinel ferrites AB_2O_4 crystallize in a cubic structure built of two face-centered cubic (fcc) sublattices. The first sublattice is formed by ferromagnetically ordered Fe^{3+} ions that occupy the tetragonal A sites of the spinel AB_2O_4 structure. The second sublattice contains ferromagnetically ordered TM^{2+} and Fe^{3+} ions which occupy the octahedral B sites. Recent experiments give evidence that thin spinel ferrite films have magnetic properties that substantially differ from those of the corresponding bulk material. For example, $NiFe_2O_4$ films with a few nanometers thickness have a saturation magnetization of about twice that of the bulk compound, depending on the growth conditions [7].
Thin spinel ferrite films are usually deposited by oxygen-assisted molecular beam epitaxy, a deposition method which allows a controlled growth down to atomic layer accuracy. In practice, however, chemical off-stoichiometry and atomic site disorder are often present in samples. As the magnetic properties are sensitively controlled by the occupancy of the transition metal ion on the lattice sites, the experimental situation with respect to the observed magnetic behavior – especially in thin films – is not yet fully understood. Also the band structure of solids is crucially dependent upon the atomic structure and site occupancy. It is therefore not surprising, that the computer simulation of the exact physical conditions in such complex systems as the AB_2O_4 ferrites is very difficult. Electronic structure calculations from first principles have estimated a smaller gap between spin-down than for spin-up band, which in tunnel experiments should result in a negative spin polarization of the tunnel current, due to the excess of spin-down electrons. In particular, for $CoFe_2O_4$ a very large exchange splitting of 1.28 eV of the conduction band was calculated. On the experimental side, however, recent studies using $CoFe_2O_4$ spin filter barriers show just very modest spin filtering at room temperature [9].

Perovskites Compared to the complex crystal structure of the ferrites, the cubic perovskite structure is relatively simple. It is therefore convenient to experimentally integrate thin perovskite layers as spin filter barriers into tunnel junctions. The perovskite structure follows the formula ABO_3, where the A atoms form the corners of the cubic cells, B atoms are located in the center and the oxygen atoms are situated in the faces' centers. In particular, it is possible to combine perovskite spin filter barriers with isostructural half-metallic ferromagnetic metals, such as $La_{2/3}Sr_{1/3}MnO_3$ (LSMO), which can be used as a spin analyzer to probe the spin filter efficiency (see Chapter 3).
Among the perovskites, $BiMnO_3$ is an established insulating and ferromagnetic oxide with a Curie temperature of 105 K. The insulating state of $BiMnO_3$ is very robust with respect to the deposition conditions. Up to now, however, no quantitative experimental determination of the exchange splitting of the conduction band has been reported. Ab initio calculations using density functional theory are somewhat contradicting in their predictions of the electronic

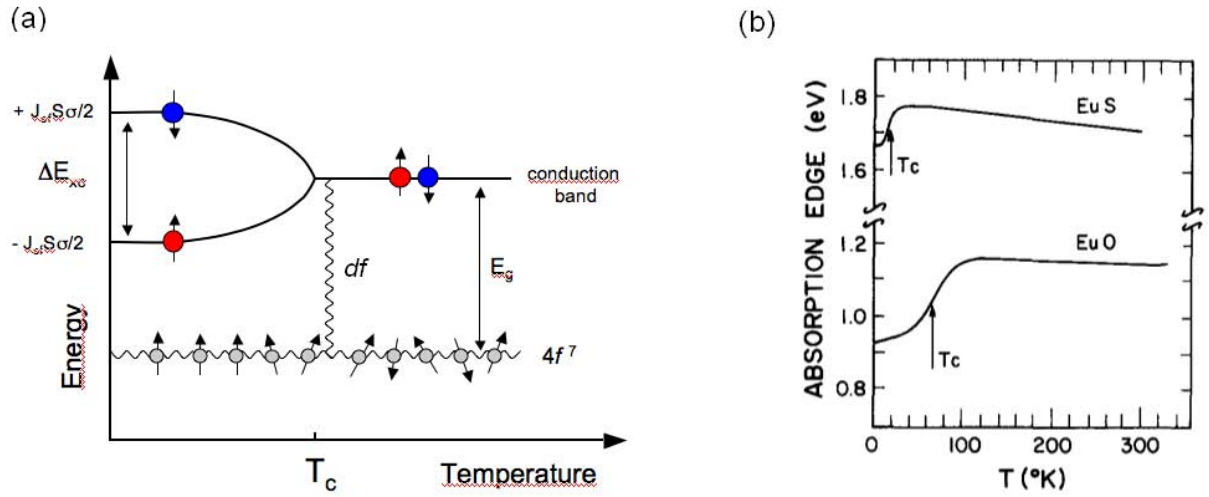


Fig. 2: *(a) Temperature dependent band structure of EuX compounds. Under the effect of magnetic exchange (J_{df}), the conduction band splits into a spin-up and spin-down sublevel below T_C. (b) Shift of the optical absorption edge of EuX compounds indicating the spin splitting of the conduction band under the effect of df exchange interaction. According to Ref. [1].*

properties, due to the difficulties modeling the highly correlated character of these materials. An exchange splitting of 0.5 eV was estimated from linear spin-density approximation (LSDA) calculations and 1.6 eV from the LSDA+U method, whereas the latter technique is commonly accepted to be more reliable to calculate band gaps [8]. In both cases, it is predicted that the bandgap is smaller for spin-up electrons. Therefore, when used as a spin filter, a $BiMnO_3$ layer should filter out spin-down electrons and produce a positively spin-polarized current.

2.4 Quantifying the spin filter effect

Evidence for the spin filter phenomenon can be obtained from the temperature dependence of the electrical resistance $R(T)$ of a magnetic tunnel barrier. Whereas in a nonmagnetic insulator the resistance increases continuously with decreasing temperature as thermal excitations are suppressed, a significant decrease in R occurs below the Curie temperature T_C of a magnetic insulator, if spin filtering is present. As discussed in Section 2.3, the conduction band splits below T_C and one spin type will tunnel preferentially. Because of the exponential dependence of the tunnel current on barrier height (recall Equation (2)), this results in a significant increase of conductivity (=decrease in resistance R) for the corresponding spin orientation with decreasing temperature. More quantitatively, the tunnel resistance R can be expressed within a simple free-electron tunnel model as

$$R_{\uparrow(\downarrow)}(T) \sim \exp\left(-d\,\Phi^{1/2}_{\uparrow(\downarrow)}(T)\right) \quad \text{with} \quad \Phi_{\uparrow(\downarrow)}(T) = \Phi_0 \pm J_{df} S\sigma(T), \tag{7}$$

where d is the barrier thickness, T is the temperature, Φ_0 is the average barrier height above T_C, J_{df} is the exchange constant, S is the spin quantum number and $\sigma(T)$ is the reduced magnetization $M(T)/M(T=0)$ of the magnetic insulator. The temperature dependence of the resistance thus scales exponentially with the magnetization of the spin filter. Therefore, ΔE_{xc} can directly be deduced from the tunnel measurement if the barrier heights Φ above and below T_C are known.

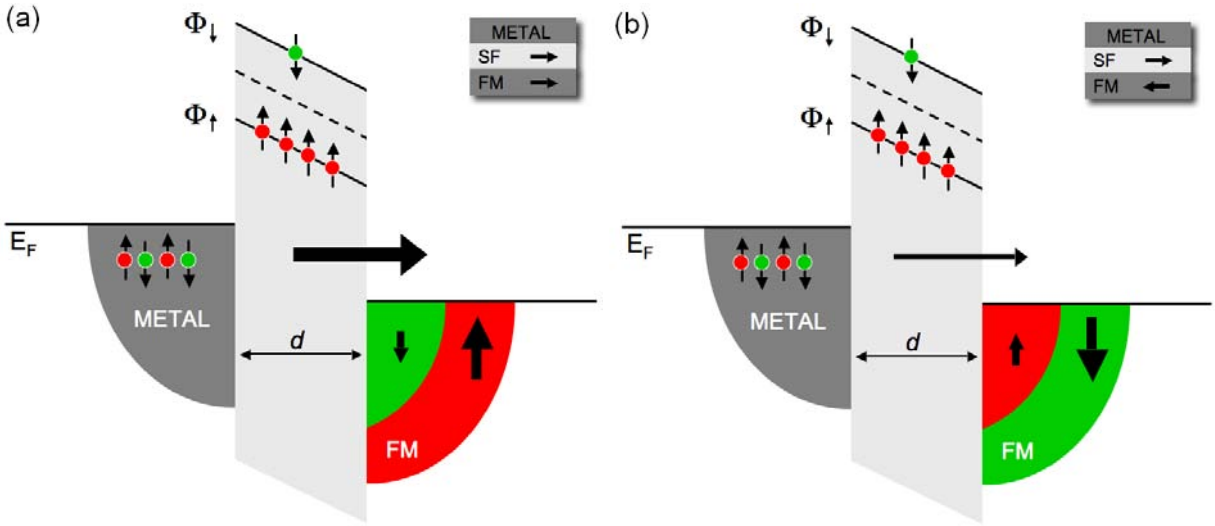


Fig. 3: *Schematic of the band diagram for a quasimagnetic tunnel junction (QMTJ). With a ferromagnetic (FM) counterelectrode, the tunnel current depends on the relative magnetization orientation between spin filter and FM. (a) For parallel alignment, a large current is measured, while for (b) antiparallel alignment, a small current results.*

The spin polarization of a tunnel current can be probed either by a superconductor (SC) or a ferromagnet (FM) as the spin detector. Superconducting tunneling spectroscopy is a standard method to determine the effective tunneling spin polarization that arises from the ferromagnetic part in FM/insulator/SC or metal/spin filter/SC junctions [2]. Typically, a superconducting aluminum (Al) electrode serves as the spin detector. If a ferromagnet is used as the spin detector, a magnetoresistance (MR) measurement can be applied to quantify the spin filtering efficiency of a magnetic tunnel barrier. The magnetic barrier functions as the spin polarizer that generates a highly polarized current, whereas a ferromagnetic counterelectrode serves as the spin analyzer. The latter collects differently the spins parallel and antiparallel to its magnetization relative to the spin filter barrier, which leads to large differences of the conductance. Therefore, the resistance of the device depends on the relative orientation of the magnetic moments of the magnetic insulating barrier and the ferromagnetic counterelectrode. Chapter 3 reviews the phenomenon of spin filter magnetoresistance and its applications in spintronic devices in more detail.

3 Spintronic devices with spin filters

3.1 Quasimagnetic tunnel junctions

The most straightforward approach to convert the spin filter effect into magnetoresistance is to integrate a magnetic insulator into a quasimagnetic tunnel junction (QMTJ). In such "spin-filter" MTJs, the spin filter tunnel barrier is sandwiched between a nonmagnetic and a ferromagnetic electrode. This combination slightly differs from conventional magnetic tunnel junctions (MTJs), in which two ferromagnetic electrodes sandwich a nonmagnetic insulator (see lecture by P. Grünberg). In both devices, however, the resistance of the junction depends on the relative alignment of the magnetization of the two magnetic layers. For a QMTJ, in the case of a parallel alignment of the spin filter barrier and the ferromagnetic electrode, a low junction

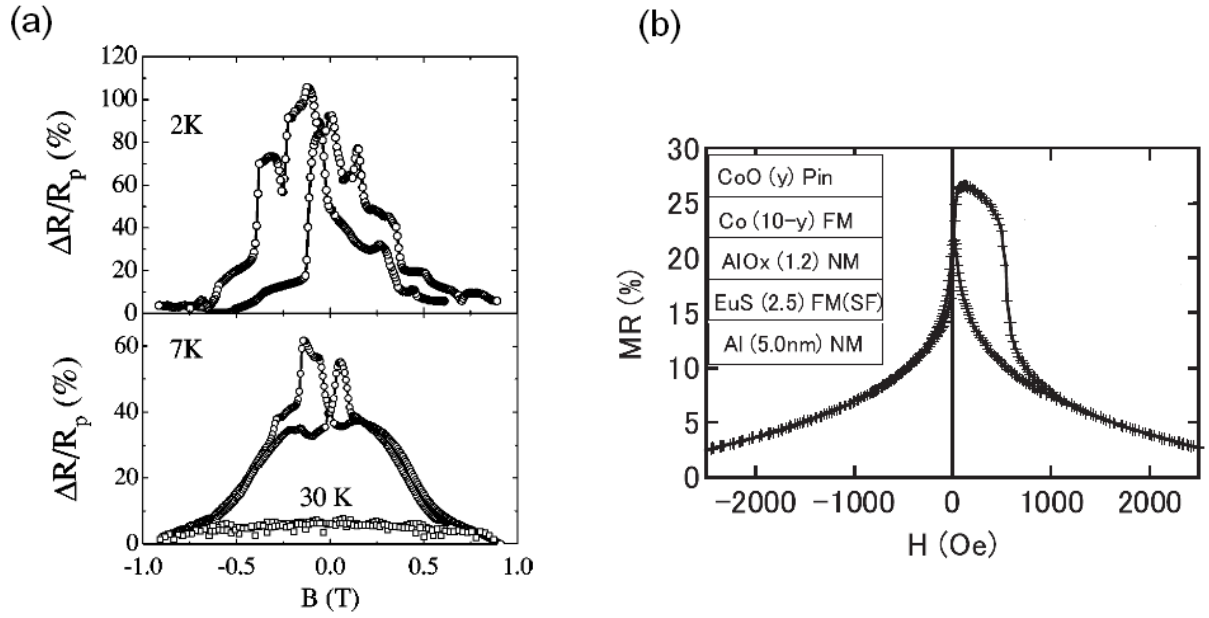


Fig. 4: *(a) Magnetoresistance (MR) as a function of magnetic field B for an Al/EuS/Gd quasi-magnetic tunnel junction (QMTJ), taken at $T = 2\,K$ (well below the EuS T_C) as well as at 7 K and 30 K (well above the EuS T_C). (b) MR of an Al/EuS/Al_2O_3/Co/CoO QMTJ recorded at $T = 4.2\,K$. An improved MR is observed if the EuS spin filter barrier and ferromagnetic electrode are magnetically decoupled by a very thin insulator.*

resistance R_p is measured. An antiparallel alignment results in a high junction resistance R_{ap}. This observation is termed "spin filter tunneling" magnetoresistance. Its characteristic quantity is the magnetoresistance (MR) ratio, which is defined – in the same manner as in conventional MTJs – as the variation of electrical resistance between the parallel (R_p) and the antiparallel (R_{ap}) state of magnetization of the two magnetic layers, i.e.

$$\mathrm{MR} = \frac{\Delta R}{R} = \frac{R_{ap} - R_p}{R_p}. \tag{8}$$

The magnitude of MR depends on the exponential dependence of the tunnel current on the barrier height, as one spin channel has a larger tunneling probability (see Section 2.1). Moreover, with a magnetic counterelectrode, we must consider the role of the spin-polarized density of states in the electrode as well. The tunnel current depends on the number of filled states in the first electrode as well as the number of available states in the second. Using one magnetic electrode, see Figure 3, the density of available states in the magnetic electrode is spin dependent, and the tunnel current will depend on the relative orientation of the filtered spins (i.e., the spin filter magnetization direction) and the electrode magnetization. As illustrated in Figure 3(a), for parallel alignment only spin up electrons tunnel through the spin filter barrier and thus they can only tunnel into majority (spin up) states in the magnetic electrode, resulting in a large current. For the antiparallel case (Figure 3(b)), the current is minimal since only the spin down states are available in the ferromagnet. Generally, this device can be considered analogous to a polarizer/analyzer optical configuration, although with non-perfect analyzer.

The magnitude of the expected "spin filter tunneling" magnetoresistance effect may be estimated within a simple two-current model, assuming spin conservation in the tunneling process

(see Sect.2.1), as

$$\mathrm{MR} = \frac{\Delta R}{R} = \frac{2P_{\mathrm{FM}}P_{\mathrm{SF}}}{1 - P_{\mathrm{FM}}P_{\mathrm{SF}}}, \tag{9}$$

where P_{FM} is the tunneling spin polarization of the ferromagnetic electrode and P_{SF} is the efficiency (polarization) of the spin filter. Eqn. (9) is basically an extension of the Jullière's formula (see lecture by D. Wortmann), from which the spin filter efficiency P_{SF} can be estimated if the spin polarization P_{FM} of the FM layer is known.

In order to observe MR, it is necessary to avoid magnetic exchange coupling between the spin filter barrier and the adjacent ferromagnetic electrode. Only if both layers show an independent magnetic switching behavior under an applied magnetic field, a well-defined TMR curve can be measured. This can be realized if the coercive fields of the spin filter barrier and FM electrode are different, so that they can reach both parallel and antiparallel alignment.

The first spin-filter experiment utilizing a QMTJ device structure was performed by LeClair *et al.* in 2002. A magnetoresistance ratio of up to 130% at $T = 2\,\mathrm{K}$ was obtained for a QMTJ with a 5 nm thick Europium Sulfide (EuS) barrier sandwiched between Aluminum (Al) and Gadolinium (Gd) electrodes as shown in Figure 4(a). Using the aforementioned simple model, the spin filter efficiency of the EuS barrier can be estimated being close to 90%. One can see from Figure 4(a) that there is considerable noise in the MR signal. This behavior is attributed to magnetic coupling between the adjacent EuS and Gd layers, which leads to instabilities in the magnetization reversal in the EuS layer upon sweeping the external magnetic field. The magnitude of TMR decreases as the temperature is raised closer to the Curie temperature T_{C} of EuS, and no magnetoresistance is observed above T_{C}.

An improved performance of a spin filter tunnel junction was realized by Nagahama *et al.*, who investigated QMTJs of the structure Al/EuS/AlO_x/Co/CoO. The AlO_x layer between the EuS barrier and the Co electrode magnetically separates the layers, whereas the antiferromagnetic CoO acts as an exchange bias layer pinning the magnetization of the Co layer. Figure 4(b) shows the MR curve of the junction at a temperature of $T = 4.2\,\mathrm{K}$ taken with a bias voltage of $V_{bias} = 5\,\mathrm{mV}$. Compared to the work on EuS junctions in Figure 4(a), the shape of the magnetoresistance curve is stable with no instability in junction resistance.

The potential of the $BiMnO_3$ perovskite as a spin filter tunnel barrier was investigated by Gajek *et al.*. In their experiments, Au-$BaMnO_3$-LSMO junctions showed a reasonable spin filter efficiency of about 22% at $T = 3\,\mathrm{K}$. A first step towards room temperature spin filtering was reported recently in a study by Ramos *et al.* with fully epitaxial spinel $CoFe_2O_3$-based tunnel junctions. The measurements reveal magnetoresistance values of -18% at 2 K and -3% at 290 K, indicating that magnetic tunnel barriers of spinel ferrites in principle could be used as spin filters over a broad temperature range.

Bias dependence in QMTJs For conventional MTJs with two ferromagnetic electrodes separated by a nonmagnetic insulator, such as Al_2O_3 or MgO, the tunnel magnetoresistance consistently decreases with increasing applied voltage. Reasons for this hallmark behavior are mechanisms like spin scattering in the barrier or excitation of spinwaves (magnons), which lead to the loss of spin information. Therefore, the operation of a conventional MTJ device is limited to low bias voltage for an optimum TMR effect. The bias dependence of a QMTJ, can show an opposite behavior. With a spin filter barrier involved, the magnetoresistance can increase with applied bias voltage.

It is the characteristic tunnel mechanism across the spin filter barrier that explains this unusual bias dependence. For a qualitative understanding, one divides the energy diagram of a QMTJ

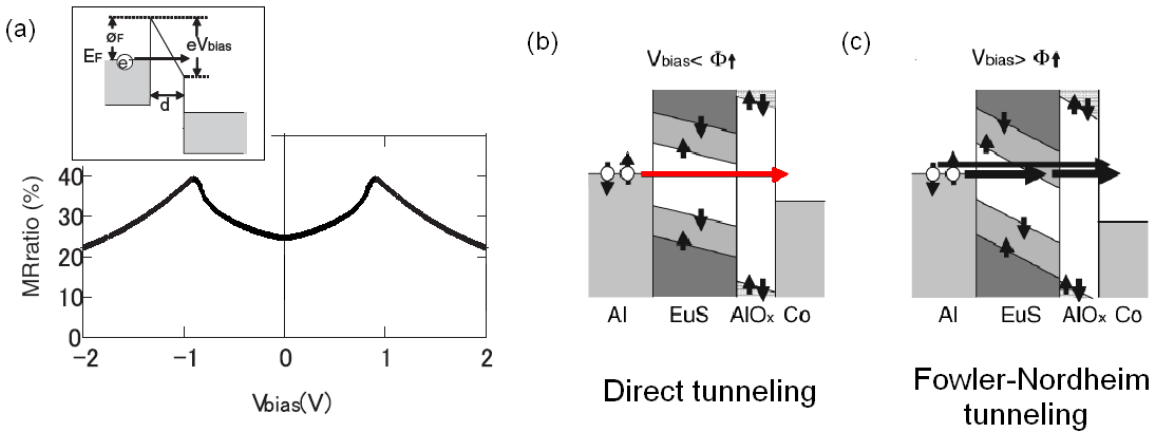


Fig. 5: *Magnetoresistance and its bias voltage dependence in QMTJs. (a) The bias dependence shows an continuous increase of the MR ratio up to high bias voltages, which is contrary to conventional magnetic tunnel junctions. It can be understood as due to different tunneling mechanisms through the spin-split conduction band of the spin filter barrier. (b) Schematic of the band diagram for a Al/EuS/AlO$_x$/Co QMTJ with $V_{bias} < \Phi_\uparrow$. Electrons must tunnel directly through both EuS and AlO$_x$ barriers. (c) Schematic band diagram for $V_{bias} > \Phi_\uparrow$. Spin-up electrons tunnel via Fowler-Nordheim tunneling through the spin-up conduction band, whereas spin-down electrons must still tunnel directly through the barrier. Adapted from Ref. [6].*

into three regions (Figure 5(b)): (i) A bias voltage may be applied which is either lower than the bottom of the spin-split conduction band, $V_{bias} < \Phi_\uparrow$, or (ii) may lie in between the spin-split conduction band, $\Phi_\uparrow < V_{bias} < \Phi_\downarrow$, or (iii) exceeds the upper conduction band edge $V_{bias} > \Phi_\downarrow$. Dependent on the bias voltage, different tunnel mechanisms are in effect. At low bias, $V_{bias} < \Phi_\uparrow$, electrons must tunnel through a trapezoidal barrier, termed as direct tunneling (DT). A different situation occurs at intermediate bias, $\Phi_\uparrow < V_{bias} < \Phi_\downarrow$, where the spin-up conduction band is below the Fermi level of the metal electrode. A narrower (spin-up) barrier is in effect, that allows spin-up electrons to tunnel via the conduction band. This mechanism is termed Fowler-Nordheim (FN) tunneling and is typically characterized by a triangular shape of the barrier and tunneling through only a part of the insulating layer. At high bias, $V_{bias} > \Phi_\downarrow$, Fowler-Nordheim tunneling occurs for both spin-up and spin-down channels, as the metallic Fermi level exceeds both barriers $\Phi_\uparrow$ and $\Phi_\downarrow$.
Fowler-Nordheim tunneling has a greater tunneling probability than direct tunneling [6]. Therefore, the spin polarization of the tunneling current increases for $V_{bias} > \Phi_\uparrow$ and causes a significant increase in magnetoresistance from where the transition from direct tunneling to Fowler-Nordheim tunneling takes place (Figure 5(a)). At even higher bias, the spin down conduction band also lowers below E_F of the metal electrode and results in a gradual reduction of MR. This special feature of a spin filter tunnel junctions allows to reach enhanced spin polarization and magnetoresistance at high bias voltages, which is not observable in conventional MTJs.

3.2 Double spin filter tunnel junctions

Whereas in quasimagnetic tunnel junctions a conventional ferromagnetic counter-electrode still is used to serve as the spin detector, one can imagine to realize magnetoresistance with spin filters without any ferromagnetic electrode involved. A device in which this idea is realized is a

double spin filter (SF) tunnel junction, which operates by solely tuning the tunneling probabilities $T_\uparrow(E) \neq T_\downarrow(E)$.
A double spin filter device consists of a tunnel junction formed by at least four layers: The outer two layers are nonmagnetic metal electrodes and function simply to provide an electrical contact to the inner layers. The inner two layers are the core part of the device and form the spin filter tunnel barrier. The key issue is that the two spin filter layers have to have unequal coercivity: One coercivity has to be chosen to be very large, so that the magnetic moment of this pinned layer stays fixed. The other layer has a smaller coercivity. This "free" layer may be switched back and forth by an external magnetic field. The junction resistance is smaller when the magnetization are parallel and larger, if they are antiparallel. This is due to the different barrier height profiles for spin up and spin down electrons as schematically sketched in Figure 6(a). When the free layer is parallel to the pinned layer, spin up electrons see a low barrier and spin down electrons a high barrier. Then, a relatively large spin up current flows, but only a very little spin down current. When the free and the pinned layers are magnetized antiparallel, both spin up and spin down electrons see a barrier profile consisting of one low barrier height section and one high barrier height section. This barrier configuration allows little spin up or spin down current flow. The magnetoresistance of a double spin filter junction depends exponentially on the relative alignment of the spin selective barrier heights of the two spin filter layers (recall Eqn.(1)). This makes the device theoretically very sensitive. A rough estimation of the MR ratio that theoretically can be expected for two spin filter barriers of few nanometer thickness is in the order of $\Delta R/R \simeq 10^5$ [11, 12]. Double spin filter barriers therefore could potentially exceed the sensitivity of conventional MTJs by at least two orders of magnitude.
The first experiment on spin filter double junctions has not been realized until very recently [13]. By combining double EuS spin filter barriers with Al electrodes, up to 60% MR was observed at a temperature of 1 K. This modest MR was attributed to imperfections of the polycrystalline structure of the layer stack and interfaces. One therefore can expect enhanced MR ratios from double SF barriers with improved crystallinity. In this work, the key issue of separating the coercive fields of both EuS layers was achieved by combining one EuS layer deposited at room temperature and the other quench-condensed at liquid nitrogen temperature ($T_{LN} = 77\,\mathrm{K}$) as shown in Figure 6(b). A thin Al_2O_3 spacer layer introduced in between the barriers was necessary to magnetically decouple both spin filter layers.
Although a double spin filter junction may lead to very sensitive MR devices, finding materials that give large signal above room temperature will be challenging. The requirements are finding a magnetic insulator with small band gaps and large exchange splitting, which can be grown as high-quality thin films. The perovskite family has the advantage that the magnetic layers can be grown relatively easily, but the disadvantage is that they are operating at most at liquid nitrogen temperatures. The second possible material system are ferrites, which can be grown epitaxially and are potentially both insulating and ferrimagnetic above room temperature.

3.3 Spin filtering in quantum wells

This section is intended to give an example of spin filter tunneling which goes beyond the standard concept of magnetic tunnel junctions. Using quantum well (QW) structures is an approach to realize spin filtering relying on the idea to select the spin character of electrons by resonant tunneling. This option becomes possible if the quantum well is made of magnetic material, and the energy levels are spin-split. The splitting of quantum well states enables one to select the resonant tunnel condition for each spin by applying the right bias voltage.

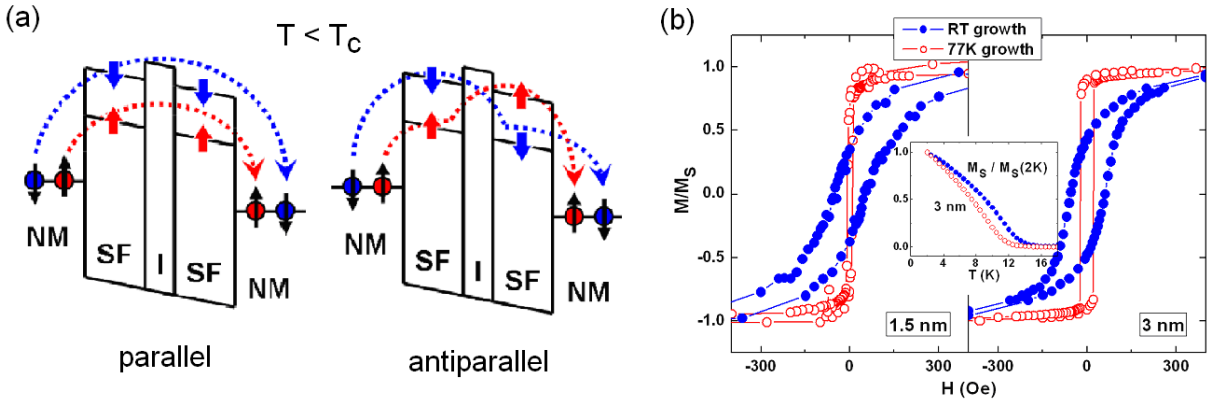


Fig. 6: *(a) Schematic illustration of the spin filter mechanism in double spin filter tunnel junctions below the Curie temperature T_C of the spin filter barrier. Dotted lines roughly indicate the barriers that the spin-up and spin-down electrons have to cross. (b) Magnetic hysteresis loops of thin EuS films deposited at 77 K and RT, film thicknesses are 1.5 nm and 3 nm, respectively. The coercivity strongly differs for the two deposition temperatures. After Ref. [13].*

A device in which spin-filter resonant tunneling can be observed is a so-called magnetic resonant tunneling diode. This structure is typically built of a fully semiconducting layer stack, in which two outermost layers are highly conductive and serve as the emitter and collector electrode. Those leads sandwich the heart of the device, that is a quantum well which is surrounded by two diamagnetic semiconducting layers serving as tunnel barriers. The well is typically made of a diluted magnetic semiconductor (DMS), in which magnetic ions are doped into conventional semiconductors to achieve ferromagnetism. If a constant external magnetic field is applied, the energy levels of the paramagnetic DMS exhibit a giant Zeeman splitting. At low temperatures, the energy levels in the well are split into spin-up and spin-down states. The amount of Zeeman splitting is controlled by the strength of the external magnetic field and can be in the order of several ten meV. Standard DMS materials typically require fields in the order of Teslas to reach the magnetic saturation, which is about one order of magnitude higher compared to “concentrated” magnetic semiconductors like the Eu chalcogenides.
Such a type of spin filtering device was realized in 2003 by Slobodskyy *et al.* [14]. They fabricated an all II-VI semiconductor resonant tunneling diode based on (Zn,Mn,Be)Se (see Figure 7(a)). The ZnBeSe layers serve as tunnel barriers and the quantum well states are formed in a ZnMnSe DMS layer. Figure 7(b) shows the current-voltage (I-V) characteristics of this resonant tunneling diode. Without an external magnetic field, a single resonant peak is present at about 125 meV. This resonance is split into two parts if a magnetic field is applied. The splitting thereby grows as a function of the magnetic field because it corresponds to the Zeeman splitting of the DMS quantum well. Each of the two peaks can be attributed to spin-up and spin-down quantum well states. One can therefore select the desired spin orientation of the current by applying a proper bias voltage to the device. The spin polarization of the current flowing through the device is thereby dependent on the external magnetic field B and the applied bias voltage V_{bias}. This experiment demonstrates the possibility of tunneling through spin resolved energy levels in semiconductors, with the special feature that the spin filter process can be voltage controlled. This idea may have potential utility for spintronic devices based on semiconductors, as will be introduced in the last section.

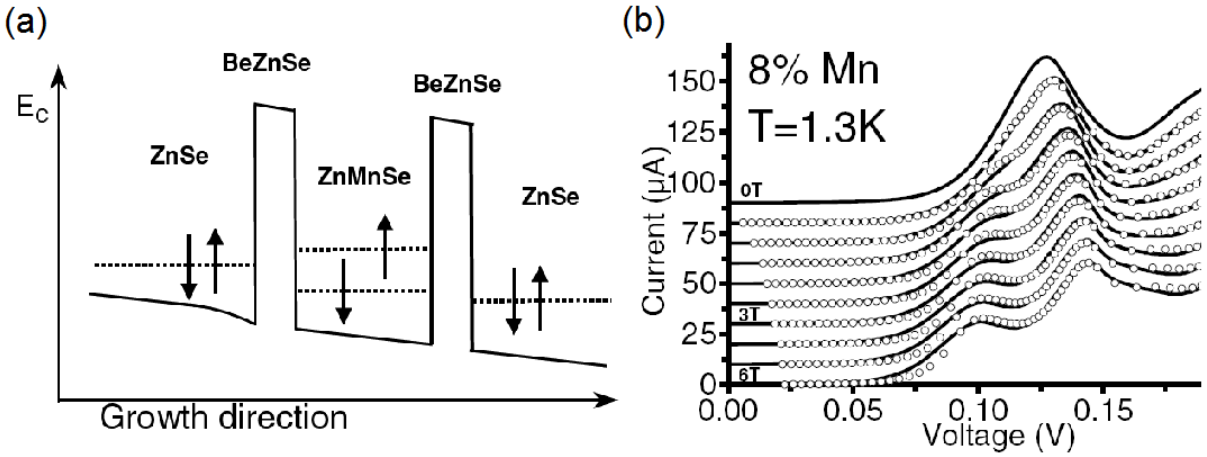


Fig. 7: *(a) Schematic view of a magnetic resonant tunneling diode (RTD) band structure under applied bias voltage after Ref. [14]. (b) Current-voltage characteristics for the RTD with ZnMnSe (Mn=8%) in the quantum well. The curves are taken from zero applied magnetic field up to B=6T.*

3.4 Spin injection through spin filters

Generating highly spin polarized tunnel currents is a very useful tool for realizing magnetoresistive effects, but also for making use of spin-polarized currents in semiconductors (SCs). Research is driven by the aim to incorporate spins into existing semiconductor technology and currently is one of the most active areas in the field of spinelectronics. The spin orientation can survive over long distances in semiconductors – for example over 100 microns in GaAs. This capability is attractive for realizing spintronic devices as, for example, spin field effect transistors or spin-light emitting diodes (spin-LEDs). In order to operate any SC-based spintronic device, the following steps have to be realized: i) generation and efficient injection of spin polarized electrons into a semiconductor, ii) coherent transport of spin through the semiconductor, and finally iii) detection of the spin. It has turned out, that the initial step – the so-called spin injection – already represents a major challenge. The earliest approach to electrically inject spins was to use a ferromagnet (FM) as the source of spin polarized carriers, but many experimental trials have been unsuccessful. A theoretical model was put forward that explains the problem encountered at the FM/SC interface: The very large spin-dependent conductivity in the FM is in contrast to a very small spin-independent conductivity in the semiconductor, such that the spin-injection coefficient is negligibly small [15]. In order to overcome this obstacle, it was proposed that the only possibility to effectively transmit spins across the FM/SC interface would be that either the current is initially almost 100% spin-polarized or that a tunneling barrier is inserted in between. The tunnel barrier would have the same low conductivity as the semiconductor, but in addition is spin selective.

One route to meet the above conditions is to make use of the spin filter effect of a magnetic tunnel barrier. An advantage is the potentially high spin filter efficiency which could produce the near 100% spin polarization that is required for an efficient spin injection. In addition, it could alleviate the problem of the strong interface sensitivity of standard tunnel barriers: As the spin filtering efficiency at first order depends on the barrier heights and not on interface density of states (recall Section 2.2), it should be more robust with respect to variations in interfacial properties. An interesting property of the Eu chalcogenides is their tunable conductivity ($\propto$ free

carrier concentration) which can be adjusted over a wide range by electron doping. In this way, the Europium compounds can be used as spin filters in the insulating state and as spin injectors when doped.
Experimental work integrating spin filters with semiconductors for injection and detection of spin polarized currents still has some way to go before catching up with theoretical models. Many problems simply arise from the growth of the according materials on semiconductors, which can lead to the formation of undesired compounds and alloys at the interface or interfacial electronic states, just to name a few. An initial step was done by studying the current transport across EuS/GaAs spin-filter/semiconductor contacts and to estimate its polarization detection efficiency [16]. The thin Schottky barrier produced by a 100nm thick doped EuS layer in contact with GaAs performs both the function of a highly polarized source as well as an insulating spin filter barrier. The exchange splitting of the EuS conduction band was determined by measuring the current-voltage (I-V) characteristics at temperatures above and below the T_C of EuS. The exchange splitting ΔE_{xc} was quantified from the the relative shift of the I-V curves towards smaller voltages for $T < T_C$, which is the signature of a lowered barrier height for electrons tunneling from EuS into the GaAs electrode. In addition, a shift of the current-voltage characteristics with temperature in the reversed bias regime (electron injection from GaAs into EuS) was observed suggesting spin filtering of the unpolarized electrons coming from GaAs. In this way, the performance of the EuS/GaAs contact as an spin injector and spin detector was probed individually. The observed effect in the current-voltage characteristics implies spin injection and detection though no spin analyzer is explicitly used.

4 Concluding remarks

Spin filter tunneling is an interesting phenomenon which has been revisited since recently for spin electronics. The potential of spin filter tunnel barriers to generate currents with very high spin polarization is essential to realize large spin-dependent effects for example in magnetic tunnel junctions, spin filter tunnel diodes or semiconductor-based devices. Unique features can be observed in devices based on spin filters, for example, enhanced magnetoresistance ratios at high bias voltages or the possibility to operate spin filtering voltage-controlled. However, until today establishing spin filtering at room temperature is still a major challenge. In either case, future studies relying on the spin filter effect will hold interesting physics ready.

References

[1] A. Mauger and C. Godart, *The magnetic, optical, and transport properties of representatives of a class of magnetic semiconductors: The europium chalcogenides* (Physics Reports **141**, 2 (51) 1986)

[2] J. S. Moodera, T. Santos and T. Nagahama, *The phenomena of spin-filter tunnelling* (J. Phys.: Condens. Matter **19** (165202) 2007)

[3] R. Schiller and W. Nolting, *Thickness dependent Curie temperatures of ferromagnetic Heisenberg films* (Sol. State Comm. **110**, (121) 1999)

[4] J. S. Moodera, X. Hao, G. A. Gibson, and R. Meservey, *Electron-spin polarization in tunnel junctions in zero applied field with ferromagnetic EuS barriers* (Phys. Rev. Lett. **61**, 5 (637) 1988)

[5] P. LeClair, J. K. Ha, H. J. M. Swagten, J. T. Kohlhepp, C. H. van de Vin, and W. J. M. de Jonge, *Large magnetoresistance using hybrid spin filter devices* (Appl. Phys. Lett. **80**, 4 (625) 2002)

[6] T. Nagahama, T. Santos, and J. S. Moodera, *Enhanced magnetotransport at high bias in quasimagnetic tunnel junctions with EuS spin filter barriers* (Phys. Rev. Lett. **99**, (016602) 2007)

[7] U. Lüders, M. Bibes, K. Bouzehouane, E. Jaquet, J.-P. Contour, S. Fusil, J.-F. Bobo, J. Fontcuberta, A. Barthélémy, and A. Fert, *Spin-filtering through ferrimagnetic $NiFe_2O_4$ tunnel barriers* (Appl. Phys. Lett. **88**, 8 (082505) 2006)

[8] M. Gajek, M. Bibes, A. Barthélémy, K. Bouzehouane, S. Fusil, M. Varela, J. Fontcuberta, and A. Fert, *Spin filtering through ferromagnetic $BiMnO_3$ tunnel barriers* (Phys. Rev. B **72**, (020406(R)) 2005)

[9] A. V. Ramos, M.-J. Guittet, J.-B. Moussy, R. Mattana, C. Deranlot, F. Petroff, and C. Gatel, *Room temperature spin filtering in epitaxial cobalt-ferrite tunnel barriers* (Appl. Phys. Lett. **91**, (122107) 2007)

[10] G. Hu and Y. Suzuki, *Negative spin polarization of Fe_2O_3 in magnetite/manganite-based junctions* (Phys. Rev. Lett. **89**, (276601) 2002)

[11] D. Worlegde and T. Geballe, *Magnetoresistive double spin filter tunnel junction* (J. Appl. Phys. **88**, (5277) 2003)

[12] A. Saffarzadeh, *Tunnel magnetoresistance in double spin filter junctions* (J. Phys.: Condens. Matter **15**, (3041) 2000)

[13] G.-X. Miao, M. Müller, and J. S. Moodera, *Magnetoresistance in double spin filter tunnel junctions with nonmagnetic electrodes and its unconventional bias dependence* (Phys. Rev. Lett. 2009)

[14] A. Slobodskyy, C. Gould, T. Slobodskyy, C. R. Becker, G. Schmidt, and L. W. Molenkamp, *Voltage-Controlled Spin Selection in a Magnetic Resonant Tunneling Diode* (Phys. Rev. Lett. **90**, (246601) 2003)

[15] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip and B. J. van Wees, *Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor* (Phys. Rev. B **62**, (4790(R)) 2000)

[16] J. Trbovic, C. Reng, P. Xiong, and S. von Molnár, *Spontaneous spin-filter effect across EuS/GaAs heterojunction* (Appl. Phys. Lett. **87**, (082101) 2005)

B 5 Spin-orbit coupling in semiconductor devices

Andreas Bringer
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Bloch wave functions including spin-orbit interaction **2**

3 Dynamics of Electrons in crystals **5**

4 Restricted geometry **7**

5 Spin manipulation by electron transport through wires **9**

A Spin-orbit matrix-elements **13**

B p-band edge states with spin-orbit coupling **13**

1 Introduction

This lecture deals with effects generated by the spin-orbit interaction in semiconductors. The influence of spin-orbit coupling on the band structure of semiconductors was recognized already in the 1950's and intensively studied, e.g. by Dresselhaus [1]. Most prominent is the splitting of the sixfold degeneracy of the p-valence states in the vicinity of the Γ-point. Like in atoms the p-states form two- and fourfold degenerate states energetically separated proportional to the spin-orbit interaction strength. With the development of epitaxial growth techniques and structuring of devices by etching and gates on nanometer-scale studies of spin-orbit effects on the transport of the electrons were taken up again, with the goal of using also the spin for signal transmission. Because of spin-orbit coupling the spin of an electron moving in an electric field sees a magnetic field and may carry out a precession motion. This is detectable via spin sensitive detectors (optical, injection into ferromagnetic material). This lecture will collect the basic theoretical ingredients to describe such effects (an extensive survey over experiments and theory is given in [2]).
In section 1 the main features introduced by spin-orbit coupling into the band structure of a semiconductor will be repeated (cf. lectures of G.Bihlmayer and Ph. Mavropoulos) and their theoretical derivation is outlined. Section 2 is dedicated to the physical consequences of the band structure when electrons in a semiconductor are influenced by external electric and magnetic fields. Their motion can be described by an *effective mass* Schrödinger equation. The parameters, m^* (effective mass), g^* (effective gyromagnetic factor of the electron spin) and effective spin-orbit coupling parameters are determined by appropriate matrix elements between band edge Bloch functions. Section 3 is occupied with the special situation of electrons moving in reduced dimension. Many hetero-structures consisting of layers of chemically different materials may trap electrons between the layers and support a 2-dimensional electron gas (*2DEG*) with the electrons moving freely parallel to the layers. To transport signals the geometry is further restricted to quasi 1-dimensional contacts. If the material is clean enough to make the mean free path of electrons of the order of the length of the contact they are called *quantum-wires*. The Schrödinger equation including spin-orbit interaction for the electron distribution in a wire is derived. In section 4 a specific example is studied. The focus is on the the dynamics of the electron spin during the passage through the wire. The phenomena of spin precession and spin polarisation are studied.

2 Bloch wave functions including spin-orbit interaction

The motion of the electrons in a crystal is governed by a periodic potential: $V_C(\vec{r}) = V_C(\vec{r}+\vec{R})$ ($\forall\ \vec{R}$, lattice vector). As a consequence the states of an electron are Bloch-states organized in bands (cf. lecture of G. Bihlmayer). This remains true, when spin-orbit coupling

$$\mathcal{H}_{SO} = \frac{\hbar}{(2m_e c)^2}\vec{\sigma}\cdot\left[\nabla V \times \left(\vec{p}+e\vec{A}\right)\right] \ , \tag{1}$$

is included. m_e is the mass of a free electron and e its charge. As required by relativity, an electric field $\vec{E} = -\nabla V/e$ is seen by the spin $\vec{S} = \hbar/2\vec{\sigma}$ ($\vec{\sigma}$ Pauli matrices, cf. lecture of G. Bihlmayer) of an electron at position $\vec{r}$ moving with velocity $\vec{v} = (\vec{p}+e\vec{A})/m_e$ in its rest frame as a magnetic field, $\mathcal{H}_{SO}$ is the Hamiltonian connected to it. $\vec{A}$ is the vector potential of an external magnetic field $\vec{B} = \nabla \times \vec{A}$ and $\vec{p}$ the canonical momentum of the electron.

Without magnetic field the Schrödinger equation including spin-orbit interaction is given by

$$\left[\frac{1}{2m_e}\vec{p}^2 + V_C + \frac{\hbar}{(2m_e c)^2}\vec{\sigma}\cdot\nabla V_C \times \vec{p}\right]\psi = \epsilon\psi \ . \tag{2}$$

In a semiconductor the Fermi-energy falls into an energy gap and the relevant states for the electrons are in the vicinity of the Γ-point (or possibly another symmetry-point) of the Brillouin-zone. It is useful to start with the solutions at the Γ-point, the lattice periodic "band-edge functions"

$$u_{0,\nu}\left(\vec{r},s\right)\ ;\ \epsilon_{0,\nu}\ :\ \nu = 0,1,2,\cdots\ ,\ s = \pm 1\ , \tag{3}$$

with

$$\langle\nu|\nu'\rangle = \frac{1}{\mathcal{V}_U}\sum_{s=\pm}\int_U d^3\vec{r}\, u_{0,\nu}u_{0,\nu'} = \delta_\nu^{\nu'}\ . \tag{4}$$

U is the Wigner–Seitz-unit-cell and $\mathcal{V}_U$ its volume.
Eq. (2) is very similar to the Schrödinger equation of a single atom, only the symmetry is reduced. In an atom the symmetry is fully spherical. With spin-orbit interaction the total angular momentum $\vec{J}$, sum of orbital momentum $\vec{L} = \vec{r}\times\vec{p}$ and spin $\vec{S}$, is a conserved quantity and classifies the states. The quantum numbers j of $\vec{J}$ are half integral, the degeneracy of the states, $2j+1$, is even. s-($l=0$)-states are twofold degenerate. For p-($l=1$)-states $\vec{L}$ and $\vec{S}$ are coupled to twofold degenerate $j = 1/2$-states and fourfold degenerate $j = 3/2$-states separated by the spin-orbit energy $\Delta_{so} = \left(\frac{\hbar}{2m_e c}\right)^2\langle V'/r\rangle$. For Eq. (2) there are only a finite number of symmetry operations defining the point group of the crystal. Most common for semiconducting crystals is tetrahedral symmetry (point group O_h). In "III-V"-compounds like GaAs the atoms of the components occupy face-centered-cubic lattice positions shifted against each other in space diagonal direction by a quarter of the lattice constant. Each atom sits in the center of a tetrahedron formed by four atoms of the other component (Fig. 1). The band-

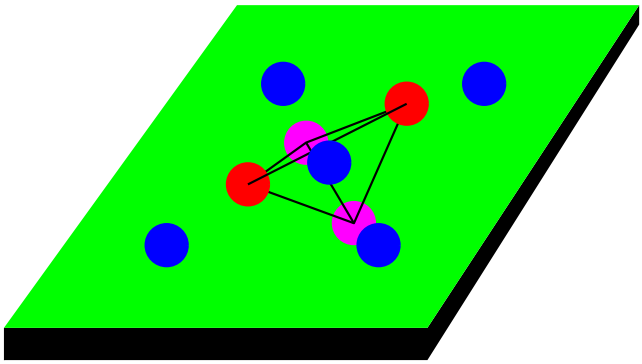


Fig. 1: *atomic positions in III-V-compounds*

edge functions of valence (v)- and conduction (c)-bands develop out of s- and p-states of the atoms (4s,4p for GaAs). Their shape is unchanged close to the nuclei because of the dominant spherically symmetric Coulomb potential. If the spin-orbit interaction is neglected, the threefold degeneracy of the p-states is not removed in tetrahedral symmetry. There are two atoms per unit cell. This leads to formation of functions with and without node between the neighbors (called *anti bonding*, *bonding* functions). In leading order the spin-orbit interaction does not change the energies of s-like states because diagonal matrix-elements $\langle\nu|\mathcal{H}_{SO}|\nu\rangle$ vanish (cf. Eq. (18)). For p-like states $u_{0,\nu}\ :\ \nu = 1,2,3$ the matrix-elements $\langle\nu|\mathcal{H}_{SO}|\nu'\rangle$ are equal for all pairs ν,ν' by symmetry (cf. Eq. (19)). The resulting eigenvalue equation has two solutions, one fourfold and one twofold degenerate (cf.Eq. (20)). The eight electrons per unit cell fill the bonding s,p - band states. The Fermi-energy is in the gap between the bonding p- and anti bonding s-band. The resulting level scheme is illustrated in Fig. 2.

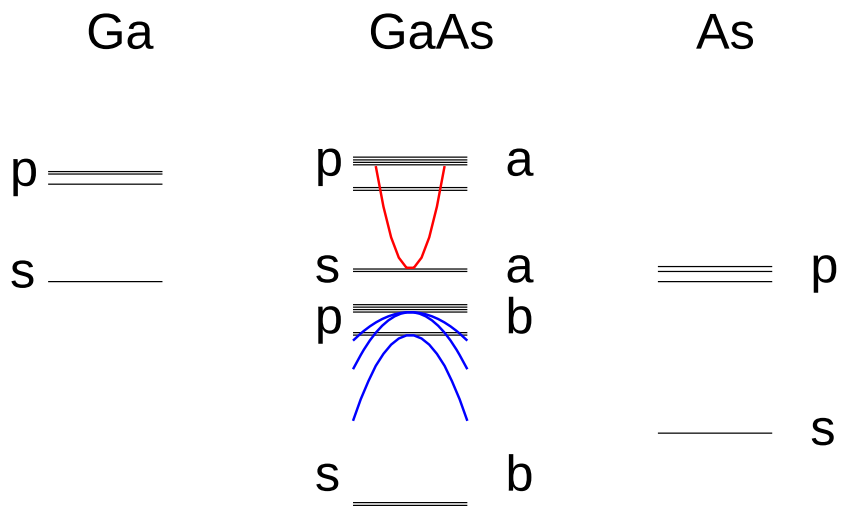


Fig. 2: *level scheme for a typical III-V-compound*

At $\vec{k} \neq 0$ the periodic part $u_{\vec{k},\nu}$ of Bloch-functions $\psi_{\vec{k},\nu} = u_{\vec{k},\nu}\ \exp(\imath\vec{k}\vec{r})$ obeys the equation

$$\left[\frac{1}{2m_e}\left(\vec{p}+\hbar\vec{k}\right)^{\,2} + V_C + \frac{\hbar}{(2m_e c)^2}\vec{\sigma}\cdot\nabla V_C \times \vec{p}\right] u_{\vec{k},\nu} =$$

$$\left[\frac{1}{2m_e}\vec{p}^{\,2} + V_C + \frac{\hbar}{m_e}\vec{k}\vec{p} + \frac{\hbar^2 k^2}{2m_e} + \frac{\hbar}{(2m_e c)^2}\vec{\sigma}\cdot\nabla V_C \times \vec{p}\right] u_{\vec{k},\nu} = \epsilon_\nu\left(\vec{k}\right) u_{\vec{k},\nu}. \qquad (5)$$

as long as $g\,k \ll 1$ (g lattice constant) perturbation expansion of the energies with respect to $\vec{k}\vec{p}$ converges. There are no corrections of the energy linear in k. All matrix-elements of $\vec{p}$ between wave-functions transforming according to the same irreducible representation of the

point group of the crystal are zero. To second order in k the energy is

$$\epsilon_\nu\left(\vec{k}\right) = \frac{\hbar^2 k^2}{2m_e} + \sum_{\mu\neq\nu} \frac{\hbar^2 \left\langle \nu \left| \vec{k}\vec{p} \right| \mu \right\rangle \left\langle \mu \left| \vec{k}\vec{p} \right| \nu \right\rangle}{m_e^2 \left(\epsilon_\nu - \epsilon_\mu\right)} \tag{6}$$

It is a linear combination of products of the components $k_x.k_y, k_z$ of $\vec{k}$ which has all symmetries of the crystal. It must be proportional to k^2. For the conduction band starting from the sa-state the dominant contributions come from the pb-states. ¡

$$\epsilon_c\left(\vec{k}\right) = \frac{\hbar^2 k^2}{2m_e}\left(1 + 2\sum_q \frac{\left|\langle sa\left|\vec{p}\right|pb_q\rangle\right|^2}{m_e\left(\epsilon_{sa}-\epsilon_q\right)} + 2\sum_d \frac{\left|\langle sa\left|\vec{p}\right|pb_d\rangle\right|^2}{m_e\left(\epsilon_{sa}-\epsilon_d\right)}\right) \tag{7}$$

(The indexes q,d indicate the splitting of the pb-states into a quartet and a doublet by spin-orbit coupling.) The bracket in Eq. 7 is larger than 1, meaning a reduction of the mass.

$$m^* = m_e\left(1 + \frac{|\mathcal{P}_q|^2}{m_e\Delta\epsilon_q} + \frac{|\mathcal{P}_d|^2}{m_e\Delta\epsilon_d}\right)^{(-1)} \tag{8}$$

Depending on the relative size of the matrix-elements and energy gaps m^* may get rather small ($.02\,m_e$ to $.07\,m_e$ in III-V compounds).

3 Dynamics of Electrons in crystals

The motion of electrons in free space is described by superpositions of plane waves. In crystals superpositions of Bloch waves are appropriate

$$\Psi = \sum_{\vec{k},\nu} \psi_{\vec{k},\nu}\chi_\nu\left(\vec{k}\right) \tag{9}$$

If the $\vec{k}$'s are concentrated around the Γ point of the Brillouin-zone, suitable Bloch-functions normalized with respect to the volume $\mathcal{V}$ of the crystal are

$$\psi_{\vec{k},\nu} = \exp\left(\imath\vec{k}\vec{r}\right) u_{0,\nu}\left(\vec{r},s\right)/\sqrt{\mathcal{V}} \tag{10}$$

formed with the band-edge functions which are independent on $\vec{k}$. With them one may separate the action of external forces from the action of the crystal-forces inside the Wigner-Seitz-cell. The crystal-forces vary on the scale of the lattice constant g, external forces on a scale S and normally $S \gg g$. The wave function Ψ may be written in the form

$$\Psi = \sum_\nu u_{0,\nu}\left(\vec{r},s\right)\Xi_\nu\left(\vec{r},s\right)$$

with a slowly varying *envelope*-function

$$\Xi_\nu(\vec{r}) = \sum_{\vec{k}} \exp(\imath\vec{k}\vec{r})\chi_\nu(\vec{k})\ .$$

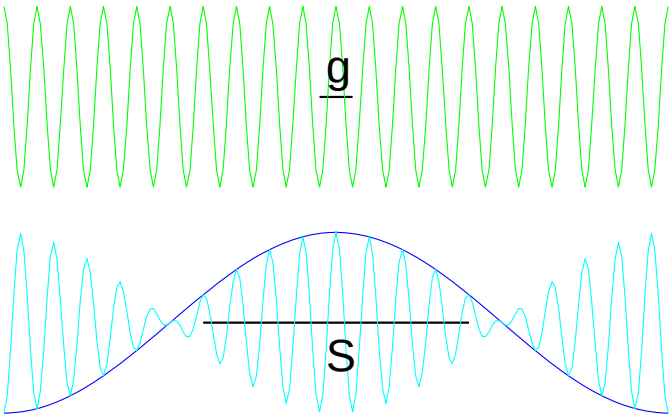


Fig. 3: *Schematic picture of the variation of u_0 (upper curve), Ξ and Ψ (lower curves blue and cyan, respectively) along a chain of crystal atoms.*

The external potentials V and $\vec{A}$ in the Schrödinger equation

$$\left[\frac{1}{2m_e}\vec{\mathcal{K}}^2 + V_C + V + g\mu_B\vec{\sigma}\cdot\left(\vec{B} + \frac{\nabla\left(V_C+V\right)}{2em_ec^2}\times\vec{\mathcal{K}}\right) - \mathcal{E}\right]\ \Psi\ =\ 0$$

($\vec{\mathcal{K}} = \vec{p} + e\vec{A}$ *kinetic* momenta) are constant across the Wigner-Seitz-cell and do not act on the u's. $\vec{p}$, V_C, $\vec{\sigma}$ do act on both the u's and Ξ's and lead to coupling of the ν's. Expansion with respect to ν yields a set of coupled differential equations for the envelope functions

$$0\ =\ \left[\epsilon_{0,\nu} - \epsilon + \frac{1}{2m_e}\left(\vec{p}+e\vec{A}\right)^2 + V + \mathcal{O}_\nu^\nu\right]\ \Xi_\nu + \sum_{\nu'\neq\nu}\mathcal{O}_\nu^{\nu'}\ \Xi_{\nu'}\ . \tag{11}$$

with coupling operators acting on the Ξ's

$$\mathcal{O}_\nu^{\nu'} = \langle\nu\left|\vec{\Pi}/m_e\right|\nu'\rangle\left(\vec{p}+e\vec{A}\right) + \langle\nu\left|\tilde{\vec{\Pi}}/m_e\right|\nu'\rangle\ \hbar\nabla V/(4m_ec^2)$$

$$+\langle\nu\left|\mu_B\vec{\sigma}\right|\nu'\rangle\cdot\left[\vec{B} + \nabla V/(2em_ec^2)\times\left(\vec{p}+e\vec{A}\right)\right]\ ,$$

defined by operators acting on the u_0's

$$\vec{\Pi} = \vec{p} + \hbar\vec{\sigma}\times\nabla V_C/(4m_ec^2)\,,\ \tilde{\vec{\Pi}} = \vec{p}\times\vec{\sigma}\,,\ \text{and}\ \vec{\sigma}.$$

Only $\vec{\sigma}$ in $\mathcal{O}$ has diagonal elements, acting on the envelope functions like an effective magnetic field with a spin-orbit contribution. The inter-band couplings may be resolved by the method of *back-folding*. For a particular state (or group of states), e.g. $\nu = c$ for conduction (or $\nu = v$ for valence) electrons, energetically separated from the other states, the reduced system

of equations with $\nu \neq c$ is a regular linear inhomogeneous system of equations. There is an operator $\mathcal{G}_\nu^{\nu'}$ which allows to express the Ξ_ν's by Ξ_c leading to

$$0 \;=\; \left[\epsilon_c - \epsilon + \frac{1}{2m_e}\left(\vec{p} + e\vec{A}\right)^2 + V + \;\mathcal{O}_c^c - \sum_{\nu,\nu' \neq c} \mathcal{O}_c^\nu \mathcal{G}_\nu^{\nu'} \mathcal{O}_{\nu'}^c\right] \Xi_c. \tag{12}$$

$\mathcal{G}$ may be expanded with respect to the off-diagonal part and a systematic expansion in terms of spin and kinetic momenta $\mathcal{K}_{x,y,z} = (p + eA)_{x,y,z}$ results. In lowest (second) order $\mathcal{G}$ is diagonal, $\mathcal{G}_\nu^{\nu'} = \delta_\nu^{\nu'}/(\epsilon_c - \epsilon_\nu)$, with the effect, that all diagonal terms get re-normalized, most important: the free electron mass m_e is replaced by an effective mass m^* and μ_B is replaced by $g^*\mu_B$ due to the non-zero commutators of the kinetic momenta $[\mathcal{K}_\alpha; \mathcal{K}_\beta] = B_\gamma$. In third order the spin-orbit-term is extended by

$$\mathcal{H}_D \propto \{\sigma_x \mathcal{K}_x, \mathcal{K}_y^2 - \mathcal{K}_z^2\} + \{\sigma_y \mathcal{K}_y, \mathcal{K}_z^2 - \mathcal{K}_x^2\} + \{\sigma_z \mathcal{K}_z, \mathcal{K}_x^2 - \mathcal{K}_y^2\} \tag{13}$$

called Dresselhaus term [1]. Here, $\{\mathcal{A}, \mathcal{B}\} = 1/2(\mathcal{A}\mathcal{B} + \mathcal{B}\mathcal{A})$ is the anti-commutator. To this order of approximation an effective mass Schrödinger equation

$$\mathcal{H}\,\Xi_c \;=\; \left[\frac{1}{2m^*}\vec{\mathcal{K}}^2 + V + g^*\mu_B\vec{\sigma}\cdot\left(\vec{B} + \frac{\nabla V}{2em_e c^2} \times \vec{\mathcal{K}}\right)\right.$$
$$\left. + \gamma_D \sum_{c.p.(x,y,z)} \{\sigma_x \mathcal{K}_x, \mathcal{K}_y^2 - \mathcal{K}_z^2\}\right] \Xi_c \;=\; \epsilon\,\Xi_c \tag{14}$$

results for the dynamics of electrons in a non degenerate band. The quantity γ_D is the material-dependent Dresselhaus parameter [2].

4 Restricted geometry

By epitaxial growth it is possible to form sandwiches of layers with different chemical compositions. Electro-chemical forces adjust the the Fermi-level and influence the energies of the electrons vertical to the layers, i.e. an external potential V acts on the electrons depending only on their position vertical to the plane of the layers. There are suitable combinations of layers, which generate *quantum-well* potentials V_{qw} as shown in Fig. 4. These potentials bind electrons in the central layer and enable the formation of a 2-dimensional free electron gas (2DEG). Inside the layer the electrons can move freely. The wave-functions in Eq. (14), $\Xi \sim \xi(z)\phi(x,y)$ factorize into functions ξ and ϕ, describing the distribution of the electrons across the quantum well and their motion in the plane, respectively. For electrons in the quantum well 2-dimensional sub-bands are formed. Their distance in energy $\delta\epsilon$ is determined by the thickness d of the active layer $\delta\epsilon \approx \hbar^2\pi^2/(2m^*d^2)$. At typical electron densities n only the lowest sub-band is occupied The Fermi energy is $\epsilon_F = \pi n\hbar^2/m^*$. If the barrier materials on both sides of the active layer are different, V_{qw} will be asymmetric with a nonzero average gradient, $\hbar^2\langle\xi\,|\nabla V_{qw}|\,\xi\rangle/(2m_e c)^2 = \alpha_R\vec{e}_z$. It defines the Rashba spin-orbit coupling constant α_R [3]. $\vec{e}_z$ is the unit vector vertical to the layer.

Electron transport may be further confined to a 1-dimensional system by preparing narrow wires with a lateral width W down to a few 10's of nm. The electrons see in addition to the quantum well potential a confining potential $V_W(x)$. As a reasonable guess, speculating

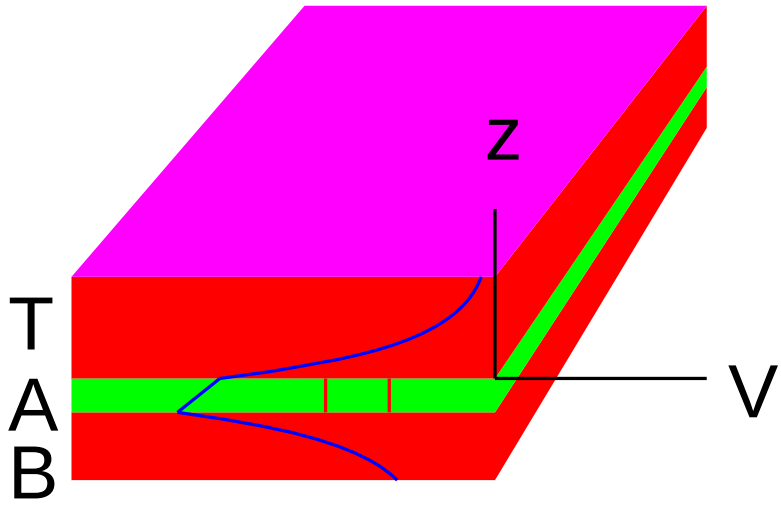


Fig. 4: *Sketch of the potential across a device consisting of three semiconducting layers. Electrons are trapped in the bound states of V_{qw}. The thickness of the active layers is typically around 5[nm] corresponding to a stack of 10 unit cells.*

on an approximately constant distribution of the charges across the wire, a *soft wall* potential $V_W = V_0/\left(\exp\left[\left(W/2-|x|\right)/\delta\right]+1\right)$ is used in the following. In a long wire the in plane part of Ξ factorizes again $\phi \sim \exp(\imath k_y y)\chi(x)$ into a plane wave factor for the motion along the wire and a spinor function χ. It determines the density and spin distribution across the wire via a second order differential equation following from Eq. (14). With a magnetic field $\vec{B} = B\vec{e}_z$ deduced from $\vec{A} = (\,0\,,\,B\,x\,,\,0\,)$ the kinetic momenta are

$$\mathcal{K}_x = \frac{\hbar\partial}{\imath\partial x}\,; \quad \mathcal{K}_y = \hbar k_y + eBx\,; \quad \mathcal{K}_z = \frac{\hbar\partial}{\imath\partial z}\,.$$

The Rashba- and Dresselhaus terms in Eq. (14) are

$$\mathcal{H}_R = \left[\sigma_x\left(\hbar k_y + eBx\right) - \sigma_y\frac{\hbar\partial}{\imath\partial x}\right]\frac{\alpha_R}{\hbar}$$

$$\begin{aligned}\mathcal{H}_D = \gamma_D\langle\xi\left|\mathcal{K}_z^2\right|\xi\rangle\left(\sigma_y\left(\hbar k_y + eBx\right) - \sigma_x\frac{\hbar\partial}{\imath\partial x}\right)\\ +\gamma_D/2\left[\sigma_x\left(\mathcal{K}_x\mathcal{K}_y^2 + \mathcal{K}_y^2\mathcal{K}_x\right) - \sigma_y\left(\mathcal{K}_y\mathcal{K}_x^2 + \mathcal{K}_x^2\mathcal{K}_y\right)\right]\,.\end{aligned} \tag{15}$$

Equation (14) is then

$$\left[\frac{1}{2m^*}\left(\mathcal{K}_x^2+\mathcal{K}_y^2\right) + V_W + \mathcal{H}_R + \mathcal{H}_D + g^*\mu_B\,B\right]\chi = \mathcal{E}\chi. \tag{16}$$

A list of values of the parameters α_R, γ_D for III-V compounds is given in [2] (Tables 6.3 and 6.6.). Eq. 16 is a system of ordinary differential equations and may be solved by any standard

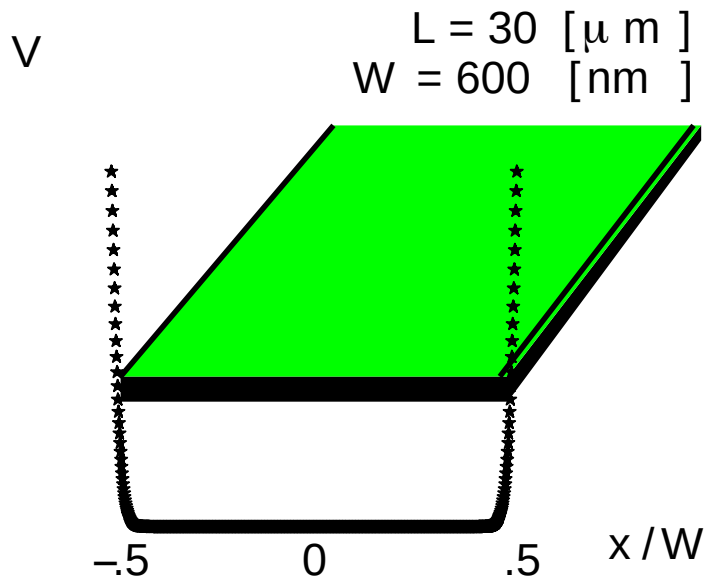


Fig. 5: *Sketch of the confining potential V_W. Typical dimensions of the wires are indicated.*

numerical differential equation solver. It should be pointed out that the full Dresselhaus term can be included. In existing treatments of the Dresselhaus term [4] only the first part - linear in $\mathcal{K}_{x,y}$ - of $\mathcal{H}_D$ was considered. This is justified as long as $k_y^2 \ll \langle \mathcal{K}_z^2 \rangle \approx 1/d^2$. However, at high enough density k_y^2 may come close to $1/d^2$ and the third order terms will contribute significantly [2]. In the following the standard definition $\beta_D = \hbar\gamma_D \langle |\mathcal{K}_z^2| \rangle$ will be used for the Dresselhaus coupling strength.
The 2-dimensional sub-band is split again into 1-dimensional sub-bands. However, the energetic separation is now $\Delta\mathcal{E} \approx \hbar^2\pi^2/(2m^*W^2)$. Typically $\Delta\mathcal{E} \ll \epsilon_F$, thus many bands cross ϵ_F. A typical χ for electrons at ϵ_F has a density distribution across the wire as shown in Fig. 6. It has 30 density peaks of almost equal height. There are 30 other χ with similar densities. The total density, sum of these is constant except for a region close to the walls. This justifies the assumption of a soft wall potential for wires.
Wire structures based on the material combination $Ga_xIn_{1-x}As$ (Gallium-Indium Arsenide), InP (Indium Phosphide) were investigated experimentally and theoretically in [5]. The experimental findings were analysed using Eq. 16. The parameters are in this case:

$$m* = 0.04\, m_e \quad \alpha_R = 5.4[meVnm] \quad \beta_D = 1[meVnm] \quad g^* = -14$$
$$n = 0.005[1/nm^2] \quad \epsilon_F = 33.[meV]$$

The following section is based on this example. The focus is on spin-orbit effects at zero magnetic field.

5 Spin manipulation by electron transport through wires

There is a one to one correspondence between a spinor

$$\chi = \begin{pmatrix} U \\ D \end{pmatrix} = \begin{pmatrix} f \\ g + i\,h \end{pmatrix}$$

defined by two complex components and a unit vector $\vec{e} = (e_x, e_y, e_z)$ defining the spin direction. Multiplying χ with the phase factor $U^*/|U|$ (physically irrelevant) gives a representation of χ with 3 real quantities f, g, h. The spin direction is then defined by

$$\vec{e}\cdot\vec{\sigma}\,\chi = \chi \quad \text{or} \quad \begin{pmatrix} e_z & e_x - \imath\, e_y \\ e_x + \imath\, e_y & e_z \end{pmatrix} \begin{pmatrix} f \\ g + i\, h \end{pmatrix} = \begin{pmatrix} f \\ g + i\, h \end{pmatrix}$$

The solution is

$$e_x = \frac{2fg}{f^2+g^2+h^2}\,, \quad e_y = \frac{2fh}{f^2+g^2+h^2}\,, \quad e_z = \frac{f^2-g^2-h^2}{f^2+g^2+h^2} \tag{17}$$

For the example state of Fig. 6 the local spin direction across the wire rotates clockwise from $+\vec{e}_z$ to $-\vec{e}_z$ in the half space $x < 0$ (cf. Fig. 7). On average it is oriented in negative x-direction.

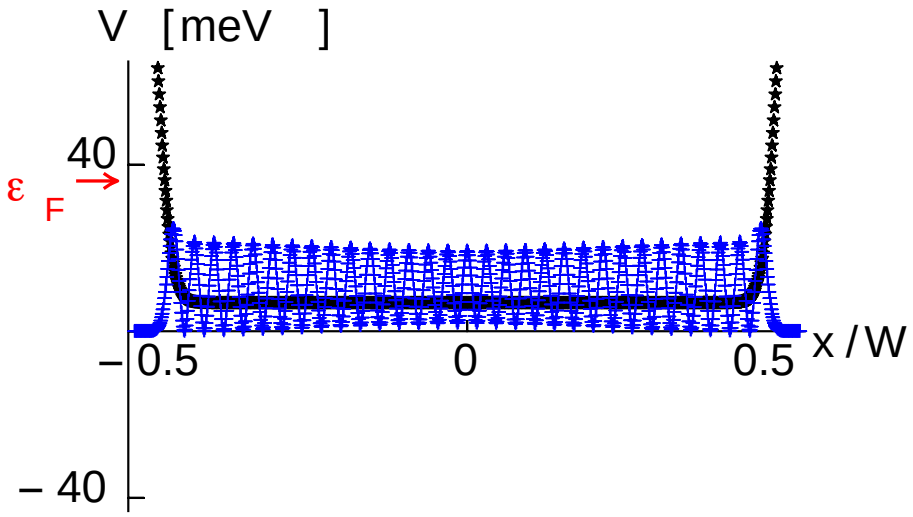


Fig. 6: *Density $|\chi|^2 = f^2 + g^2 + h^2$ of a spinor at ϵ_F, $k_y = 50/W$.*

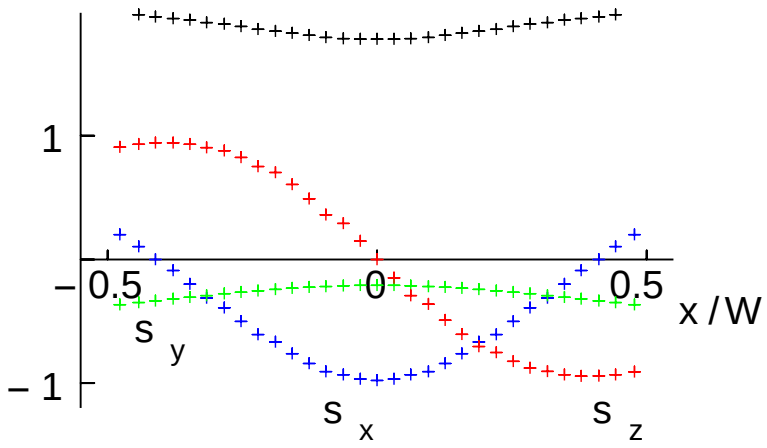


Fig. 7: *Spin direction Eq. 17 of χ Fig. 6.*

(The small y-component is due to the Dresselhaus-term. With the Rashba term alone all states would have the spins oriented in the (x, z)-plane.)

To each solution χ of Eq. 16 exists a solution $\tilde{\chi}$ with the same energy but inverted spin structure and slightly different wave number (for the example state $\delta k = -1.8/W$). This is in line with the general symmetry arguments discussed in the lectures of G. Bihlmayer and Ph. Mavropoulos. By superpositions of χ and $\tilde{\chi}$ states can be prepared with spins pointing in a specified direction at a given position in the wire. Along the wire χ and $\tilde{\chi}$ propagate with different phase velocities. The spin rotates along the y-direction. The length for one complete revolution is $\lambda_s = 4\pi/\delta k$. Based on this observation a spin transistor device has been proposed [6]. A

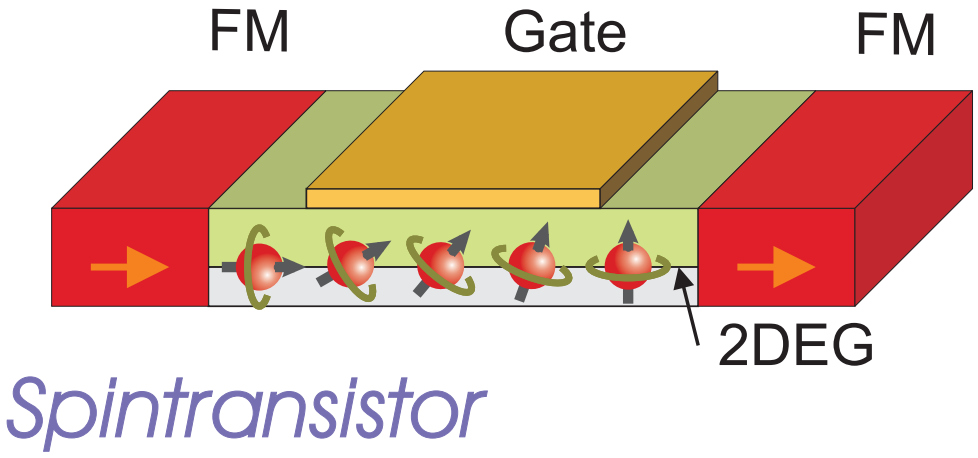


Fig. 8: *Datta-Das Spin-transistor [6]*

spin-polarised current is injected by a spin polarized source. The spins rotate along the wire and can be detected by a polarised drain. The rotation length λ_s can be manipulated by a gate and enable signal transmission. The problem here is that many spin-paired states cross ϵ_F. λ_s is different for all pairs. Even if an electron was injected with a well defined spin direction it would be a superposition of many states which rotate differently. The information on the original spin direction will soon be lost, a simple spin relaxation mechanism (cf. lecture of Ph. Mavropoulos).

The strength of the spin-orbit interaction increases with k_y as seen from Eq. (14). Neighboring bands are degenerate at $k_y = 0$ for a wire with left-right symmetry. With increasing k_y they have the tendency to cross each other (cf. Fig. 9) These crossings are resolved by $\mathcal{H}_{R,D}$, such

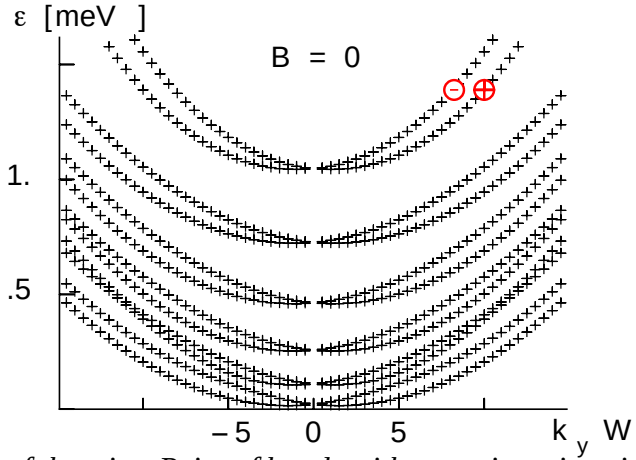


Fig. 9: *Band structure of the wire. Pairs of bands with opposite spin orientation indicated.*

that the spin orientation is transferred from the energetically higher to the lower state. A series

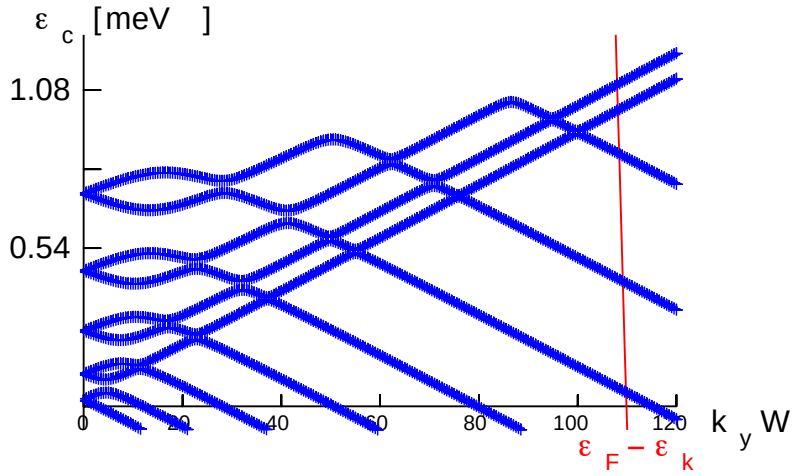


Fig. 10: *confinement energy ϵ_c, i.e. the total energy minus the kinetic energy in the y-direction, of the lowest bands*

of crossings appears along the k_y-axis (cf. Fig. 10). As a consequence at ϵ_F bands with the highest k_y will have the same spin orientation. For the case of square-well confinement the first eight bands have equal spin orientation The spin polarised part (red in Fig. 11) amounts to

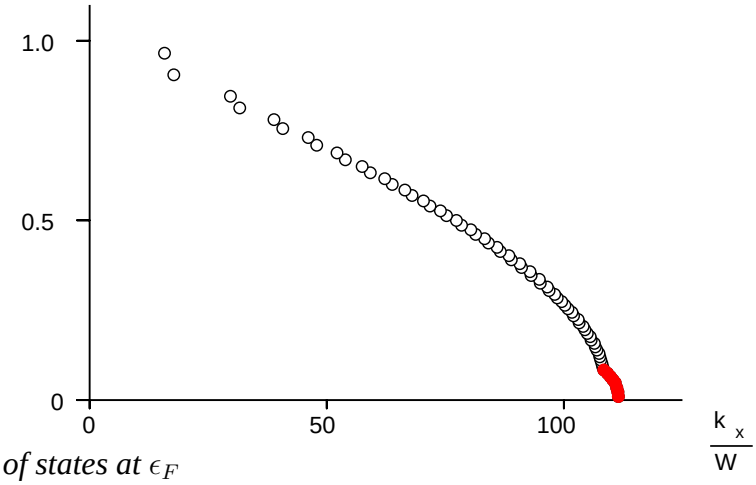


Fig. 11: *Density of states at ϵ_F*

approximately 10% of the total density of states at ϵ_F. With an appropriately charged gate these states could be filtered out.

Appendices

A Spin-orbit matrix-elements

The diagonal matrix-elements of the spin-orbit interaction $\langle\nu|\mathcal{H}_{SO}|\nu\rangle$ vanish. The defining integrals of $\nabla V \times \vec{p}$, e.g. for the z-component ($\vec{r} = (x, y, z)$):

$$\int_U d^3\vec{r}\, u_{0,\nu} \left(\frac{\partial}{\partial x} V_C \frac{\partial}{\partial y} u_{0,\nu} - \frac{\partial}{\partial y} V_C \frac{\partial}{\partial x} u_{0,\nu} \right) =$$

$$\int_U d^3\vec{r}\, V_C \left(\frac{\partial}{\partial x} \left[u_{0,\nu} \frac{\partial}{\partial y} u_{0,\nu} \right] - \frac{\partial}{\partial y} \left[u_{0,\nu} \frac{\partial}{\partial x} u_{0,\nu} \right] \right) = 0 \qquad (18)$$

are zero, as seen by partial integration. The integral over the boundary of the Wigner-Seitz cell U is zero because of the periodicity of the functions in the integrand.
Many off-diagonal matrix-elements $\langle\mu|\mathcal{H}_{SO}|\nu\rangle$ either vanish because of symmetry or may be transformed into one another by symmetry operations. A set of degenerate p-states has three functions $u_{0,x}, u_{0,y}, u_{0,z}$ which transform like the components of $\vec{r}$. Then e.g. for the pair $u_{0,x}, u_{0,z}$

$$\int_U d^3\vec{r}\, u_{0,x} \left[\nabla V \times \vec{p}\right]_z u_{0,z} = \int_U d^3\vec{r}\, u_{0,x} \frac{\hbar}{\imath} \left(\frac{\partial}{\partial x} V_C \frac{\partial}{\partial y} u_{0,z} - \frac{\partial}{\partial y} V_C \frac{\partial}{\partial x} u_{0,z} \right) =$$

(via $\langle x, y, z\rangle \to \langle -x, -y, z\rangle$ -symmetry)

$$- \int_U d^3\vec{r}\, u_{0,x} \frac{\hbar}{\imath} \left(\frac{\partial}{\partial x} V_C \frac{\partial}{\partial y} u_{0,z} - \frac{\partial}{\partial y} V_C \frac{\partial}{\partial x} u_{0,z} \right) = - \int_U d^3\vec{r}\, u_{0,x} \left[\nabla V \times \vec{p}\right]_z u_{0,z}$$

Matrix-elements with two equal indexes are zero. For the pair $u_{0,x}, u_{0,y}$

$$\int_U d^3\vec{r}\, u_{0,x} \left[\nabla V \times \vec{p}\right]_z u_{0,y} = \int_U d^3\vec{r}\, u_{0,x} \frac{\hbar}{\imath} \left(\frac{\partial}{\partial x} V_C \frac{\partial}{\partial y} u_{0,y} - \frac{\partial}{\partial y} V_C \frac{\partial}{\partial x} u_{0,y} \right) =$$

(partial integration and [x,y]-symmetry)

$$- \int_U d^3\vec{r}\, u_{0,y} \frac{\hbar}{\imath} \left(\frac{\partial}{\partial x} V_C \frac{\partial}{\partial y} u_{0,x} - \frac{\partial}{\partial y} V_C \frac{\partial}{\partial x} u_{0,x} \right) = \int_U d^3\vec{r}\, u_{0,y} \left[\nabla V \times \vec{p}\right]_z u_{0,x} \qquad (19)$$

All matrix-elements $\left\langle \alpha \right| \left]\nabla V \times \vec{p}\right]_\beta \left|\gamma\right\rangle$ with pairwise different indexes are equal. They vanish if there is an inversion center in the crystal. Inversion symmetry transforms the integrands into their negatives. In crystals without inversion symmetry (e.g. II-VI and III-V compounds) they may be different from zero and define $\Delta_{so} = \frac{\hbar}{(2m_e c)^2} \left\langle \alpha \right| \left[\nabla V \times \vec{p}\right]_\beta \left|\gamma\right\rangle$, the analog Δ_{so} of atom (cf. Sect. 2).

B p-band edge states with spin-orbit coupling

$[\nabla V \times \vec{p}]$ generates the matrices

$$\mathcal{M}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\imath\Delta_{so} \\ 0 & \imath\Delta_{so} & 0 \end{pmatrix} \mathcal{M}_y = \begin{pmatrix} 0 & 0 & -\imath\Delta_{so} \\ 0 & 0 & 0 \\ -\imath\Delta_{so} & 0 & 0 \end{pmatrix} \mathcal{M}_z = \begin{pmatrix} 0 & -\imath\Delta_{so} & 0 \\ \imath\Delta_{so} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Scalar product with $\vec{\sigma}$ extends these to a 6×6 matrix

$$\mathcal{M}_S = \begin{pmatrix} \mathcal{M}_z & \mathcal{M}_x - \imath\mathcal{M}_y \\ \mathcal{M}_x + \imath\mathcal{M}_y & \mathcal{M}_z \end{pmatrix}$$

The eigenvalue equation for the energy ϵ of the p-edge levels is (ϵ_0 : energy without spin-orbit coupling and $\mathcal{U}$: 6×6-unity matrix)

$$\det\left[(\epsilon_0 - \epsilon)\mathcal{U} + \mathcal{M}_S\right] = \left((\epsilon_0 - \epsilon + \Delta_{so})^2 (\epsilon_0 - \epsilon - 2\Delta_{so})\right)^2 = 0$$

has two solutions: $\epsilon = \epsilon_0 + \Delta_{so}$, with 2 and $\epsilon = \epsilon_0 - 2\Delta_{so}$ with 4 eigenvectors. The corresponding spinors are

$$\begin{pmatrix} u_{0,x} - \imath\, u_{0,y} \\ -\, u_{0,z} \end{pmatrix} \begin{pmatrix} u_{0,z} \\ u_{0,x} + \imath\, u_{0,y} \end{pmatrix}$$

$$\begin{pmatrix} u_{0,x} + \imath\, u_{0,y} \\ 0 \end{pmatrix} \begin{pmatrix} -2\, u_{0,z} \\ u_{0,x} + \imath\, u_{0,y} \end{pmatrix} \begin{pmatrix} u_{0,x} - \imath\, u_{0,y} \\ 2\, u_{0,z} \end{pmatrix} \begin{pmatrix} 0 \\ u_{0,x} - \imath\, u_{0,y} \end{pmatrix} \qquad (20)$$

.

References

[1] G. Dresselhaus. Spin-orbit coupling effects in zinc-blende structures. *Phys. Rev.*, 100:580, 1955.

[2] R. Winkler. *Spin orbit coupling effects in two-dimensional electron and hole systems*. Springer–Verlag, Berlin, Heidelberg, New York, 2003.

[3] Yu.A. Bychkov and E. I. Rashba. Oscillatory effects and the magnetic susceptibility of carriers in inversion layers. *Journal of Physics C (Solid State Physics)*, 17(33):6039–6045, 1984.

[4] Shengli Zhang, Run Liang, Erhu Zhang, Lei Zhang, and Yachao Liu. Magnetosubbands of semiconductor quantum wires with Rashba and Dresselhaus spin-orbit coupling. *Physical Review B (Condensed Matter and Materials Physics)*, 73(15):155316, 2006.

[5] Th. Schäpers, V. A. Guzenko, A. Bringer, M. Akabori, M. Hagedorn, and H. Hardtdegen Spin-orbit coupling in GaxIn1-xAs/InP 2-dimensional electron gases and quantum wire structures *Semiconductor Science and Technology* Expected online publication: March 2009 (print: April 2009)

[6] S. Datta and B. Das. Electronic analog of the electro-optical modulator. *Appl. Phys. Lett.*, 56:665–667, 1990.

B 6 Spin relaxation in nonmagnetic metals and semiconductors

Phivos Mavropoulos
Institut für Festkörperforschung
and Institute for Advanced Simulation
Forschungszentrum Jülich, D-52425 Jülich

Contents

1 Introduction **2**

2 Phenomenological aspects of spin relaxation and spin dephasing **2**

3 Experimental methods for detecting spin relaxation and dephasing **4**

4 Mechanisms of spin relaxation and dephasing **6**

5 Spin-orbit coupling **9**
5.1 Short introduction 9
5.2 Basics of spin-orbit coupling in semiconductors 11
5.3 Systems with inversion symmetry 11
5.4 Systems without inversion symmetry 14

6 Elliott-Yafet mechanism **15**

7 D'yakonov-Perel' mechanism **19**

8 Spin-flip scattering due to impurity spin-orbit coupling **22**

A Notation and symbols **24**

1 Introduction

In the field of spintronics [1, 2] it is obviously highly important to gain control of spin populations, i.e., ensembles of spin-polarized carriers (electrons or holes). In paramagnetic materials and in the absence of external fields a spin population represents a non-equilibrium state, which after some characteristic *spin relaxation time* reaches equilibrium by becoming depolarized. If, during this time, the spin population has diffused over a characteristic distance, then this distance is called the *spin relaxation length*.
Spin relaxation has been studied for over half a century, long before the emergence of spintronics. A number of experimental techniques, discussed in section 3, give direct or indirect information on the relaxation time and length. Although phenomenologically one tries to boil down all spin relaxation phenomena to a few parameters, defined in section 2, the related physics is very rich, encompassing many effects and mechanisms which can vary with material, temperature, etc. A summary of these mechanisms is given in section 4. However, perhaps the most common source of depolarization is the spin-orbit coupling (introduced in section 5). This manifests itself through a number of mechanisms, reviewed in sections 6-8.
The present manuscript is meant to be an introduction to the physics of spin relaxation in paramagnetic metals and semiconductors. It should be noted, however, that spin relaxation in magnetic materials is also a very active field of research. In the that case, the equilibrium state is spin-polarized, self-sustained by the exchange interaction; the term "spin relaxation" refers again to the return to the equilibrium state after some distortion of the polarization. Furthermore, spin relaxation in quantum dots is currently a subject of increasing importance and intense research in view of their relevance in quantum computing. For a presentation of this subject, the reader is referred to the manuscript by Carola Meyer.

2 Phenomenological aspects of spin relaxation and spin dephasing

Assume that a paramagnetic sample is subject to an external magnetic field $\vec{B}(t) = B_0\,\hat{z} + \vec{B}_1(t)$ that has a static z-component B_0 and a transverse (x-y) oscillating component, $\vec{B}_1(t)$. After relaxation, an initial spin population density $\vec{s}(\vec{r}, t=0)$ will obtain an equilibrium value $s_0\,\hat{z}$ (the transverse field is considered weak and rapidly oscillating so that the equilibrium value is along the z direction). The magnetization is given in terms of the spin density as $M(\vec{r}, t) = \gamma\,\vec{s}(\vec{r}, t)$, with γ the gyromagnetic ratio (see also Appendix A). The phenomenological equations describing the spin dynamics are the Bloch equations with an additional diffusion term, represented by a diffusion constant D:

$$\frac{\partial s_x(\vec{r},t)}{\partial t} = \gamma\,(\vec{s}\times\vec{B})_x - \frac{s_x(\vec{r},t)}{\tau_2} + D\nabla^2 s_x(\vec{r},t) \quad (1)$$

$$\frac{\partial s_y(\vec{r},t)}{\partial t} = \gamma\,(\vec{s}\times\vec{B})_y - \frac{s_y(\vec{r},t)}{\tau_2} + D\nabla^2 s_y(\vec{r},t) \quad (2)$$

$$\frac{\partial s_z(\vec{r},t)}{\partial t} = \gamma\,(\vec{s}\times\vec{B})_z - \frac{s_z(\vec{r},t) - s_0}{\tau_1} + D\nabla^2 s_z(\vec{r},t) \quad (3)$$

In these equations, two characteristic times appear: the *relaxation time* τ_1, associated with the z-component of $\vec{s}$, and the *dephasing time* τ_2, associated with the transverse components. These

depend, in general, on the magnetic field. Under weak magnetic fields, if carriers in the system are electrons (e.g. in metals or n-doped semiconductors) and the system is isotropic, one has $\tau_1 = \tau_2$; in this case, following Žutić et al. [1], we will denote both τ_1 and τ_2 by τ_s. In anisotropic materials, however, there is no guarantee that $\tau_1 = \tau_2$, but it has been shown that $\tau_2 \leq 2\tau_1$ [3].

The physical significance of τ_1 is that $1/\tau_1$ gives the rate for the spin system to reach equilibrium by exchanging energy and/or momentum with the lattice. Particularly on application of a magnetic field, the initially unpolarized electron system must exchange energy with the lattice, coming from the Zeeman splitting of the electron levels, to acquire the final polarization $s_0\,\hat{z}$. In the absence of an external field energy exchange is also possible (e.g. if the spin population is produced at high-energy states by optical pumping), but there is always a momentum relaxation that accompanies the spin relaxation (we will return to this in the description of the Elliott-Yafet and D'yakonov-Perel' mechanisms). $1/\tau_1$ is thus the equilibration rate of the diagonal components of the spin-density matrix [2].

On the other hand, τ_2 is associated with the equilibration of the non-diagonal elements of the spin-density matrix [2]. If a spin population with perpendicular orientation with respect to the magnetic field is injected in the sample, then this will start precessing about the magnetic field with a frequency that depends on the g-factor. However, the g-factor itself depends on crystal momentum and energy. Therefore there is bound to be a dispersion in the precession frequency, coming from the spread of the initial population in k-space. On top of that come collisions that further spread out the ensemble in k-space. Thus, $1/\tau_2$ is the rate at which an initially polarized ensemble, peaked around some particular spin direction perpendicular to the magnetic field, will lose its phase coherence in spin space ending up in an unpolarized ensemble.

In the phenomenological approach to spin relaxation it is frequently implied that there is a thermal "semi-equilibrium" separately among spin-up ($\uparrow$) and spin-down ($\downarrow$) electrons, characterized e.g. by different Fermi levels $E_\mathrm{F}^\uparrow$ and $E_\mathrm{F}^\downarrow$, which in a first approximation do not interact with each other. In a second approximation, however, they do interact, and ($E_F^\uparrow$, $E_F^\downarrow$) are taken as time-dependent parameters that ultimately converge to a common Fermi level E_F. This argument was physically justified by Overhauser [4] by noting that the characteristic spin-flip times are long in comparison to the characteristic collision times which lead to momentum and energy relaxation. In the same publication, Overhauser also set forth to calculate the time dependence of $E_F^\uparrow$ and $E_F^\downarrow$ due to several mechanisms, and from these the spin density, and showed that a relaxation time approximation ($\partial\vec{s}/\partial t = -\vec{s}/\tau_\mathrm{s}$ in the absence of fields) is justified.

We close this section by commenting on the notion of spin-flip length Λ_sf vs. spin relaxation length L_s [5]. An electron in a crystal, having an average Fermi velocity v_F, can undergo many collisions at impurities, phonons, etc., that change its momentum $\vec{p}$, before a spin-flipping collision. If the mean free path between collisions is L_p, then the momentum relaxation time is $\tau_\mathrm{p} = L_\mathrm{p}/v_\mathrm{F}$. If N collisions are needed before a spin flip, then the spin relaxation time is $\tau_\mathrm{s} = N\tau_\mathrm{p}$, and the electron has travelled a total length of $\Lambda_\mathrm{sf} = NL_\mathrm{p} = L_\mathrm{p}\tau_\mathrm{s}/\tau_\mathrm{p}$. However, due to the random collisions this length has been travelled in a random "zig-zag" motion, and is therefore not the total distance from the point of origin. In the absence of an external electric field, the total distance from the point of origin is determined by the electron random walk and is therefore proportional to $\sqrt{N}$: $L_\mathrm{s} = L_\mathrm{p}\sqrt{(1/3)\tau_\mathrm{s}/\tau_\mathrm{p}} = \sqrt{(1/3)L_\mathrm{p}\Lambda_\mathrm{sf}} = \sqrt{(1/3)L_\mathrm{p}v_\mathrm{F}\tau_\mathrm{s}} = \sqrt{D\tau_\mathrm{s}}$. This is the spin relaxation length, also called spin diffusion length. $D = L_\mathrm{p}v_\mathrm{F}/3$ is the diffusion constant, with the factor $1/3$ coming from averaging in three spatial directions.

3 Experimental methods for detecting spin relaxation and dephasing

We now summarily discuss some experimental techniques for detecting spin relaxation and dephasing time and length. Among these there exist spectroscopic techniques, real-space, voltage or magnetocurrent measuring techniques, and combined real-space real-time spectroscopic methods.

- *Conduction electron spin resonance (CESR)* is a technique based on the resonant absorption of electromagnetic radiation. Under the influence of a static external magnetic field B, the electron levels split by an energy $\Delta E = g\mu_{\rm B}B$. This is actually also a phenomenological definition of the g-factor. Under incident electromagnetic radiation of frequency ν, one obtains a resonance by varying the field B up to the point $h\nu = g\mu_{\rm B}B$. The position of the resonance gives information on the g-factor, while the width of the resonance gives information on the spin relaxation time [6, 7]. (Due to orbital-momentum effects, the g-factor differs from the free-electron $g_0 = 2.0023\cdots$. The deviation can be particularly large in some semiconductors; e.g., for GaAs it varies between $g = -0.47$ and $g = -0.33$ for temperatures between 0 and 300 K [8].)

- *Nuclear polarization* measurements can be used [9] to derive the electron spin population s in semiconductors, and from this τ_1. These experiments rest on the hyperfine interaction of electron spins with nuclear spins $\vec{I}_{\rm N}$ at positions $\vec{R}_{\rm N}$, $H_{\rm hf} \sim \sum_{\vec{R}_{\rm N}} \vec{I}_{\rm N} \cdot \vec{s}\,\delta(\vec{r} - \vec{R}_{\rm N})$ due to which the nuclear spins are polarized in the presence of polarized electrons (actually the nuclear magnetic moments react to the magnetic field created by polarized electrons on contact). Because the nuclear gyromagnetic ratio is small, the interaction is weak and the timescale of these experiments can be of the order of several hours.

- *Polarized photoluminescence* [10] refers to the emission of circularly polarized light due to the recombination of spin polarized electrons with holes, after optical pumping (i.e., creation of spin-polarized electron population by absorption of circularly polarized light [9]). This method is used in semiconductors. From the degree of circular polarization of the emitted light, together with selection rules, the spin polarization of the recombining electrons is derived (in the absence of spin polarization, the emitted light is not circularly polarized). In early experiments [10] only a lower bound to the spin relaxation time could be determined if the recombination time was known. However, recent techniques allow for time-resolved measurements with picosecond accuracy, and the full dynamics of conduction electron polarization is disclosed [11]. Polarized photoluminescence is also used to detect the efficiency of electrical spin injection experiments. Polarized electrons injected from a ferromagnet into the semiconductor conduction band traverse the semiconductor and recombine with unpolarized holes in a quantum well at the far side of the junction. Again the degree of spin polarization can be calculated by the polarization of the emitted light. Here, knowledge of $\tau_{\rm s}$ and of the electron drift velocity in the semiconductor can provide information on the polarization at the injection point [12].

- *Faraday and Kerr rotation* are phenomena related to the change of angle of linear polarization of light when it is transmitted through (Faraday) or reflected by (Kerr) a magnetized sample. Application of an external magnetic field can bring a splitting of the

chemical potentials for spin up and spin down, which is gradually equilibrated by electron spin-flips resulting in a magnetization; during this equilibration, the time-dependent magnetization is probed by Kerr rotation, revealing τ_s [13]. Alternatively, spin populations can be pumped by circularly polarized laser and later read-out by the Faraday or Kerr rotation of linearly polarized laser. By virtue of the high (sub-picosecond) resolution of pump-probe experiments, these measurements can reveal the propagation of spin populations in real space and real time [14, 15].

- *Electrical spin injection and detection in "non-local" experiments.* This type of experiments was first introduced by Johnson and Silsbee [16]. A spin-polarized current is injected into a non-magnetic material (metal or semiconductor) from a ferromagnetic contact, say at position $x = 0$, and travels to the opposite electrode at some position $x < 0$ (see figure 1). However, the non-magnetic material extends also in the direction $x > 0$, and part of the spin population can diffuse also in that direction. Then one has a so called "pure spin current" at $x > 0$, i.e., spin current without charge current. The spin diffusion causes a chemical potential imbalance in the two spins, $\mu_\uparrow(x) \neq \mu_\downarrow(x)$, which can be detected as a voltage difference by a second "read-out" ferromagnetic contact at $x > 0$. The difference $\mu_\uparrow(x) - \mu_\downarrow(x) \sim \exp(-x/L_s)$ drops with distance and becomes negligible after the spin relaxation length. Thus by changing the position of the read-out contact, the spin relaxation length can be identified as the distance where there is no more voltage difference.

- *Transit-time, spin-injection magnetocurrent experiments.* In case of electrical spin injection and detection in semiconductors (i.e., via a ferromagnetic contacts), the experimental setup can consist of a four-terminal device with spin-valve junctions for injection and detection and an accelerating voltage V_a through the semiconductor (see figure 2). Such a device was used by Appelbaum and co-workers [17, 18] for demonstrating spin injection in Si. The injected spin population precesses with frequency $\omega = \gamma B$ due to an external magnetic field B, and the collector current I_c oscillates with the accumulated precession angle $\omega\tau_\mathrm{tr}$: $I_\mathrm{c} = I_0 + \Delta I_0(\tau_\mathrm{tr})\cos(\omega\tau_\mathrm{tr}) = I_0 + \Delta I_0(\tau_\mathrm{tr})\cos(\gamma B\tau_\mathrm{tr})$. The transit time through the semiconductor, τ_tr, can be tuned by V_a and evaluated by the period of oscillation as a function of B. On the other hand, the magnetocurrent $\Delta I_0(\tau_\mathrm{tr})$ (difference between maximum and minimum collector current) depends on the ratio τ_tr/τ_1. By fitting an exponential decay to data for various transit times, $\Delta I_0(\tau_\mathrm{tr}) \sim \exp(\tau_\mathrm{tr}/\tau_1)$, the spin relaxation time can be found [18].

- *Giant magnetoresistance* (GMR) values depend on the spin diffusion length of the ferromagnet and of the (metallic) nonmagnetic spacer. Within the Valet-Fert model [19], GMR in current-perpendicular-to-plane geometry can be calculated by using the spin diffusion length as an input parameter (among others). By fitting the Valet-Fert model to measured GMR values, the spin relaxation length can be extracted [5].

- *Weak antilocalization* experiments are based on the change of resistance by application of a magnetic field in two- or one-dimensional structures (ultrathin metallic films, two-dimensional electron gas, nanowires etc.). The effect is due to a change from destructive to constructive interference of electron paths upon application of the field, as electrons are scattered by impurities. The change in resistance is related to the phase relaxation length of the electrons; spin-orbit coupling at the impurities gives additional contributions [20] that can be experimentally observed.

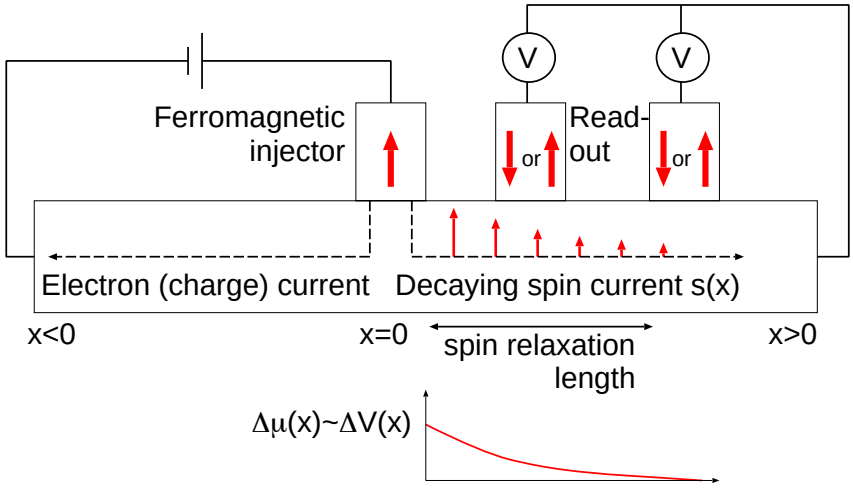


Fig. 1: *Schematic description of the non-local measurement experiment for the spin relaxation length. The chemical potential difference, $\Delta\mu(x) = \mu_\uparrow(x) - \mu_\downarrow(x)$, is read out as a voltage difference when switching the magnetization of read-out contacts at various x. At x larger than the spin-relaxation length, the potential difference becomes negligible.*

Depending on material, temperature, doping, etc., the spin relaxation times can vary by orders of magnitude. For example, a collection of data [21] in n-doped GaAs at low temperatures (4 K) shows a spread of τ_s between 0.01 and 100 ns, depending on donor concentration. In pure Si, $\tau_s = 200$ ns was reported at 85 K and 65 ns at 150 K [18]. Among metals, probably the longest spin relaxation time is found in Li, which is chracterized by very low spin-orbit coupling. Depending on impurity concentration (Li samples between 99% and 99.9% pure) Feher and Kip [6] report τ_s between 3 and 300 ns at low temperatures.

4 Mechanisms of spin relaxation and dephasing

There are numerous phenomena and mechanisms that lead to spin relaxation. Which mechanism applies to a certain material can depend on whether the material is a metal or a semiconductor, on the strength of spin-orbit coupling, on the presence of inversion symmetry, on the shape of the Fermi surface, on the gap with, etc. The type of mechanism that applies is not only material-dependent, but also sample-dependent (sample size, type of impurities), and also dependent on the experimental details (temperature, density of excited spins). Here we summarize a number of possible spin relaxation mechanisms. In the next sections, after a brief introduction to spin-orbit coupling, we focus on its effect to spin relaxation via three mechanisms: the Elliott-Yafet and D'yakonov-Perel' mechanisms and spin-flip scattering at impurities.

- *The Elliott-Yafet mechanism* [22, 3] is most important in crystals with structural inversion symmetry, where the Bloch states have predominantly up or down spin. It is based on momentum scattering at phonons or impurities. The with a small but non-zero probability the scattered electron can end up at a state of the band structure with opposite spin. We discuss it in more detail in section 6.

- *The D'yakonov-Perel' mechanism* [23] applies in semiconductors without inversion symmetry and semiconductor heterostructures. It is based on the appearance of a spin-orbit-

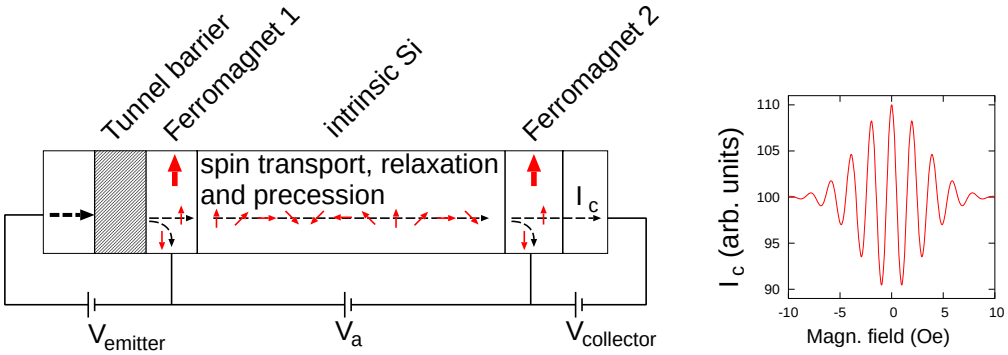


Fig. 2: *Left: Schematic description of the four-terminal spin injection device of Appelbaum and co-workers [17, 18]. In the emitter circuit, only hot electrons (high above $E_{\rm F}$) are able to penetrate the barrier. Due to the spin-dependent band structure of Ferromagnet 1, spin-down electrons are scattered and lose energy, while spin-up electrons can traverse without significant attenuation and enter the semiconductor. There they are accelerated by the voltage $V_{\rm a}$ to a drift velocity $v_{\rm d}$, their spin precesses due to an external magnetic field, and they reach the collector circuit after transit time $\tau_{\rm tr}$. Here, Ferromagnet 2 allows only spin-up electrons to enter the collector circuit and be detected. The collector current $I_{\rm c}$ depends on the spin polarization (magnitude and direction) of the incident electrons at the second interface of the semiconductor spacer: $I_c = I_0 + \Delta I_0 \cos(\gamma B \tau_{\rm tr})$. Right: Schematic view of oscillations of collector current due to spin precession caused by an external magnetic field. The damping at high fields is due to the Hanle effect.*

induced, $\vec{k}$-dependent spin quantization axis, around which electrons precess. Momentum scattering to another $\vec{k}$ changes the precession axis, so that in the end all information about the spin phase is lost. We discuss this in more detail in section 7.

- *Spin-flip scattering due to impurity spin-orbit coupling* can obviously also lead to spin relaxation. We focus on this subject in section 8.

- *The Bir-Aronov-Pikus mechanism* [24] is due to exchange interaction of electrons with holes. The interaction hamiltonian is of the form $H = A\,\vec{\boldsymbol{S}} \cdot \vec{J}\,\delta(\vec{r}_{\rm e} - \vec{r}_{\rm h})$, where A is a parameter describing the exchange strength, $\vec{\boldsymbol{S}}$ is the electron spin operator, $\vec{J}$ the hole angular momentum operator, and $\vec{r}_{\rm e}$, $\vec{r}_{\rm h}$ the positions of electron and hole. This exchange interaction can lead to spin-flip scattering with cross-section (and $\tau_{\rm s}$) differing according to temperature, hole concentration, etc. [1]. The Bir-Aronov-Pikus mechanism applies to semiconductors.

- *Exchange coupling among spin-polarized conduction electrons* can enhance the spin relaxation time in semiconductors. This was observed by Stich et al. [25] by photoexciting different populations of spin-polarized electrons in a two-dimensional electron gas system. As the initial degree of spin polarization was varied between 0 and 30%, a significant increase of $\tau_{\rm s}$ from ~ 25 ps to ~ 200 ps was observed. Calculations [25] including a Hartree-Fock term in the electron-electron interaction are consistent with experiment, contrary to the case when the Hartree-Fock term is neglected; thus the importance of the exchange coupling is demonstrated.

- *Hyperfine interaction with nuclear spins* can be the most important source of spin relaxation in semiconductors at low temperatures, when other degrees of scattering are frozen. This is especially true for the case shallow donor states (i.e., impurity bound states close to the conduction band), which are occupied at low temperatures. The hyperfine interaction hamiltonian, $H_{\rm hf} = \frac{8\pi}{3}\frac{\mu_0}{4\pi}\mu_{\rm B}\hbar\,\gamma_{\rm N}\sum_{\vec{R}_{\rm N}}\vec{I}_{\rm N}\cdot\vec{s}\;\delta(\vec{r}-\vec{R}_{\rm N})$, is weak because nuclear gyromagnetic ratios $\gamma_{\rm N}$ are orders of magnitude smaller than the electronic one, therefore the corresponding spin relaxation times can be long.

- *The Hanle effect*, manifest in transport experiments, is the dephasing of spin populations when precessing in a magnetic field during propagation. Due to diffusion, different electrons traverse the sample at slightly different times $\tau_{\rm tr}$. The accumulated total angle of precession for each electron is $\phi = \omega\tau_{\rm tr}$, where ω is the precession frequency. A spread $\Delta\tau_{\rm tr}$ leads to a spread in $\Delta\phi = \omega\Delta\tau_{\rm tr}$. As ω is proportional to the magnetic field B, if B is large enough $\Delta\phi$ becomes comparable to ϕ and the observed spin polarization is suppressed [18]. This suppression at high fields is shown schematically in figure 2 (right).

- *Inelastic electron-electron scattering in synergy spin-orbit coupling* can lead to a significant reduction of the spin relaxation time as the initial spin population energy rises above the Fermi level. A recent calculation [26] based on the GW approximation including spin-orbit coupling shows a reduction of $\tau_{\rm s}$ by up to three orders of magnitude in metals (depending on the material) if the initial spin is at a modest 0.3 eV above the Fermi level; such energies are common in magnetic junctions due to the bias voltage.

- *Electron-electron scattering at magnetic impurities* can lead to relaxation of the spin current by the following mechanism: a propagating spin-up electron hops on a magnetic impurity site, temporarily paying Coulomb energy due to the higher occupation of the impurity; this energy is regained when an electron hops off the impurity. If the second electron is of spin-down type, the current spin-polarization has changed sign, while the impurity has undergone a spin flip.

In general, more than one mechanisms for spin relaxations are simultaneously present. If they can be considered as independent (this is in many cases, but not always, a correct assumption), and if each one of them is characterized by a relaxation time τ_i, then relaxation rates can be added and the combined relaxation time is given by [4]:

$$\frac{1}{\tau_{\rm s}} = \sum_i \frac{1}{\tau_i}. \tag{4}$$

An additional mechanism that affects the spin dephasing time is *motional narrowing*. It is the inhibition of phase relaxation caused by randomly fluctuating forces [1]. In the presence of a constant Zeeman field the phase of a spin changes by $\Delta\phi = \omega\Delta t$ over time t. If, on the other hand, the field is rapidly and randomly fluctuating then the phase will perform a random walk depending on the correlation time of the fluctuations $\tau_{\rm c}$. The average accumulated phase will thus be proportional to the square root of the steps of the random walk, $\sqrt{t/\tau_{\rm c}}$, and to the accumulated phase at each step, $\omega\,\tau_{\rm c}$: $\Delta\phi \sim \sqrt{t/\tau_{\rm c}}\,(\omega\,\tau_{\rm c}) = \omega\sqrt{t\,\tau_{\rm c}}$ [1]. As $\tau_{\rm c}$ becomes shorter (the fluctuations become more rapid), $\Delta\phi$ decreases: the spin does not have enough time to accumulate phase in any direction, dephasing is inhibited, and the spin dephasing time increases. This effect is important for instance in the D'yakonov-Perel' mechanism.

An example of different mechanisms being important at different temperatures in the same material can be seen in the experiments by Lepine [27] (Lepine's results are conveniently summarized in figure IV.8 in the review by Fabian et al. [2]). Lepine considered spin relaxation in Si doped with P, which provides a donor state 45 meV below the conduction band edge. Three temperature regimes are revealed and discussed [27, 1, 2]. For $T < 50$ K, the electrons are trapped in donor bound states, and the hyperfine interaction is responsible for the relaxation; here, $\tau_{\rm s} \sim 10$ ns. As the temperature rises to $50 < T < 75$ K, $\tau_{\rm s}$ increases because of motional narrowing of the trapped electrons due to scattering with excited electrons at the conduction band. At $75 < T < 150$ K, also the first excited donor state becomes populated and contributes to the hyperfine-induced relaxation; due to motional narrowing, $\tau_{\rm s}$ continues to increase, up to a peak at 120-150 K (at the peak value, the measured $\tau_{\rm s}$ is between 90 and 30 ns, strongly dependent on donor concentration which ranges between 7.4×10^{14} and $8 \times 10^{16}/{\rm cm}^3$; high concentration leads to low $\tau_{\rm s}$). Above 150 K the conduction band is richly populated and the Elliott-Yafet mechanism sets in, causing a decrease of $\tau_{\rm s}$.

5 Spin-orbit coupling

Spin-orbit coupling, in synergy with momentum scattering, is probably the most common source of spin relaxation. Therefore we shortly introduce the effect and its consequences on the band structure, before discussing the spin-orbit induced relaxation mechanisms. For further information, the reader is referred to the manuscripts of G. Bihlmayer and A. Bringer in this volume, or to any book on quantum mechanics (more specialized are, e.g. the books by Rose [28] and Strange [29]).

5.1 Short introduction

The spin-orbit coupling describes the coupling of the electron spin moment to internal or external electric fields. In a classical picture, it can be understood by realizing that an electric field, Lorentz-transformed to the frame of reference of a moving electron, contains a magnetic field component that couples to the electron magnetic moment. However, quantum mechanically it is not obvious how to define a reference frame moving with the electron. In quantum mechanics, the spin-orbit coupling follows from the Dirac equation as a relativistic effect. On simplifying the Dirac equation to the Schrödinger equation with relativistic corrections added to the hamiltonian, the term describing spin-orbit coupling is

$$H_{\rm soc} = \frac{e\hbar}{4m^2c^2}\vec{E}\cdot(\vec{p}\times\vec{\boldsymbol{\sigma}}). \tag{5}$$

Here, $\vec{E} = \vec{\nabla}V(\vec{r})$ is the electric field, with $V(\vec{r})$ the electrostatic potential. From this equation it is evident that the spin-orbit hamiltonian, in contrast to a usual potential, is non-local in nature, as it includes a derivative of the wavefunction via the linear momentum operator $\vec{p}$.
In the case of a central potential, $V(\vec{r}) = V(r)$, the spin-orbit hamiltonian can be rewritten after some manipulations in the form

$$H_{\rm soc} = \frac{e\hbar}{4m^2c^2}\frac{1}{r}\frac{dV(r)}{dr}\,\vec{L}\cdot\vec{\boldsymbol{\sigma}} = \xi(r)\,\vec{L}\cdot\vec{\boldsymbol{S}} \tag{6}$$

whence a coupling of the electron spin with the angular momentum, $\vec{L} = \vec{r}\times\vec{p}$, becomes evident ($\xi(r)$ is a shorthand notation for $(\hbar/2m^2c^2)(1/r)dV/dr$, and $\vec{\boldsymbol{S}} = \vec{\boldsymbol{\sigma}}/2$ is the spin

operator). In atoms, in principle one must account for the interaction of the electron spin with the electric field generated by all other electrons, plus the nucleus. In practice, however, mean-field equations are in many cases adequate, meaning that an average potential is taken, e.g. the one found within density-functional calculations. This mean-field approximation is valid partly due to the fact that the most important contribution in Eq. (6) comes from the strong nuclear electric field, from positions r close to the nucleus—in the vicinity of the nucleus, Eq. (6) gives $(1/r)dV/dr \sim -Z|e|/r^3$ (Z is the atomic number). This is also the reason that heavy atoms show, as a rule, stronger spin-orbit coupling than light atoms. Furthermore, the appearance of $\vec{L}$ in the interaction implies that s-electrons are immune to spin-orbit coupling. The interaction is expected to be especially strong for p orbitals that are closer to the nucleus than d or f orbitals[1]—the term $1/r^3$ close to the nucleus outweighs the higher orbital momentum of d and f orbitals.

In solids it is obvious that the potential cannot be spherically symmetric, even within mean-field theory. However, in most cases an atom-site-dependent orbital momentum is defined, and one accepts as an approximation a spin-orbit hamiltonian of the form $\sum_n (1/r)dV(\vec{r}+\vec{R}_n)/dr$, where $\vec{R}_n$ are the atom positions and $\vec{r}$ is confined in the atomic cell. Again the physical argument for this approximation is that the spin-orbit potential is spherical where it is strong (close to the nucleus), and negligibly weak in the interstitial region, where non-spherical terms appear.

It is mathematically convenient, but also insightful, to separate the term $\vec{L}\cdot\vec{\boldsymbol{S}} = \frac{1}{2}\vec{L}\cdot\vec{\boldsymbol{\sigma}}$ appearing in Eq. (6) in the following manner:

$$\vec{L}\cdot\vec{\boldsymbol{\sigma}} = L_z\boldsymbol{\sigma}_z + \frac{1}{2}(L_+\boldsymbol{\sigma}_- + L_-\boldsymbol{\sigma}_+). \tag{7}$$

Here, $L_\pm = L_x \pm iL_y$ are raising and lowering operators of the z-component of the angular momentum (and similarly for the spins). Thus, denoting the angular momentum eigenvalues by l (for L^2) and m (for L_z), action of L_+ on a state $|l,m\rangle$ yields a state $|l,m+1\rangle$, unless m has the highest possible value, $m=l$, for which $L_+|l,m=l\rangle = 0$ (similarly for lowering m by L_-). For the electron spin, s can only take the values $\pm 1/2$ (representing the spinor wavefunctions $|\uparrow\rangle = \binom{1}{0}$ and $|\downarrow\rangle = \binom{0}{1}$), so that the only non-vanishing results are $\boldsymbol{\sigma}_+|\downarrow\rangle \sim |\uparrow\rangle$, $\boldsymbol{\sigma}_-|\uparrow\rangle \sim |\downarrow\rangle$. In Eq. (7), the first term is spin-conserving, while the second is spin-flipping and most important for spin relaxation. We see, e.g., that action of the operator $\vec{L}\cdot\vec{\boldsymbol{\sigma}}$ on a wavefunction $|l,m\rangle|\uparrow\rangle$ adds an admixture of $|l,m+1\rangle|\downarrow\rangle$ due to the term $L_+\boldsymbol{\sigma}_-$. This means that, starting by a pure-spin, pure-angular-momentum state, the action of the spin-orbit coupling produces a beating between higher and lower m and up and down spin. I.e., the s and the m are not constants of motion. However, the beating does not change the total angular momentum l, which remains a constant of the motion. Also, a spin lowering is accompanied by a raising of m, so that the total z-component m_j of the total angular momentum j, is not changed. These rules are elegantly proven by observing that the squared orbital angular momentum, L^2, and the total angular momentum, $\vec{J} = \vec{L} + \frac{1}{2}\vec{\boldsymbol{\sigma}}$, commute with the spin-orbit hamiltonian: $[L^2, H_{\rm soc}] = [\vec{J}, H_{\rm soc}] = 0$. There are two exceptions to the spin-beating: the spin-up state of highest m and the spin-down state of lowest m are eigenstates of the spin-orbit hamiltonian, because application of the spin-flip term on these gives zero (for example, starting from $|l,m=l\rangle|\uparrow\rangle$ one can neither raise the spin, in order to lower m, nor raise m, in order to lower s). However, such states are unlikely to appear as eigenstates in solids (except in cases with special symmetry), because hybridization of wavefunctions of neighbouring atoms will create an admixture of different orbitals.

[1] p wavefunctions start as $\psi(r) \sim r$, while d and f as r^2 and r^3, respectively.

5.2 Basics of spin-orbit coupling in semiconductors

From the above discussion it follows that the basis $|l,m\rangle|s\rangle$, although valid, should be better replaced by a basis consisting of eigenstates of the spin-orbit operator (at least for the single atom). These can be constructed by appropriate linear combinations. Actually, due to the crystal environment of solids, the commonly used basis in the absence of spin-orbit coupling is not $|l,m\rangle$, but linear combinations of the m states that correspond to oriented orbitals (such as p_x, p_y, p_z) but carry no z-component of the angular momentum.[2] In semiconductors, an appropriate basis that accounts for the spin-orbit coupling is given then (up to normalization prefactors) by the following linear combinations [1]:

$$
\begin{array}{llllll}
|0,0\rangle\,|\uparrow\rangle & = |s\rangle\,|\uparrow\rangle & (j=\frac{1}{2}, m_j=\frac{1}{2}) & \Gamma_6 & \text{(a)} \\
|0,0\rangle\,|\downarrow\rangle & = |s\rangle\,|\downarrow\rangle & (j=\frac{1}{2}, m_j=-\frac{1}{2}) & \Gamma_6 & \text{(b)} \\
|1,1\rangle\,|\downarrow\rangle - |1,0\rangle|\uparrow\rangle & = (|p_x\rangle + i\,|p_y\rangle)\,|\downarrow\rangle - |p_z\rangle\,|\uparrow\rangle & (j=\frac{1}{2}, m_j=-\frac{1}{2}) & \Gamma_7 & \text{(c)} \\
|1,-1\rangle\,|\uparrow\rangle + |1,0\rangle\,|\downarrow\rangle & = (|p_x\rangle - i\,|p_y\rangle)\,|\uparrow\rangle + |p_z\rangle\,|\downarrow\rangle & (j=\frac{1}{2}, m_j=-\frac{1}{2}) & \Gamma_7 & \text{(d)} \\
|1,1\rangle\,|\uparrow\rangle & = (|p_x\rangle + i\,|p_y\rangle)\,|\uparrow\rangle & (j=\frac{3}{2}, m_j=\frac{3}{2}) & \Gamma_8 & \text{(e)} \\
|1,1\rangle\,|\downarrow\rangle + 2|1,0\rangle\,|\uparrow\rangle & = (|p_x\rangle + i\,|p_y\rangle)\,|\downarrow\rangle + 2|p_z\rangle\,|\uparrow\rangle & (j=\frac{3}{2}, m_j=\frac{1}{2}) & \Gamma_8 & \text{(f)} \\
|1,-1\rangle\,|\uparrow\rangle - 2|1,0\rangle\,|\downarrow\rangle & = (|p_x\rangle - i\,|p_y\rangle)\,|\uparrow\rangle - 2|p_z\rangle\,|\downarrow\rangle & (j=\frac{3}{2}, m_j=-\frac{1}{2}) & \Gamma_8 & \text{(g)} \\
|1,-1\rangle\,|\downarrow\rangle & = (|p_x\rangle - i\,|p_y\rangle)\,|\downarrow\rangle & (j=\frac{3}{2}, m_j=\frac{3}{2}) & \Gamma_8 & \text{(h)}
\end{array}
\tag{8}
$$

Expressions (8a,b) are used to describe the conduction s-band, while (8c-h) are used to describe the valence p-band, including heavy holes, light holes, and the split-off band. The notation ($\Gamma_{6,7,8}$) refers to the irreducible representation of the states.

A typical band structure of a semiconductor of the diamond structural type (e.g., Ge) or zinc-blende type (e.g., GaAs) is shown schematically in figure 3. The most severe effect of spin-orbit coupling is the splitting of the valence band top and the admixture of states of different spins.[3] The former has consequences for the temperature-dependent hole population, while the latter is of major interest for spin transport: the mixed-spin character of some of the Γ_8 states, in particular (8f and g), means that hole spin polarization holes has a very short lifetime. Therefore, experimental attempts towards spin transport in semiconductors are focused on electrons.

A further effect of spin-orbit coupling in semiconductors is the spin splitting of the conduction band, which, however, is much smaller and only present if there is no inversion symmetry in the lattice. We discuss this in section 5.4.

5.3 Systems with inversion symmetry

Although the spin-orbit coupling is expected to lift of degeneracies, in many metallic systems (e.g., alkali and noble metals) it is clear that there is a single branch of the Fermi surface, indicating a double (spin-up–spin-down) degeneracy of the bands. This is due to the presence of space-inversion symmetry in these systems; the same is true for the conduction band degeneracy of diamond-structure semiconductors (as is Si and Ge). A comprehensive discussion of the

[2] For example, $|l=1, m=1\rangle = (|p_x\rangle + i\,|p_y\rangle)/\sqrt{2}$, $|l=1, m=0\rangle = |p_z\rangle$, $|l=1, m=-1\rangle = (|p_x\rangle - i\,|p_y\rangle)/\sqrt{2}$.

[3] The magnitude of the splitting Δ_0 reflects the spin-orbit coupling strength of the p-states. As the atomic number increases, Δ_0 is enhanced. E.g., $\Delta_0 = 0.044$ eV for Si, 0.295 eV for Ge, 0.341 eV for GaAs, and 0.75 eV for GaSb. Note that in III-V and II-VI semiconductors, the p-states of the valence band edge are centered at the anion, therefore the anion atomic number is most important for Δ_0.

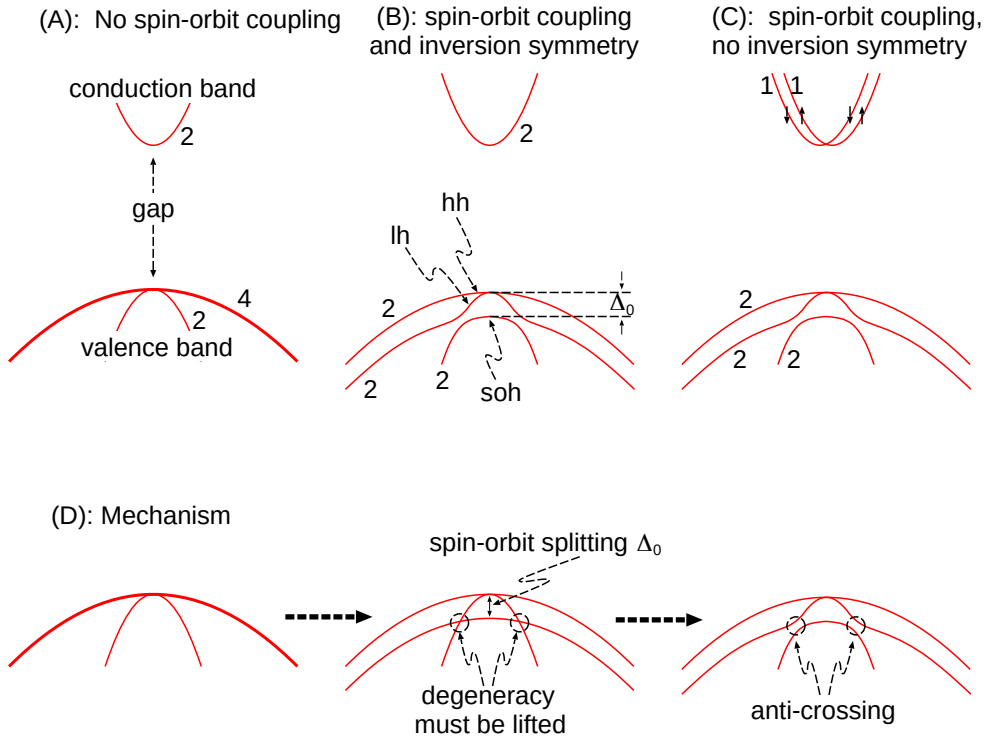


Fig. 3: *Schematic description of the effect of spin-orbit coupling on the direct-gap semiconductor valence and conduction bands (in semiconductors of tetrahedral geometry). Numbers indicate the degeneracy of each band (including spin degeneracy).* Δ_0 *is the spin-orbit splitting; "lh" stands for light holes, "hh" for heavy holes, and "soh" for holes of the split-off band. (A): Band structure in the absence of spin-orbit coupling. (B): Spin-orbit coupling is present, but the structure has inversion symmetry (diamond structure). Then a splitting of the valence band occurs, corresponding to orbitals of the type (8e-h) for the heavy and light holes and type (8c,d) for the split-off holes. (C): In the absence of inversion symmetry (e.g., zinc-blende structure or at semiconductor interfaces), the conduction band also splits as described by the Dresselhaus or Rashba hamiltonians. The arrows on the conduction band indicate the* $\vec{k}$*-dependent spin quantization axis. (D): Mechanism leading to the valence band splitting. In a first step, the* Γ_7*-type and* Γ_8*-type orbitals split with respect to each other. However, crossings that appear between them at finite* k *must lead to hybridization, leading to anti-crossings (right).*

subject, based on an analysis of the action of time-reversal and space-inversion operators in the presence of the spin-orbit hamiltonian, is given by Yafet [3].
Although the band degeneracy is not lifted, the hamiltonian eigenstates $\psi_{\vec{k}}$ cannot be assigned a single spin direction any more. Nevertheless, in most cases it is possible to choose $\psi_{\vec{k}}$ so that they are predominantly of spin-up or spin-down character. Following Fabian and Das Sarma [30], we denote these states by $\psi_{n\vec{k}\uparrow}$ if they are of spin-up character on the average, i.e., if $\langle\psi_{n\vec{k}\uparrow}|\boldsymbol{S}_z|\psi_{n\vec{k}\uparrow}\rangle > 0$, and by $\psi_{n\vec{k}\downarrow}$ if $\langle\psi_{n\vec{k}\downarrow}|\boldsymbol{S}_z|\psi_{n\vec{k}\downarrow}\rangle < 0$. Elliott [22] has discussed the spin-down admixture in $\psi_{n\vec{k}\uparrow}$ (and vice-versa) via a perturbation-theory approach as follows. Suppose that, in the absence of spin-orbit coupling, $\psi_{n\vec{k}\uparrow} = \chi_{n\vec{k}}|\uparrow\rangle$. On switching on spin-orbit coupling, transitions to other states $n'\vec{k}$ have to be accounted for ($\vec{k}$ does not change, because the spin-orbit potential is lattice-periodic). In first-order perturbation theory we have:

$$\begin{aligned}
\psi_{n\vec{k}\uparrow} &= |\chi_{n\vec{k}}\rangle|\uparrow\rangle + \\
&\quad \sum_{n'\neq n}\left[\frac{\langle\uparrow|\langle\chi_{n'\vec{k}}|\xi\vec{L}\cdot\vec{\boldsymbol{S}}|\chi_{n\vec{k}}\rangle|\uparrow\rangle}{E_{n\vec{k}}-E_{n'\vec{k}}}|\chi_{n'\vec{k}}\rangle|\uparrow\rangle + \frac{\langle\downarrow|\langle\chi_{n'\vec{k}}|\xi\vec{L}\cdot\vec{\boldsymbol{S}}|\chi_{n\vec{k}}\rangle|\uparrow\rangle}{E_{n\vec{k}}-E_{n'\vec{k}}}|\chi_{n'\vec{k}}\rangle|\downarrow\rangle\right] \\
&= |\chi_{n\vec{k}}\rangle|\uparrow\rangle + \\
&\quad \sum_{n'\neq n}\left[\frac{\langle\chi_{n'\vec{k}}|\xi L_z\boldsymbol{S}_z|\chi_{n\vec{k}}\rangle}{E_{n\vec{k}}-E_{n'\vec{k}}}|\chi_{n'\vec{k}}\rangle|\uparrow\rangle + \frac{1}{2}\frac{\langle\chi_{n'\vec{k}}|\xi\vec{L}_+|\chi_{n\vec{k}}\rangle}{E_{n\vec{k}}-E_{n'\vec{k}}}|\chi_{n'\vec{k}}\rangle|\downarrow\rangle\right]
\end{aligned} \tag{9}$$

where Eq. (7) has been used to separate the spin-conserving and spin-flip contributions, spin angular momenta have been raised and lowered by $\boldsymbol{S}_\pm$, and spinor-orthogonality relations have been employed ($\langle s|s'\rangle = \delta_{ss'}$). The last term gives the spin-down admixture in $\psi_{n\vec{k}\uparrow}$ and is usually small. However, if the energy difference $E_{n\vec{k}} - E_{n'\vec{k}}$ becomes small, higher-order perturbation theory has to be employed, and the admixture can be large.
The perturbed Bloch wavefunction is thus written in the form

$$\psi_{n\vec{k}\uparrow}(\vec{r}) = \Big(a_{n\vec{k}}(\vec{r})|\uparrow\rangle + b_{n\vec{k}}(\vec{r})|\downarrow\rangle\Big)e^{i\vec{k}\cdot\vec{r}} \tag{10}$$

where $a_{n\vec{k}}$ and $b_{n\vec{k}}$ are the lattice-periodic parts. An analogous relation to Eq. (10) holds for $\psi_{n\vec{k}\downarrow}(\vec{r})$. However, the combination of time-reversal and space-inversion symmetry dictates that the lattice-periodic parts of the spin-down wavefunction for $+\vec{k}$ are related to the ones of the spin-up wavefunction for $-\vec{k}$ [3]:

$$\psi_{n;+\vec{k}\downarrow}(\vec{r}) = \Big(a^*_{n;-\vec{k}}(\vec{r})|\downarrow\rangle - b^*_{n;-\vec{k}}(\vec{r})|\uparrow\rangle\Big)e^{+i\vec{k}\cdot\vec{r}} \tag{11}$$

If we denote by $E_{n\vec{k}\uparrow,\downarrow}$ the energy eigenvalues corresponding to the (non-pure-spin) eigenfunctions $\psi_{n\vec{k}\uparrow,\downarrow}(\vec{r})$, then

$$E_{n;+\vec{k}\uparrow} = E_{n;-\vec{k}\downarrow} = E_{n;+\vec{k}\downarrow} \tag{12}$$

with the first equation following from time-reversal symmetry (also known as Kramers degeneracy) and the second by space-inversion symmetry; this is the degeneracy discussed in the beginning of the present section.
Usually, but not always, $a_{n\vec{k}}$ and $b_{n\vec{k}}$ are such that $|b_{n\vec{k}}|/|a_{n\vec{k}}| \ll 1$. Actually, the magnitude of $b_{n\vec{k}}$ is determined by the smallest among the energy differences $E_{n\vec{k}} - E_{n'\vec{k}}$, since these appear in the denominator in Eq. (9). Calling this Δ, we have the rule-of-thumb estimation:

$$|b_{n\vec{k}}|/|a_{n\vec{k}}| \sim \langle\xi\rangle/\Delta. \tag{13}$$

where $\langle\xi\rangle$ is a parameter resulting from averaging the spin-orbit coupling strength $\xi(r)$. Relations (10-13) are central to the Elliott-Yafet theory for spin relaxation, as we shall discuss in section 6.

5.4 Systems without inversion symmetry

In the absence of space inversion symmetry, the band degeneracy given by the second equation in (12) is lifted. This was realized early [31] for semiconductors of the zinc-blende type, where the two atoms of the unit cell are inequivalent, or semiconductor heterostructures [32] and two-dimensional electron gas where there are built-in average electric fields $E\,\hat{z}$ as well as inhomogeneities of the effective mass $m^*(z)$ (z is the direction of growth of the heterostructure). In these cases, a remarkable effect takes place. The semiconductor conduction band splits in two branches, but in such a way that each k-vector in the band structure is associated with its own spin quantization axis. The effect is modelled by the hamiltonian

$$H = \frac{\hbar^2}{2m^*}k^2 + \frac{\hbar}{2}\vec{\boldsymbol{\sigma}}\cdot\vec{\Omega}(\vec{k}\,). \tag{14}$$

The first part is the usual parabolic dispersion relation of the conduction band with effective mass m^*, while the second part includes the spin. At each $\vec{k}$, a spin quantization axis $\vec{\Omega}(\vec{k}\,)$ is defined, with the absolute value $|\vec{\Omega}(\vec{k}\,)|$ determining the strength of the effect. The term $\vec{\Omega}(\vec{k}\,)\cdot\vec{\boldsymbol{\sigma}}$ is a non-diagonal 2×2 matrix in spin space; diagonalization yields the two energy eigenvalues of the spin-split band, with opposite spin eigenvalues (one in the direction $\vec{\Omega}$ and one in the direction $-\vec{\Omega}$):

$$E_{\vec{k}\uparrow,\downarrow} = \frac{\hbar^2}{2m^*}k^2 \pm \frac{1}{2}\hbar|\vec{\Omega}(\vec{k}\,)| \tag{15}$$

The Kramers degeneracy, i.e., the degeneracy due to time-inversion symmetry, is still present, yielding $E_{\vec{k}\uparrow} = E_{-\vec{k}\downarrow}$ (with $\uparrow$ and $\downarrow$ defined with respect to the local spin axis), which also implies that $\vec{\Omega}$ is an odd function of $\vec{k}$: $\vec{\Omega}(-\vec{k}\,) = -\vec{\Omega}(\vec{k}\,)$.
$\vec{\Omega}(\vec{k}\,)$ is material-specific, depending on the spin-orbit coupling strength, the band gap, the proximity to the interface etc. Within perturbation theory, it is seen that hamiltonian (14) arises from the interaction of the conduction band, which is of s character, with the spin-split p bands. Here we present only two cases, and we refer the reader to [1] and [2] for a broader discussion.

- The k^3-Dresselhaus term is present in bulk semiconductors (III-V or II-VI) of the zinc-blende structure. In this case we have [31, 33]

$$\vec{\Omega}(\vec{k}) = \frac{\alpha\hbar^2}{(2m^*E_{\rm g})^{1/2}}\left(k_x(k_y^2-k_z^2)\,\hat{x} + k_y(k_z^2-k_x^2)\,\hat{y} + k_z(k_x^2-k_y^2)\,\hat{z}\right) \tag{16}$$

  where $E_{\rm g}$ is the band gap width and α is a material-dependent parameter (e.g., for GaAs $\alpha = 0.07$).

- The Bychkov-Rashba term [34, 35] is present in asymmetric quantum well heterostructures, in deformed bulk systems, and in the two-dimensional electron gas (see also lecture notes by Gustav Bihlmayer and Andreas Bringer). In this case,

$$\vec{\Omega}(\vec{k}) = \alpha_{\rm BR}\left(\vec{k}\times\hat{z}\right) \tag{17}$$

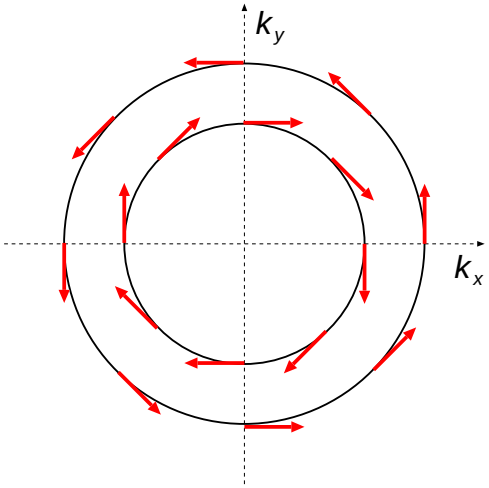


Fig. 4: *Constant-energy lines of the two-dimensional electron gas under the influence of the Rashba hamiltonian. Arrows indicate the $\vec{k}$-dependent spin quantization axis.*

where $\hat{z}$ is considered in the direction of growth in quantum wells (or, equivalently, perpendicular to the plane in the two-dimensional electron gas). Here, $\vec{k}$ is a two-dimensional vector.

Schematically, the resulting spin-split conduction band structure is shown in figure 3C. The spin-structure of the resulting constant-energy line in the case of the Rashba hamiltonian is shown in figure 4.
In recent years, splitting of the Rashba type has been discovered also in the surface states of metallic and semimetallic systems [e.g., Ag and Au(111), Bi(111), or surface alloys Bi/Ag(111)]. Since surface states are confined to the vicinity of crystal surfaces, they have many similarities to a two-dimensional electron gas. It is interesting that the Rashba parameter and the splitting of surface states in metals can be gigantic compared to semiconductor heterostructures. E.g., for an GaSb/InAs heterostructure $\alpha_{\mathrm{BR}} = 0.09$ eVÅ [36] while for Au(111) $\alpha_{\mathrm{BR}} = 0.33$ eVÅ [37]. Taking into account that the Rashba splitting is proportional to k [see Eq. (17)], in Au, where k is determined by the Fermi energy, the splitting is of the order of 100 meV, while in semiconductor heterostructures, where the conduction band is very close to the Brillouin zone center, it is much smaller. For a comprehensive description of the Rashba effect at metal surfaces we refer to [38].

6 Elliott-Yafet mechanism

We turn now to a more detailed description of one of the spin relaxation mechanisms, namely the Elliott-Yafet mechanism. The Elliott-Yafet mechanism is present in metals or semiconductors with space inversion symmetry, and is based on the form (10) and (11) of the spin-dependent wavefunctions together with momentum scattering.
Consider an electron occupying a predominantly spin-up state of the form (10): $\psi_{\vec{k}\uparrow} = \Big(a_{\vec{k}}|\uparrow\rangle + b_{\vec{k}}|\downarrow\rangle\Big)e^{i\vec{k}\cdot\vec{r}}$ (we drop the band index n as it is irrelevant for this discussion). After a momentum-scattering event at an impurity or a phonon, there is an amplitude that the electron is scattered in

another spin-up state, $\psi_{\vec{k}'\uparrow}(\vec{r})$, but also in the corresponding spin-down state, of the form (11): $\psi_{\vec{k}'\downarrow} = \left(a^*_{-\vec{k}'}|\downarrow\rangle - b^*_{-\vec{k}'}|\uparrow\rangle\right)e^{i\vec{k}'\cdot\vec{r}}$. If one neglects the spin-orbit coupling during scattering (e.g., if the spin-orbit coupling of the impurity is insignificant), then the spin-conserving and spin-flip scattering probabilities, $P^{\uparrow\uparrow}_{\vec{k}\vec{k}'}$ and $P^{\uparrow\downarrow}_{\vec{k}\vec{k}'}$ are, in the Born approximation,

$$\begin{aligned} P^{\uparrow\uparrow}_{\vec{k}\vec{k}'} &= \left|\langle\psi_{\vec{k}\uparrow}|\delta H|\psi_{\vec{k}'\uparrow}\rangle\right|^2 \\ &= \left|\langle a_{\vec{k}}\, e^{i\vec{k}\cdot\vec{r}}|\delta H|a_{\vec{k}'}\, e^{i\vec{k}'\cdot\vec{r}}\rangle + \langle b_{\vec{k}}\, e^{i\vec{k}\cdot\vec{r}}|\delta H|b_{\vec{k}'}\, e^{i\vec{k}'\cdot\vec{r}}\rangle\right|^2 \qquad (18) \\ P^{\uparrow\downarrow}_{\vec{k}\vec{k}'} &= \left|\langle\psi_{n\vec{k}\uparrow}|\delta H|\psi_{\vec{k}'\downarrow}\rangle\right|^2 \\ &= \left|-\langle a_{n\vec{k}}\, e^{i\vec{k}\cdot\vec{r}}|\delta H|b^*_{-\vec{k}'}\, e^{i\vec{k}'\cdot\vec{r}}\rangle + \langle b_{\vec{k}}\, e^{i\vec{k}\cdot\vec{r}}|\delta H|a^*_{-\vec{k}'}\, e^{i\vec{k}'\cdot\vec{r}}\rangle\right|^2 \qquad (19) \end{aligned}$$

where δH is the perturbation that causes the scattering [formally, neglecting the spin-orbit coupling during scattering means that $\langle\uparrow|\delta H|\downarrow\rangle = \langle\downarrow|\delta H|\uparrow\rangle = 0$, which was taken into account in Eq. (18,19)]. The important conclusion from Eq. (18,19) is that the spin-flip probability includes factors of the order $|a|^2|b|^2$, while the spin-conserving probability includes factors of the order $|a|^4$. According to Eq. (13), the following order-of-magnitude estimation follows for the spin-flip compared to the spin-conserving probability:

$$\frac{P^{\uparrow\downarrow}_{\vec{k}\vec{k}'}}{P^{\uparrow\uparrow}_{\vec{k}\vec{k}'}} \sim \left(\frac{\langle\xi\rangle}{\Delta}\right)^2. \qquad (20)$$

The value of $(\langle\xi\rangle/\Delta)^2$ varies very much among metals [39], but it usually is very small; e.g., taking the alkali metals from lighter to heavier (increasing $\langle\xi\rangle$), in Li it is of the order of 10^{-10}, in Na 10^{-5}, in K 10^{-4}, in Rb 10^{-3} and in Cs 10^{-2}. This means that in Li one spin-flip happens every 10^{10} momentum scattering events, while in Cs one expects one spin-flip in every 100 scattering events. However, there are also extreme cases: e.g., for Au, which has a very strong spin-orbit coupling, $(\langle\xi\rangle/\Delta)^2 = 0.8$ [39].

After a Fermi surface averaging of $P^{ss'}_{\vec{k}\vec{k}'}$, yielding the averaged probabilities $P^{ss'}$ the spin relaxation time can be connected to the momentum relaxation time τ_p. Since the scattering probability $P = P^{\uparrow\uparrow} + P^{\downarrow\downarrow} + P^{\uparrow\downarrow} + P^{\downarrow\uparrow}$ is related to τ_p as $P = 1/\tau_\mathrm{p}$, while the spin-flip probability $P^{\uparrow\downarrow}$ is related to the spin relaxation time as $P^{\uparrow\downarrow} = 1/\tau_\mathrm{s}$, it is expected, in view of relation (20), that $\tau_\mathrm{s}^{-1} \propto \tau_\mathrm{p}^{-1}$, with a proportionality constant of the order of $(\langle\xi\rangle/\Delta)^2$. However, in many systems the band structure has a complex form, and Δ (and thus b) can strongly vary on the Fermi surface. Then the Fermi surface average of $P^{\uparrow\downarrow}_{\vec{k}\vec{k}'}$ gives an average value of $|b|$, $\langle b^2\rangle$. Realizing that the $|a|$ is of the order of unity, we obtain $\tau_\mathrm{s}^{-1} \sim \langle b^2\rangle\ \tau_\mathrm{p}^{-1}$.

On the other hand, $\langle b^2\rangle$ can be related to the deviation of the g-factor from the free-electron value g_0, $\delta g = g - g_0$. To justify this, we note that δg is related to the expectation value of the angular momentum [3] upon application of a magnetic field B along z: $(\delta g + g_0)\mu_\mathrm{B}B = E_{\vec{k}\uparrow} - E_{\vec{k}\downarrow} = 2\mu_\mathrm{B}B\langle\psi_{\vec{k}\uparrow}|(L_z + g_0\boldsymbol{S}_z)|\psi_{\vec{k}\uparrow}\rangle$ (the last step following from perturbation theory), whence $\delta g = 2\langle\psi_{\vec{k}\uparrow}|L_z|\psi_{\vec{k}\uparrow}\rangle$. Also by perturbation theory [22] it is found that the matrix element of L_z is proportional to $b_{\vec{k}}$. Averaging over the the Fermi surface gives at the end $\delta g^2 \sim \langle b^2\rangle$. This leads to the Elliott relation,

$$\frac{1}{\tau_\mathrm{s}} \sim \frac{\delta g^2}{\tau_\mathrm{p}}, \qquad (21)$$

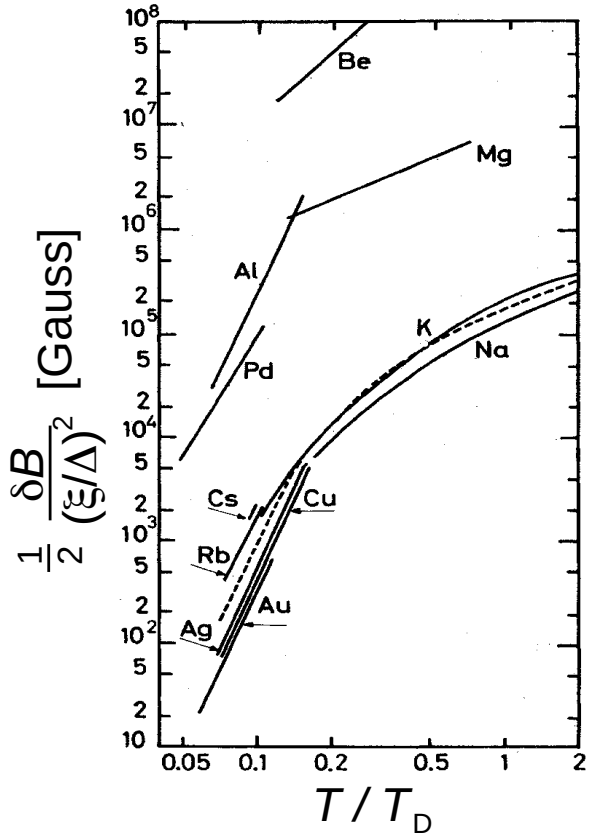


Fig. 5: *Spin relaxation time (measured via conduction-electron spin resonance linewidth, $\frac{1}{2}\delta B$) versus temperature for several metals, as reported by Monod and Beuneu in ref. [39]. $T_{\rm D}$ is the Debye temperature.*

which gives a rough, order-of-magnitude estimate of $\tau_{\rm s}$.

A second expression, that gives a rough estimate of $\tau_{\rm s}$ as a function of temperature, is the Yafet relation, which relates the relaxation time $\tau_1(T)$ to the resistivity $\rho(T)$:

$$\frac{1}{\tau_1(T)} \sim \frac{\langle b^2 \rangle}{\rho(T)}. \tag{22}$$

A hand-waving argument supporting relation (22) is that the resistivity is related to the momentum relaxation rate $1/\tau_{\rm p}$, so that the Yafet relation is expected due to the Elliott relation. However, especially at low temperatures more detailed considerations are needed. These were carried out by Yafet [3]. Considering the electron-phonon matrix elements, and taking into account the change in spin-orbit coupling strength due to the lattice distortion by phonons, he showed that the spin relaxation rate has a T^5 behavior at low temperatures in metals, just as is known for the resistivity.

Experimental results on the temperature dependence of $\tau_{\rm s}$ in metals were collected and analysed by Monod and Beuneu [39, 40] (see figure 5). Actually, $1/\tau_{\rm s}$ was deduced from measurements of the conduction-electron spin resonance linewidth, $\frac{1}{2}\delta B$, normalized to the predicted value of $\langle b^2 \rangle$, $(\langle \xi \rangle / \Delta)^2$; the temperature was devided the Debye temperature $T_{\rm D}$, to normalize for the amount of phonon scattering. It was found that the data for all monovalent metals fall on the same curve, while very strong deviations were seen for polyvalent metals. The explanation for

this was suggested by Silsbee and Beuneu [41], and later refined by Fabian and Das Sarma [30, 42], who also performed a quantitative calculation for Al. The central point is that the relation $\langle b^2 \rangle \sim (\langle \xi \rangle / \Delta)^2$ (with some average value of Δ) is valid in monovalent metals, where the whole Fermi surface is contained in the first Brillouin zone, but not for polyvalent metals, where the Fermi surface crosses the Brillouin zone boundary. At such crossing points, two branches of the Fermi surface could meet causing a degeneracy. However, the degeneracy is lifted due to the presence of the periodic crystal potential $V(\vec{r})$. The degeneracy lifting is proportional to the Fourier component of V, $V_{\vec{G}}$ ($\vec{G}$ is a reciprocal lattice vector), and therefore the magnitude of Δ is approximately $V_{\vec{G}}$, which can be small[4] making $b_{\vec{k}}$ large for those particular k-points. The same is true if there are more than one branches of the Fermi surface that approach each other or cause accidental degeneracies within the first Brillouin zone (as in transition metals). It follows that there are "hot spots" on the Fermi surface that strongly enhance the value of $\langle b^2 \rangle$ compared to the estimate $\langle b^2 \rangle \sim (\langle \xi \rangle / \Delta)^2$. For example, calculations for Al [30] show that, while the hot spots amount only to a small portion of the Fermi surface, they are enough to increase $\langle b^2 \rangle$ by two orders of magnitude. Then the spin relaxation rate is much higher. This is why polyvalent metals deviate from the "regular" Monod-Beuneu curve (figure 5).

The Elliott-Yafet mechanism plays a significant role for spin relaxation also in semiconductors. Here, the band structure close to the conduction band minimum is easier to model than in polyvalent metals, and one can be based on semi-analytical approximations for the wave functions (such as the $\vec{k} \cdot \vec{p}$-approximation). Approximate analytical results are therefore possible. For example, Chazalviel [43] was able to derive the following expression for the spin relaxation time of electrons occupying a state of energy E in III-V semiconductor conduction bands (E is measured from the conduction band edge):

$$\frac{1}{\tau_{\rm s}(E)} = A \left(\frac{\Delta_0}{E_{\rm g} + \Delta_0} \right)^2 \left(\frac{E}{E_{\rm g}} \right)^2 \frac{1}{\tau_{\rm p}(E)} \tag{23}$$

Here $E_{\rm g}$ is the semiconductor gap, Δ_0 the valence-band spin-orbit splitting (see figure 3), and A a parameter which depends on the type of scattering mechanism (phonons, impurities, etc). Depending on the energy distribution of a spin population, this expression should be averaged over energies to obtain the mean spin-relaxation rate. What is important to recognize in this expression is that the relaxation rate increases with spin-orbit coupling strength (expressed by Δ_0) but decreases with the gap width.

As regards the temperature dependence of the spin relaxation rate in non-degenerate semiconductors, $1/\tau_{\rm s}(T) \sim T^2/\tau_{\rm p}(T)$, i.e., $1/\tau_{\rm s}(T)$ varies according to the scattering mechanism [1], as $\tau_{\rm p}(T)$ depends on the scattering mechanism; e.g., for scattering by charged impurities, $1/\tau_{\rm p} \sim T^{-3/2}$, i.e., $1/\tau_{\rm s}(T) \sim T^{1/2}$. In metals, on the other hand, $1/\tau_{\rm s}(T)$ is expected to obey the T^5-law at low temperatures (as shown by Yafet [3]), while $1/\tau_{\rm s}(T) \sim T$ at higher temperatures.[5]

[4] If $\Delta \sim V_{\vec{G}}$ becomes small, first-order perturbation theory is not enough, and the expansion (9) should be taken to higher order.

[5] The T^5-law is derived from the assumption of phonon scattering; however, in reality one expects $1/\tau_{\rm s}(T) \sim T^5 +$ const., where the constant comes from impurity scattering that is most important at low temperatures and is related to the residual resistivity of metals.

7 D'yakonov-Perel' mechanism

Principles

A mechanism of quite different nature is the one discovered by D'yakonov and Perel' [23], which actually leads to dephasing, rather than relaxation. In the previous section we saw that in the case of the Elliott-Yafet mechanism increased momentum scattering leads to increased spin relaxation. On the contrary, when the D'yakonov-Perel' mechanism prevails, increased momentum scattering can lead to a decrease of linger spin lifetimes. The reason for this, as we will discuss, is motional narrowing.

Consider a system without inversion symmetry, e.g., a III-V semiconductor or a semiconductor heterostructure. The spin-orbit coupling is described by a hamiltonian of the type (14). According to the considerations of D'yakonov and Perel' [23], the degeneracy-lifting term $(\hbar/2)\vec{\boldsymbol{\sigma}} \cdot \vec{\Omega}(\vec{k}\,)$ can be viewed as a $\vec{k}$-dependent Larmor precession frequency $\Omega(\vec{k}\,) = |\vec{\Omega}(\vec{k}\,)|$ around the axis defined by the direction of $\vec{\Omega}(\vec{k}\,)$. To see this, note that the splitting (15) of the two spin eigenstates by $\hbar\Omega$ causes a time-dependent difference in phase of the two eigenstates, $\psi_{\vec{k}\uparrow}$ and $\psi_{\vec{k}\downarrow}$, by $\delta\varphi = \Omega t$. If at time $t = 0$ a wavefunction ψ at $\vec{k}$ has spin along some arbitrary axis, different than $\vec{\Omega}(\vec{k}\,)$, it will be partly projected to the $\uparrow$ and partly to the $\downarrow$ eigenfunction, $\psi(0) = a|\uparrow\rangle + b|\downarrow\rangle$. After time t it will have evolved to $\psi(t) = e^{i(\hbar k^2/2m^*)t}\,[e^{i\Omega t/2}\,a\,|\uparrow\rangle + e^{-i\Omega t/2}\,b\,|\downarrow\rangle]$. The time-dependent expectation value of the electron spin, $\vec{s}(t) = \langle\psi(t)|\vec{\boldsymbol{\sigma}}|\psi(t)\rangle$, will precess around the direction $\vec{\Omega}$ with frequency Ω. The concept of precession is, of course, only valid if the energy spread of the wavepacket ψ is larger than the spin-splitting $\hbar\Omega$, otherwise the electron experiences two states of opposite spin but with the same energy, and there can be no precession (see figure 6, right).

Continuing with the D'yakonov and Perel' idea, at some random time t_1 a momentum scattering will occur from $\vec{k}$ to $\vec{k}'$. The electron spin $\vec{s}(t_1)$, which is assumed not to change during scattering, will start precessing around the new spin axis, $\vec{\Omega}(\vec{k}')$. As the random scattering continues, the memory of the initial spin direction will be lost. The idea is shown schematically in figure 6.

We now consider the effect of an increased random momentum-scattering rate $\tau_{\rm p}^{-1}$. We saw that, after each momentum scattering event, the spin precession axis $\vec{\Omega}$ changes. Since the scattering is random (obeying some probability distribution $P_{\vec{k}\vec{k}'}$), $\vec{\Omega}$ changes randomly; and since the time δt between subsequent scattering events is also random (determined again by $P_{\vec{k}\vec{k}'}$), the accumulated angle is also not fixed. Effectively, the electron spin is experiencing a randomly changing force, and the accumulated angle after time t is subject to a random walk in spin space. The correlation time of this walk, $\tau_{\rm c}$, is proportional to the momentum relaxation time $\tau_{\rm p} \sim \langle P^{-1}\rangle$ (where $\langle P^{-1}\rangle$ is loosely understood as an appropriate average of $P^{-1}_{\vec{k}\vec{k}'}$). Therefore the total accumulated phase (spin angle) after time t is proportional to the square-root of the number of random-walk steps, $\sqrt{N} = \sqrt{t/\tau_{\rm p}}$, and to the change in phase at each step, $\langle\Omega\rangle\,\tau_{\rm p}$: $\delta\varphi = (\langle\Omega\rangle\,\tau_{\rm p})\sqrt{t/\tau_{\rm p}} = \langle\Omega\rangle\,\sqrt{t\,\tau_{\rm p}}$ (where, loosely again, $\langle\Omega\rangle$ is an average of $\vec{\Omega}(\vec{k}\,)$). If we define the spin relaxation time as the time needed for $\delta\varphi \sim 1$, we arrive at the result $1/\tau_{\rm s} = \langle\Omega\rangle^2\,\tau_{\rm p}$: rapid momentum scattering (small $\tau_{\rm p}$) leads to slower relaxation. This is precisely the mechanism of motional narrowing: rapid, random changes in the "force" cancel each other out; the spin does not have enough time to precess in any particular direction. Curiously, the spin diffusion length resulting from the D'yakonov-Perel' mechanism does not depend on $\tau_{\rm p}$. This can be seen by writing $L_{\rm s} = \sqrt{D\tau_{\rm s}}$, with the diffusion constant $D \sim \tau_{\rm p}$ and $1/\tau_{\rm s} = \langle\Omega\rangle^2\,\tau_{\rm p}$.

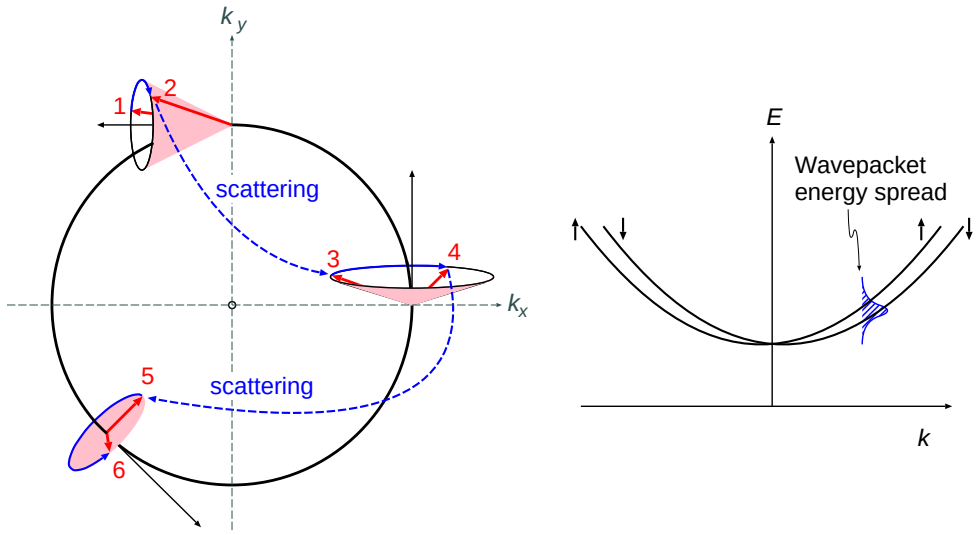


Fig. 6: *Left: D'yakonov-Perel' mechanism. A spin initially at "1" is precessing around the $\vec{k}$-dependent spin quantization axis $\vec{\Omega}(\vec{k}\,)$. After some time it is momentum-scattered from "2" to "3" without change of the spin direction during scattering. However, it now starts to precess around a new axis. After a few random scattering events, the information on the initial spin direction is lost. In this picture, elastic scattering is assumed; the large circle represents a constant-energy line of a two-dimensional electron gas system. Right: The D'yakonov-Perel' concept is meaningful if the energy spread of the wavepacket is larger than the band separation $\hbar\Omega$.*

Boltzmann transport equation

The qualitative considerations presented so-far can be elaborated on and quantified by working out expressions for $\langle P\rangle$ and $\langle\Omega\rangle$ and using an appropriate transport equation. We now shortly discuss a transport equation of the Boltzmann-type, following ref. [2]. The Boltzmann equation provides a semi-classical approach to electronic transport phenomena. Within this approach, an electron wavepacket is treated as a classical particle which moves in the material under the influence of external forces—electric and magnetic fields, or temperature and chemical potential gradients. It is assumed to have a definite crystal momentum $\vec{k}$, but also a well-defined trajectory $\vec{r}(t)$. During collisions at impurities, phonons, etc., a transition probability $P_{\vec{k}\vec{k}'}$ is assumed, which is derived from quantum mechanics. The concept is meaningful if the collisions are rare enough that a wavepacket can be formed between them that is localized enough in real space, so that the notion of a trajectory is meaningful, but also in $\vec{k}$-space, so that the electron crystal momentum is well-enough defined. Also, the scattering events should be random, so that at the end the scattering probabilities can be added up instead of the amplitudes, i.e., correlated scattering events and interference are neglected by the formalism. Under these assumptions, the Boltzmann equation is essentially a continuity (i.e., particle-conservation) equation for the electron distribution function $f_{\vec{k}}(\vec{r};t)$ in phase space $(\vec{r},\vec{k})$:

$$\frac{\partial f}{\partial t}+\frac{d\vec{r}}{dt}\cdot\vec{\nabla}_{\vec{r}}f+\frac{d\vec{k}}{dt}\cdot\vec{\nabla}_{\vec{k}}f=\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{24}$$

The term on the right-hand side describes the change in occupation of point $(\vec{r},\vec{k})$ due to collisions, depends on $P_{\vec{k}\vec{k}'}$, and must account for the fact that scattering from $\vec{k}$ to $\vec{k}'$ is only possible

if $\vec{k}$ is occupied and $\vec{k}'$ empty. In absence of collisions, current conservation requires that the left-hand side vanishes. By identifying the time derivatives of $\vec{r}$ as the group velocity $\vec{v}_{\vec{k}}$ and of $\vec{k}$ as the acceleration by external electric and magnetic fields due to Newton's second law, $\hbar d\vec{k}/dt = e(\vec{E} + \vec{v}_{\vec{k}} \times \vec{B})$, it is seen that the Boltzmann equation describes the response of the distribution function to external stimuli. At equilibrium, and in the absence of external fields, f is just the Fermi-Dirac function.

The next step is to include the spin distribution in the Boltzmann equation quantum-mechanically. This is be done by introducing the spin-density matrix at $t = 0$, for an electron at $\vec{k}$, resolved with respect to an initial "global" reference frame as $\boldsymbol{\rho}_{\vec{k}}(0) = |\uparrow\rangle f_{0\uparrow}\langle\uparrow| + |\downarrow\rangle f_{0\downarrow}\langle\downarrow| = \frac{1}{2}(f_{0\uparrow} + f_{0\downarrow})\mathbf{1} + \frac{1}{2}(f_{0\uparrow} - f_{0\downarrow})\boldsymbol{\sigma}_z$. (In an arbitrary frame of reference the spin-density matrix is of the form $\rho\,\mathbf{1} + \vec{s}\cdot\vec{\boldsymbol{\sigma}}$, where ρ is the particle density and $\vec{s}$ the spin density.) The spin-density matrix will evolve in time in accordance to the Heisenberg relation:

$$\frac{d\boldsymbol{\rho}_{\vec{k}}(t)}{dt} = \frac{i}{\hbar}[H_{\text{spin}}, \boldsymbol{\rho}_{\vec{k}}(t)] + \frac{\partial \boldsymbol{\rho}_{\vec{k}}(t)}{\partial t}. \tag{25}$$

Here, $H_{\text{spin}} = \frac{\hbar}{2}\vec{\boldsymbol{\sigma}} \cdot \vec{\Omega}(\vec{k})$ is the spin-part of the hamiltonian (14) that causes the spin precession. The last term, $\partial\boldsymbol{\rho}_{\vec{k}}(t)/\partial t$, represents the "explicit" time dependence of $\boldsymbol{\rho}$, due to terms other than the spin hamiltonian, and is given in analogy to Eq. (24) by the continuity equation together with a collision term.[6] The result is:

$$\frac{\partial \boldsymbol{\rho}_{\vec{k}}(t)}{\partial t} - \frac{i}{\hbar}[H_{\text{spin}}, \boldsymbol{\rho}_{\vec{k}}(t)] + \frac{1}{\hbar}e(\vec{E} + \vec{v}_{\vec{k}} \times \vec{B}) \cdot \vec{\nabla}_{\vec{k}} \boldsymbol{\rho}_{\vec{k}}(t) + \vec{v}_{\vec{k}} \cdot \vec{\nabla}_{\vec{r}} \boldsymbol{\rho}_{\vec{k}}(t) = \left(\frac{\partial \boldsymbol{\rho}_{\vec{k}}(t)}{\partial t}\right)_{\text{coll}} \tag{26}$$

Due to the term $[H_{\text{spin}}, \boldsymbol{\rho}_{\vec{k}}(t)]$, the initial distribution $\boldsymbol{\rho}_{\vec{k}}(0)$ that was diagonal in spin obtains non-diagonal terms, $\rho_{\vec{k};ss'}(t)$.

Next we focus on the collision term, assuming that the scattering is elastic and spin-conserving. The squared modulus of the scattering amplitude from $\vec{k}$ to $\vec{k}'$ is, as we defined earlier, $P_{\vec{k}\vec{k}'}$. However, on top of this, one has to account for that $\vec{k}$ must be occupied and $\vec{k}'$ empty (due to the Pauli principle). This results in the expression $\sum_{\vec{k}'} P_{\vec{k}\vec{k}'}\,\rho_{\vec{k};ss'}(1 - \rho_{\vec{k}';ss'})$ for the loss rate from $\vec{k}$ to all $\vec{k}'$.[7] But the collision term must also include the gain rate due to scattering from all $\vec{k}'$ into $\vec{k}$, which is $\sum_{\vec{k}'} P_{\vec{k}'\vec{k}}\,\rho_{\vec{k}';ss'}(1 - \rho_{\vec{k};ss'})$. Due to the principle of detailed balance, $P_{\vec{k}\vec{k}'} = P_{\vec{k}'\vec{k}}$. Summing up we obtain

$$\left(\frac{\partial \boldsymbol{\rho}_{\vec{k}}(t)}{\partial t}\right)_{\text{coll}} = \sum_{\vec{k}'} P_{\vec{k}\vec{k}'}(\boldsymbol{\rho}_{\vec{k}'} - \boldsymbol{\rho}_{\vec{k}}) \tag{27}$$

At the end the observable spin density can be found by taking the trace of the spin-density matrix with the spin operator: $\vec{s}_{\vec{k}}(\vec{r}, t) = \text{Tr}(\frac{\hbar}{2}\vec{\boldsymbol{\sigma}}\boldsymbol{\rho}_{\vec{k}}(\vec{r}, t))$. Thus, the kinetic equation for the spin density is:

$$\frac{\partial \vec{s}_{\vec{k}}(\vec{r}, t)}{\partial t} - \vec{\Omega}(\vec{k}) \times \vec{s}_{\vec{k}} + \frac{1}{\hbar}e\left[(\vec{E} + \vec{v}_{\vec{k}} \times \vec{B}) \cdot \vec{\nabla}_{\vec{k}}\right] \vec{s}_{\vec{k}} + \left[\vec{v}_{\vec{k}} \cdot \vec{\nabla}_{\vec{r}}\right] \vec{s}_{\vec{k}} = \sum_{\vec{k}'} P_{\vec{k}\vec{k}'}(\vec{s}_{\vec{k}} - \vec{s}_{\vec{k}'}) \tag{28}$$

Expressions (26) and (28) are complicated integrodifferential equations that can be solved numerically by iterative methods.

[6] From an alternative aspect the Boltzmann equation (24) can be written for $\boldsymbol{\rho}$ but with an extra term added to account for the change due to the Heisenberg relation. The result is the same.

[7] Note that here the matrix elements appear, not the full matrices $\boldsymbol{\rho}_{\vec{k}}$ and $\boldsymbol{\rho}_{\vec{k}'}$.

In the absence of external fields, and assuming a homogeneous spin distribution (created e.g. by an initial homogeneous photoexcitation of a spin population in the material), we have $\vec{v}_{\vec{k}} \cdot \vec{\nabla} \boldsymbol{\rho} = 0$ and the Boltzmann equation (26) becomes

$$\frac{\partial \boldsymbol{\rho}_{\vec{k}}(t)}{\partial t} - \frac{i}{\hbar}[H_{\text{spin}}, \boldsymbol{\rho}_{\vec{k}}(t)] = \sum_{\vec{k}'} P_{\vec{k}\vec{k}'}(\boldsymbol{\rho}_{\vec{k}'} - \boldsymbol{\rho}_{\vec{k}}) \tag{29}$$

Under some simplifying approximations, this equation can be solved analytically, giving an estimate for the spin relaxation time. Here we present the basic ideas, and refer to [2] for a thorough derivation.

First, the matrix $\boldsymbol{\rho}_{\vec{k}}$ is decomposed in a spherically symmetric part $\langle \boldsymbol{\rho} \rangle$ (averaged over $\vec{k}$, assuming an isotropic dispersion relation $\hbar^2 k^2/2m$) plus a remaining anisotropic part, $\boldsymbol{\rho}_{1\,\vec{k}}$ (for which $\langle \boldsymbol{\rho}_{1\,\vec{k}} \rangle = 0$): $\boldsymbol{\rho}_{\vec{k}} = \langle \boldsymbol{\rho} \rangle + \boldsymbol{\rho}_{1\,\vec{k}}$. The anisotropic part is very rapidly brought to a semi-equilibrium value, due to momentum relaxation. Also, $\boldsymbol{\rho}_{1\,\vec{k}}$ turns out to be small, of the order of H_{spin}, so that terms containing $H_{\text{spin}}\boldsymbol{\rho}_{1\,\vec{k}}$ are dropped, simplifying the analysis. First, a solution for the semi-equilibrium value of $\boldsymbol{\rho}_{1\,\vec{k}}$ is found, which is then used to find the time-dependence of $\langle \boldsymbol{\rho} \rangle$. Taking the cubic Dresselhaus term (16) in the spin hamiltonian, the following result is obtained [2] for τ_{s}:

$$\frac{1}{\tau_{\text{s}}} = \frac{32}{105}\frac{1}{\gamma_3}\,\tau_{\text{p}}\,\alpha^2 \frac{E_{\vec{k}}^3}{\hbar^2 E_{\text{g}}}. \tag{30}$$

Here, γ_3 is a measure of the momentum relaxation time relative to the randomization time of $\vec{\Omega}(\vec{k}\,)$ (the two cannot be very different, since $\vec{\Omega}(\vec{k}\,)$ is randomized by momentum scattering, but the exact value of γ_3 depends on the details of the scattering mechanism).

Final remarks

In bulk systems, the D'yakonov-Perel' mechanism is most important for wide-gap semiconductors. Furthermore, the related spin dephasing increases strongly with temperature. For degenerate semiconductors [where the scattering is is practically only at E_{F}, so that we can set $E_{\vec{k}} = E_{\text{F}}$ in Eq. (30)], the temperature dependence of τ_{s} follows the one of τ_{p}; for example, for charged defects, $1/\tau_{\text{s}} \propto 1/\tau_{\text{p}} \propto T^{3/2}$ [1]. In non-degenerate semiconductors, an averaging that includes the Fermi distribution function is needed. An in-depth analysis of the D'yakonov-Perel' mechanism, including the effect of external magnetic fields has been performed by Pikus and Titkov [44]. Finally we note that, although in this section we mentioned phonons and impurities as sources of scattering, electron-electron collisions can also be a serious source of scattering especially at high energies, and can lead to spin relaxation of the D'yakonov-Perel type [45].

8 Spin-flip scattering due to impurity spin-orbit coupling

When electrons scatter off impurity atoms, there is an amplitude for spin flip merely due to the spin-orbit coupling of the impurity. We will call this *on-site* spin flip, to distinguish it from the *band-structure-type* spin flip of the Elliott-Yafet or D'yakonov-Perel' type. This mechanism for spin relaxation cannot be considered as independent from the Elliott-Yafet or D'yakonov-Perel' mechanisms in the spirit of Eq. (4), because interference effects can occur.[8] However,

[8] For example, a spin is flipped at an impurity, but flipped again by the Elliott-Yafet mechanism, in a single step. Then one must add the corresponding amplitudes, instead of adding the probabilities as Eq. (4) suggests.

under certain conditions, the impurity-induced spin-orbit coupling will be the main source of spin relaxation. These conditions are: (i) The host spin-orbit coupling should be small enough compared to the impurity spin-orbit coupling, so that, at each impurity-scattering event, the band-structure-type spin-flip is negligible compared to the on-site spin flip. (ii) The temperature should be low enough that phonon-induced band-structure-type of mechanism should also be negligible—actually, since the phonon-induced band-structure-type spin relaxation scales with the number of phonons, while impurity scattering scales with the concentration of impurities, the low-temperature limit depends on the impurity concentration. (iii) The energy of the initial spin population should not be too high, otherwise electron-electron scattering can dominate.

Furthermore, impurities are usually considered as independent scatterers, meaning that the scattering rate is assumed to increase linearly with impurity concentration, and scattering rates by different types of impurities are also additive. This is mostly a good approximation, but can break down at low dimensions (e.g., quantum wires or ultrathin films), at extremely low temperatures, or if the impurities are not statistically distributed (e.g., if they tend to cluster). In these cases, correlated scattering events can become important.

On-site spin-flip scattering has been calculated within density-functional theory for s-p impurities in Mg [46, 47] and in Cu [48]. The spin-orbit coupling of the host was neglected in both works. The calculations were done in two steps. First, neglecting the impurity spin-orbit coupling, the the self-consistent impurity potential $V(r)$ and the l-dependent scattering solution of the Schrödinger equation, $R_l(r;E)$ were found. Then the additional, spin-orbit related scattering was expressed [46] by the (j,l)-dependent scattering matrix $\delta t_{jl}^{\rm soc}$, which was found within first-order perturbation theory (Born approximation) as

$$\delta t_{jl}^{\rm soc}(E_{\rm F}) = \int R_l^2(r;E_{\rm F}) \left\{ \begin{array}{c} l \\ -l-1 \end{array} \right\} \xi(r) r^2 dr, \tag{31}$$

with $\xi(r) = e\hbar/(4m^2c^2)(1/r)dV(r)/dr$ [see Eq. (6)]. The factors l and $-l-1$ are used for $j = l+1/2$ and $j = l-1/2$, respectively. By using the t-matrix the spin-orbit cross section was calculated and found to compare well to results from weak antilocalization experiments. Note that the spin-flip part of the spin-orbit cross section, $\sigma_{\rm sf}$, is related to the spin relaxation time by $\sigma_{\rm sf} = xV_{\rm at}/(v_{\rm F}\tau_{\rm s})$ where x is the impurity concentration, $V_{\rm at}$ the atomic volume and $v_{\rm F}$ the Fermi velocity [48].

It was found [46, 47] that, along the s-p series (Cu-Kr and Ag-Xe), the spin-orbit cross section first increases strongly, peaks towards the end and finally drops. This behavior was interpreted via the increase of the spin-orbit coupling strength with the atomic number together with the decrease of the p-orbital matrix element of $\xi(r)$, ξ_p, at $E_{\rm F}$ towards the end of the series (as the p electrons become fully occupied and are transferred to energies below $E_{\rm F}$).

In reference [48] the values of $\tau_{\rm p}(\vec{k})$ and $\tau_{\rm s}(\vec{k})$ were analyzed as functions of the position of $\vec{k}$ on the Cu Fermi surface. A clear correlation between the two was found, except for the case of Ga impurities. The interpretation was that the Ga impurity is mostly an s-wave scatterer, and s-electrons experience no spin-orbit coupling. An important conclusion of this work is that a fully self-consistent treatment of the scattering problem, including charge relaxation around the impurity, is essential for the correct description of the spin relaxation time.

Acknowledgments

I would like to thank Gustav Bihlmayer for clarifications on the spin-orbit splitting of surface states in noble metals and Carola Meyer for discussions on the hyperfine interaction. In preparing the present manuscript I was highly benefited from the review articles of Žutić, Fabian and Das Sarma [1] and Fabian et al. [2].

Appendix

A Notation and symbols

Matrices are presented in boldface: $\boldsymbol{M}$. Vectors are designated by an arrow: $\vec{B}$. Unit vectors are designated by a "hat": $\hat{x}$, $\hat{y}$, $\hat{z}$. The Pauli matrix vector is

$$\begin{aligned} \vec{\boldsymbol{\sigma}} &= \boldsymbol{\sigma}_x\,\hat{x} + \boldsymbol{\sigma}_y\,\hat{y} + \boldsymbol{\sigma}_z\,\hat{z} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\hat{x} + \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\hat{y} + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\hat{z} \end{aligned}$$

The spin operator is denoted by $\vec{\boldsymbol{S}} = \frac{\hbar}{2}\,\vec{\boldsymbol{\sigma}}$, while the angular momentum operator by $\vec{L} = \vec{r}\times\vec{p}$; $\vec{r}$ and $\vec{p}$ are the position and momentum operators. For the crystal momentum the symbol $\vec{k}$ is used.
The electron mass is denoted by m (the effective mass by m^*), the electron charge by $e < 0$, Planck's constant is $h = 2\pi\hbar$, and the speed of light in vacuum is c. Time is denoted by t. The symbol τ is used to denote characteristic times (e.g. relaxation time), while T is reserved for the absolute temperature; k_{B} is the Boltzmann constant. The symbol for the electronic gyromagnetic ratio (ratio of magnetic moment to angular momentum) is $\gamma = g\,\mu_{\mathrm{B}}/\hbar$, where $\mu_{\mathrm{B}} = |e|\hbar/2mc$ is the Bohr magneton and g the electron g-factor; $g_0 = 2.0023\cdots$ is the free-electron g-factor. The magnetic permeability of vacuum is denoted by μ_0. E_{g} is the band gap energy and E_{F} the Fermi level. Frequencies are denoted by ω or Ω.

References

[1] I. Žutić, J. Fabian, and S. Das Sarma, *Spintronics: Fundamentals and applications*, Rev. Mod. Phys. **76**, 323 (2004).

[2] J. Fabian, A. Matos-Abiague, C. Ertler, P. Stano, and I. Žutić, *Semiconductor spintronics*, Acta Physica Slovaca **57**, 565 (2007).

[3] Y. Yafet, *g-Factors and spin-lattice relaxation of conduction electrons*, in Solid State Physics, Vol. **14**, edited by F. Seitz and D. Turnbull (Academic, New York, 1963), p.2.

[4] A.W. Overhauser, *Paramagnetic relaxation in metals*, Phys. Rev. **89**, 689 (1953).

[5] J. Bass and W.P. Pratt Jr., *Spin-diffusion in metals and alloys, and spin-flipping at metal/metal interfaces: an experimentalist's critical review*, J. Phys.: Condens. Matter **19**, 183201 (2007).

[6] G. Feher and A.F. Kip, *Electron Spin Resonance Absorption in Metals. I. Experimental* Phys. Rev. **98**, 337 (1955).

[7] F.J. Dyson, *Electron Spin Resonance Absorption in Metals. II. Theory of Electron Diffusion and the Skin Effect*, Phys. Rev. **98**, 349 (1955).

[8] M. Oestreich and W.W. Rühle, *Temperature dependence of the electron Landé g-factor in GaAs*, Phys. Rev. Lett. **74**, 2315 (1995).

[9] G. Lampel, *Nuclear dynamic polarization by optical electronic saturation and optical pumping in semiconductors*, Phys. Rev. Lett. **20**, 491 (1968).

[10] R.R. Parsons, *Band-to-band optical pumping in solids and polarized photoluminescence*, Phys. Rev. Lett. **23**, 1152 (1969).

[11] S. Oertel, J. Hübner, and M. Oestreich, *High temperature electron spin relaxation in bulk GaAs*, Appl. Phys. Lett. **93**, 132112 (2008).

[12] B.T. Jonker, G. Kioseoglou, A.T. Hanbicki, C.H. Li, and P.E. Thompson, *Electrical spin-injection into silicon from a ferromagnetic metal/tunnel barrier contact*, Nature Physics **3**, 542 (2007).

[13] A. Y. Elezzabi, M. R. Freeman, and M. Johnson, *Direct Measurement of the Conduction Electron Spin-Lattice Relaxation Time T_1 in Gold*, Phys. Rev. Lett. **77**, 3220 (1996).

[14] J.M. Kikkawa and D. D. Awschalom, *Lateral drag of spin coherence in gallium arsenide*, Nature **397**, 139 (1999).

[15] B. Beschoten, E. Johnston-Halperin, D. K. Young, M. Poggio, J.E. Grimaldi, S. Keller, S.P. DenBaars, U.K. Mishra, E.L. Hu, and D.D. Awschalom, *Spin coherence and dephasing in GaN*, Phys. Rev. B **63**, 121202 (2001).

[16] M. Johnson and R. H. Silsbee, *Interfacial charge-spin coupling: Injection and detection of spin magnetization in metals*, Phys. Rev. Lett. **55**, 1790 (1985).

[17] I. Appelbaum, B. Huang, and D.J. Monsma, *Electronic measurement and control of spin transport in silicon*, Nature **447**, 295 (2007).

[18] B. Huang, D.J. Monsma, and I. Appelbaum, *Coherent spin transport through a 350 micron thick silicon wafer*, Phys. Rev. Lett. **99**, 177209 (2007).

[19] T. Valet and A. Fert, *Theory of the perpendicular magnetoresistance in magnetic multilayers*, Phys. Rev. B **48**, 7099 (1993).

[20] S. Hikami, A. I. Larkin, and Y. Nagaoka, *Spin-Orbit Interaction and Magnetoresistance in the Two Dimensional Random System*, Prog. Theor. Phys. **63**, 707 (1980).

[21] R.I. Dzhioev, K.V. Kavokin, V.L. Korenev, M.V. Lazarev, B.Y. Meltser, M.N. Stepanova, B.P. Zakharchenya, D. Gammon, and D.S. Katzer, 2002, *Low-temperature spin relaxation in n-type GaAs*, Phys. Rev. B **66**, 245204 (2002).

[22] R.J. Elliott, *Theory of the effect of spin-orbit coupling on magnetic resonance in some semiconductors* Phys. Rev. **96**, 266 (1954).

[23] M.I. D'yakonov, and V.I. Perel', *Spin relaxation of conduction electrons in noncentrosymmetric semiconductors*, Fiz. Tverd. Tela **13**, 3581 (1971) [Sov. Phys. Solid State **13**, 3023 (1972)].

[24] G.L. Bir, A.G. Aronov, and G.E. Pikus, *Spin relaxation of electrons due to scattering by holes*, Zh. Eksp. Teor. Fiz. **69**, 1382 (1975) [Sov. Phys. JETP **42**, 705 (1976)].

[25] D. Stich, J. Zhou, T. Korn, R. Schulz, D. Schuh, W. Wegscheider, M. W. Wu, and C. Schüller, *Effect of initial spin polarization on spin dephasing and the electron g-factor in a high-mobility two-dimensional electron system*, Phys. Rev. Lett. **98**, 176401 (2007).

[26] V.P. Zhukov, E.V. Chulkov, and P.M. Echenique, *First-principle approach to the study of spin relaxation times of excited electrons in metals*, Phys. Stat. Sol. (a) **205**, 1296 (2008).

[27] D.J. Lepine, *Spin resonance of localized and delocalized electrons in phosphorus-doped silicon between 20 and 30 K*, Phys. Rev. B **2**, 2429 (1970).

[28] M.E. Rose, *Elementary Theory of Angular Momentum* (Dover Publications, 1995).

[29] P. Strange, *Relativistic quantum mechanics with applications in condensed matter and atomic physics* (Cambridge University Press, 1998).

[30] J. Fabian and S. Das Sarma, *Spin relaxation of conduction electrons in polyvalent metals: theory and a realistic calculation*, Phys. Rev. Lett. **81**, 5624 (1998).

[31] G. Dresselhaus, *Spin-orbit coupling effects in zinc blende structures*, Phys. Rev. **100**, 580 (1955).

[32] E.I. Rashba, *Properties of semiconductors with an extremum loop. 1. Cyclotron and combinational resonance in a magnetic field perpendicular to the plane of the loop*, Fiz. Tverd. Tela **2**, 1109 (1960) [Sov. Phys. Solid State **2**, 1224 (1960)].

[33] M.I. D'yakonov and V.I. Perel', *Spin orientation of electrons associated with the interband absorption of light in semiconductors*, Zh. Eksp. Teor. Fiz. **60**, 1954 (1971) [Sov. Phys. JETP **33**, 1053 (1971)].

[34] Y.A. Bychkov and E. I. Rashba, *Oscillatory effects and the magnetic-susceptibility of carriers in inversion-layers*, J. Phys. C **17**, 6039 (1984).

[35] Y.A. Bychkov and E. I. Rashba, *Properties of a 2D electron-gas with lifted spectral degeneracy*, Pis'ma Zh. Eksp. Teor. Fiz. **39**, 66 [JETP Lett. **39**, 78 (1984)].

[36] J. Luo, H. Munekata, F.F. Fang, and P.J. Stiles, *Effects of inversion asymmetry on electron energy band structures in GaSb/InAs/GaSb quantum wells*, Phys. Rev. B **38**, 10142 (1988).

[37] S. LaShell, B. A. McDougall, and E. Jensen, *Spin Splitting of an Au(111) Surface State Band Observed with Angle Resolved Photoelectron Spectroscopy*, Phys. Rev. Lett. **77**, 3419 (1996).

[38] G. Bihlmayer, Y.M. Koroteev, P.M. Echenique, E.V. Chulkov, and S. Blügel, *The Rashba-effect at metallic surfaces*, Surf. Sci. **600**, 3888 (2006).

[39] P. Monod and F. Beuneu, *Conduction-electron spin flip by phonons in metals: Analysis of experimental data*, Phys. Rev. B **19**, 911 (1979).

[40] F. Beuneu and P. Monod, *The Elliott relation in pure metals*, Phys. Rev. B **18**, 2422 (1978).

[41] R.H. Silsbee, and F. Beuneu, *Model calculation of the frequency and temperature dependence of the electron-spin-resonance linewidth of aluminum*, Phys. Rev. B **27**, 2682 (1983).

[42] J. Fabian and S. Das Sarma, *Phonon-induced spin relaxation of conduction electrons in aluminum*, Phys. Rev. Lett. **83**, 1211 (1999).

[43] J.-N. Chazalviel, *Spin relaxation of conduction electrons in n-type indium antimonide at low temperature*, Phys. Rev. B **11**, 1555 (1975).

[44] G.E. Pikus and A.N. Titkov, in *Optical Orientation*, Modern Problems in Condensed Matter Science Vol. 8, p. 109, edited by F. Meier and B.P. Zakharchenya (North Holland, Amsterdam, 1984).

[45] M.M. Glazov and E.L. Ivchenko, *Precession spin relaxation mechanism caused by frequent electron-electron collisions*, Pis'ma v Zhurnal Éksperimental' noi i Teoreticheskoi Fiziki **75**, 476 (2002) [JETP Letters **75**, 403 (2002)].

[46] N. Papanikolaou, N. Stefanou, P. H. Dederichs, S. Geier, and G. Bergmann, *First-principles calculations of the spin-orbit scattering cross section of sp impurities in Mg*, Phys. Rev. Lett. **69**, 2110 (1992).

[47] S. Geier, G. Bergmann, N. Papanikolaou, N. Stefanou, and P.H. Dederichs, *Observation of a resonance in the spin-orbit scattering of 5(s,p) impurities in Mg and Cu*, Solid State Commun. **87**, 471 (1993).

[48] D.V. Fedorov, P. Zahn, M. Gradhand, and I. Mertig, *First-principles calculations of spin relaxation times of conduction electrons in Cu with nonmagnetic impurities*, Phys. Rev. B **77**, 092406 (2008).

B 7 Galvanomagnetic Transport: from Hall Effect to AMR

P. S. Bechthold
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH
e-mail: p.bechthold@fz-juelich.de

Contents

1 Introduction **2**

2 Phenomenology of Galvanomagnetic Transport **4**

3 Hall Effect **6**

4 Lorentz-Magnetoresistance **11**

5 The Boltzmann Equation **17**

6 Resistivity of Ferromagnets **20**

7 Anisotropic Magnetoresistance (AMR) **24**

8 Conclusion **31**

A Derivation of Equation (32) **31**

1 Introduction

Galvanomagnetic transport denotes the transport properties of metals and semiconductors under the simultaneous action of electric and magnetic fields. If we apply an electric field $\mathbf{E}$ to a homogeneous metal or semiconductor it will generate a current that on a wide range of conditions will obey Ohm's law , which on a microscopic level states a linear relationship between the local current density $\mathbf{j}$ and the field $\mathbf{E}$:

$$\mathbf{j} = \sigma \mathbf{E}. \tag{1}$$

σ is the electrical conductivity. Equivalently, we could have written

$$\mathbf{E} = \rho \mathbf{j}. \tag{2}$$

where ρ is the resistivity, the inverse of the conductivity. More general, a charged particle moving in an electromagnetic field is subject to the Lorentz force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{3}$$

where q is the charge of the particle, $\mathbf{E}$ is an accelerating electric field, $\mathbf{v}$ is the velocity of the particle and $\mathbf{B}$ is the magnetic flux density or magnetic induction. Thus, a free electron moving with a constant velocity $\mathbf{v}$ (electric field $\mathbf{E} = \mathbf{0}$) perpendicular to a homogeneous field $\mathbf{B}$ will be forced into a circular motion according to

$$\frac{mv^2}{r} = |e|vB. \tag{4}$$

Here $|e|$ is the modulus of the electron charge, r is the radius of the trajectory and mv^2/r is the modulus of the centripetal force. The time of circulation is $T = 2\pi r/v$ which defines the cyclotron frequency

$$\omega_c = \frac{2\pi}{T} = \frac{eB}{m}, \tag{5}$$

which for nonrelativistic particles does not depend on the particle velocity and the radius of the trajectory. This equation remains valid for conduction electrons in solids, as well, if we replace the free-electron mass m by the cyclotron mass m_c. [1]

Now, if we apply an electric field and perpendicular to it a magnetic field to a conducting solid the vector of the current density will no longer be parallel to the electric field, at least for a short time (see above). In general, the presence of a magnetic field will generate an anisotropy so that the current-density vector will deviate from the direction of the applied electric field. We then have to define a tensorial relationship between the current density and the applied fields which for the vector components - referred to a right handed orthonormal reference system (x, y, z) gives:

$$j_i = \sigma_{ij}^0 E_j + \alpha_{ijk} E_j B_k + \beta_{ijkl} E_j B_k B_l + higher\ order\ terms \quad (i, j, k, l = x, y, z). \tag{6}$$

[1] In case of spherical energy surfaces (Fermi surfaces) the cyclotron mass is identical to the effective mass. In more general cases it can be calculated from the effective mass tensor [1]:

$$m_{ij}^*(\mathbf{k}) = [\frac{1}{\hbar^2} \frac{\partial^2 \epsilon(\mathbf{k})}{\partial k_i \partial k_j}]^{-1}.$$

Here we have used Einstein's summation convention which says that the sum has to be taken over repeated indices on the right hand side of the equation. The extra index 0 at σ_{ij}^0 will become clear below. The tensor components $\sigma_{ij}^0, \alpha_{ijk}$ and β_{ijkl} are material constants and are subject to symmetry restrictions which depend on the crystallographic symmetry of the sample material [2]. This looks quite complicated, but if there are no second order or higher order terms in the electric field components we can – as is usually done – include all the magnetic field contributions into a generalized definition of the electric conductivity. The price, we have to pay, is that now the conductivity tensor will depend on the magnetic field.

$$j_i = \sigma_{ij}(\mathbf{B})E_j. \tag{7}$$

With these conditions the σ_{ij}^0 components in eq. 6 correspond to $\sigma_{ij}(\mathbf{B} = \mathbf{0})$. Using the Onsager relations [3] we get the symmetry restriction

$$\sigma_{ij}(\mathbf{B}) = \sigma_{j\,i}(-\mathbf{B}). \tag{8}$$

That is, the conductivity tensor is symmetric with respect to a simultaneous reversal of the magnetic induction. For the reverse relation (2) we now get:

$$E_j = \rho_{j\,i}(\mathbf{B})j_i, \tag{9}$$

where $\underline{\underline{\rho}}$ is the inverse tensor of $\underline{\underline{\sigma}}$.
The components of both these tensors are related by

$$\rho_{ii} = (\sigma_{jj}\sigma_{kk} - \sigma_{jk}\sigma_{kj})/\Delta(\sigma);\ \rho_{ij} = (\sigma_{ik}\sigma_{kj} - \sigma_{ij}\sigma_{kk})/\Delta(\sigma) \tag{10}$$

where $\Delta(\sigma)$ is the determinant of σ_{ij}.
Similarly,

$$\sigma_{ii} = (\rho_{jj}\rho_{kk} - \rho_{jk}\rho_{kj})/\Delta(\rho);\ \sigma_{ij} = (\rho_{ik}\rho_{kj} - \rho_{ij}\rho_{kk})/\Delta(\rho). \tag{11}$$

In the absence of a magnetic field, σ_{ij} and $\rho_{\,ij}$ are symmetrical, $\sigma_{ij} = \sigma_{j\,i}$, and this implies that a sytem of orthogonal axes can be found, with respect to which they are diagonal. When a field $\mathbf{B}$ is present the tensors are generally not symmetrical and each of them requires all nine components for a complete specification.
In the following we will only consider constant currents, i.e. DC currents. Moreover, we will assume that our sample is at a constant temperature. We will ignore the Joule heating and therefore exclude any temperature gradients. We will also restrict the discussion to the low-field limit throughout this contribution. In mathematical terms this situation is described by $\omega_c\tau \ll 1$ where τ is the scattering time of the charge carriers. This means that the charge carriers are scattered many times by phonons, impurities or defects before they can complete a full cyclotron orbit. In this case we do not need to worry about Landau level quantization or any quantum oscillations that occur in the high field limit $\omega_c\tau \gg 1$, and may lead, for example, to the quantum Hall effect.
The remainder of this contribution is organized as follows. In Chapter 2 we will start with a phenomenological description of the galvanomagnetic transport based on the classical equations of motion. In the subsequent chapters we will deal with the galvanomagnetic behaviour of non-magnetic materials and first discuss the ordinary Hall effect and then the Lorentz magnetoresistance. In this course we will soon realize that the classical description is insufficient and that a quantum mechanical model based on the Pauli principle (Fermi statistics) and electronic

band structure – the so called semiclassical model – will be needed to understand and interpret the experimental observations. Within this framework we will then derive an equation for the electrical conductivity with the aid of the Boltzmann equation. Finally, we will study what changes when the materials are ferromagnetic themselves and carry an internal magnetization. We will finish this contribution with a basic description of the anisotropic magnetoresistance (AMR) and its applications.

2 Phenomenology of Galvanomagnetic Transport

The first attempt to calculate the electrical conductivity was made by Drude prior to the development of the quantum theory [1, 4].
His model is based on the assumption of a free electron gas moving in the confinement of a metal block, interacting with the metal ions only through elastic collisions. Today we know that Drudes model is only valid within certain limits as we shall also see below. But much of his Ansatz remains applicable when certain quantities like velocities, the carrier masses and the results are interpreted in terms of band structure theory [2].
Drude's starting point was the classical equation of motion:

$$m\frac{dv}{dt} + \gamma v = -e\ E, \tag{12}$$

where he introduced a velocity dependent damping ($\gamma = m/\tau$) due to scattering processes. v is the drift velocity of the electrons which decays with a relaxation time τ when the accelerating field is switched off. The superimposed thermal motion of the carriers averages to zero. In the stationary state we have $dv/dt = 0$ and we get

$$v = -\frac{e\tau}{m}\ E = -\mu E. \tag{13}$$

Thus the drift velocity is proportional to the accelerating field where the quantity

$$\mu = \frac{e\tau}{m} \tag{14}$$

is called the charge carrier mobility. With a charge carrier density n we then get a current density

$$j = -env = en\mu E. \tag{15}$$

Together with Ohm's law

$$\mathbf{j} = \sigma\mathbf{E}. \tag{16}$$

we finally get the conductivity

$$\sigma = en\frac{e\tau}{m} = en\mu. \tag{17}$$

The conductivity is proportional to the charge, the density, and the mobility of the charge carriers. It increases with the relaxation time which is a measure of the time between scattering events.

[2] Sommerfeld extended his model to a quantum mechanical free electron gas by introducing the Pauli principle through the Fermi-Dirac distribution function. This was later further extended by introducing the lattice periodicity caused by a weak periodic potential and leading to the band structure model with Bloch functions, reciprocal lattices, Brillouin zones etc [1].

Now we extend equation (12) by introducing an additional magnetic field:

$$m\frac{d\mathbf{v}}{dt} + \frac{m}{\tau}\mathbf{v} = -e\ (\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{18}$$

In the stationary state $d\mathbf{v}/dt = 0$ we then get

$$\mathbf{v} = -\frac{e\tau}{m}\ (\mathbf{E} + \mathbf{v} \times \mathbf{B}) = -\mu(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{19}$$

The current density now becomes:

$$\mathbf{j} = -en\mathbf{v} = en\mu\mathbf{E} - \mu(\mathbf{j} \times \mathbf{B}) = \sigma_0\mathbf{E} - \mu(\mathbf{j} \times \mathbf{B}) \tag{20}$$

where $\sigma_0 = \sigma(B = 0)$. Without loss of generality we may assume that $\mathbf{B}$ is aligned with the z-direction of our coordinate sytem. Then we can write equation (20) in terms of components:

$$\begin{aligned} j_x &= \sigma_0 E_x - \mu B_z j_y \\ j_y &= \sigma_0 E_y + \mu B_z j_x \\ j_z &= \sigma_0 E_z \end{aligned} \tag{21}$$

or since $\mu B_z = \tau\frac{e}{m}B_z = \tau\omega_c$

$$\begin{aligned} \sigma_0 E_x &= j_x - \tau\omega_c j_y \\ \sigma_0 E_y &= \tau\omega_c j_x + j_y \\ \sigma_0 E_z &= j_z \end{aligned} \tag{22}$$

or in matrix notation ($\rho_0 = 1/\sigma_0$):

$$\begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \rho_0 \begin{pmatrix} 1 & (\tau\omega_c) & 0 \\ -(\tau\omega_c) & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \rho_{ij}(B_z) j_j \tag{23}$$

or

$$\begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \rho_0 \begin{pmatrix} 1 & (\mu B_z) & 0 \\ -(\mu B_z) & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \rho_{ij}(B_z) j_j \tag{24}$$

and

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \frac{\sigma_0}{1 + (\tau\omega_c)^2} \begin{pmatrix} 1 & -(\tau\omega_c) & 0 \\ (\tau\omega_c) & 1 & 0 \\ 0 & 0 & 1 + (\tau\omega_c)^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \sigma_{ij}(B_z) E_j \tag{25}$$

or

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \frac{\sigma_0}{1 + (\mu B_z)^2} \begin{pmatrix} 1 & -(\mu B_z) & 0 \\ (\mu B_z) & 1 & 0 \\ 0 & 0 & 1 + (\mu B_z)^2 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \sigma_{ij}(B_z) E_j. \tag{26}$$

We see now that the presence of a magnetic field renders the resistivity and the conductivity anisotropic quantities even for an isotropic material. The conductivities and resistivities have changed in the (x, y)-plane. The consequences of this will be discussed in the next chapters.

3 Hall Effect

This effect was discovered by Edwin H. Hall during his PhD-work and was published in 1879 [5, 6, 7]. Let us assume a non-magnetic slab of a conducting material with parallel planes, length l, width b and thickness d like the one depicted in Fig. 1 and a coordinate system as indicated in the figure. We apply an electric field parallel to the long axis and a magnetic field perpendicular to the slab. When we switch on the electric field the charge carriers will initially be deflected sideways by the Lorentz force, the charges will get accumulated at one side, until the generated electric forces completely balance the Lorentz force. At steady state conditions the current will flow parallel to the applied electric field and we can measure a "Hall"-voltage U_H between the sample sides. The stationary-state condition is that the Lorentz force and the transverse electric force cancel each other, i.e.:

$$F_y = -e(\mathbf{v} \times \mathbf{B})_y - eE_y = ev_x B_z - eE_y = 0, \tag{27}$$

where v_x is the drift velocity of the charges and $E_y = U_H/b$ is the Hall field. When we assume that only electrons are the charge carriers – like in metals or n-doped semiconductors at suffiently low temperatures – the current density in the x-direction is $j_x = -en_e v_x = i/(b \cdot d)$. Here, n_e is the electron density of the sample and i is the total current. Therefore we get

$$U_H = E_y \cdot b = -\frac{1}{n_e e} j_x B \cdot b = -\frac{1}{n_e e} i\ B/d = R_H i\ B/d. \tag{28}$$

R_H is called Hall coefficient. Thus, by measuring the current i, the field $\mathbf{B}$, the thickness d of the slab and the Hall voltage U_H we can determine the charge carrier density n_e of our sample. What we have obtained here is the standard textbook derivation of the Hall field [8]. We could have obtained this result also from our equations (23) - (26) by putting the transverse current $j_y = 0$. We will now use this relation to calculate the current density in our sample. With $j_y = 0$ from equation (26) we get $(\mu B_z)E_x + E_y = 0$. Thus the longitudinal current density gets:

$$j_x = \frac{\sigma_0}{1 + (\mu B_z)^2}(E_x - (\mu B_z)E_y) = \frac{\sigma_0}{1 + (\mu B_z)^2}(E_x + (\mu B_z)^2 E_x) = \sigma_0 E_x. \tag{29}$$

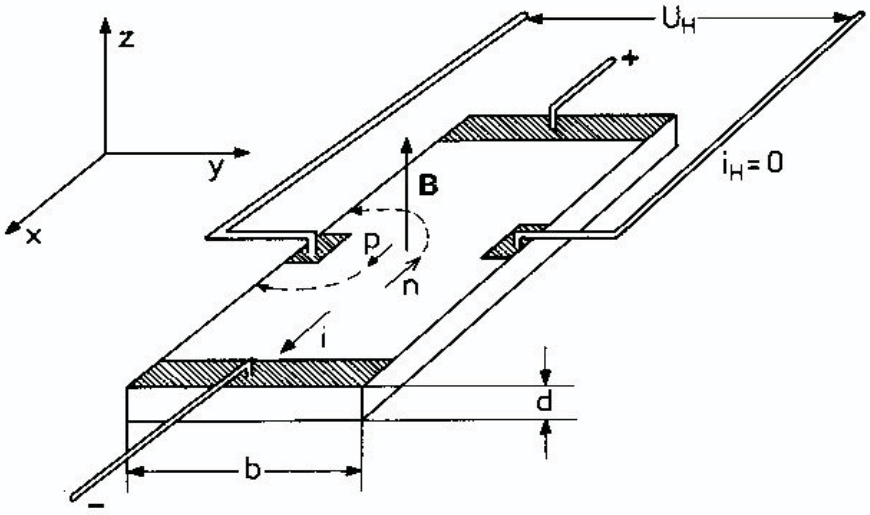


Fig. 1: *Scheme of a Hall arrangement [8].*

Despite the anisotropy in equation (26) the resistivity of the longitudinal current has not changed. Experimentally, however, one might observe changes in the resistance. One simple reason is in the velocity distribution of the charge carriers [9, 10]. This means that the Lorentz force is slightly different for different charge carriers, while the action of the Hall field is always the same. Accordingly, the Lorentz forces are compensated on average only. Individual charges might still be slightly deflected from their ideal trajectories. The longer path lengths then lead to an increase of the resistivity. Anyway, to measure significant changes in the resistance – the so called magnetoresistance – we have to prevent the formation of the Hall field. How this can be done will be discussed in the next section.

To get a significant Hall signal the sample should be thin and the charge-carrier density should not be too large. That is why semiconductors are the preferred materials in Hall sensors.

In p-doped semiconductors the Lorentz force on holes is directed toward the same side of the sample, because of their positive charge and opposite drift velocity. Thus, if we consider a p-doped semiconductor at low temperature, where mainly holes are the charge carriers, the electric field E_y and with it the polarity of the voltage U_H and the Hall coefficient will change sign. Therefore, from the sign of the voltage and the Hall coefficient we will be able to distinguish the charge type of the main carriers.

While monovalent metals – particularly the alkalis – nicely obey this simple classical rule others strongly deviate. The deviations observed in ferromagnetic metals will be addressed below. Some divalent group IIA and group IIB metals even show a reversed sign of the Hall coefficient [3]. The reason is found in their electronic structure, which has to be treated quantum mechanically. The atoms of the group II metals all exhibit a fully occupied s^2 electronic subshell. Thus, one might rather expect a weak van der Waals like bond rather than a metallic bond. Dimers and small clusters of these materials are indeed very weakly bound. They become metallic in character at larger particle sizes, because the s-subshell finally hybridizes with the respective p-subshell. Thus, at the Fermi level one expects an almost completely filled valence band with a negative curvature resulting in a negative effective electron mass. This contribution can be replaced by the compensating contribution of holes – now with positive effective mass.

A more rigorous treatment of positive charge carriers in metals can be obtained from a detailed investigation of the Fermi surface. We want to illustrate this with the aid of Fig. 2. It shows schematically the Harrison construction of a fictitious square reciprocal lattice with Fermi spheres that exceed the first Brillouin zone. The rule that Fermi surfaces should cut the zone boundary at right angles is ignored here for clarity. According to the Pauli principle only a small fraction of electrons in the Fermi sphere can contribute to the charge transport. They are located in an energy range of the order of kT around the Fermi energy. In a semiclassical approach we may consider an electron as a wave packet moving in reciprocal space [1, 12]. Its velocity in k-space is the group velocity of this wave packet:

$$\mathbf{v}(k) = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{1}{\hbar}\frac{\partial \varepsilon(\mathbf{k})}{\partial \mathbf{k}} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\varepsilon(\mathbf{k}) \tag{30}$$

Its momentum is $\mathbf{p} = \hbar\mathbf{k}$. The equation of motion of the electron in a magnetic field is then

$$\hbar\frac{d\mathbf{k}}{dt} = -e(\mathbf{v}(\mathbf{k}) \times \mathbf{B}). \tag{31}$$

From these equations we learn that the electron moves perpendicular to $\mathbf{B}$ and $\mathbf{v}(\mathbf{k})$ in k-space. Moreover, $\mathbf{v}(\mathbf{k})$ points in the direction of a gradient of a constant energy surface which is the

[3] In this situation, the simple Drude model fails.

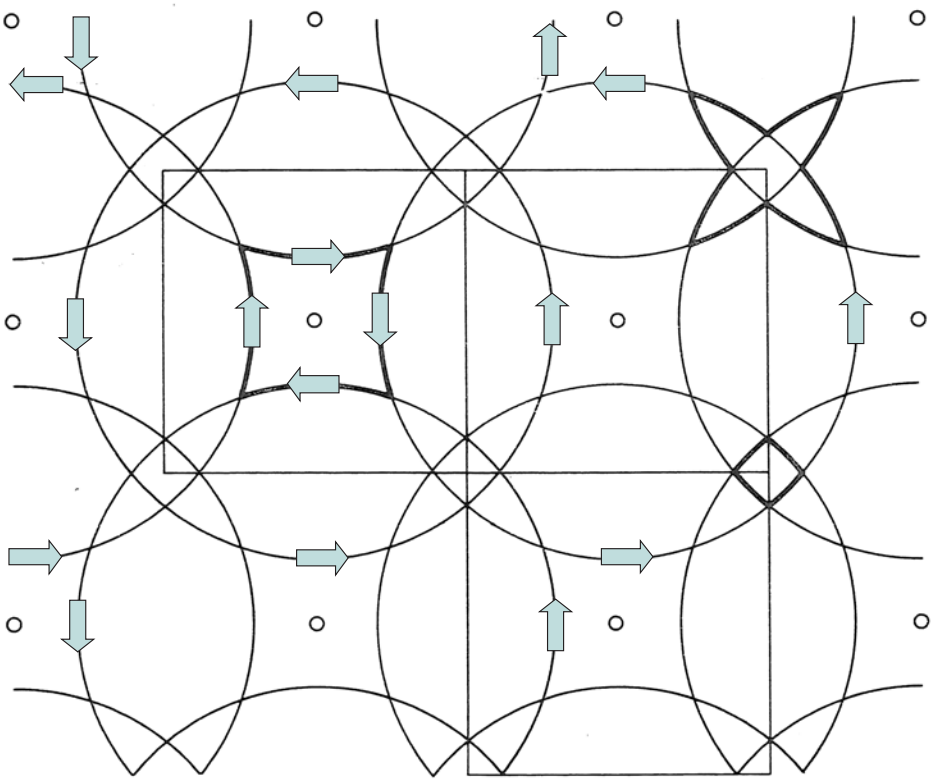


Fig. 2: *Simplified scheme of a reciprocal lattice with Fermi spheres that exceed the size of the first Brillouin zone. A square hole pocket is formed that can contribute a positive part to the Hall coefficient of a metal or semiconductor. See text for details. [11].*

direction perpendicular to the surface. In other words: in k-space the electron moves tangential to a constant energy surface which is the Fermi surface or in the simplified picture of fig. 2 the Fermi sphere. Let us assume that all the electrons move counterclockwise around the Fermi surface as indicated by the arrows in the figure. Then we see that at the surface of the accentuated square the electrons move clockwise around the square. This is precisely the behaviour of positive charges which make a positive contribution to the Hall coefficient. Thus, the square is a hole pocket. This picture is also consistent with the imagination of an almost completely filled band, because if we fill up the Fermi spheres with more electrons the Fermi spheres will expand and the hole pocket will shrink. When the spheres completely overlap the hole pocket will disappear. The respective band is then filled up and will no longer contribute to the charge transport. This simple example shows that the Fermi surfaces of materials need to be studied very carefully if Hall coefficients are to be determined.

From this simple consideration we must conclude that the charge transport in the quantum mechanical treatment deviates drastically from the classical picture. In the classical picture the current is equally carried by all the electrons. In the quantum mechanical picture only a small fraction of the electrons in the vicinity of the Fermi level will contribute to the conductivity, all moving approximately with the Fermi velocity. The rest is blocked by the Pauli principle. The classical picture assumes that the electrons collide with ion cores. In quantum mechanics, if the ion cores are kept in their perfect lattice positions there will be no scattering at all. This is because the electrons are described by Bloch waves. They are eigenfunctions of the system and therefore they are time independent. Thus, quantum mechanically scattering can occur only

due to lattice imperfections, i. e. phonons, impurities and defects, including surfaces. In the quantum picture the conduction phenomena occur close to the Fermi surface, and despite the defects mainly the differences of the Fermi surfaces are determining the conduction behavior of the materials.
Why then can we use classical formulas at all? This point will be addressed in chapter 5 on the Boltzmann equation, which takes into account the Fermi-Dirac distribution and leads to the same conductivity formula as the classical picture. With this in mind we have to check all the results deduced from the classical approach for their consistency with quantum mechanics. It is an interesting aspect of this consideration that the classical formulas could only be saved by introducing positive charge carriers into quantum mechanics.
With these remarks we come back to the Hall effect. When both type of charge carriers contribute to the Hall effect the equation for the Hall coefficient must be modified. In general the signal will contain contributions from both electrons and holes and the sign of the Hall coefficient determines whether electrons or holes are the dominant charge carriers. This is particularly the case for semiconductors at elevated temperatures, where (due to intrinsic conduction) electrons and holes simultaneously contribute to the charge carriers. In the presence of both charge carriers we then have to deal with an ambipolar current and the valence and conduction band simultaneously. In this case not only the carrier densities (n_e for electrons and n_h for holes), but also the drift mobilities (μ_e and μ_h) of the carriers become involved. In general, the charge carriers will have different mobilities in both bands, usually $\mu_h < \mu_e$. With highly doped samples at sufficiently low temperatures essentially only one type of charge carriers will contribute to the conduction. In such cases one can get back to the single-band picture discussed above. The following equation allows us to calculate the Hall voltage. Its somewhat lengthy derivation is given in the appendix.

$$U_H = E_H \cdot b = \frac{n_h\mu_h^2 - n_e\mu_e^2}{e(n_h\mu_h + n_e\mu_e)^2} j_x B_z \cdot b = \frac{n_h - n_e(\mu_e/\mu_h)^2}{e(n_h + n_e\mu_e/\mu_h)^2}\, iB/d. \tag{32}$$

Here the ratio of the mobilities has been introduced. If electrons and holes had the same density and the same mobility, they would completely compensate each other. In reality this is not the case. The Hall field will change sign depending whether there are more electrons or more holes accumulated by the Lorentz force. U_H changes sign at $n_h = n_e(\mu_e/\mu_h)^2$ rather than at the intrinsic concentration $n_h = n_e$. This equation is widely used to characterize Hall coefficients of semiconductors and metals.
With these results we can now interpret the temperature behaviour of the Hall coefficient of InSb as given in Fig. 3 [13, 9, 14, 15] . InSb is a low band gap semiconductor $\varepsilon_g = 0.18eV$ and an important sensor material. The figure shows the modulus of the Hall coefficient as a function of the inverse temperature for various degrees of doping (temperature from right to left given in the upper legend). The letters A and B refer to n-doped samples, the numbers 1 to 4 to p-doped samples with increasing concentration of the dopend in this order. The sample denoted by V is the purest. The concentration of the dopant varies from $10^{13}/cm^3$ (curve V) to $2 \times 10^{17}/cm^3$ (curve 4) [13, 15]. As can be seen at low temperatures the modulus of the Hall coefficient decreases with increasing doping due to the increasing carrier density. For a given doping level it is largely constant at low temperatures, because of the fixed carrier concentration supplied by the dopant. This is true also for the sample V, which seems to have a small concentration of free electrons due to some remaining impurities. When the temperature increases the growing number of electron-hole pairs finally leads to an exponential decrease of the Hall coefficient as drastically demonstrated for the pure sample. Since electrons exhibit

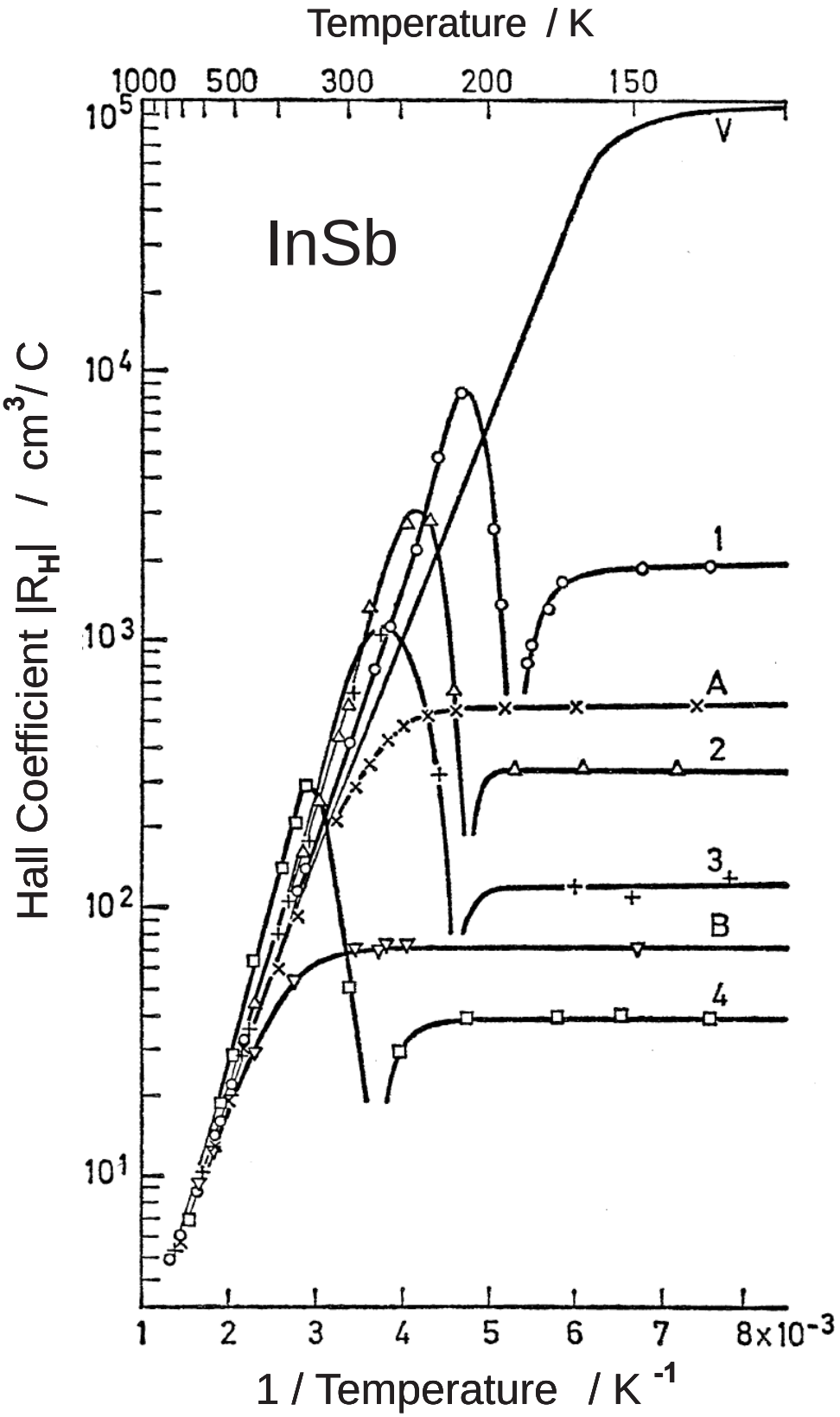


Fig. 3: *Modulus of the Hall coefficient of InSb at various degrees of doping as a function of the inverse temperature. [9, 13, 14, 15].*

a larger mobility the n-doped samples – as well as the pure sample – have a negative Hall coefficient and gradually approach the line of intrinsic conduction (curves V, A and B). In contrast, the initially positive Hall coefficient of the p-doped samples changes sign when the density of intrinsic carriers reaches a certain level. Then, it even slightly overshoots the intrinsic conduction line and finally approaches it from the other side. Naturally, the crossing of the zero line occurs at lower temperatures for samples with initially lower p-dopend concentration.
So far, we have treated all carriers alike except for the sign. In reality, they may have an energy distribution, anisotropic masses and different relaxation times. In such cases, it might be necessary to include correction factors to the simple equations given above. These factors are usually of the order of one [15, 16].
From the sign of the Hall coefficient we learn whether electrons or holes are the primary charge carriers. For a single band from the measurement of the Hall coefficient we can determine the density and the charge sign of the carriers. From the equation

$$\mu = R_h \sigma$$

and a simultaneous determination of the conductivity, we can also identify the carrier mobility. Van der Pauw [17] has described procedures to avoid and minimize experimental errors in Hall measurements. Today the Hall effect is widely used to either measure magnetic fields with calibrated Hall sensors or to determine carrier densities and mobilities of carriers in semiconductors.

4 Lorentz-Magnetoresistance

Just like the ordinary Hall effect, the Lorentz magnetoresistance (also called ordinary magnetoresistance (OMR)) occurs in all conducting materials. It describes the change of the material's resistivity when an external magnetic field is applied.
Depending on the direction of the magnetic field with respect to the current flow we distinguish between longitudinal and transverse magnetoresistance. The transverse effect is the more important. From our equations (23) - (26) we can see that the conductivity and the resistivity may change in the (x, y)-plane when a magnetic field is applied in the z-direction perpendicular to the film. To realize this situation we have to make use of the transverse current generated by the Lorentz force. In the Hall measurements we have eliminated the transverse current (on the average, see above) by building up an opposing electric field. As a result the resistivity of the thin film did not change with the magnetic field applied, at least in our simplified approach. Now we want to make use of the transverse current itself. To do so we have to avoid or to shortcut the opposing electric field. Suppose we have shortened the Hall field in Fig. 1 and apply an electric field in the x-direction $E_x = E_0$ and $E_y, E_z = 0$, we get from eqs. (23), (24): $E_x = \rho_0(j_x + (\mu B_z) j_y$. In addition, we have $E_y = -(\mu B_z) j_x + j_y = 0$ or $j_y = (\mu B_z) j_x$. For the electric field component in x-direction E_x this gives

$$E_x = \rho_0(1 + (\mu B_z)^2) j_x. \tag{33}$$

The resistance grows quadratically with the applied field. The relative increase of the resistance is usually measured by :

$$\frac{\rho(B) - \rho_0}{\rho_0} = \frac{\Delta\rho}{\rho_0} = (\mu B_z)^2 \tag{34}$$

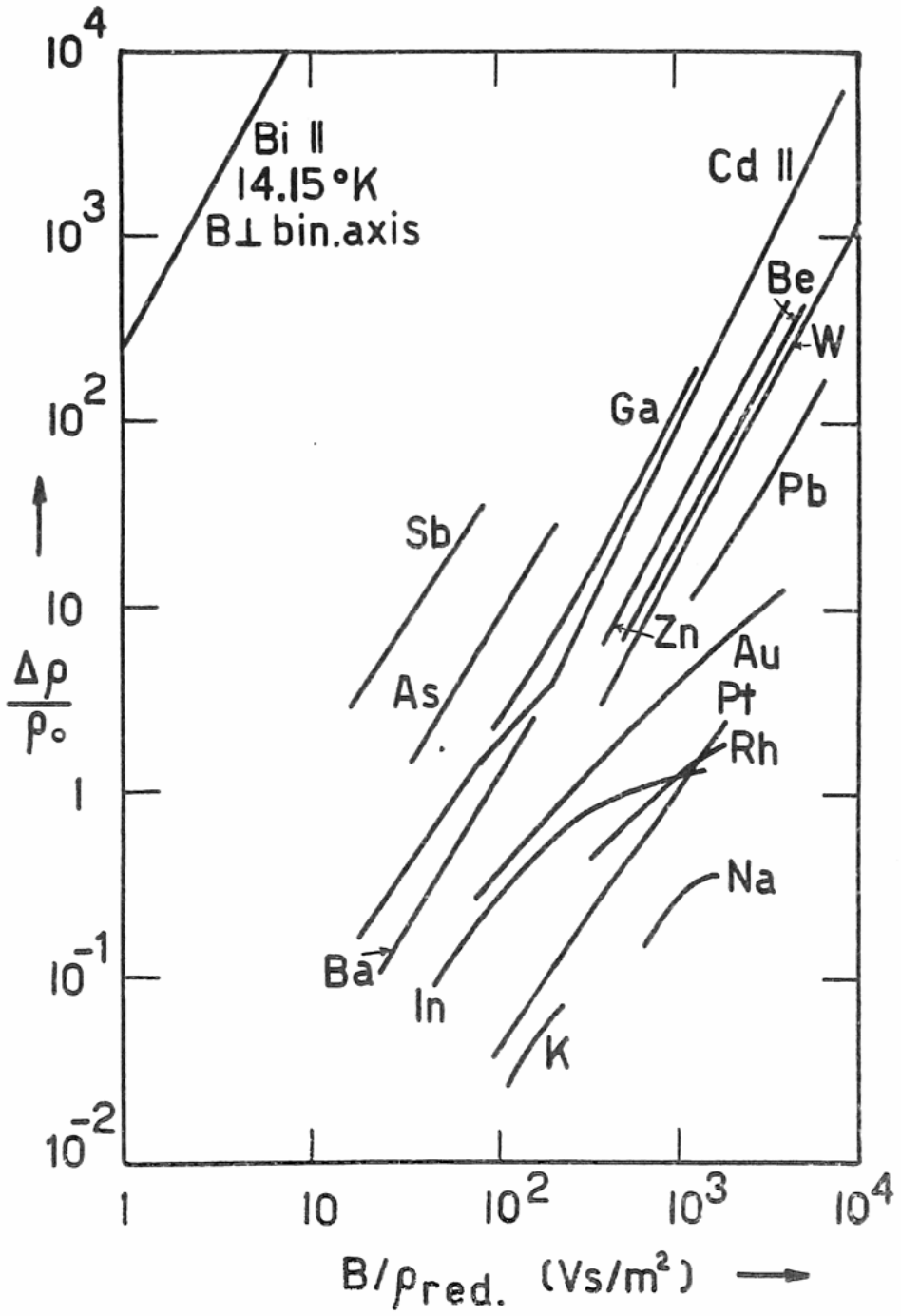


Fig. 4: *Reduced Kohler diagram where $\rho_{red} = \rho(T)/\rho(\Theta)$ and Θ is the Debye temperature of the metal. [20].*

It is this quantity that is usually called the “Lorentz magnetoresistance” or “ordinary magnetoresistance” (OMR) . The effect is qualitatively easy to understand. In between two subsequent scattering events the Lorentz force deflects the electrons on their way to the counterelectrode. This increases the electron path way and therefore the average number of collisions with phonons, impurities and defects. This consequently increases the resistance.

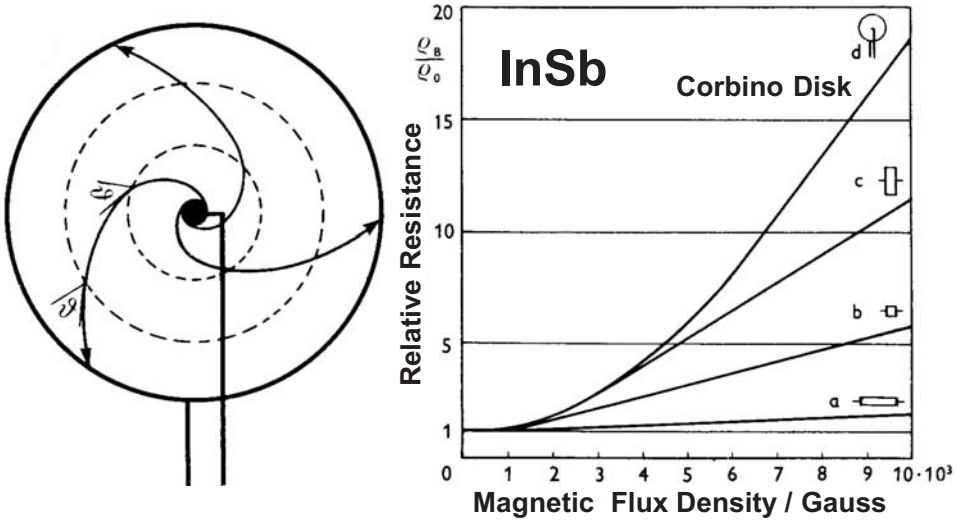


Fig. 5: *Left: Principle scheme of a Corbino disk. Right: Relative resistance as a function of magnetic induction for samples of n-doped InSb of equal purity, but different geometrical shape. The upper curve is that of a Corbino disk. Lower curves are for plain samples with differing length/width ratios (l/w =1/3, 1/1 and 10/1 from above). This demonstrates the influence of the geometrical factor. [9, 10].*

In simple metals in the low field limit it follows the scaling law

$$\frac{\rho(B)-\rho_0}{\rho_0} = \frac{\Delta\rho}{\rho_0} = const(B/\rho_0)^2, \tag{35}$$

i.e., for a given metal all measurements should follow the same parabola when $\Delta\rho/\rho_0$ is plotted versus B/ρ_0, independent of temperature or purity of the sample. This is known as Kohler's rule [4] [11, 19, 20, 21] . In our free electron approach this is evident, because $\mu B \propto \tau B$ and the scattering time is inversely proportional to ρ_0 for a given metal. Kohler's derivation is more rigorous, however, because it relies on the Boltzmann equation and thus includes the Pauli exclusion principle. Fig. 4 demonstrates that a variety of elementary metals nicely follows this rule. The Kohler rule is valid only, when all the participating charge carriers exhibit the same scattering time τ, i.e. the same microscopic scattering mechanisms. Deviations from Kohler's rule therefore indicate, that different scattering mechanisms might be at play in a sample.
In normal metals the Lorentz magnetoresistance is a small effect and has no technological applications. The only exception is the semimetal Bi, which exhibits $\sim 18\%$ magnetoresistance in a transverse field of 0.6 T [22]. Therefore, in early applications of the magnetoresistive effect a Bi-spiral was used to measure magnetic fields [23]. Later it was found that InSb shows an even larger Lorentz magnetoresistance [10], and this, indeed, led to technological applications (see below) .
Experimentally, the most effective way to avoid the formation of a Hall field is to use a Corbino disk (Fig. 5) [10, 18] . It consists of two concentric circular electrodes with the metallic or semi-

[4]The rule can even be extended to higher field ranges, when the quadratic field dependence is replaced by a material dependent function f:

$$\frac{\Delta\rho}{\rho_0} = f(B/\rho_0) \tag{36}$$

where, of course, f is parabolic in the low field limit.

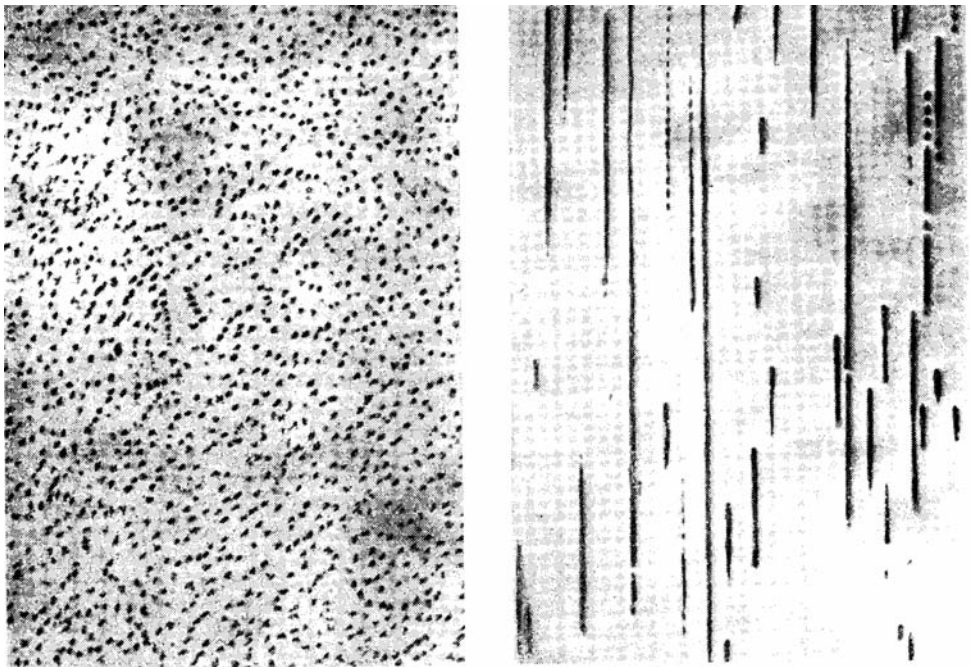


Fig. 6: *Micrographs of polished sections of an oriented InSb-NiSb eutectic as used in commercial Lorentz magnetoresistors. The samples were cut perpendicular and parallel, respectively, to the NiSb needles. The needles have diameters of about* $1\mu m$*. The length is up to about* $50\mu m$*. When oriented perpendicular to the applied electric field, the needles will shortcut the Hall field. Inside the needles the current will flow parallel to the needle axis, outside the needles it gets deflected by the Lorentz force [14, 24, 25].*

conducting material, e.g. InSb, in between. The externally applied electric field is then pointing in a radial direction and the equipotential lines are concentric circles, as well. Therefore, a balancing transverse Hall field cannot build up in this geometry. The trajectory of the current flow is a logarithmic spiral. The current meets every equipotential line at the same angle. The angle ϑ is defined by the ratio of the transverse to radial current density: $tan\vartheta = j_t/j_r$ and is called the Hall angle. With an applied magnetic field the resistance grows due to the magnetic field induced lengthening of the trajectories. Fig. 5 also shows the quadratic field dependence of the relative magnetoresistance of InSb for the Corbino disk and three other InSb samples of identical purity, but different length/widths ratio (1:3, 1:1 and 1:10). This demonstrates that the resistivity also depends on sample geometry [9, 10]. The latter effect will not be further dicussed here.

The Corbino Disk, while occasionally used in scientific studies is not suited for sensor applications, because it does not provide a large enough resistance to be useful for sensor applications. For larger resistances one has to elongate the current path length. This can be easily achieved by meandering a narrow stripe of material but then one has to meet precautions to avoid the occurance of the Hall field. The elegant solution to this is the "Feldplatte" also known as MDR (Magnetic field Dependent Resistor) . It was invented in the 1960's and since 1965 is available as a commercial product. It contains a meandering thin layer ($8-25\mu$m) of InSb which is doped with 1.8 % NiSb . At this concentration InSb and NiSb form an eutectic in which small needles of NiSb (Ø1μm, length up to 50μm) grow inside the InSb. At favourable growth conditions these needles orient parallel to the crystallographic growth front (e.g. in a Zone melting pro-

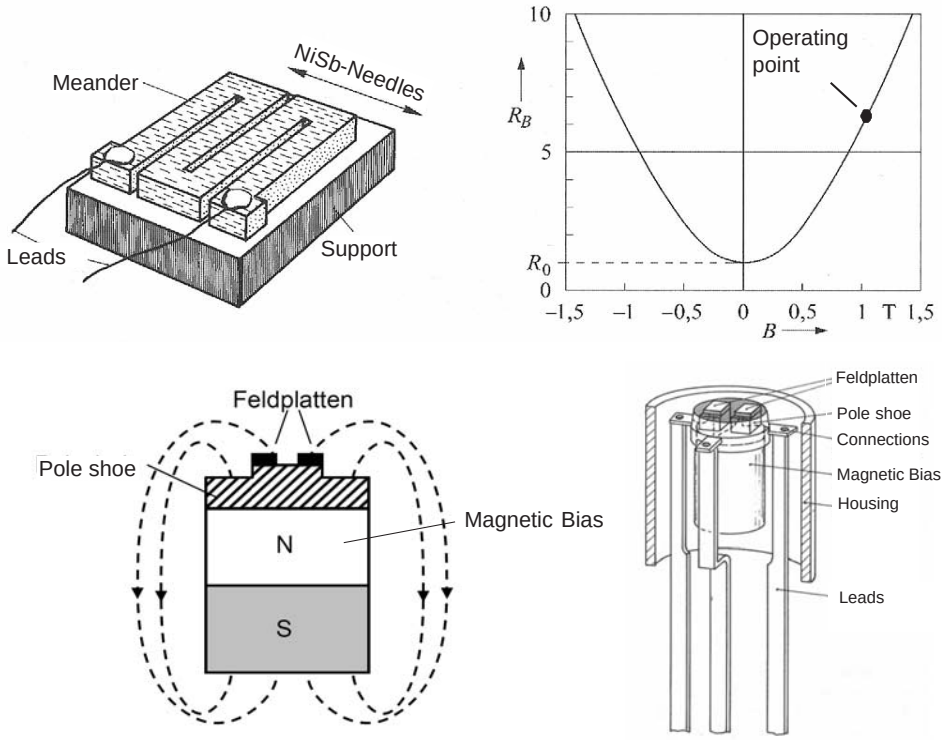


Fig. 7: *Upper left: Scheme of a meandering magnetoresistor based on the InSb-NiSb eutectic with the NiSb needles oriented perpendicular to the direction of the applied electric field. Upper right: the response curve of such a device. Lower left: Principle of magnetic bias for two "Feldplatten" combined in a bridge circuit. Lower right: Arrangement of two "Feldplatten" in a commercial device. [14, 24, 25, 26].*

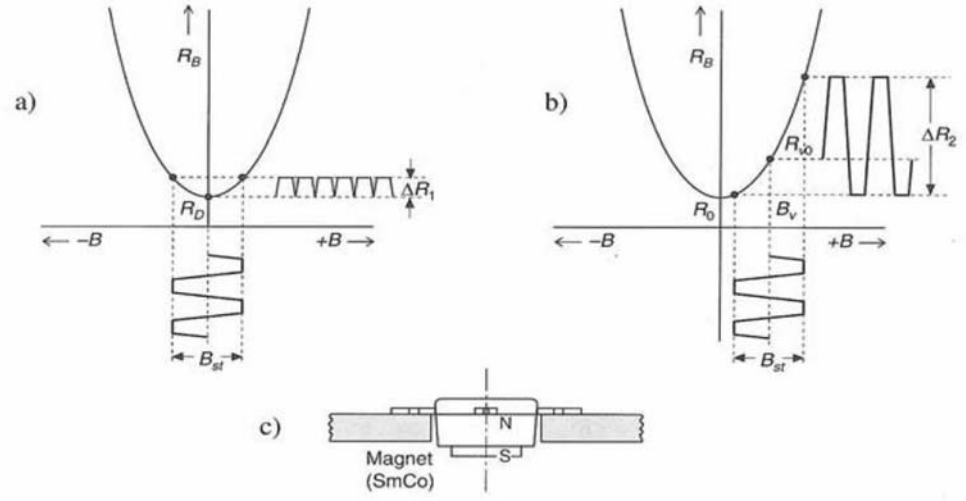


Fig. 8: *(a) Response of an unbiased Lorentz magnetoresistor. Positive and negative field directions cannot be distinguished. (b) Response of a magnetically biased magnetoresistor with the operating point shifted to the right in the figure. Now the field direction can be identified. Simultaneously the sensitivity is increased. (c) Mounting of a biased "Feldplatte" [25, 26].*

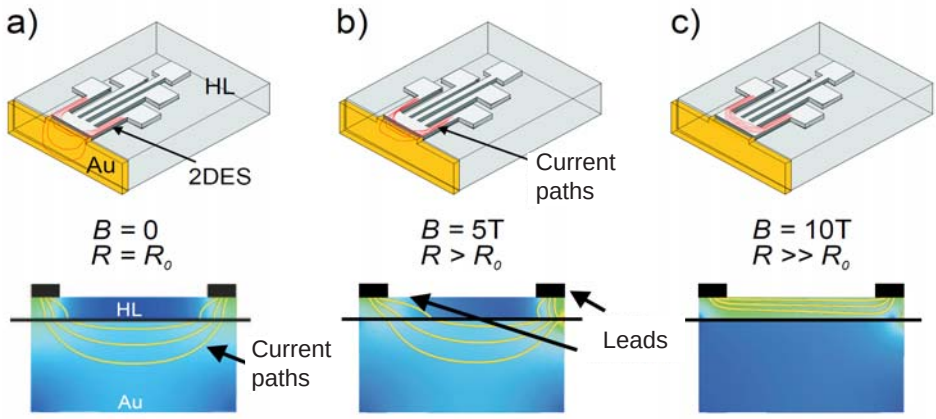


Fig. 9: *Simulation of the EMR-effect. At $B = 0$ the current passes preferentially through the gold contacts. With increasing magnetic field the current is more and more forced into the semiconducting layer (HL). This leads to an increase of the magnetoresistance [28].*

cess)(fig. 6) [14, 24, 25]. The conductivity of the NiSb is larger than that of InSb by two orders of magnitude. When oriented perpendicular to the applied electric field the needles, therefore, shorten the Hall field (fig. 7). Due to the meandering the resistance may be changed from a few Ω to some $k\Omega$. One obstacle of the "Feldplatte" is its temperature sensitivity. Commercial devices, therefore contain two of these resistors in a bridge circuit (figs. 7). One as a reference for temperature compensation. Because of the B^2 dependence of the resistivity, the "Feldplatte" cannot initially distinguish the direction of an applied magnetic field. Therefore, commercial devices are biased by a permanent hard magnetic layer (figs. 7 and 8). Then the field direction can be identified. It also increases the sensitivity due to the slope of the B^2 parabola and allows an adjustment of the operating point to the field strength desired. "Feldplatten" have been widely used as automotive sensors and even in space vehicles, but are currently more and more replaced by other magnetoresistive sensors. One reason is found in the difficulties to embed them into integrated circuits.

Nevertheless, InSb has recently regained new interest, since it was realized that it is possible to achieve drastic increases of the magnetoresistance simply by choosing favourable geometric arrangements of the resistor material and the metallic electrodes [27, 28, 29]. The arrangement is such that at $B = 0$ the current flows preferentially through the metallic leads. An increasing magnetic field more and more forces the current to flow through the semiconductor thereby increasing the resistance of the device. An example of a computer simulation of such an arrangement is shown in fig. 9 [28]. This effect is now called extraordinary magnetoresistance (EMR) . It is believed to possess great technological potential for future applications.

5 The Boltzmann Equation

With the aid of the Boltzmann equation we will now sketch how to derive an equation for the conductivity of our system. In chapter 2 we have discussed the charge carrier motion in terms of the classical Drude model. We have already seen that this is not in complete agreement with experimental findings. We will therefore discuss the transport properties in a more realistic band structure approach.
In the band picture the electrons are considered as independent particles moving in a weak perfectly periodic potential. Quantum mechanical solutions to this single particle model are the Bloch wave functions which have completely incorporated the interaction with the lattice potential. In the strictly periodic potential where the ion cores of the metal occupy their ideal periodic positions the Bloch electrons are not scattered because the Bloch waves are the eigenfunctions of the system. Thus scattering processes can occur only at deviations from the strict lattice periodicity, i.e. at lattice phonons, impurities, defects like dislocations, and at surfaces. Phonons are the dominant scatterers at higher temperatures, whereas defects and impurities become important at low temperatures where the lattice vibrations are frozen out. In a solid at thermal equilibrium the single particle energy levels are occupied according to the Fermi- Dirac distribution function.

$$f(E) = \frac{1}{1 + exp[(E - E_F)/k_B T]}. \tag{37}$$

Therefore, most of the electrons cannot contribute to interaction processes because they are hindered by the Pauly principle. Only those electrons with energies close enough to the Fermi level can take part.
In the so called semiclassical approach an electron is described by a wave packet in k-space consisting of a superposition of Bloch wave functions centered at a certain wave vector $\mathbf{k}$. The drift velocity of the electron is then given by the group velocity of this wave packet [1].

$$\mathbf{v}(k) = \frac{1}{\hbar}\nabla_k \varepsilon_{\mathbf{k}} = \frac{1}{\hbar}\frac{\partial \varepsilon(\mathbf{k})}{\partial \mathbf{k}} \tag{38}$$

$\mathbf{v}(\mathbf{k})$ is parallel to the gradient of the surface of constant energy in k-space, i.e. it is not in general parallel to the wave vector. The electron wave packet is extended in r- and in k-space. The Δr and Δk values thereby have to obey the uncertainty relation $\Delta r \Delta k \approx 1$. The spread of the wavepacket is assumed to be small with respect to the dimensions of the Brillouin zone. Then in the direct lattice its extension is large compared to the dimensions of the unit cell. For completely free electrons the energy band is a parabola in k-space $\varepsilon(\mathbf{k}) = \frac{\hbar^2 k^2}{2m}$ and we get $\mathbf{v}(\mathbf{k}) = \frac{\hbar \mathbf{k}}{m}$.
If we consider $\hbar\mathbf{k}$ as the momentum of the electron in k-space the equation of motion with an applied Lorentz force is:

$$\frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar}(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{39}$$

An external electric field will accelerate the electrons and therefore change the energy distribution. The small distortions due to phonons or due to impurities scatter the electrons with the tendency to restore equilibrium.
We will now describe the effect of the externally applied fields and the scattering processes on the distribution function of the electrons and seek the variation of the distribution function with time. For simplicity we will assume that only one single band is involved.

We consider a volume in phase space and consider the flow of particles in and out. The change in the distribution function is:

$$df/dt = \partial f/\partial t + (\partial f/\partial \mathbf{r})(\partial \mathbf{r}/\partial t) + (\partial f/\partial \mathbf{k})(\partial \mathbf{k}/\partial t) = (\partial f/\partial t)_{coll} \tag{40}$$

where the last term is the change of the function due to collisions. We can rewrite this equation in the form:

$$df/dt = \partial f/\partial t + \dot{\mathbf{k}} \cdot \nabla_k f + \dot{\mathbf{r}} \cdot \nabla_r f = \frac{\partial f}{\partial t}|_{coll}. \tag{41}$$

This is the Boltzmann equation. It is the starting point for the discussion of the transport phenomena. Introducing the above values for $\mathbf{v_k}$ and $d\mathbf{k}/dt$ gives

$$\partial f/\partial t + \mathbf{v}(\mathbf{k})(\partial f/\partial \mathbf{r}) - (e/\hbar)(\mathbf{E} + \mathbf{v} \times \mathbf{B})(\partial f/\partial \mathbf{k}) = (\partial f/\partial t)_{coll} \tag{42}$$

Since we have assumed that there are no thermal gradients in our system we get $\partial f/\partial \mathbf{r} = 0$. To proceed further we have to specify the collision term. This is quite complicated in general [30, 31]. Here we will use the relaxation time approximation . Thereby it is assumed that the deviation of the distribution function from the equilibrium distribution function f_0 (the Fermi-Dirac function) is small:

$$f(\mathbf{k}) = f_0(\mathbf{k}) + f_1(\mathbf{k}). \tag{43}$$

Further it is assumed that the collision term can be expressed as

$$(\partial f/\partial t)_{coll} = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau} = -\frac{f_1(\mathbf{k})}{\tau}. \tag{44}$$

where τ is the relaxation time. It determines the rate of return to the equilibrium distribution when the external field is switched off because then we have

$$(\partial f/\partial t) = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau}. \tag{45}$$

The general solution to this is

$$f(t) = f_0 + [f(0) - f_0]e^{-t/\tau} \tag{46}$$

where $f(0)$ is the distribution at the time when the fields are switched off. Thus the non-equilibrium distribution decays exponentially towards the equilibrium distribution when the driving fields are switched off. Introducing the collision term (44) into the Boltzmann equation (42) we get

$$\partial f/\partial t - (e/\hbar)(\mathbf{E} + \mathbf{v} \times \mathbf{B})(\partial f/\partial \mathbf{k}) = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau} \tag{47}$$

Since we are interested in a stationary state we have $\partial f/\partial t = 0$. We will here also omit the influence of the magnetic field.[5] We are the then left with:

$$-(e/\hbar)\mathbf{E}(\partial f/\partial \mathbf{k}) = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau} \tag{48}$$

which can be rewritten as

$$f(\mathbf{k}) = f_0(\mathbf{k}) + (e/\hbar)\tau\mathbf{E}(\partial f/\partial \mathbf{k}) \tag{49}$$

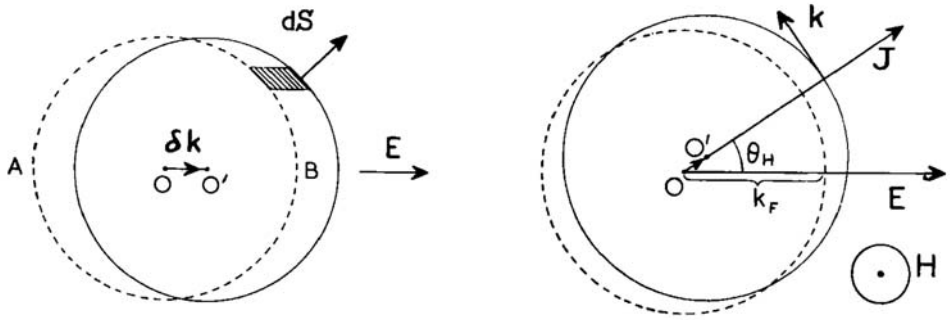


Fig. 10: *Displacement of the Fermi sphere, (a) by an electric field, (b) by an electric field and a magnetic field normal to the page [12].*

This equation can be solved iteratively by replacing $\partial f/\partial \mathbf{k}$ by $\partial f_0/\partial \mathbf{k}$ in the first step [8]

$$f(\mathbf{k}) \approx f_0(\mathbf{k}) + (e/\hbar)\tau \mathbf{E}(\partial f_0/\partial \mathbf{k}). \tag{50}$$

The right side may be considered as the first elements of a Taylor series of a function

$$f(\mathbf{k}) = f_0(\mathbf{k} + (e/\hbar)\tau \mathbf{E}). \tag{51}$$

Therefore the effect of a weak electric field in k-space is simply a shift of the Fermi sphere by an amount $\delta k = (e/\hbar)\tau E$ in the direction of the electric field (Fig. 10(a)). A magnetic field applied in addition will rotate the displaced Fermi sphere by the Hall angle about the direction of the magnetic field (Fig. 10(b)) [12].
The carrier density is given by integration over k-space

$$n = \frac{2}{8\pi^3} \int f(\mathbf{k}) d\mathbf{k}. \tag{52}$$

The factor 2 is due to the two spin directions spin $\uparrow$ and spin $\downarrow$, the factor $1/(2\pi)^3$ due to the k-space integration [32]. With that the current density gets:

$$\mathbf{j} = -\frac{e}{4\pi^3} \int \mathbf{v}(\mathbf{k}) f(\mathbf{k}) d\mathbf{k}. \tag{53}$$

This can be written

$$\mathbf{j} = -\frac{e}{4\pi^3} \left(\int \mathbf{v}(k) f_0(\mathbf{k}) d\mathbf{k} + \int \mathbf{v}(\mathbf{k}) f_1(\mathbf{k}) d\mathbf{k} \right) \tag{54}$$

where the first term $\int \mathbf{v}(\mathbf{k}) f_0(\mathbf{k}) d\mathbf{k} = 0$ because of the antisymmetric integrand. At zero applied electric field there is no net current flow. Introducing the above value for

$$f_1 = (e/\hbar)\tau \mathbf{E}(\partial f_0/\partial \mathbf{k}) = (e/\hbar)\tau \mathbf{E}(\partial f_0/\partial \varepsilon)(\partial \varepsilon/\partial \mathbf{k}) = e\tau \mathbf{E}\mathbf{v}(\mathbf{k})(\partial f_0/\partial \varepsilon) \tag{55}$$

gives

$$\mathbf{j} = -\frac{e^2 \mathbf{E}}{4\pi^3} \int \tau \mathbf{v}(\mathbf{k}) \mathbf{v}(\mathbf{k}) (\partial f_0/\partial \varepsilon) d\mathbf{k}. \tag{56}$$

[5] a more general treatment including the magnetic field is given in the literature [9, 11, 12, 15, 30, 31]

For the conductivity tensor this results in

$$\underline{\underline{\sigma}} = \frac{e^2}{4\pi^3} \int \tau \mathbf{v}(\mathbf{k})\mathbf{v}(\mathbf{k})(-\partial f_0/\partial\varepsilon)d\mathbf{k}. \tag{57}$$

Now consider a metal. The volume element in k-space $d\mathbf{k}$ can be replaced by

$$d\mathbf{k} = d\mathbf{S}\frac{d\mathbf{k}}{d\varepsilon}d\varepsilon = \frac{d\mathbf{S}d\varepsilon}{d\varepsilon/d\mathbf{k}} = \frac{d\mathbf{S}d\varepsilon}{|\nabla_{\mathbf{k}}\varepsilon|} \tag{58}$$

where dS is a surface element of the Fermi sphere (see fig. 10) and the gradient is perpendicular to the Fermi surface. With this equation (57) can be written as

$$\underline{\underline{\sigma}} = \frac{e^2}{4\pi^3} \int \frac{\tau \mathbf{v}(\mathbf{k})\mathbf{v}(\mathbf{k})}{|\partial\varepsilon/\partial\mathbf{k}|}(-\partial f_0/\partial\varepsilon)d\mathbf{S}d\varepsilon. \tag{59}$$

The derivative of the Fermi-Dirac distribution function can be approximated by a δ-function which gives

$$\underline{\underline{\sigma}} = \frac{e^2}{4\pi^3\hbar} \int_{FS} \frac{\tau \mathbf{v}(\mathbf{k})\mathbf{v}(\mathbf{k})}{v} d\mathbf{S}. \tag{60}$$

Now the integration is over the Fermi surface. For isotropic or cubic materials the conductivity tensor has only diagonal elements which are all identical. We can then replace $v_x v_x = v_y v_y = v_z v_z = v^2/3$ and get

$$\underline{\underline{\sigma}} = \frac{e^2}{4\pi^3\hbar} \int_{FS} \frac{\tau v}{3} dS. \tag{61}$$

Using for the Fermi Sphere the relations $\int_{FS} k_F dS = 4\pi k_F^3,\ v_F = \hbar k_F/m$ and for the carrier density $n = \frac{1}{4\pi^3}\frac{4\pi}{3}k_F^3$ we get

$$\sigma = \frac{ne^2\tau}{m}. \tag{62}$$

This is precisely the relation that we deduced initially from the Drude model. But note: The physical picture behind it is here completely different. Here only electrons at the Fermi energy are involved. Presumably it is this formal agreement that is responsible for the success of the classical model at all.

So far, we have studied the influence of an external magnetic field on the conductivity or resistivity of a nonmagnetic material. Now we will investigate what happens when the material is ferromagnetic itself.

6 Resistivity of Ferromagnets

Up to now we have studied the galvanomagnetic effects that were generated in nonmagnetic materials by an externally applied magnetic field. When dealing with ferromagnetic materials the sample will show an intrinsic magnetization M which is of course expected to contribute to the Hall effect as well as the magnetoresistance.

The Hall effect, indeed, includes an additional contribution that is directly proportional to the magnetization of the material but it is often much larger than what might be expected when

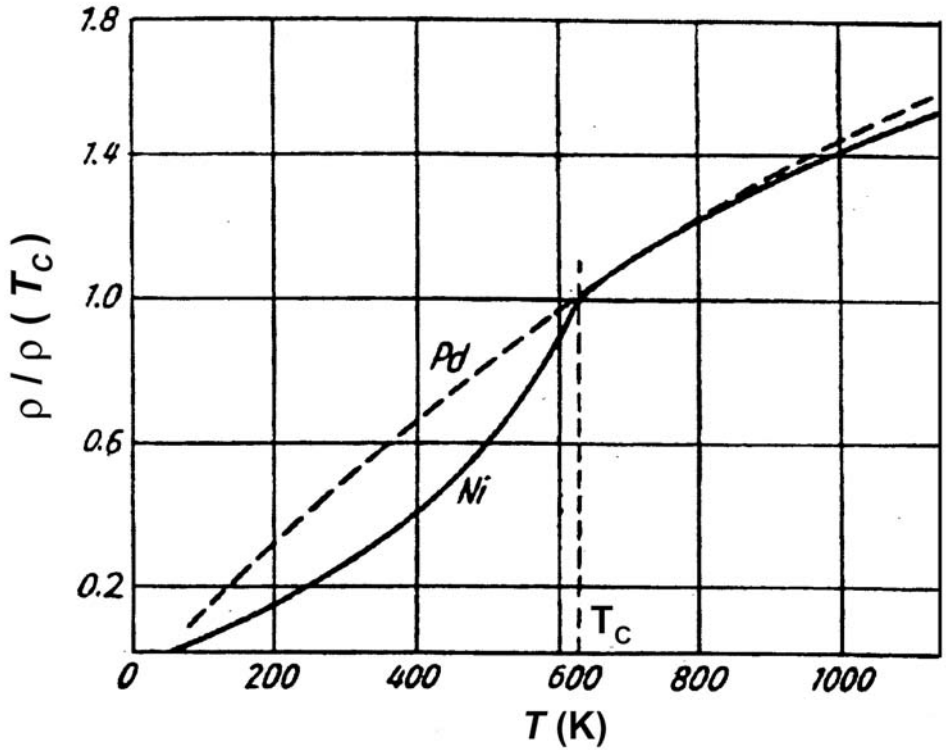


Fig. 11: *Temperature dependence of the relativ resistivities of Ni and Pd normalized to their values at T_c of Ni ($T_c = 631K$)[39].*

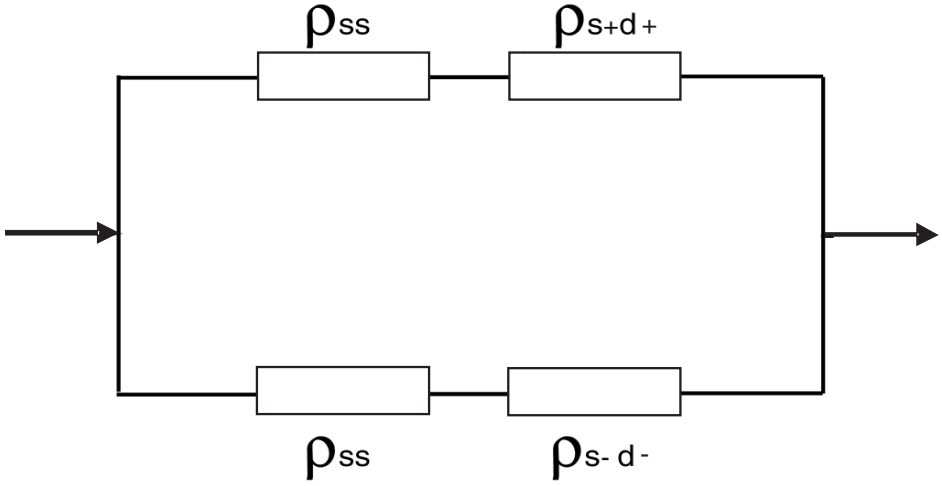


Fig. 12: *Scheme of resistivities due to Mott's two current model. Spin up and spin down channels are signed + and -, respectively.*

$\mu_0 M$ is simply added to the externally applied magnetic field. The effect is therefore called the "anomalous Hall effect". The Hall resistivity may then be written [6]

$$\rho_H = R_0 \mu_0 H + R_s \mu_0 M, \tag{63}$$

where $R_0 \mu_0 H$ is the contribution of the ordinary Hall effect and $R_s \mu_0 M$ is the anomalous

[6] There are different notations used in the literature.

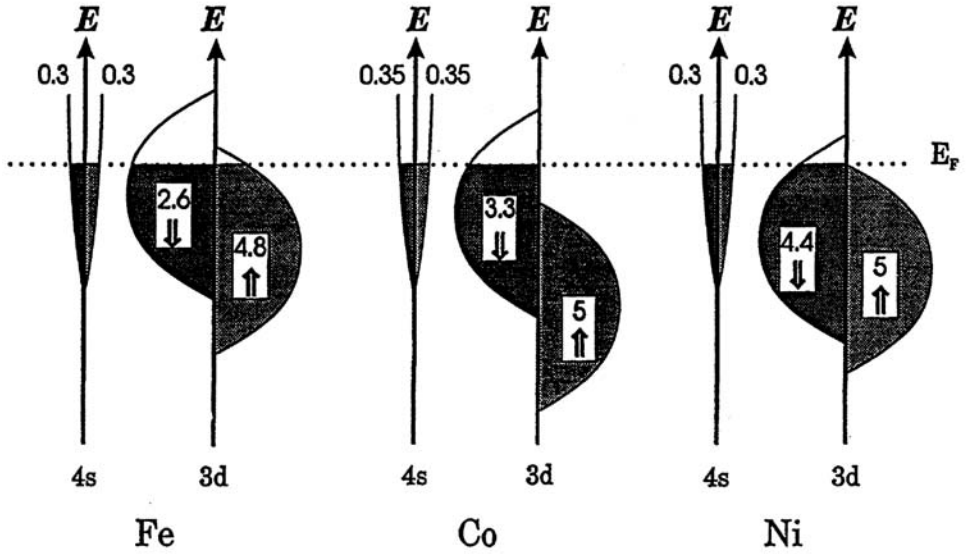


Fig. 13: *Scheme of the densities of states in the sp- and d-bands of ferromagnetic Fe, Co and Ni. The occupation numbers of electrons in the down-spin and up-spin bands are also shown [40, 41].*

contribution. R_S is usually much larger than the ordinary constant R_0. Since M saturates at high magnetic fields the anomalous Hall effect also saturates at high fields. Although the effect has already been discovered more than a century ago by Edwin Hall himself [33], it is still not fully understood. The anomalous Hall effect (AHE) is known to be a consequence of spin-orbit coupling, but the details of the mechanisms that contribute are still subject to controversial discussions.

The anomalous Hall effect has gained new interest because it is also related to the recently discovered Spin-Hall effect. Therefore, two extra contributions are devoted to both these effects in this spring school [34, 35] and therefore we will skip the discussion here.

The other effect caused by the intrinsic magnetization is the anisotropic magnetoresistance (AMR) which was discovered in 1856 by William Thomson the later Lord Kelvin [36]. Before we go into the details of this effect,however, we want to discuss an anomaly that can occur in a ferromagnet without an external magnetic field applied. Fig. 11 shows a comparison of the relativ resistivities of Pd and Ni as a function of temperature [31, 37, 38, 39]. The data are normalized to their values at the Curie temperature of Ni ($631K$). Pd is situated below Ni in the periodic table. Therefore, their electronic structures are quite similar but Pd is paramagnetic at all temperatures while Ni shows the ferromagnetic phasetransition when the temperature decreases below $631K$. Therefore the curves suggest that the differences are due to the formation of a spontaneous magnetization in Ni which is associated with an exchange splitting of the Ni d-states. To explain this behavior Sir N. F. Mott suggested a two current model [42]. It formally splits up the current into two spin channels, one for spin up electrons and the other one for spin down electrons as schematically sketched in Fig. 12. In his approach Mott assumedindependent spin-up and spin-down currents, i.e. the spin information is conserved in the scattering processes. The current is mainly carried by the s-electrons due to their small effective mass (large curvature of the s-bands). The d electrons with their much bigger effective mass (flat d-bands) can contribute only little to the conductivity. The electrons can undergo $s-s$ and $s-d$ scattering transitions where the $s-d$ transitions contribute the most to the resistivity. In

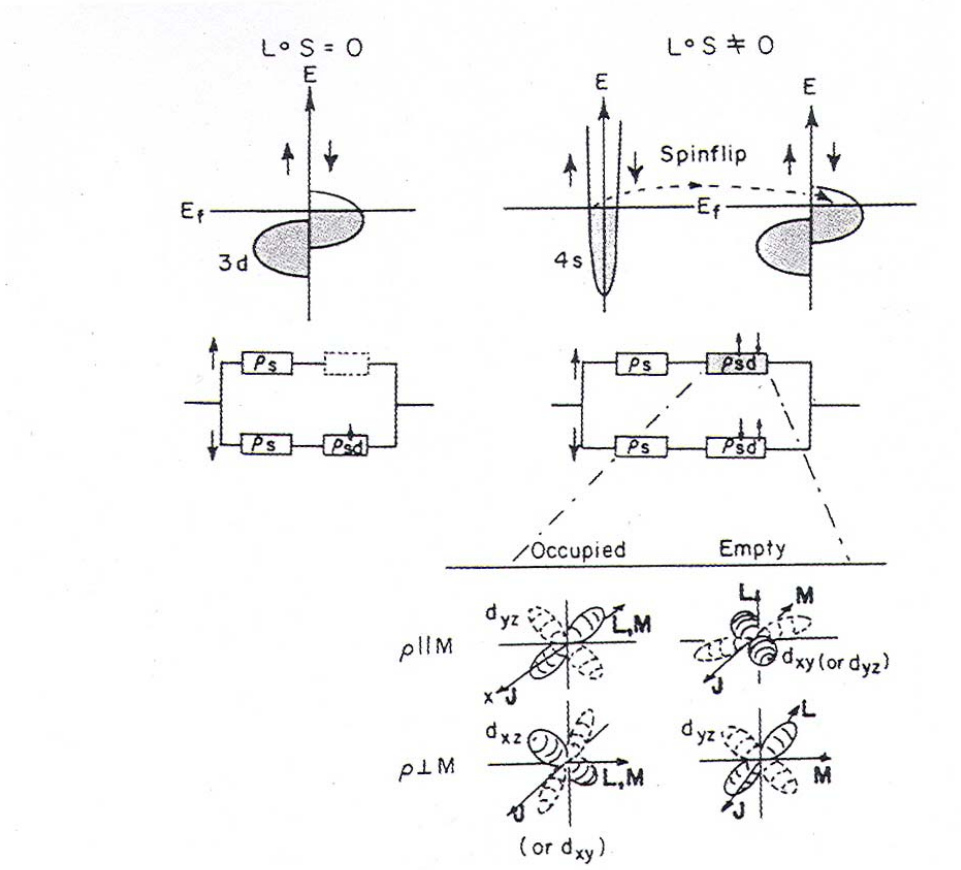


Fig. 14: *(Left: Density of states ($3d$) of a strong ferromagnet without spin-orbit coupling and the spin-separated resistivity contributions according to Mott's two-current model. In the absence of spin-orbit coupling only $s-d$ scattering processes in the spin-down channel are allowed. Right: Inclusion of spin-orbit coupling opens up the possibility of spin-flip transitions in the $s-d$ channels. As a consequence, also the spin-up channel will now contribute to the conductivity. [52].*

the paramagnetic phase there is no difference in the resistivity of the two spin channels but in the ferromagnetic phase the system develops a spin-dependent asymmetry. Specifically in Ni the majority-spin d-states get completely filled and these states are no longer available for scattering events (Fig. 13) - at least in the absence of spin-orbit coupling. This causes a reduction of the resistivity of the majority spin channel and a reduction of the total resistivity.
A. Fert and I.A. Campbell have later refined Mott's model by allowing also spin-flip scattering due to spin-orbit coupling [43, 44, 45, 46, 47]. The spin-orbit coupling adds a spin-dependent componenet to the scattering potential. In order to understand why spin-orbit coupling may be responsible for spi-flip processes, we have to consider the respective quantum mechanical operator $\mathbf{L}\cdot\mathbf{S}$. It can be written with the aid of ladder operators (raising and lowering operators) in the form

$$\mathbf{L}\cdot\mathbf{S} = L_xS_x + L_yS_y + L_zS_z = L_zS_z + (L^+S^- + L^-S^+)/2 \tag{64}$$

where $L^\pm = L_x \pm iL_y$ and $S^\pm = S_x \pm iS_y$. Applying $L^\pm$ to a wavefunction has the effect of raising or lowering the m_l quantum number of that wave function

$$L^\pm\psi(m_l) \longrightarrow \psi(m_l \pm 1). \tag{65}$$

Likewise S^+ and S^- can raise or lower the m_s quantum number, i. e. flip a spin. The operator (64) to the wavefunction acts first on the spin and then on the orbital part of the wavefunction and thereby introduces spin-flips between different m_l values, which opens the possibility for additional spin-mixing $s-d$ transitions. In a simple qualitative picture, the spin-orbit coupling enables different transition channels. First of all, s spin-up ($s^\uparrow$) electrons can now scatter into d spin-down hole states ($d^\downarrow$). As a second important mechanism, also transitions of the type $d^\uparrow \rightarrow s^\uparrow$ become possible, which thereby create unoccupied $d^\uparrow$ states. These empty states open up further channels for spin-flip and non spin-flip $s-d$ scattering. However, the scattering probability depends on the $\mathbf{k}$-vector of the electrons and the orbital m_l into which the electron is scattered. This situation is sketched in the bottom, right part of fig. 14, illustrating selected orbitals, which contribute to the resistivity ρ for the two geometries with current flowing parallel and perpendicular to the magnetization M, respectively.
Next we discuss the influence of an external magnetic field on the anisotropic magnetoresistance.

7 Anisotropic Magnetoresistance (AMR)

In ferromagnetic samples the resistance depends on the orientation of the magnetization with respect to the direction of the electric current. Usually the resistivity is larger when the current and the magnetization are parallel and smaller when they are perpendicular. Fig. 15 shows typical examples how the resistivity changes when an external magnetic field is applied [48]. Initially the resistivity increases when an external field is applied parallel to the current and decreases when it is applied perpendicular. These initial effects are due to the reorientation of magnetic domains by the applied field. A field of a few Oersted magnitude is usually sufficient to achieve saturation. Above saturation a slow increase or decrease of both resistivities is observed. The increase can simply be explained by the normal magnetoresistance induced by the Lorentz force, the decrease is attributed to the so called spin disorder resistivity. This contribution becomes particularly importent when the Curie temperature is approached and depends on the scattering of conduction electrons into the exchange split d-states [48]. To obtain values of $\rho_\parallel$ and $\rho_\perp$ which are independent of the externally applied field the measured curves are extrapolated to $\mathbf{B}=0$ as indicated in the figures. The difference between $\rho_\parallel(\mathbf{B}=0)$ and $\rho_\perp(\mathbf{B}=0)$ is called spontaneous resistivity anisotropy. It disappears above T_c because it is associated with the spontaneous internal magnetization. The resistivity anisotropy is often normalized to the average resistivity ρ_0 which is defined as

$$\rho_0 = \rho_{average} = (\frac{1}{3}(\rho_\parallel(B=0) + \frac{2}{3}\rho_\perp(B=0)), \tag{66}$$

where $\rho_\parallel(B=0)$ and $\rho_\perp(B=0))$ are the extrapolated values. This ratio is called the anisotropic magnetoresistivity ratio or coefficient. Note, that the average resistivity defined above is not in general identical with the resistivity of the field free demagnetized sample [52].
The anisotropic magnetoresistance is caused by anisotropic scattering of the charge carriers by spin-orbit coupling. The resistance varies with the orientation of the d-orbitals parallel or perpendicular to the magnetic field.
Now consider a thin homogeneous strip of a magnetoresistive material, e. g. permalloy, with a large aspect ratio, i.e. its length is much larger than its width. If the sample thickness exceeds a certain value (a few $10\ nm$) we can expect the magnetization to lie in the plane of the film.

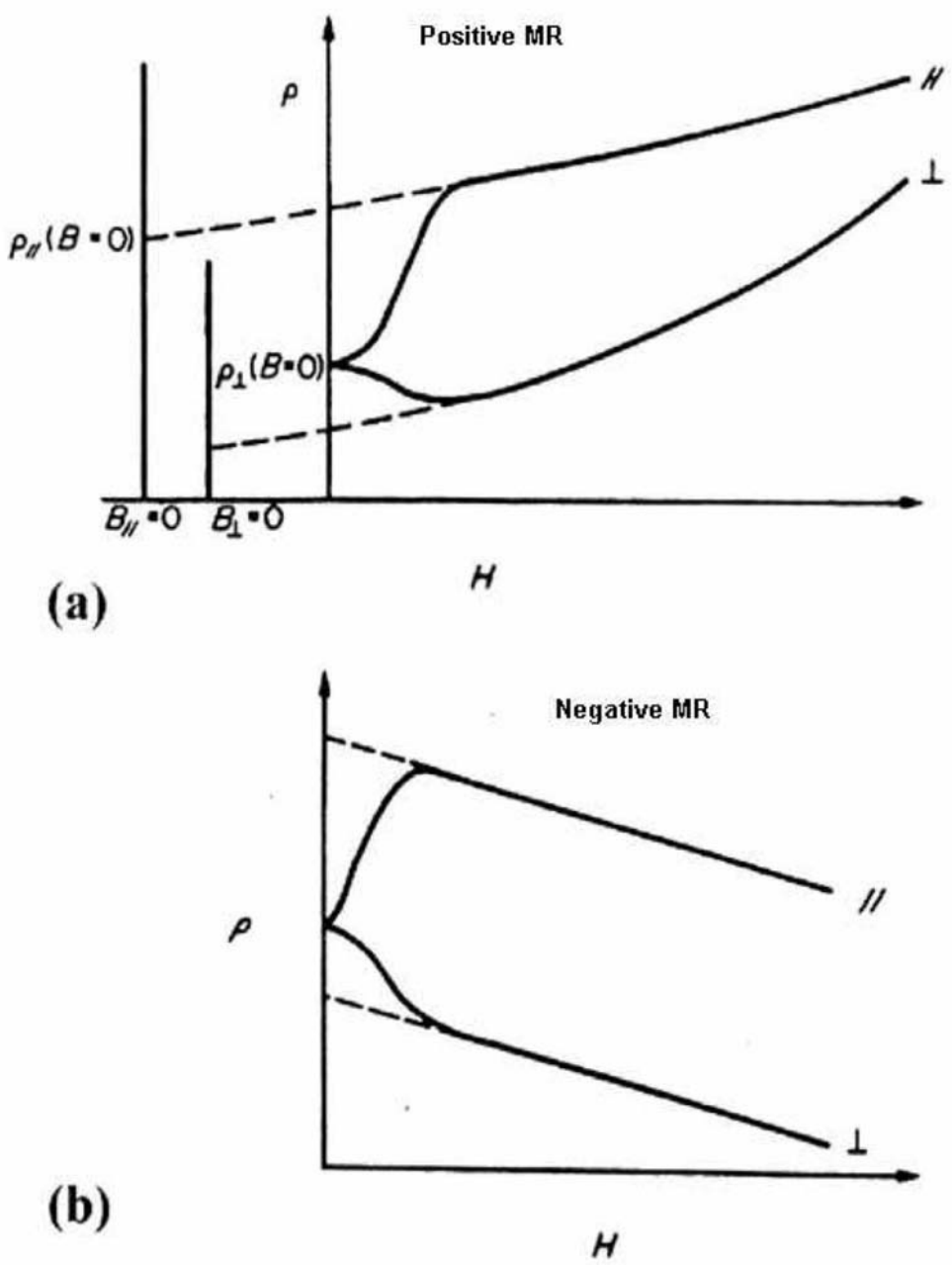


Fig. 15: *schematic resistivity change for ferromagnets. The extrapolated resistivities are also shown. After saturation of the magnetization the resistivity either increases (a): normal positive magnetoresistance or (b) decreases due to the reduction of spin-disorder. [48].*

To minimize the stray field the magnetization will orient parallel to the long axis of the film which therefore is the easy axis.[7] We identify this axis with the z'-axis of an orthonormalized coordinate sytem and the perpendicular in plane direction with the x'-axis.[8] In this case we can

[7] The easy axis of a sample is the axis along which the sample is most readily magnetized.

[8] In the following we will use two coordinate systems: one (x', y', z') in which the magnetization is fixed with the z' axis. The other one $(x, y = y', z)$ which is fixed to the sample geometry can be rotated about the common y-axis by an arbitrary angle φ (see Figs. 16 and 17).

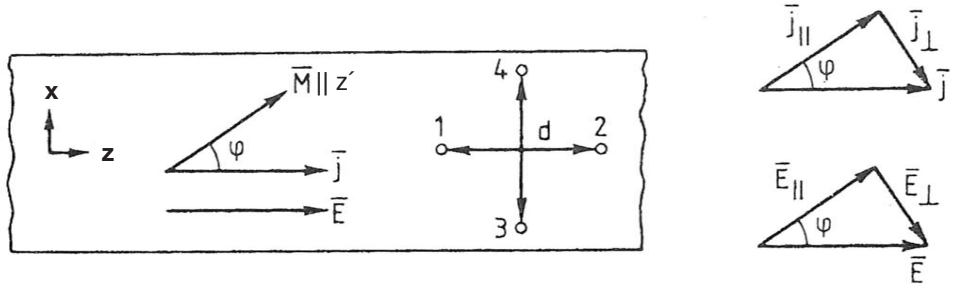


Fig. 16: *Geometric arrangement for the calculation of anisotropic magnetoresistance and planar Hall effect components as described in the text [54, 49].*

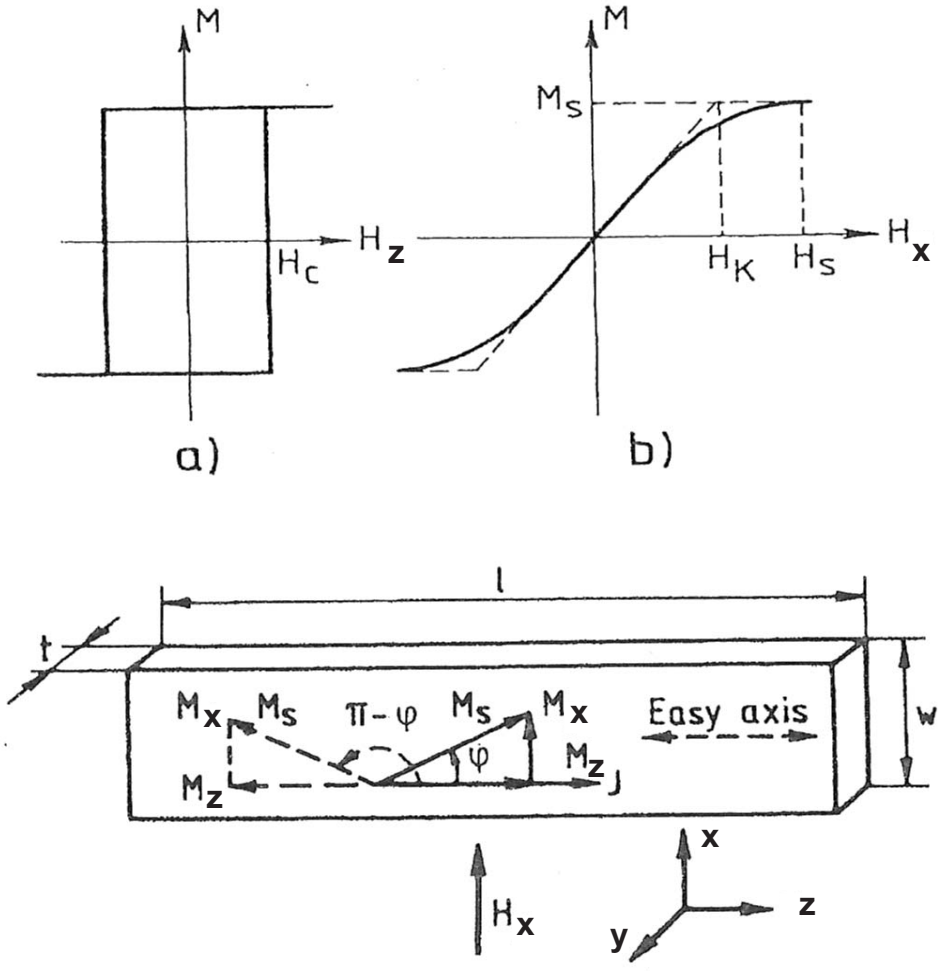


Fig. 17: *Idealized easy axis (a) and hard axis (b) magnetization curves of an in plane magnetized anisotropic magnetic thin film as shown in the lower figure. The z-axis (long axis) is the easy axis and the x-axis the hard axis. The rotation of the magnetization upon application of an external magnetic field parallel to the hard axis is indicated. [54, 49].*

easily guess the relation between an applied electric field and the current density achieved. It is:

$$\begin{pmatrix} E_{x'} \\ E_{y'} \\ E_{z'} \end{pmatrix} = \begin{pmatrix} \rho_\perp & -\rho_H & 0 \\ \rho_H & \rho_\perp & 0 \\ 0 & 0 & \rho_\parallel \end{pmatrix} \begin{pmatrix} j_{x'} \\ j_{y'} \\ j_{z'} \end{pmatrix} \quad \text{or} \quad \mathbf{E}' = (\rho_{i'j'}(\mathbf{M}))\mathbf{j}' \tag{67}$$

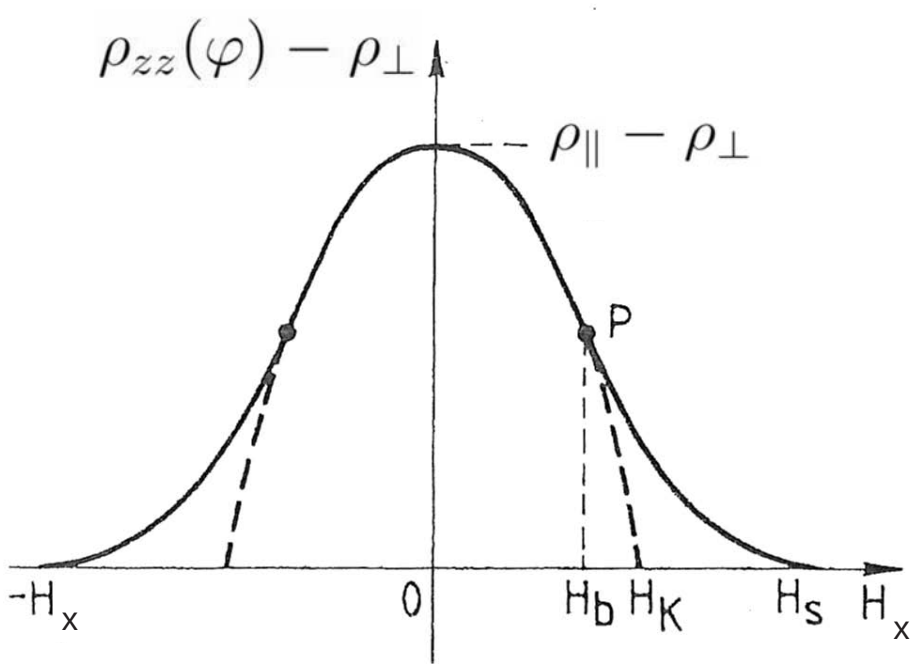


Fig. 18: *Response of a permalloy thin-film magnetoresistor as described in the text [54, 49].*

where $\rho_\parallel$ and $\rho_\perp$ are obviously in accordance with the description given above. ρ_H is the resistivity corresponding to the anomalous Hall effect which we will not further consider. However, in this form the resistivity tensor is only valid when the magnetization of the sample is fixed to the easy axis. For a detailed analysis of the AMR we need a more general form of the tensor where the current density $\mathbf{j}'$ and the magnetization $\mathbf{M}$ include an arbitrary angle φ in the (x', z')-plane. We derive this form by a similarity transformation of the AMR-matrix in the following way:
We assume that the coordinate system defined above is fixed with the magnetization: $M \parallel z'$. We rotate the sample and with it $\mathbf{E}'$ and $\mathbf{j}'$ by an angle φ and define a new rotated coordinate system (x, y) so that the new orientation of the long sample axis is parallel to the new z-axis together with the applied field $\mathbf{E}$ and the current density $\mathbf{j}$. This situation is indicated in Fig. 16. For the equation (67) to remain valid we have to rotate $\mathbf{E}$ and $\mathbf{j}$ back to the old system:

$$\mathbf{E}' = R(\varphi)\mathbf{E} \quad \text{and} \quad \mathbf{j}' = R(\varphi)\mathbf{j} \tag{68}$$

where $R(\varphi)$ is the transformation matrix that provides the backrotation. Thus from equation (67) we get

$$\mathbf{E}' = R(\varphi)\mathbf{E} = (\rho_{i'j'}(\mathbf{M}))R(\varphi)\mathbf{j} = (\rho_{i'j'}(\mathbf{M}))\mathbf{j}'. \tag{69}$$

where $(\rho_{i'j'}(\mathbf{M}))$ is the resistivity tensor defined above. Using the associativity of matrix multiplication we therefore get the AMR-matrix in the transformed coordinates:

$$\rho_{ij}(\mathbf{M}) = R^{-1}(\varphi)(\rho_{i'j'}(\mathbf{M}))R(\varphi)$$

$$= \begin{pmatrix} cos\varphi & 0 & sin\varphi \\ 0 & 1 & 0 \\ -sin\varphi & 0 & cos\varphi \end{pmatrix} \begin{pmatrix} \rho_\perp & -\rho_H & 0 \\ \rho_H & \rho_\perp & 0 \\ 0 & 0 & \rho_\parallel \end{pmatrix} \begin{pmatrix} cos\varphi & 0 & -sin\varphi \\ 0 & 1 & 0 \\ sin\varphi & 0 & cos\varphi \end{pmatrix}$$

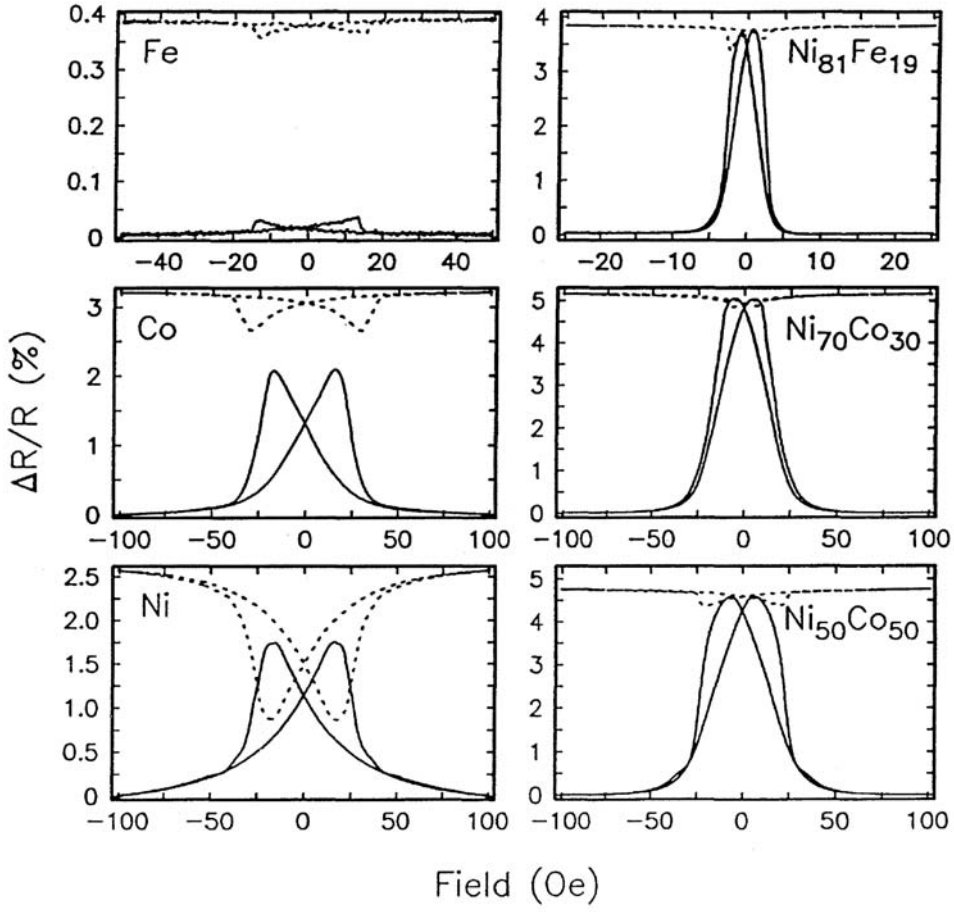


Fig. 19: *Examples of the anisotropic magnetoresistance effect in sputtered polycrystalline films of Fe, Co, Ni and $Ni_{81}Fe_{19}$ (permalloy), $Ni_{70}Co_{30}$ and $Ni_{50}Co_{50}$. The full and dotted lines correspond to magnetic field applied orthogonal and parallel to the current respectively in the plane of the films. The films in each case are ≈ 1000 Å thick [40].*

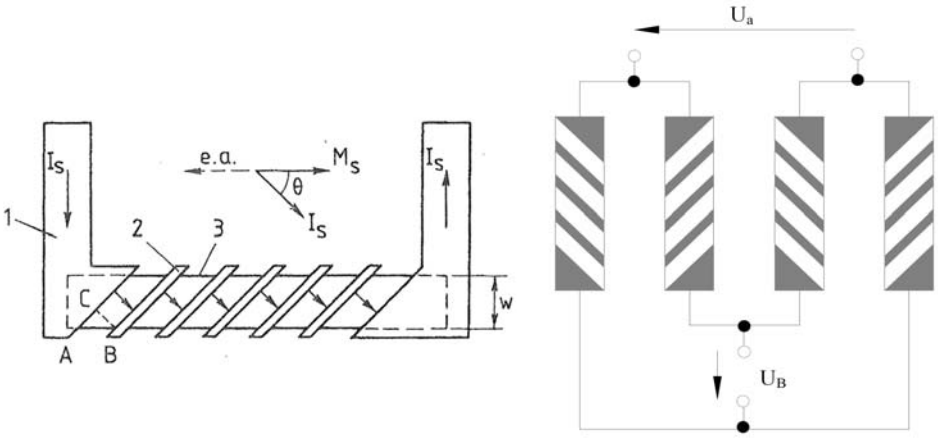


Fig. 20: *left: Scheme of a permalloy (3) barber-pole magnetoresistor with canted conductor strips(2). Magnetization and current direction is indicated. right: Schematic Wheatstone-bridge arrangement of four barber poles. [49].*

$$= \begin{pmatrix} \rho_\perp cos^2\varphi + \rho_\parallel sin^2\varphi & -\rho_H cos\varphi & -\rho_\perp sin\varphi cos\varphi + \rho_\parallel cos\varphi sin\varphi \\ \rho_H cos\varphi & \rho_\perp & -\rho_H sin\varphi \\ -\rho_\perp sin\varphi cos\varphi + \rho_\parallel cos\varphi sin\varphi & \rho_H sin\varphi & \rho_\perp sin^2\varphi + \rho_\parallel cos^2\varphi \end{pmatrix} \quad (70)$$

Thus for the electric field applied in the z-direction we have

$$E_z = \rho_{zz}(\varphi) j_z \quad (71)$$

where

$$\rho_{zz}(\varphi) = \rho_\perp sin^2\varphi + \rho_\parallel cos^2\varphi = \rho_\perp + (\rho_\parallel - \rho_\perp) cos^2\varphi. \quad (72)$$

This is often rewritten as

$$\rho_{zz}(\varphi) = (\frac{1}{3}\rho_\parallel + \frac{2}{3}\rho_\perp) + (\rho_\parallel - \rho_\perp)(cos^2\varphi - \frac{1}{3}) \quad (73)$$

where $\rho_0 = (\frac{1}{3}\rho_\parallel + \frac{2}{3}\rho_\perp)$ is the average resistivity defined above. The resistivity coefficient is then defined as

$$\frac{\rho_{zz}(\varphi) - \rho_0}{\rho_0} = \frac{\Delta\rho}{\rho_0} = \frac{(\rho_\parallel - \rho_\perp)}{(\frac{1}{3}\rho_\parallel + \frac{2}{3}\rho_\perp)}(cos^2\varphi - \frac{1}{3}). \quad (74)$$

By applying an electric field in the z direction we not only create a current j_z in the direction of the applied field but also generate an electric field perpendicular to this current.

$$E_x = \rho_{xz}(\varphi) j_z \quad (75)$$

where

$$\rho_{xz}(\varphi) = (\rho_\parallel - \rho_\perp) cos\varphi sin\varphi = (\rho_\parallel - \rho_\perp)\frac{1}{2} sin2\varphi \quad (76)$$

This effect is called the "planar Hall effect" or "pseudo Hall effect" and is believed to have some potential for applications with micro- or nanostructured spintronic devices [50, 51]. It is zero when the magnetization is parallel or perpendicular to the easy axis.

In the following we will calculate the AMR response of a sample under somewhat idealized conditions. For instance we will ignore any demagnetizing effects and assume idealized hysteresis conditions. In accordance with our above premisis we consider a ferromagnetic thin film (e.g. permalloy) with a large aspect ratio. We also assume that the magnetization is in the film plane. Without a magnetic field applied the magnetization direction is parallel to the easy axis of the film which again we identify with the long z-axis of our specimen. If a magnetic field is varied parallel to this easy axis one observes a rectangular hysteresis (Fig. 17) [49]. On the contrary, if the external magnetic field is varied along the coplanar hard axis the magnetization in this direction will show a linear slope and finally saturate. The Magnetic moment of the film rotates towards the hard axis. During this procedure the film is divided into elongated antiparallel domains, whose magnetizations rotate at an angle φ for the parallel ones and $\pi - \varphi$ for the antiparallel ones when the vertical field H_x is applied in the film plane [49] (Fig. 17). Neglecting demagnetization effects we have the relation:

$$\frac{M_x}{M_s} = sin\varphi = \frac{H_x}{H_K} \quad \text{for} \quad -H_K \leq H_x \leq H_K. \quad (77)$$

Here M_x is the x-component of the magnetization, M_s is the saturation magnetization H_x is the applied magnetic field and H_k is the magnetization where saturation sets in. Introducing this result into equation (72) gives:

$$\rho_{zz}(\varphi) = \rho_\perp + (\rho_\parallel - \rho_\perp)(1 - (\frac{H_x}{H_K})^2). \quad (78)$$

The graph of this equation is an upside down parabola with the maximum at $H_x = 0$ and it ends on the abscissa where the H_x value reaches the saturation point H_K. This graph is plotted in Fig. 18. Experimentally it is found that the function more smoothly approaches the abscissa (solid line in the figure). This discrepancy is attributed to the demagnetizing fields that we have ignored so far. As a consequence saturation is not reached at H_K but at $H_s = H_K + H_d$ where H_d is the demagnetizing field. It is clear from the curve that in the vicinity of zero field ($H_x = 0$) the magnetoresistor is of low sensitivity and due to its quadratic response also highly nonlinear. Moreover it cannot detect the polarity of the applied external field. For sensor applications it is therefore desirable to shift the operating point to the inflection point of the curve (Fig. 18) where the highest sensitivity can be achieved and the response is approximately linearized. At this point the orientation of the field can also be detected. The shift of the operation point can either be achieved by a biasing field of a nearby permanent magnet (just as in the case of the Lorentz magnetoresistance described above) or by use of a so called barber-pole structure to be discussed below. The sensitivity is optimal when the angle between the magnetization and the current direction is kept at about 45^o. This principle has been used in AMR-read heads which have been used for a few years in hard disk technology before they were replaced by the GMR based spin-valve read heads. Details on AMR-read head technology are found in the literature [52, 53, 54, 55]. Fig. 19 shows magnetoresistance runs for thin films of $Fe, Co, Ni,$ permalloy $Ni_{81}Fe_{19}$, and two nickel-cobalt alloys [40]. Except for Fe and Ni the sensitivity is larger when the magnetization is perpendicular to the current (solid lines) rather than parallel (dotted lines). The effect is in the range of a few percent. The curves show a clear hysteresis. Obviously the samples do not follow our idealized magnetization curve and show a little coercivity which corresponds to the position of the two maxima. The magnetization cannot follow the external field immediately. The occurance of the hysteresis is attributed to an easy axis dispersion in the literature [49]. This means that the sample easy axis is a macroscopic average, but locally inside the sample the easy axes might show in slightly different directions in different domains. During the magnetization procedure the sample splits up into a multidomain structure with longitudinal domains parallel to the average easy axis. The switching behavior of these domains is responsible for the hysteresis. It was also reported that a slight vertical remanence could be observed when the vertical field was reduced to zero.

As pointed out above the optimum operating point of an AMR-sensor is obtained when the current and the magnetization include an angle of about 45^o. There exists an alternative technique to applying a bias field, namely to force the current into an angle of 45^o with respect to the magnetic easy axis. This is done in a barber-pole structure (Fig. 20). The name stems from the similarity of the structure with barber poles which since the middle ages have been used as signatures of barber shops and are sometimes still used today. The barber pole sensor consists of a permalloy magnetoresistive film with its easy axis parallel to its length. It is covered by conductor strips which are canted at 45^o. The strips are up to 50 times better conductors than the magnetoresistive layer. This forces the equipotential lines inside the magnetoresistor parallel to the strips so that the current flow in the magnetoresistor is perpendicular to the strips, i. e. at 45^o with respect to the magnetization. Miniaturized commercial sensors combine four such barberpoles in a Wheatstone bridge arrangement. A typical chip size is $1.6mm \times 1.6mm$ at resistance of $1.7K\Omega$. They are used in contactless angular or linear position measurements. They are characterized by high sensitivity and stability and high reliability. AMR-sensors are very sensitiv and operate in the $\approx 1nT$ to $1mT$ range in contrast to Hall sensors which operate in the $\approx 0.1mT$ to $100T$ range [56, 57].

8 Conclusion

In this contribution we have discussed galvanomagnetic transport in nonmagnetic and ferromagnetic materials. All the effects are bulk effects and since they are at the very basis of spintronics, it is necessary to consider them in many spintronic measurements and devices. Due to their high reliability the effects have been and are still partially used in technology, but are also slowly replaced by other spintransport effects, such as giant or tunneling magnetoresistance. We have also shown that quantum mechanical techniques are necessary to properly interpret the eletrical transport phenomena. From the theoretical point of view particularly the influence of spin-orbit coupling needs further detailed analysis. For reasons of clarity, we have taken a somewhat simplified point of view in our discussion of the effects, e.g. considering only the influence of a single or at most two bands. In reality even more bands need to be taken into account. Moreover, there are additional effects, which become important in confined geometries like domain wall resistivity or domain wall pinning, which can superimpose the described behavior and are dealt with in other contributions of this Spring School.

Appendix

A Derivation of Equation (32)

In the following calculation we neglect the fact that the conductivity and with it the currents and velocities in the (x,y)-plane are reduced by factors $1/(1+(\mu_{e,h}B_z)^2) = 1/(1+(\tau_{e,h}\omega_c^{(e,h)})^2)$, respectively [9], because these factors approximately cancel during the calculation and for weak magnetic fields are close to 1, anyway: $(\tau\omega_c \ll 1)$.
Then to calculate the Hall field E_H we start with the Lorentz force:

$$F_L = q(\mathbf{v}\times\mathbf{B}) = q\begin{pmatrix} v_yB_z - v_zB_y \\ v_zB_x - v_xB_z \\ v_xB_y - v_yB_x \end{pmatrix}. \tag{79}$$

Since at stationary conditions we will have current flowing only in the x direction, i.e. $\mathbf{v} = (v_x, 0, 0)$, and as $\mathbf{B} = (0, 0, B_z)$ the active component of the Lorentz force is

$$F_{L,y} = -qv_xB_z, \tag{80}$$

where for electrons $q = -e$ and $v_x = -|v_x| = v_x^{(e)} < 0$ and for holes $q = +e$ and $v_x = v_x^{(h)} > 0$. Thus, the resulting Lorentz force is

$$F_{L,y}^{(e)} = -(-e)(v_x^{(e)}B_z) = ev_x^{(e)}B_z \quad \text{for electrons} \tag{81}$$

and

$$F_{L,y}^{(h)} = -(e)(v_x^{(h)}B_z) = -ev_x^{(h)}B_z \quad \text{for holes.} \tag{82}$$

Neglecting the quadratic terms in equations (25) and (26) and by comparison with equation (13) we write $v_x^{(e)} = -\mu_e E_x$ and $v_x^{(h)} = \mu_h E_x$. Including the action of the Hall field E_H we get:

$$F_y^{(e)} = -e(E_H - v_x^{(e)}B_z) = -e(E_H + \mu_e E_x B_z) \quad \text{for the electrons} \tag{83}$$

[9] compare with equations (25) and (26)

and

$$F_y^{(h)} = e(E_H - v_x^{(h)} B_z) = e(E_H - \mu_h E_x B_z) \quad \text{for the holes.} \tag{84}$$

Putting these forces into the respective equations of motion gives

$$m_e \frac{dv_y^{(e)}}{dt} + \frac{m_e}{\tau_e} v_y^{(e)} = -e(E_H + \mu_e E_x B_z) \quad \text{for the electrons} \tag{85}$$

and

$$m_h \frac{dv_y^{(h)}}{dt} + \frac{m_h}{\tau_h} v_y^{(h)} = e(E_H - \mu_h E_x B_z) \quad \text{for the holes.} \tag{86}$$

At steady state the time derivatives are zero and therefore the currents in the y-direction for the electrons and for the holes are given by

$$j_y^{(e)} = -en_e v_y^{(e)} = en_e \mu_e (E_H + \mu_e E_x B_z) \tag{87}$$

and

$$j_y^{(h)} = en_h v_y^{(h)} = en_h \mu_h (E_H - \mu_h E_x B_z). \tag{88}$$

Once equilibrium is reached, there will be no net current in the y-direction:

$$j_y^{(e)} + j_y^{(h)} = 0. \tag{89}$$

Here it is ***not*** required that the force generated by the Hall field E_H cancels the Lorentz forces for the electrons and holes separately. However the currents in the y-direction cancel each other. The total current in the y-direction then gets

$$j_y^{(e)} + j_y^{(h)} = 0 = en_e \mu_e (E_H + \mu_e E_x B_z) + en_h \mu_h (E_H - \mu_h E_x B_z) \tag{90}$$

and thus

$$(n_e \mu_e + n_h \mu_h) E_H = (n_h \mu_h^2 - n_e \mu_e^2) E_x B_z. \tag{91}$$

Finally, replacing $E_x = j_x / \sigma_{ambipolar} = j_x / [e(n_e \mu_e + n_h \mu_h)]$ we derive the Hall voltage:

$$U_H = E_H \cdot b = \frac{n_h \mu_h^2 - n_e \mu_e^2}{e(n_h \mu_h + n_e \mu_e)^2} j_x B_z \cdot b = \frac{n_h - n_e(\mu_e/\mu_h)^2}{e(n_h + n_e \mu_e/\mu_h)^2} \, iB/d \tag{92}$$

where the ratio of the mobilities has been introduced.

References

[1] N. W. Ashcroft and N. D. Mermin, Solid State Physics (Saunders College, 1976).

[2] J. F. Nye, Physical Properties of Crystals, At The Clarendon Press, Oxford (1972)

[3] L. Onsager, Phys. Rev. **37**, 405 (1931); ibid. **38**,2265(1931)

[4] P. Drude,Annalen der Physik **1**, 566 (1900); ibid **3**, 369 (1900)

[5] E.H. Hall: "On a New Action of the Magnet on Electric Currents". Am. J. Math. **2**, p.287-292 (1879)

[6] C. M. Hurd, The Hall Effect in Metals and Alloys, Plenum Press, New York (1972)

[7] C. L. Chien, C.R. Westgate, The Hall Effect and its Applications, Plenum Press, New York (1980)

[8] H. Ibach, H. Lüth, Solid State Physics, Springer, Berlin,Heidelberg, New York (2003)

[9] O. Madelung, Grundlagen der Halbleiterphysik, Springer Verlag, Berlin, Heidelberg, New York (1970)

[10] H. Weiss und H. Welker, Zur transversalen magnetischen Widerstandsänderung von InSb, Z. Phys. **138**, 322 (1954).

[11] F. J. Blatt, Physics of Electronic Conduction in Solids, Mc Graw Hill, New York (1968)

[12] J. M. Ziman, Electrons in Metals, A short guide to the Fermi surface,Contemp. Phys. **3**, 241; ibid. 321; ibid. 401; ibid. **4**, 1; ibid. 81 (1962); J. M. Ziman, Principles of the Theory of Solids, Cambridge, At the University Press (1969)

[13] O. Madelung und H. Weiss, Die elektrischen Eigenschaften von Indiumantimonid II, Z. Naturforschg. **9a**, 527 (1954)

[14] H.Weiss, Structure and Application of Galvanomagnetic Devices, Pergamon Press, Oxford (1969)

[15] K. Seeger, Semiconductor Physics, 5th edition, Springer Verlag, Berlin, Heidelberg (1991)

[16] P.Y. Yu, M.Cardona, Fundamentals of Semiconductors, 3rd Edition, Springer Verlag, Berlin, Heidelberg, New York (2001)

[17] L. J. van der Pauw: A method of measuring specific resistivity and Hall effect of discs of arbitrary shape. Philips Research Reports **13**, (1) 1 (1958); L.J. van der Pauw, A method of measuring the resistivity and Hall coefficient on lamellae of arbitrary shape. Philips Technical Review **20**, 220 (1958); http://www.ecse.rpi.edu/ schubert/More-reprints/

[18] O.M. Corbino, Elektromagnetische Effekte, die von der Verzerrung herrühren, welche ein Feld an der Bahn der Ionen in Metallen hervorbringt, Phys. Z. **12**, 561 (1911)

[19] M. Kohler, Annalen der Physik **424**, 211 (1938); ibid **441**, 18 (1950) and references therein.

[20] J.-P. Jan, Galvanomagnetic and Thermomagnetic Effects in Metals, in: F. Seitz, D. Turnbull (Edts) Solid State Phys. Vol. **5**, p. 1, Academic Press, New York (1975)

[21] P.L. Rossiter, Metals and Alloys, Conductivity in, Encyclopedia of Applied Physics, Vol. **10**, p.163, VCH Publishers (1994)

[22] P. K. Siwach, H. K. Singh, and O. N. Srivastava, J. Phys.: Condens. Matter **20**, 273201 (2008)

[23] G. Bublitz, Messung magnetischer Gleichfelder mit Wismutspiralen und Wismutplatten, ATM **5**, 391 (1937)

[24] H. Weiss, The "Feldplatte" a new semiconductor magnetoresistance device, IEEE Transactions on Magnetics **2**, 540 (1966)

[25] H. Weiß, Galvanomagnetische Bauelemente, Festkörper Probleme **7**, 200-216, Springer Verlag, Berlin, Heidelberg (1967); H. Welker, Chemie-Ing.-Techn. **37**, 916 (1965)

[26] Fig. taken from: K. Mertens, http://www.et.fh-muenster.de/persons/mertens/se_folien/sekap4.pdf

[27] S.A. Solin and L.R. Ram-Mohan, Geometry Driven Magnetoresistance, p. 2979 in: H. Kronmüeller, S. Parkin (Eds.), Handbook of Magnetism and Advanced Magnetic Materials, Vol. **5**: Spintronics and Magnetoelectronics, John Wiley & Sons (2007); S. A. Solin, T. Thio, D. R. Hines and J. J. Heremans, Science **289**, 1530 (2000); S. A. Solin, D. R. Hines and A. C. H. Rowe, Appl. Phys. Lett. **80**, 4012 (2002).

[28] M. Holz, O. Kronenwerth, D. Grundler, Magnetoresistance of semiconductor-metal-hybrid structures: The effect of material parameters and contact resistance. Phys. Rev. **B 67**, 195312 (2003); Appl. Phys. Lett. **86**, 072513 (2005); Appl. Phys. Lett. **87**, 172501 (2005); M. Hoener, O. Kronenwerth, Ch. Heyn, D. Grundler, and M. Holz, J. Appl. Phys. **99**, 036102 (2006); C. H. Möller, O. Kronenwerth, Ch. Heyn, and D. Grundler, Appl. Phys. Lett. **84**, 3343 (2004); Fig.: by M. Holz In: Matthias Hoener, Optimierung von Extraordinary-Magnetoresistance-Magnetfeldsensoren aus mikrostrukturierten Metall-Halbleiter-Hybridstrukturen, Diplomarbeit, Universität Hamburg (2004); www.physik.tu-berlin.de/cluster/Dipl.20062004.pdf

[29] Chuan-Bing Rong, Hong-Wei Zhang, Ji-Rong Sun, and Bao-Gen Shen, Enhanced extraordinary magnetoresistance in the semiconductor-metal hybrid structure with three current leads, Appl. Phys. Lett. **89**, 052503 (2006)

[30] O. Madelung, Introduction to Solid State Theory, Springer Verlag, Berlin, Heidelberg, New York (1978) [The original German Version appeared in three volumes: O. Madelung, Festkoerpertheorie I-III, Springer Verlag Berlin, Heidelberg, New York (1972)];

[31] J. M. Ziman, Electrons and Phonons, Clarendon Press, Oxford (2001)

[32] see ref. 1 Chapter 2

[33] E. H. Hall, Philos. Mag. **10**, 301 (1880); **12**, 157 (1881)

[34] Y. Mokrousov, this springschool

[35] A. Fert, this springschool

[36] W. Thomson, On the electrodynamic qualities of metals - effects of magnetization on the electric conductivity of nickel and of iron, Proc. Royal Soc. London **8**, 546-555 (1857)

[37] J. G. G. Conybeare, Proc. Phys. Soc. **49**, 29 (1937)

[38] H. H. Potter, Proc. Phys. Soc. **49**, 671 (1937)

[39] R. Gross, A. Marx, Spintronics; http://www.wmi.badw.de/teaching/Lecturenotes/

[40] S.S.P. Parkin, 2.4 Giant Magnetoresistance and Oscillatory Interlayer Coupling in Polycrystalline Transition Metal Multilayers, in: B. Heinrich and J.A.C. Bland (Eds.) Ultrathin Magnetic Structures II, Springer-Verlag Berlin Heidelberg 1994

[41] J. Mathon: Contemp. Phys. **32**, 143 (1991)

[42] N. F. Mott, Proc. Roy. Soc. **153**, 699(1936) and Adv. Phys. **13**, 325(1964)

[43] I. A. Campbell, A. Fert, A. R. Pomeroy, Phil. Mag **15**, 977 (1967)

[44] A. Fert, I. A. Campbell, Phys. Rev. Lett. **21**, 1190 (1968)

[45] I. A. Campbell, A. Fert, O.Jaoul, J. Phys. C, Metal Phys. Suppl. **1**, S95 (1970)

[46] A. Fert, I. A. Campbell, J.Phys. F., Metal Phys. **6**, 849 (1976)

[47] I. A. Campbell, A. Fert, in Ferromagnetic Materials, Wohlfarth (Ed.), Vol **3**, chapter 9, p.747, North Holland Publ. Comp., Amsterdam (1982)

[48] A. Fert, C. Vouille, 30. IFF-Ferienkurs, Magnetische Schichtsysteme, Forschungszentrum Jülich 1999

[49] P. Ciureanu, Magnetoresistive Sensors, Chapter 5 in: P. Ciureanu and S. Middelhoek (Edts.), Thin Film Resistive Sensors, IOP Publishing (1992)

[50] L. Ejsing, M. F. Hansen, A. K. Menon, H. A. Ferreira, D. L. Graham, and P. P. Freitas, Appl. Phys. Lett. **84**, 4729 (2004)

[51] Y. S. Huang, C. C. Wang, A. O. Adeyeye, and D. Tripathy, J. Appl. Phys. **99**, 08C508 (2006)

[52] R. C. O'Handley, Modern Magnetic Materials, Principles and Applications, John Wiley & Sons, New York (2000)

[53] R. P. Hunt, IEEE Trans. MAG-**7**, 150 (1971)

[54] D. A. Thompson, L. T. Romankiw, and A. F. Mayadas, IEEE Trans. MAG-**11**, 1039 (1975)

[55] T. R. McGuire, R. L Potter: IEEE Trans. Mag. MAG-11, 1018 (1975)

[56] R. Steiner Vanha, H. Baltes, Lateral and Vertical Hall Microsystems,in: H. Baltes, W.Göpel, J.Hesse (Eds.), Sensors, Vol. **8** Update, Wiley-VCH Verlag, Weinheim (2001)

[57] Slawomir Tumanski, Thin Film Magnetoresistive Sensors, IOP Publishing, Bristol (2001)

C 1 The standard model of spin injection

Jaroslav Fabian[1] and Igor Žutić[2]

[1]Institute for Theoretical Physics
University of Regensburg, Regensburg, Germany

[2]Department of Physics
State University of New York at Buffalo, Buffalo,
NY 14260, USA

Contents

1 Introduction **2**

2 Simple model of spin injection **3**

3 Spin-polarized transport: concepts and definitions **4**

4 The standard model of spin injection: F/N junction **11**

5 Nonequilibrium resistance and spin bottleneck **14**

6 Transparent and tunnel contacts, conductivity mismatch **15**

7 Silsbee-Johnson spin-charge coupling **16**

8 Spin injection in $F/N/F$ junctions **18**

9 Nonlocal spin-injection geometry: Johnson-Silsbee spin injection experiment. **22**

1 Introduction

The generation of nonequilibrium electron spin, as well as the nonequilibrium spin itself, in electronic materials (metals and semiconductors), is called *spin accumulation*.[1] The most important techniques for spin accumulation are electrical spin injection, optical spin orientation, and spin resonance. By electrical spin injection, or simply *spin injection*, we mean spin accumulation by injecting spin-polarized electrons from one material to another, by electric current. The source material could be a ferromagnetic metal, for example Fe, in which there is a difference in the densities of spin up and spin down electrons. Such a difference is characterized by a spin *polarization*. In the ferromagnet the spin polarization exists in equilibrium. In contrast, if electrons from the ferromagnet are injected into a nominally nonmagnetic metal, say, Al, the resulting spin polarization in Al is a nonequilibrium one: spin accumulates in Al. Another possibility is an electrical spin injection between two nonmagnetic materials, say Al and Cu. If one of the materials has a nonequilibrium spin, electric current can lead to spin accumulation in the other material. Electrical spin injection is the main topic of these lecture notes.

The two other techniques for spin accumulation historically preceded spin injection. Optical orientation is a process of generating nonequilibrium spin optically, by exposing the material to a circularly polarized light. The angular momentum of the photons is transferred to the electron spin. Optical orientation is most effective in direct band semiconductors such as GaAs. The historically first technique for investigation nonequilibrium spin has been electron spin resonance. Application of a magnetic field splits the spin up and spin down electron states (Zeeman splitting) with a corresponding equilibrium spin polarization. A microwave radiation[2] can induce transitions between the spin-split states, generating nonequilibrium spin. The spin resonance technique has been used in metals and semiconductors. There are other ways to generate spin accumulation, typically much less efficient as with the three ways mentioned above. One example is the spin Hall effect, in which electric current leads to a separation of spin up and spin down electrons at the edges parallel to the current flow. Another possibility is to first accumulate nuclear spin in the lattice ions; electron spins can be then polarized via the hyperfine interaction.

The standard model of spin injection originates from the proposal of Aronov [1] who suggested the possibility of electrical spin injection from a ferromagnetic to a nonmagnetic conductor. The thermodynamics of spin injection has been developed by Johnson and Silsbee, who also formulated a drift-diffusion transport model for spin transport across ferromagnet/nonmagnet (F/N) interfaces [2, 3]. This model has been shown to be essentially equivalent to the standard model as presented here [4, 5]. The theory of spin injection was further developed in [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. In particular the presentation of Rashba [14, 15] has inspired the formulation of the standard model of spin injection in the reviews [4, 5] which these lecture notes follow and extend. These reviews should be consulted for original references and examples of experimental results.

[1] By spin in spin injection is meant a spin ensemble, rather than an individual electron spin.

[2] Microwave photons have energies matching the electron Zeeman splitting which is typically $0.01 - 1$ meV's, in fields of order tesla. Radio waves are typically used for nuclear spin resonance.

2 Simple model of spin injection

Perhaps the simplest model of spin injection considers a steady flow of a spin-polarized electric current from a ferromagnet to a nonmagnetic conductor. The ferromagnet has an electron spin polarization P_0; for the present purposes P_0 is the relative difference between the "relevant" densities of spin up and spin down electrons. More specific definitions of the term are given later. In a typical ferromagnetic metal P_0 is 10–50%. In nonmagnetic metals the spin polarization at equilibrium vanishes.
Calling the ferromagnetic conductor F and the nonmagnetic conductor N, we have a simple F/N junction. We wish to answer the following question:

Given the equilibrium spin polarization P_0 in the ferromagnet, what is the spin accumulation in the nonmagnetic conductor if electric current j flows through the junction?

In order to answer this question, we need to know how much spin per unit time arrives from F to N. The simplest answer would be $j_{s0}/(-e)$, where the spin current

$$j_{s0} = P_0 j, \tag{1}$$

as the spins are attached to the electrons flowing through the interface. We can take this value as a very rough estimate of what to expect. What Eq. 1 neglects is the possibility of spin accumulation in the ferromagnet. As we will see later, spin indeed accumulates in the ferromagnet, strongly modifying the above estimate for j_{s0}. Another simplification we made is to suppose that the spin is preserved during crossing the interface. This approximation is actually quite good and will be used in the standard model as well.
Knowing the spin current at the interface, we can focus on the N region. What happens to the spin which crosses the interface? Unlike charge, spin is not conserved. Spin relaxes to the equilibrium value (which is zero in N) due to spin-flip scattering and other spin-randomizing processes. As a result, the motion of the spin in the presence of spin current will be diffusive.[3]
For the spin density $s(x)$ in the N region we can then write a diffusion equation

$$\frac{d^2 s}{dx^2} = \frac{s}{L_s^2}, \tag{2}$$

where L_s is the spin diffusion length in the nonmagnetic conductor. In terms of diffusivity D and the spin relaxation time τ_s the spin diffusion length is given as

$$L_s = \sqrt{D\tau_s}. \tag{3}$$

The diffusion equation has a general solution,

$$s(x) = s_0 e^{-x/L_s}, \tag{4}$$

where $s_0 = s(0)$ is the spin density at the interface, $x = 0$. Above we applied the physical condition that $s(\infty) = 0$.
What remains is to connect the spin density s_0 with the spin current j_{s0}. Since the transport of spin is diffusive, the spin current is

$$j_s = (-e) \times -D\frac{ds}{dx}. \tag{5}$$

[3] In general, the motion will be a combination of drift and diffusion. At reasonable electric fields driving the electric current the drift is much smaller than diffusion and can be neglected.

Note that we define the spin current as the electric current corresponding to the spin flow—that is why the multiplication by $-e$ above. At $x = 0$, using Eq. 4, we obtain

$$j_s = \frac{-eD}{L_s} s_0 e^{-x/L_s}. \tag{6}$$

Assuming that the spin current is continuous across the interface (spin relaxation is absent there), $j_s(0) = j_{s0}$, we find

$$s_0 = j_{s0} \frac{L_s}{-eD}. \tag{7}$$

The full spin density profile in N is given by

$$s(x) = j_{s0} \frac{L_s}{-eD} e^{-x/L_s}. \tag{8}$$

The total amount of accumulated spin is

$$s_{\rm acc} = \int_0^\infty s(x)dx = \frac{j_{s_0} L_s^2}{-eD} = \frac{j_{s0}}{-e}\tau_s. \tag{9}$$

In effect, the spin is pumped into the N region. The steady state is achieved by spin relaxation: The more pumping and the less spin relaxation, the higher is the spin accumulation.[4]

3 Spin-polarized transport: concepts and definitions

Quasichemical potentials. In thermodynamic equilibrium the chemical potential η throughout the electronic system is uniform, determining the electron density

$$n_0(\eta) = \int d\epsilon g(\epsilon) f_0(\epsilon), \tag{10}$$

where $g(\epsilon)$ is the electronic density of states at the energy ϵ and f_0 is the equilibrium Fermi-Dirac distribution function at a given temperature T,

$$f_0(\epsilon) = \frac{1}{\exp(\epsilon - \eta)/k_B T + 1}. \tag{11}$$

In the presence of an electrostatic potential $\phi(x)$ giving rise to electric current due to the electric field $E = -\nabla\phi$ inside the conductor, the chemical potential is no longer uniform (the system is no longer an equilibrium one):

$$\eta \to \eta + e\mu(x), \tag{12}$$

where the space dependent addition $\mu(x)$ is the *quasichemical potential*. Since typically the momentum relaxes on length scales smaller than the variation of ϕ, we can assume the local nonequilibrium electron distribution function to be only energy dependent,

$$f(\epsilon, x) = f_0[\epsilon - e\phi(x) - \eta - e\mu(x)]. \tag{13}$$

Then the nonequilibrium electron density is

$$n(x) = \int d\epsilon g(\epsilon) f(\epsilon, x) = n_0(\eta + e\mu + e\phi). \tag{14}$$

[4]Think of inflating a raptured balloon: the more you blow and the tinier is the hole the bigger the balloon gets. The rapture symbolizes spin relaxation.

Local charge neutrality. We make the assumption that charge does not accumulate inside the conductor under bias ϕ. This is an excellent approximation for metals and highly doped (degenerate) semiconductors. On the other hand, charge can be injected and accumulated in nondegenerate semiconductors due to the large screening length. For such cases the standard spin injection model does not apply. The local charge neutrality means that

$$n(x) = n_0. \tag{15}$$

This gives the general condition,

$$\mu(x) = -\phi(x). \tag{16}$$

The quasichemical potential fully balances the electrostatic potential.

Electric current. The electric current comprises the drift current, proportional to the electric field $E = -\nabla\phi$, and the diffusion current, proportional to the gradient of the electron density ∇n:

$$j = \sigma E + eD\nabla n. \tag{17}$$

The two proportionality parameters are conductivity σ and diffusivity, D. Due to charge neutrality the diffusion current is absent. We will keep it in the discussion as diffusion will be present in the spin flow. Using Eq. 14, we write

$$\nabla n = \frac{\partial n_0}{\partial \eta} e\nabla\phi + \frac{\partial n_0}{\partial \eta} e\nabla\mu. \tag{18}$$

Substituting to Eq. 17 gives

$$j = \left(-\sigma + e^2 D\frac{\partial n_0}{\partial \eta}\right)\nabla\phi + e^2 D\frac{\partial n_0}{\partial \eta}\nabla\mu. \tag{19}$$

There are two important consequences of this equation. First, if the chemical potential is uniform, $\nabla\mu = 0$, the current has to vanish. This gives the condition on the conductivity,

$$\sigma = e^2 D\frac{\partial n_0}{\partial \eta}, \tag{20}$$

known as the Einstein relation. To a good approximation $\partial n_0/\partial\eta = g(\eta)$, where $g(\eta)$ is the electron density of states at the Fermi level. Second, using the Einstein relation, the electric current is expressed through the quasichemical potential only,

$$j = \sigma\nabla\mu. \tag{21}$$

This equation generalizes the familiar $j = \sigma E$ to situations with diffusive currents. The gradient of μ carries information on both drift and diffusion.
In a steady state, the continuity of the electric current requires that

$$\nabla j = 0, \tag{22}$$

that is, the current is uniform. We can also identify the total increase of the quasichemical potential across the system with applied voltage. Indeed, for a uniform system of length L integration of Eq. 21 gives

$$\Delta\mu = \frac{L}{\sigma} j = \mathcal{R} j, \tag{23}$$

where $\mathcal{R}$ is the electric resistance of the system.[5]

[5]We consider conductors of a unit area cross section. For rectangular conductors of cross-sectional area S all the resistances that appear in this article should be divided by S.

Contact resistance. At sharp contacts the chemical potential need not be continuous. Instead of Eq. 21 we write

$$j = \Sigma \Delta\mu, \tag{24}$$

in which Σ is the *contact conductance* and $\Delta\mu$ is the increase of the chemical potential across the interface. The contact electrical resistance is

$$\mathcal{R}_c = \frac{1}{\Sigma_c}. \tag{25}$$

Problem. Consider two conductors, A and B, forming a junction with contact resistance $\mathcal{R}_c$. The conductivities of A and B are σ_A and σ_B. Integrate Eq. 21 for each conductor, and apply the condition of the electric current continuity together with Eq. 24 to obtain j as a function of the applied voltage. What is the total junction resistance? The standard model of spin injection goes in the same spirit as this exercise.

Spin density and spin polarization. Consider a conductor with the electron density n. This density comprises the densities of spin up and spin down electrons:

$$n = n_\uparrow + n_\downarrow. \tag{26}$$

We define the spin density as

$$s = s_\uparrow - s_\downarrow. \tag{27}$$

A relative difference between the spin up and spin down densities is the spin polarization of the density,

$$P_n = \frac{s}{n}. \tag{28}$$

We add the label n to stress that we speak about the density spin polarization. For a general spin-resolved quantity X, we will have

$$P_X = \frac{X_\uparrow - X_\downarrow}{X_\uparrow + X_\downarrow}, \tag{29}$$

and call it the "spin polarization of X."

Spin accumulation. Let us allow for different densities of states $g_\uparrow$ and $g_\downarrow$ at the Fermi level, as well as different quasichemical potentials $\mu_\uparrow$ and $\mu_\downarrow$ for spin up and spin down electrons. The equilibrium chemical potential η is the same for both spin species.[6] Then

$$n_\uparrow(x) = n_{\uparrow 0}\left(\eta + e\mu_\uparrow + e\phi\right) \approx n_{\uparrow 0} + \frac{\partial n_{\uparrow 0}}{\partial \eta}\left(e\mu_\uparrow + e\phi\right), \tag{30}$$

$$n_\downarrow(x) = n_{\downarrow 0}\left(\eta + e\mu_\downarrow + e\phi\right) \approx n_{\downarrow 0} + \frac{\partial n_{\downarrow 0}}{\partial \eta}\left(e\mu_\downarrow + e\phi\right), \tag{31}$$

[6]The energy can flow between spin up and down electrons leading to a common temperature. Similarly, spin-flip processes lead to exchange of particles among the two spin pools, giving a unique equilibrium chemical potential.

where we have expanded the nonequilibrium densities assuming that $\mu + \phi$ is much smaller than the equilibrium chemical potential η; this is a good approximation as it is the electrons close to the Fermi level that contribute to spin accumulation in degenerate conductors. Since $\partial n_0/\partial\eta = g$, we find

$$n_\uparrow(x) = n_{\uparrow 0} + g_\uparrow e\mu_\uparrow + g_\uparrow e\phi, \tag{32}$$
$$n_\downarrow(x) = n_{\downarrow 0} + g_\downarrow e\mu_\downarrow + g_\downarrow e\phi. \tag{33}$$

The local charge neutrality, $n_\uparrow + n_\downarrow = n_0$, then leads to the condition,

$$g(\mu + \phi) + g_s\mu_s = 0, \tag{34}$$

where

$$g = g_\uparrow + g_\downarrow \tag{35}$$
$$g_s = g_\uparrow - g_\downarrow. \tag{36}$$

For nonmagnetic conductors $g_s = 0$, recovering Eq. 16. From Eqs. 32 and 33, using the charge neutrality Eq. 34, we obtain for the spin density

$$s = s_0 + eg_s(\mu + \phi) + eg\mu_s = s_0 + 4e\mu_s\frac{g_\uparrow g_\downarrow}{g}. \tag{37}$$

Here we denoted the quasichemical and *spin quasichemical*[7] potentials

$$\mu = (\mu_\uparrow + \mu_\downarrow)/2, \tag{38}$$
$$\mu_s = (\mu_\uparrow - \mu_\downarrow)/2. \tag{39}$$

The accumulated nonequilibrium spin δs defined by

$$s = s_0 + \delta s, \tag{40}$$

is then

$$\delta s = 4e\frac{g_\uparrow g_\downarrow}{g}\mu_s. \tag{41}$$

Both the nonequilibrium spin density δs and the spin quasichemical potential μ_s are often termed *spin accumulation*.

Charge and spin currents. Charge current is the total electric current carried by spin up and spin down electrons,

$$j = j_\uparrow + j_\downarrow. \tag{42}$$

By contrast, spin current is the difference between the electric currents carried by spin up and spin down electrons:

$$j_s = j_\uparrow - j_\downarrow. \tag{43}$$

The two spin components of the electric current are given by

$$j_\uparrow = \sigma_\uparrow\nabla\mu_\uparrow, \tag{44}$$
$$j_\downarrow = \sigma_\downarrow\nabla\mu_\downarrow. \tag{45}$$

[7]Beware of factors of "2". In the literature μ_s is sometimes defined by the plain difference $\mu_\uparrow - \mu_\downarrow$.

We have labeled the conductivities and the quasichemical potentials with the corresponding spin index. In nonmagnetic conductors $\sigma_\uparrow = \sigma_\downarrow$. Let us introduce the charge and spin conductivities as follow:

$$\sigma = \sigma_\uparrow + \sigma_\downarrow, \tag{46}$$
$$\sigma_s = \sigma_\uparrow - \sigma_\downarrow. \tag{47}$$

The electric charge and spin currents become

$$j = \sigma\nabla\mu + \sigma_s\nabla\mu_s, \tag{48}$$
$$j_s = \sigma_s\nabla\mu + \sigma\nabla\mu_s. \tag{49}$$

For a nonmagnetic conductor $\sigma_s = 0$ and the charge and spin currents decouple; the charge current is driven by the gradient of the quasichemical potential while the spin current is driven by the gradient of the spin accumulation. In a ferromagnetic conductor $\sigma_s \neq 0$ and a gradient in spin accumulation can cause a charge current. Similarly, a gradient in the quasichemical potential alone would cause a spin current.

Current spin polarization. The spin polarization of the electric current P_j is defined according to Eq. 29,

$$P_j = \frac{j_\uparrow - j_\downarrow}{j_\uparrow + j_\downarrow} = \frac{j_s}{j}. \tag{50}$$

Extract $\nabla\mu$ from Eq. 48,

$$\nabla\mu = \frac{1}{\sigma}\left(j - \sigma_s\nabla\mu_s\right), \tag{51}$$

and substitute into Eq. 49:

$$j_s = \sigma_s\nabla\mu + \sigma\nabla\mu_s = P_\sigma j + 4\frac{\sigma_\uparrow\sigma_\downarrow}{\sigma}\nabla\mu_s. \tag{52}$$

Here P_σ is the conductivity spin polarization,

$$P_\sigma = \frac{\sigma_\uparrow - \sigma_\downarrow}{\sigma_\uparrow + \sigma_\downarrow} = \frac{\sigma_s}{\sigma}. \tag{53}$$

The spin and charge currents are coupled through P_σ. Finally, the current spin polarization is

$$P_j = \frac{j_s}{j} = P_\sigma + \frac{1}{j}4\nabla\mu_s\frac{\sigma_\uparrow\sigma_\downarrow}{\sigma}. \tag{54}$$

In nonmagnetic conductors $P_\sigma = 0$ and spin current is due to the gradient in spin accumulation only.

Spin-polarized currents in contacts. The above formalism can be rewritten for contacts with discrete jumps of the quasichemical potentials. Following Eq. 24, the spin-resolved currents are

$$j_\uparrow = \Sigma_\uparrow\Delta\mu_\uparrow, \tag{55}$$
$$j_\downarrow = \Sigma_\downarrow\Delta\mu_\downarrow. \tag{56}$$

Defining the contact charge and spin conductances as

$$\Sigma = \Sigma_\uparrow + \Sigma_\downarrow, \tag{57}$$
$$\Sigma_s = \Sigma_\uparrow - \Sigma_\downarrow, \tag{58}$$

we can write

$$j = \Sigma\Delta\mu + \Sigma_s\Delta\mu_s, \tag{59}$$
$$j_s = \Sigma_s\Delta\mu + \Sigma\Delta\mu_s. \tag{60}$$

Going through similar steps as above of Eq. 54, we obtain for the spin current polarization in the contact

$$P_{jc} = P_\Sigma + \frac{1}{j}\frac{\Delta\mu_s(0)}{R_c}. \tag{61}$$

Here

$$P_\Sigma = \frac{\Sigma_\uparrow - \Sigma_\downarrow}{\Sigma_\uparrow + \Sigma_\downarrow} = \frac{\Sigma_s}{\Sigma}, \tag{62}$$

is the contact spin conductance polarization and

$$R_c = \frac{\Sigma}{4\Sigma_\uparrow\Sigma_\downarrow}, \tag{63}$$

is the effective contact resistance,[8] determining the drop of the spin accumulation across the contact; R_c is a quarter of the series resistance of the spin up and spin down contact resistances. In a spin unpolarized contact $R_c = \mathcal{R}_c = 1/\Sigma$.

Diffusion of spin accumulation. In nonmagnetic systems it is sufficient to use the continuity of the charge current, Eq. 22, to find the profile of the quasichemical potential $\mu(x)$. In the presence of spin polarization, we need a continuity condition for the spin current as well; the continuity of the charge current remains unchanged: $\nabla j = 0$. Since, unlike charge, spin is not conserved, the continuity equation for the spin current is

$$\nabla j_s = e\frac{\delta s}{\tau_s}, \tag{64}$$

where δs is the deviation of the spin density from its equilibrium value: $s = s_{\rm eq} + \delta s$. The divergence of the spin current is proportional to the rate of spin relaxation $1/\tau_s$, with τ_s denoting the spin relaxation time. On one hand,

$$\nabla j_s = e\frac{\delta s}{\tau_s} = 4e^2\mu_s\frac{g_\uparrow, g_\downarrow}{g}\frac{1}{\tau_s} \tag{65}$$

where we used Eq. 41 for δs. On the other hand, Eq. 52 gives

$$\nabla j_s = \nabla\left(P_\sigma j + \nabla\mu_s\frac{4\sigma_\uparrow\sigma_\downarrow}{\sigma}\right) = 4\frac{\sigma_\uparrow\sigma_\downarrow}{\sigma}\nabla^2\mu_s. \tag{66}$$

[8] This is the first of a series of effective resistances which appear in the spin injection problem. To distinguish them from the corresponding electrical resistances we use calligraphic symbols for the latter.

Comparing the two we get the following diffusion equation for spin accumulation:

$$\nabla^2 \mu_s = \frac{\mu_s}{L_s^2}, \tag{67}$$

where the generalized spin diffusion length L_s is

$$L_s = \sqrt{\overline{D}\tau_s}, \tag{68}$$

and the generalized diffusivity

$$\overline{D} = \frac{g}{g_\uparrow/D_\downarrow + g_\downarrow/D_\uparrow}. \tag{69}$$

In a nonmagnetic conductor $\overline{D} = D$. Representative spin relaxation times τ_s in nonmagnetic metals and semiconductors are nanoseconds, and spin diffusion lengths micrometers. In ferromagnetic conductors these quantities are smaller by several orders of magnitude.

Spin-charge coupling. Let us write Eq. 51 as

$$\nabla \mu = \frac{j}{\sigma} - P_\sigma \nabla \mu_s, \tag{70}$$

and integrate it over a homogeneous region of a conductor:

$$\Delta \mu = j\mathcal{R} - P_\sigma \Delta \mu_s, \tag{71}$$

where $\mathcal{R}$ is the electrical resistances of the region. Consider a homogeneous ferromagnetic conductor of length $L \gg L_s$, stretching from $x = -L$ to $x = 0$. Assume that at $x = 0$ there is a spin accumulation $\mu_s(0)$. Applying the above equation gives

$$\mu(0) - \mu(-L) = j\mathcal{R} - P_\sigma \mu_s(0), \tag{72}$$

where the conductor's electric resistance is $\mathcal{R} = L/\sigma$ and we assumed absence of spin accumulation at $x = -L$. In a nonmagnetic conductor $P_\sigma = 0$ and the increase of the quasichemical potential is due to the charge current flow only. In a ferromagnetic conductor the increase is also due to the spin accumulation. In an open circuit ($j = 0$) the increase in the quasichemical potential is

$$\mu(0) - \mu(-L) = -P_\sigma \mu_s(0), \tag{73}$$

This increase defines the electromotive force (emf) per unit charge[9] generated by the spin accumulation in the ferromagnetic conductor. Similarly, we can calculate the corresponding drop in the electric potential,

$$\phi(-L) - \phi(0) = (P_g - P_\sigma)\mu_s(0), \tag{74}$$

where we used the local neutrality condition, Eq. 34. The density of states spin polarization is

$$P_g = \frac{g_\uparrow - g_\downarrow}{g_\uparrow + g_\downarrow} = \frac{g_s}{g}. \tag{75}$$

Equation 74 is an example of spin-charge coupling: The presence of a spin accumulation in a conductor with an equilibrium spin polarization, a nonequilibrium voltage drop (electromotive force) develops. Electrostatic detection of the voltage drop then allows to extract the magnitude of the spin accumulation.

[9]The spin accumulation at first generates spin diffusion and the connected electron flow—since we are dealing with a ferromagnet. In the open circuit a balancing electric field develops preventing unlimited buildup of charges at the two ends of the conductor. The resulting emf is the work done by the source of the spin accumulation in bringing the electrons through the conductor against the built-up electric field.

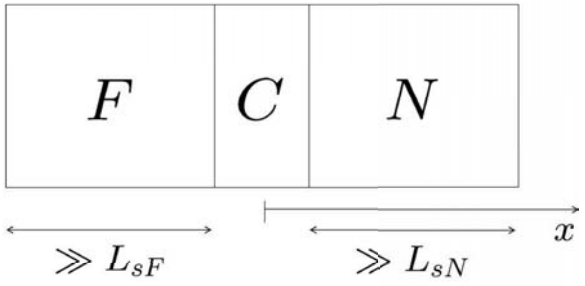


Fig. 1: *The F/N junction above comprises a ferromagnetic conductor F, a nonmagnetic conductor N, as well as the contact C between them at $x = 0$. It is assumed that the widths of the F and N regions are much larger (we call their size "∞") than the corresponding spin diffusion length.*

4 The standard model of spin injection: F/N junction

We pose the following question:

Knowing the equilibrium materials parameters of a ferromagnet (F), a nonmagnetic conductor (N), as well as the properties of the contact (C) between them, what is the spin current polarization and spin accumulation in N, in the presence of electric current j?

The scheme of the F/N junction we consider is in Fig. 1. The spin current polarization at the contact is termed *spin injection efficiency*. We denote it as P_j. To obtain P_j we need to consider spin-polarized transport separately in the three regions: F, C, and N. The solutions for the transport equations will then be connected by suitable continuity conditions. We also add labels F, C, and N to the quantities pertaining to the three regions.

Ferromagnetic conductor. The ferromagnetic conductor occupies the region $(-\infty, 0)$. The spin accumulation profile is given by the solution of the diffusion equation, Eq. 67, as

$$\mu_{sF} = \mu_{sF}(0)e^{x/L_{sF}}. \tag{76}$$

We have applied the condition that there is no spin accumulation at $x = -\infty$: $\mu_{sF}(-\infty) = 0$. This condition is well satisfied if the length of the ferromagnet, indicated by "∞", is much larger than the spin diffusion length L_{sF}. From the above we have

$$\nabla\mu_{sF}(0) = \frac{\mu_{sF}(0)}{L_{sF}}. \tag{77}$$

Substituting to Eq. 54 we obtain the spin current polarization in the F region of the contact

$$P_{jF}(0) = P_{\sigma F} + \frac{1}{j}\frac{\mu_{sF}(0)}{R_F}, \tag{78}$$

where we denote

$$R_F = \frac{\sigma_F}{4\sigma_{F\uparrow}\sigma_{F\downarrow}}L_{sF}, \tag{79}$$

the effective resistance of the ferromagnet; R_F is a quarter of the serial resistance of the spin up and spin down resistances of a piece of the ferromagnet of length L_{sF}. We stress that R_F is *not* the actual resistance of the F region $\mathcal{R}_F$, which is

$$\mathcal{R}_F = \frac{\text{“}\infty\text{”}}{\sigma_\uparrow + \sigma_\downarrow}, \tag{80}$$

given as a parallel resistance of the two spin channels over the entire size “∞” of the ferromagnet. The two resistances, R_F and $\mathcal{R}_F$ can be very different!

Nonmagnetic conductor. In a nonmagnetic conductor the transport and materials parameters are spin independent and all the equilibrium polarizations, such as P_σ or P_g, vanish. The profile of the spin accumulation is

$$\mu_{sN} = \mu_{sN}(0)e^{-x/L_{sF}}, \tag{81}$$

satisfying the boundary condition $\mu_{sN}(\infty) = 0$. We then have

$$\nabla\mu_{sN}(0) = -\frac{\mu_{sN}(0)}{L_{sN}}, \tag{82}$$

and the spin current polarization at the contact

$$P_{jN}(0) = -\frac{1}{j}\frac{\mu_{sN}(0)}{R_N}, \tag{83}$$

where

$$R_N = \frac{L_{sN}}{\sigma_N}, \tag{84}$$

is the effective resistance of the N region; R_N is the resistance of a piece of a conductor of size L_{sN}. Again, R_N can be very different from the actual electric resistance of the N region, $\mathcal{R}_N$.

Contact region. The contact region is described by Eq. 61. For our F/N contact the spin current polarization is

$$P_{jc} = P_\Sigma + \frac{1}{j}\frac{\mu_{sN}(0) - \mu_{sF}(0)}{R_c}. \tag{85}$$

Spin injection efficiency. We have three equations for the spin current polarizations, Eqs. 78, 83, 85, in three different regions. We assume that spin is conserved across the contact. As a consequence, the spin current (and thus spin current polarization) is continuous there:

$$P_j \equiv P_{jF}(0) = P_{jN}(0) = P_{jc}. \tag{86}$$

Solving this straightforward algebraic problem leads to the important expression for the spin injection efficiency:

$$P_j = \frac{R_F P_{\sigma F} + R_c P_\Sigma}{R_F + R_c + R_N} = \langle P_\sigma \rangle_R. \tag{87}$$

This equation is one of the main results of the standard model of spin injection. The spin injection efficiency is the weighted average of the equilibrium spin conductance polarizations of the system; the weight is the relative effective resistance.

Problem. *F/F junction.* Calculate the spin injection efficiency for a F/F junction of two different ferromagnets. Show that $P_j = \langle P_\sigma \rangle_R$ still holds.

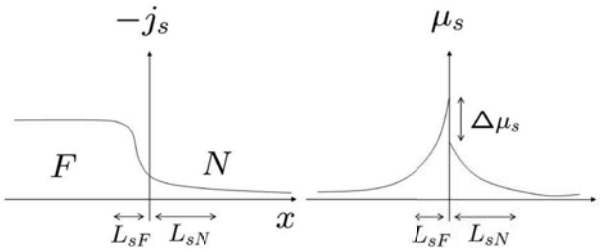


Fig. 2: *Sketch of the spatial profile of the spin current j_s and the spin quasichemical potential μ_s in an F/N junction in the spin injection regime. While the spin current is continuous throughout the junction, the spin quasichemical potential experiences a jump at the contact.*

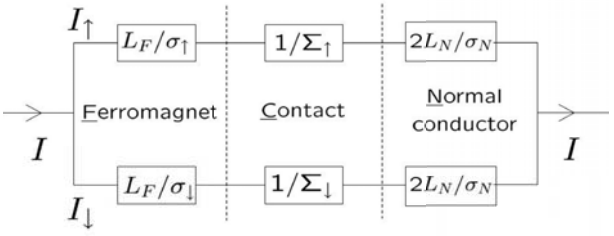


Fig. 3: *Equivalent circuit of a F/N junction in the spin injection regime. The electric current splits into spin up and spin down channels, each with a series of three effective resistances as indicated.*

Spin injection and spin extraction. Knowing P_j we can calculate the spin accumulation in the N region,

$$\mu_{sN}(0) = -jP_jR_N, \tag{88}$$

and the corresponding spin density polarization,

$$P_n(0) = \frac{s(0)}{n} = e\mu_{sN}(0)\frac{g_N}{n} = -jeR_N\frac{g_N}{n}P_j. \tag{89}$$

Since the spin polarization is proportional to the electric current, the electric spin injection is a realization of spin pumping. In a typical spin injection experiment electrons flow from F to N, so that $j < 0$. In this case $P_n(0)$ has the same size as P_j and we speak of *spin injection*. If the electric current is reversed, $j < 0$, electrons from N flow into F. Now $P_n(0)$ has the opposite sign to P_j and we speak of *spin extraction*. For a positive P_j, for example, more spin up than spin down electrons are transported through the contact, leaving a negative spin density in the N region. A sketch of the profile of the spin current and spin quasichemical potential across an F/N junction is shown in Fig. 2.

Equivalent circuit. The standard model of spin injection can be formulated by a simple equivalent circuit model, shown in Fig. 3. The model is a parallel circuit with spin up and spin down channels. Each region is characterized by the corresponding effective resistance.

> **Problem.** Calculate $I_s = I_\uparrow - I_\downarrow$ from the equivalent circuit model and show that $P_j = I_s/I$ agrees with Eq. 87.

Problem. Formulate the equivalent circuit model for a F/F junction.

Problem. *F/N/N junction*. Consider electrical spin injection in an $F/N/N$ junction in which the two N regions are different (say, GaAs and Si). Calculate the spin injection efficiency at the N/N interface. What is the spin accumulation at both sides of this interface? Sketch the profile of the spin accumulation across this junction.

5 Nonequilibrium resistance and spin bottleneck

In the absence of spin accumulation the resistance of the F/N junction is $\mathcal{R}_F + \mathcal{R}_N + \mathcal{R}_c$. Spin accumulation leads to an additional positive resistance $\delta\mathcal{R}$ so that the increase of the quasichemical potential (which generates the emf) is

$$\mu_N(\infty) - \mu_F(-\infty) = (\mathcal{R}_F + \mathcal{R}_N + \mathcal{R}_c + \delta\mathcal{R})j. \tag{90}$$

Let us apply Eq. 71 to the three regions, F, C, and N, successively:

$$\mu_F(0) - \mu_F(-\infty) = j\mathcal{R}_F - P_{\sigma F}\mu_{sF}(0), \tag{91}$$
$$\mu_N(0) - \mu_F(0) = j\mathcal{R}_c - P_\Sigma\left[\mu_{sN}(0) - \mu_{sF}(0)\right], \tag{92}$$
$$\mu_N(\infty) - \mu_N(0) = j\mathcal{R}_N. \tag{93}$$

We have used that $\mu_{sF}(-\infty) = 0$. Summing up the above equations gives for the nonequilibrium resistance

$$\delta\mathcal{R} = -(P_{\sigma F} - P_\Sigma)\mu_{sF}(0) - P_\Sigma\mu_{sN}(0). \tag{94}$$

Expressing the spin quasichemical potentials at $x = 0$ in terms of the spin injection efficiency, see Eqs. 78 and 83,

$$\mu_{sF}(0) = jR_F(P_j - P_{\sigma F}), \tag{95}$$
$$\mu_{sN}(0) = -jR_N P_j, \tag{96}$$

we get

$$\delta\mathcal{R} = -P_\Sigma(P_j - P_\Sigma)R_c - P_{\sigma F}(P_j - P_{\sigma F})R_F. \tag{97}$$

Using the expression for P_j in Eq. 87, we obtain the final result

$$\delta\mathcal{R} = \frac{R_N(P_\Sigma^2 R_c + P_{\sigma F}^2 R_F) + R_F R_c(P_{\sigma F} - P_\Sigma)^2}{R_F + R_c + R_N} > 0. \tag{98}$$

The nonequilibrium resistance is always positive!

Problem. Obtain the nonequilibrium resistance $\delta\mathcal{R}$ from the equivalent circuit model in Fig. 3, as $\delta\mathcal{R} = \mathcal{R} - L_{sF}/\sigma_F - L_{sN}/\sigma_N$.

What is the reason behind the additional positive resistance due to spin accumulation? As the nonequilibrium spin piles up in the ferromagnet and the spin-polarizing contact region, the spin diffusion there pushes the electrons against the flow of the electric current. Indeed, the electric current brings electrons from the spin-polarized region to the nonmagnetic conductor, while the spin diffusion in the ferromagnet and the contact drives them back to the ferromagnet. This *spin bottleneck effect* causes the additional electrical resistance of the junction.

6 Transparent and tunnel contacts, conductivity mismatch

Two important cases are analyzed: transparent and tunnel contacts.

Transparent contacts. By transparent contacts we mean the condition

$$R_c \ll R_N, R_F. \tag{99}$$

This is the case of usual ohmic contacts between two metals or degenerate semiconductors. Using our results for the F/N junction, a transparent contact is characterized by the spin efficiency

$$P_j = \frac{R_F}{R_F + R_N} P_{\sigma F}. \tag{100}$$

For metals σ_F is usually somewhat less than σ_N, as $L_{sN} \gg L_{sF}$. We then get

$$P_j \approx (\sigma_N/\sigma_F)(L_{sF}/L_{sN}). \tag{101}$$

If N is a semiconductor while F is a metal, so that $\sigma_N \ll \sigma_F$, the spin injection efficiency is greatly reduced. This inefficiency of the spin injection from a ferromagnetic metal to a nonmagnetic semiconductor via a transparent contact is known as the *conductivity mismatch* problem, since it comes from the greatly different conductivities of the two regions of the junctions.
The nonequilibrium resistance of a transparent contact is

$$\delta\mathcal{R} = \frac{R_N R_F}{R_N + R_F} P_{\sigma F}^2. \tag{102}$$

Again, since typically R_N is greater than R_F,

$$\delta\mathcal{R} \approx R_F P_{\sigma F}^2 = \frac{L_{sF}}{\sigma_F} P_{\sigma F}^2. \tag{103}$$

In the extreme limit of the conductivity mismatch, the nonequilibrium resistance will be negligible as compared to the usual electrical junction resistance which will be dominated by $\mathcal{R}_N$.

Tunnel contacts. By tunnel contacts we mean

$$R_c \gg R_N, R_F. \tag{104}$$

The contact dominates the electric properties of the junction. The spin injection efficiency for a tunnel contact is

$$P_j \approx P_\Sigma. \tag{105}$$

The contact also dominates the spin injection efficiency. The conductance mismatch in tunnel contacts plays no role and spin injection from a ferromagnetic metal to a nonmagnetic semiconductor can be highly efficient.
The nonequilibrium resistance of a tunnel contact is

$$\delta\mathcal{R} = R_N P_\Sigma^2 + R_F (P_{\sigma F} - P_\Sigma)^2. \tag{106}$$

This is in general much less than the electric resistance of the contact, $\mathcal{R}_c$.

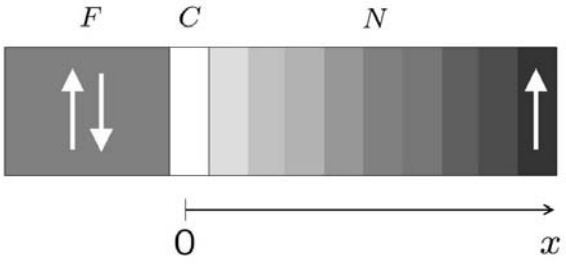


Fig. 4: *Scheme of the Silsbee-Johnson spin-charge coupling. A spin source at the far right of the N region drives spin by diffusion towards the spin-polarizing contact and ferromagnet. The proximity of the nonequilibrium spin and the equilibrium spin polarization gives rise to an emf in the open circuit (or electric current when the circuit is closed).*

Problem. *Spin accumulation in transparent and tunnel junctions*. Calculate the spin accumulation $\mu_{sN}(0)$ and the spin density polarization $P_{\sigma N}(0)$ in a transparent and a tunnel F/N junction. What is the spin density polarization in the conductivity mismatch problem? Can it be significant?

Problem. *Equivalent circuit of the conductivity mismatch problem*. Draw the equivalent circuit for the conductivity mismatch problem of a transparent F/N junction and use it to explain the spin injection inefficiency.

7 Silsbee-Johnson spin-charge coupling

Driving electric current through a F/N interface generates spin accumulation by the process of spin injection. The Silsbee-Johnson spin charge coupling is the inverse of spin injection: emf develops by the presence of a spin accumulation in the proximity of a ferromagnetic conductor. We will analyze the coupling in an open F/N junction, that is in the absence of electric current ($j = 0$), under the condition of $\mu_{sN}(\infty) \neq 0$ which models a source of nonequilibrium spin far in the nonmagnetic region. The scheme is shown in Fig. 4.
The induced emf is the increase of the quasichemical potential across the junction,

$$\mathrm{emf} = \mu_N(\infty) - \mu_F(-\infty). \tag{107}$$

The charge neutrality and the physical condition that $\mu_{sF}(-\infty) = 0$ guarantee that the emf can be detected as a drop of the electric voltage:

$$\mathrm{emf} = \mu_N(\infty) - \mu_F(-\infty) = \phi_F(-\infty) - \phi_N(\infty). \tag{108}$$

Our strategy is to first express the quasichemical potential increase in terms of the spin accumulations at the contact, and then use the spin current continuity at the contact as well as the diffusion of the spin accumulation to find the spin accumulations.
In the absence of electric current we can apply Eq. 71 to F, C, and N regions sequentially:

$$\mu_F(0) - \mu_F(-\infty) = -P_{\sigma_F}\mu_{sF}(0), \tag{109}$$

$$\mu_N(0) - \mu_F(0) = -P_\Sigma\left[\mu_{sN}(0) - \mu_{sF}(0)\right], \tag{110}$$

$$\mu_N(\infty) - \mu_N(0) = 0. \tag{111}$$

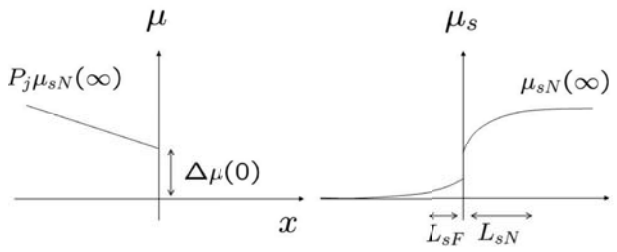


Fig. 5: *Sketch of the spatial profile of the quasichemical potential μ (left) and the spin quasichemical potential μ_s (right) in a F/N junction in the spin-charge coupling regime.*

In the nonmagnetic conductor there is no voltage drop associated with the presence of spin accumulation if $j = 0$. Summing up the above equations gives

$$\mathrm{emf} = \mu_N(\infty) - \mu_F(-\infty) = \mu_{sF}(0)(P_\Sigma - P_{\sigma F}) - \mu_{sN}(0)P_\Sigma. \tag{112}$$

In the N region, due to the presence of spin accumulation at the far right, the spin accumulation diffusion profile is

$$\mu_{sN}(x) = \mu_{sN}(\infty) + [\mu_{sN}(0) - \mu_{sN}(\infty)]\, e^{-x/L_{sN}}, \tag{113}$$

as can be verified by direct substitution to the diffusion equation, Eq. 67. To calculate the spin current at $x = 0$ in the N region we need the gradient,

$$\nabla\mu_{sN}(0) = -\frac{1}{L_{sN}}\left[\mu_{sN}(0) - \mu_{sN}(\infty)\right]. \tag{114}$$

We are now ready to calculate the spin currents at the interface, for the three regions. Equation 52 gives

$$j_{sN}(0) = -\frac{1}{R_N}\left[\mu_{sN}(0) - \mu_{sN}(\infty)\right], \tag{115}$$

$$j_{sF}(0) = \frac{1}{R_F}\mu_{sF}(0), \tag{116}$$

$$j_{sc} = \frac{1}{R_c}\left[\mu_{sN}(0) - \mu_{sF}(0)\right]. \tag{117}$$

Assuming that the three spin currents are equal,

$$j_s \equiv j_{sF}(0) = j_{sc} = j_{sN}(0), \tag{118}$$

we obtain

$$j_s = \frac{\mu_{sN}(\infty)}{R_F + R_c + R_N}. \tag{119}$$

The emf is then found from

$$\mathrm{emf} = (P_{\sigma F}R_F + P_\Sigma R_c)j_s, \tag{120}$$

which gives the spin-charge coupling in the final form

$$\mathrm{emf} = P_j\mu_{sN}(\infty), \tag{121}$$

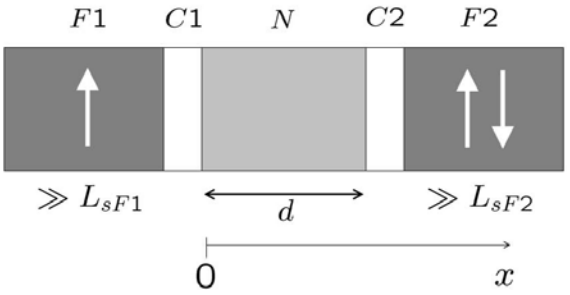


Fig. 6: *Sketch of an $F/N/F$ junction. The magnetization of the $F2$ conductor can be up or down, giving parallel and antiparallel configurations.*

where P_j is the spin injection efficiency of the junction, given in Eq. 87. The spin-charge coupling allows electrical detection of spin accumulation. Placing a ferromagnetic electrode over a conductor with nonequilibrium spin generates emf in the open circuit, or electric current if the junction is part of a closed circuit. The spin accumulation can be generated electrically (see the section on the nonlocal geometry) or by other means (optically or by spin resonance). Figure 5 shows the profile of the quasichemical potentials across the junction.

The origin of the spin-charge coupling can be traced to the presence of the spin current in the ferromagnet. If $P_\sigma \neq 0$ the spin current would also induce electric current. In an open circuit there is instead a balancing emf induced.

> **Problem.** Sketch the spatial profiles of μ, μ_s, and j_s in the F/N junction in the Silsbee-Johnson spin-charge coupling regime.

8 Spin injection in $F/N/F$ junctions

The same technique which we applied to study the spin injection in F/N junctions is applicable to more general structures. We will use it to analyze the spin injection in $F/N/F$ junctions. By independent switching of the orientations of the magnetizations of the two F regions, the junctions can be in the parallel ($\uparrow\uparrow$) or antiparallel ($\uparrow\downarrow$) configurations. We will in particular be interested in the difference of the junction electrical resistance for antiparallel and parallel configurations,

$$\Delta\mathcal{R} = \delta\mathcal{R}^{\uparrow\downarrow} - \delta\mathcal{R}^{\uparrow\uparrow}. \tag{122}$$

This difference contains only contributions of the respective nonequilibrium resistances.

Spin injection efficiencies. The described geometry is shown in Fig. 6. The width of the nonmagnetic conductor is d. The two ferromagnetic layers are labeled $F1$ and $F2$. We start with a generic asymmetric configuration in which $F1$ and $F2$ are different. Going through similar steps as in the F/N junction, we find that the spin current polarizations in the ferromagnets at

$x = 0$ and $x = d$ are

$$P_{jF1}(0) = P_{\sigma F1} + \frac{1}{j}\frac{\mu_{sF1}(0)}{R_{F1}}, \quad (123)$$

$$P_{jF2}(d) = P_{\sigma F2} - \frac{1}{j}\frac{\mu_{sF2}(d)}{R_{F2}}. \quad (124)$$

Similarly, at the two contacts we have

$$P_{jc1} = P_{\Sigma 1} + \frac{1}{j}\frac{\Delta\mu_s(0)}{R_{c1}}, \quad (125)$$

$$P_{jc2} = P_{\Sigma 2} + \frac{1}{j}\frac{\Delta\mu_s(d)}{R_{c2}}. \quad (126)$$

In contrast to the F/N junction, the N region is of finite width d. Considering the quasichemical potentials $\mu_{sN}(0)$ and $\mu_{sN}(d)$ as yet unknown boundary conditions, the solution to the diffusion equation 67 is

$$\mu_{sN}(x) = \mu_{sN}(d)\frac{\sinh(x/L_{sN})}{\sinh(d/L_{sN})} - \mu_{sN}(0)\frac{\sinh\left[(x-d)/L_{sN}\right]}{\sinh(d/L_{sN})}. \quad (127)$$

By evaluating $\nabla\mu_{sN}(0)$ and $\nabla\mu_{sN}(d)$ from the above equation, we obtain for the spin current polarizations in the nonmagnetic region,

$$P_{jN}(0) = \frac{1}{jR_N}\frac{1}{\sinh(d/L_{sN})}\left[\mu_{sN}(d) - \mu_{sN}(0)\cosh(d/L_{sN})\right], \quad (128)$$

$$P_{jN}(d) = \frac{1}{jR_N}\frac{1}{\sinh(d/L_{sN})}\left[\mu_{sN}(d)\cosh(d/L_{sN}) - \mu_{sN}(0)\right]. \quad (129)$$

The above equations for the spin current polarizations need to be supplemented by the continuity conditions for the spin currents at the two contacts:

$$P_{j1} \equiv P_{jF1}(0) = P_{jc1} = P_{jN}(0), \quad (130)$$

$$P_{j2} \equiv P_{jF2}(d) = P_{jc2} = P_{jN}(d). \quad (131)$$

The algebraic system is now complete and we can solve it to obtain the spin injection efficiencies P_{j1} and P_{j2} at the two junctions $F1/N$ and $N/F2$:

$$P_{j1} = P^0_{j1}R_1\frac{R_N\coth(d/L_{sN}) + R_{c2} + R_{F2}}{D_0} + P^0_{j2}\frac{R_2R_N}{D_0\sinh(d/L_{sN})}, \quad (132)$$

$$P_{j2} = P^0_{j2}R_2\frac{R_N\coth(d/L_{sN}) + R_{c1} + R_{F1}}{D_0} + P^0_{j1}\frac{R_1R_N}{D_0\sinh(d/L_{sN})}. \quad (133)$$

Here

$$D_0 = R_N^2 + (R_{c1} + R_{F1})(R_{c2} + R_{F2}) + R_N(R_{c1} + R_{F1} + R_{c2} + R_{F2})\coth(d/L_{sN}), \quad (134)$$

and P^0_{j1} and P^0_{j2} are the spin injection efficiencies of the individual junctions giving by Eq. 87; similarly R_1 and R_2 are the two effective junction resistances:

$$R_1 = R_{F1} + R_{c1} + R_N, \quad R_2 = R_{F2} + R_{c2} + R_N. \quad (135)$$

For a thick N region, if $d \gg L_{sN}$, we recover the spin injection efficiencies of the individual junctions: $P_{j1} \approx P^0_{j1}$ and $P_{j2} \approx P^0_{j2}$, as expected for spin uncoupled contacts. In the opposite limit of a thin N, if $d \ll L_{sN}$,

$$P_{j1} = P_{j2} = \frac{P^0_{j1}R_1 + P^0_{j2}R_2}{R_{c1} + R_{F1} + R_{c2} + R_{F2}}. \tag{136}$$

The spin injection efficiencies are a weighted mixture of the efficiencies of the two individual junctions. Finally, for tunnel contacts, such as $R_c \gg R_N$, R_F, and $R_N(L_s/a)$, we recover the limit of independent junctions, $P_{j1} \approx P^0_{j1}$ and $P_{j2} \approx P^0_{j2}$.

Nonequilibrium resistance. In order to find the value of the nonequilibrium resistance due to the spin bottleneck, we need to find the increase of the quasichemical potential $\Delta\mu$ across the junction:

$$\mathrm{emf} = \mu_{F2}(\infty) - \mu_{F1}(-\infty) = j\mathcal{R} + j\delta\mathcal{R}, \tag{137}$$

where

$$\mathcal{R} = \mathcal{R}_{F1} + \mathcal{R}_{c1} + \mathcal{R}_N + \mathcal{R}_{c2} + \mathcal{R}_{F2}, \tag{138}$$

is the electrical resistance of the junction in the absence of spin accumulation.
Let us apply Eq. 71 to the five regions of the $F/N/F$ junction: $F1$, $C1$, N, $C2$, and $F2$. In this sequence, the regional increases of the quasichemical potential are

$$\begin{aligned}
\mu_{F1}(0) - \mu_{F1}(-\infty) &= j\mathcal{R}_{F1} - P_{\sigma F1}\mu_{sF1}(0), && (139)\\
\mu_N(0) - \mu_{F1}(0) &= j\mathcal{R}_{c1} - P_{\Sigma 1}\Delta\mu_s(0), && (140)\\
\mu_N(d) - \mu_N(0) &= j\mathcal{R}_N, && (141)\\
\mu_{F2}(d) - \mu_N(d) &= j\mathcal{R}_{c2} - P_{\Sigma 2}\Delta\mu_s(d), && (142)\\
\mu_{F2}(\infty) - \mu_{F2}(d) &= j\mathcal{R}_{F2} + P_{\sigma F2}\mu_{sF2}(d). && (143)
\end{aligned}$$

We have used that $\mu_{sF1}(-\infty) = \mu_{sF2}(\infty) = 0$. Summing these equations up we extract

$$j\delta\mathcal{R} = -P_{\sigma F1}\mu_{sF1}(0) - P_{\Sigma 1}\Delta\mu_s(0) - P_{\Sigma 2}\Delta\mu_s(d) + P_{\sigma F2}\mu_{sF2}(d). \tag{144}$$

Expressing the spin chemical potentials in terms of the spin injection efficiencies P_{j1} and P_{j2},

$$\begin{aligned}
\mu_{sF1}(0) &= (P_{j1} - P_{\sigma F1})jR_{F1}, && (145)\\
\Delta\mu_s(0) &= (P_{j1} - P_{\Sigma 1})jR_{c1}, && (146)\\
\Delta\mu_s(d) &= (P_{j2} - P_{\Sigma 2})jR_{c2}, && (147)\\
\mu_{sF2}(d) &= (P_{\sigma F2} - P_{j2})jR_{F2}, && (148)
\end{aligned}$$

we find for the nonequilibrium resistance

$$\delta R = -P_{\sigma F1}(P_{j1} - P_{\sigma F1})R_{F1} - P_{\Sigma 1}(P_{j1} - P_{\Sigma 1})R_{c1} - P_{\Sigma 2}(P_{j2} - P_{\Sigma 2})R_{c2} - P_{\sigma F2}(P_{j2} - P_{\sigma F2})R_{F2}. \tag{149}$$

Resistance difference $\Delta\mathcal{R}$. Denoting as

$$\Delta P_j \equiv P_j^{\uparrow\downarrow} - P_j^{\uparrow\uparrow}, \tag{150}$$

$$\Delta^+ P_j \equiv P_j^{\uparrow\downarrow} + P_j^{\uparrow\uparrow}, \tag{151}$$

the difference and the sum of the spin injection efficiencies for antiparallel and parallel magnetizations of the two ferromagnets, we find (assuming that the spin efficiencies are positive, for example)

$$\Delta\mathcal{R} = -(R_{F1}P_{\sigma F1} + R_{c1}P_{\Sigma 1})\Delta P_{j1} - (R_{F2}P_{\sigma F2} + R_{c1}P_{\Sigma 2})\Delta^+ P_{j2}. \tag{152}$$

From Eqs. 132 and 133 we find

$$\Delta P_{j1} = -2\frac{R_2 R_N}{D_0 \sinh(d/L_{sN})} P_{j2}^0, \tag{153}$$

$$\Delta^+ P_{j2} = -2\frac{R_1 R_N}{D_0 \sinh(d/L_{sN})} P_{j1}^0. \tag{154}$$

With that we finally get our desired result,

$$\Delta\mathcal{R} = \frac{4R_1 R_2}{D_0 \sinh(d/L_{sN})} R_N |P_{j1}^0 P_{j2}^0|. \tag{155}$$

As expected, $\Delta\mathcal{R}$ vanishes exponentially if $d \gg L_{sN}$, as the differences between parallel and antiparallel cases diminish. On the other hand, for $d \ll L_{sN}$,

$$\Delta\mathcal{R} \approx \frac{4R_1 R_2}{R_{F1} + R_{c1} + R_{c2} + R_{F2}} P_{j1}^0 P_{j2}^0. \tag{156}$$

Transparent contacts. Put $R_{c1} = R_{c2} = 0$ and consider the interesting case of a thin N layer, $d \ll L_{sN}$. For simplicity assume the same ferromagnets, $R_{F1} = R_{F2}$. Then

$$\delta\mathcal{R} \approx 2R_F P_{\sigma F}^2. \tag{157}$$

Problem. Analyze $\Delta\mathcal{R}$ in Eq. 156 in the conductivity mismatch regime, $R_N \gg R_c, R_F$.

Tunnel contacts. Suppose now that the most resistive regions are the contacts and the N region is thin, $d \ll L_{sN}$. Assuming a symmetric junction, the nonequilibrium resistance difference is

$$\Delta\mathcal{R} \approx \frac{2R_c P_\sigma^2}{1 + (R_c/R_N)(d/2L_{sN})}. \tag{158}$$

The spin accumulation detection by ΔR will be most sensitive if

$$dR_c \ll L_{sN} R_N, \tag{159}$$

as then $\Delta\mathcal{R} \approx R_c$ and the resistance change is maximized. Let us find the physical meaning of the above inequality by invoking the definition of R_N, the diffusion length L_{sN}, and the Einstein relation:

$$1 \gg \frac{dr_c}{r_N L_{sN}} = \frac{d}{L_{sN}^2} r_c \sigma_N = \frac{d}{D_N \tau_{sN}} r_c e^2 g_N D_N \approx e^2 (dg_N) \frac{1}{\Sigma_c} \frac{1}{\tau_s}. \tag{160}$$

Expressing the tunnel conductance Σ_c through an effective tunneling probability per unit time, $P_{\rm tunnel}$,

$$\Sigma_c = e(dg_N)P_{\rm tunnel}, \tag{161}$$

and introducing the dwell time

$$\tau_{\rm dwell} = 1/P_{\rm tunnel}, \tag{162}$$

we come to the conclusion that the spin accumulation detection in $F/N/F$ tunnel junctions is most efficient if

$$\tau_{\rm dwell} \ll \tau_{sN}. \tag{163}$$

In words, the dwell time of the electrons in between the two tunnel barriers (the average time the electron spends in the N region) must be much smaller than the spin relaxation time.

Problem. *$N/F/N$ junction.* Calculate the spin efficiency P_j and the nonequilibrium resistance $\delta\mathcal{R}$ for a symmetric $N/F/N$ junction.
a) Show that in the limit of a thin F layer ($d \ll L_{sF}$)

$$P_j = \frac{R_c P_\Sigma}{R_c + R_N}, \tag{164}$$

$$\delta\mathcal{R} = \frac{2R_c R_N P_\Sigma^2}{R_c + R_N}. \tag{165}$$

b) Verify that in the limit of a thick F layer ($d \gg L_{sF}$) the spin injection efficiency P_j reduces to its value for a single F/N junction, and that $\delta\mathcal{R}$ of an $N/F/N$ junction is twice the nonequilibrium resistance of the individual F/N junctions.

9 Nonlocal spin-injection geometry: Johnson-Silsbee spin injection experiment.

In the $F/N/F$ junction studied in the previous section the electric current flows through both contacts. As the current often brings spurious effects from the point of view of spin detection, especially in the presence of an external magnetic field (the Hall effect or anisotropic magnetoresistance), it is important to consider spin injection geometries in which the spin detection circuit is open. We have already met one example of an open circuit spin detection: the Silsbee-Johnson spin-charge coupling. This scheme can be naturally extended to include a spin injection contact, giving what is called a *nonlocal spin-injection geometry* (as the injection and detection circuits are independent) or the *Johnson-Silsbee spin injection experiment*, after the original spin injection scheme.
Our goal is to answer the following question:

Suppose electric current drives spin injection in the spin injection circuit $F1/N$ as indicated in Fig. 7. What is the emf in the open $F2/N$ junction?

The two ferromagnetic electrodes are on the top of a nonmagnetic conductor, separated by spin-polarizing contacts. Spin is injected into N from $F1$. While the electric current flows in the closed circuit formed by $F1/N$, the spin current flows also towards the spin detection circuit $N/F2$. The charge and spin flows are indicated in Fig. 8. For spin the contact $F1/N$ acts as a

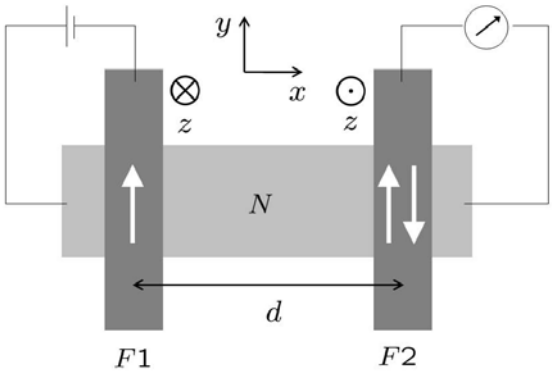


Fig. 7: *Nonlocal geometry for spin injection and detection. The $F1/N$ circuit is closed, the $F2/N$ is open; the electrode $F1$ acts as a spin source, $F2$ as a spin drain. The axes labels are indicated. The directions of the z axis are opposite for the two junctions.*

spin source, while $N/F2$ as a spin sink. The source and sink will appear as special boundary conditions for the spin transport in N. The axes labels are defined in Fig. 7.
We need to be a little careful with this geometry since in principle we are now dealing with a two (if not three) dimensional problem. Nevertheless, the problem can be decoupled to one-dimensional ones if we assume, realistically, that the dimensions of the ferromagnetic electrodes are much greater than the spin diffusion length in the ferromagnets. Indeed, a representative L_{sF} would be on the order of 10 nm or so. In that case we can consider the spin current in $F1$ and $F2$ one-dimensional, along z. On the other hand, we assume that the contact dimensions between F and N, as well as the thickness of N, are much smaller than the spin diffusion length in the non-magnetic conductor, so that the spin current in N can be considered one-dimensional as well. Typically L_{sN} would be more than 1 μm. In most other cases one would need to set up a two-dimensional drift-diffusion problem.[10]
With the above physical restrictions, the quantities labeled N vary along x, while those of F_1 and F_2 along z, as indicated in Figs. 7 and 8. For example, if we write $\mu_{sF2}(0)$ we mean $\mu_{sF2}(z=0)$, the value of the spin quasichemical potential in $F2$ at the place of contact with N. Any variation of μ_{sF2} along y or x is insignificant, occurring at the contact edges only.
We now apply the boundary conditions for the spin quasichemical potentials μ_s at infinities:

$$\mu_{sN}(\pm\infty) = \mu_{sF1}(-\infty) = \mu_{sF2}(\infty) = 0. \tag{166}$$

Let us consider each junction separately.

Spin injector: $F1/N$ junction. The distribution of the spin currents is shown in Fig. 8. In $F1$ the spin accumulation has the profile

$$\mu_{sF1}(z) = \mu_{sF1}(0)e^{z/L_{sF2}}, \tag{167}$$

giving for the spin current at the contact, using Eq. 52

$$j_{sF1}(0) = jP_{\sigma F1} + \frac{1}{R_{F1}}\mu_{sF1}(0). \tag{168}$$

[10]Suppose, for example, that the thickness of N would be much greater than L_{sN}. Then the spin injected from $F1$ would diffuse not only left and right, along x, but also down, along z, forming a complicated diffusion profile. If the contact would be point-like, the surface of an equal spin density would be a semisphere.

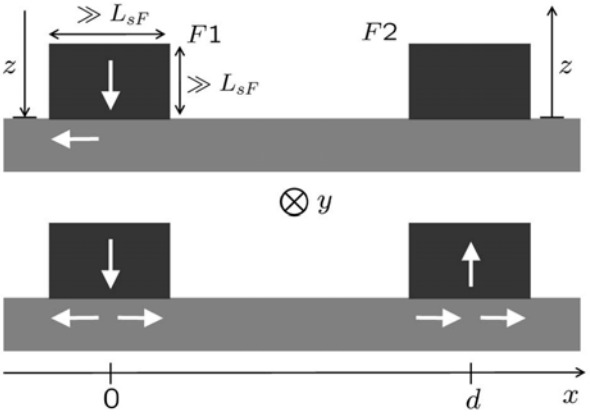


Fig. 8: *Cross view of the nonlocal geometry. The upper graph indicates the flow of electrons, the lower graphs shows the flow of spins. From the point of view of the spin flow in N, the injector plays a role of a spin source, while the detector acts as a spin sink, taking away some of the spin current.*

The spin current through the spin-polarizing contact $C1$ is

$$j_{sc1} = jP_{\Sigma 1} + \frac{1}{R_{c1}}\left[\mu_{sN}(0) - \mu_{sF1}(0)\right]. \tag{169}$$

To obtain the spin current in N, we need to know the profile of the spin quasichemical potential μ_{sN}. Treating the chemical potential at $x = 0$ and $x = d$ as yet unknown, the profile is given by Eq. 127 for $x > 0$. The profile in the whole N region is

$$\mu_{sN}(x \le 0) = \mu_{sN}(0)e^{x/L_{sN}}, \tag{170}$$

$$\mu_{sN}(0 < x < d) = \mu_{sN}(d)\frac{\sinh(x/L_{sN})}{\sinh(d/L_{sN})} - \mu_{sN}(0)\frac{\sinh\left[(x-d)/L_{sN}\right]}{\sinh(d/L_{sN})}, \tag{171}$$

$$\mu_{sN}(x \ge d) = \mu_{sN}(d)e^{-x/L_{sN}}. \tag{172}$$

The spin current is not continuous at $x = 0$, due to the presence of the spin source:

$$j_{sN}(0+) = \frac{1}{R_N}\left[-\mu_{sN}(0)\coth\left(d/L_{sN}\right) + \frac{\mu_{sN}(d)}{\sinh\left(d/L_{sN}\right)}\right], \tag{173}$$

$$j_{sN}(0-) = \frac{1}{R_N}\mu_{sN}(0). \tag{174}$$

The continuity of the spin current at the contact requires that[11]

$$j_{sN}(0+) = j_{sN}(0-) + j_{sc1} = j_{sN}(0-) + j_{sF1}. \tag{175}$$

Using the above algebraic system, we find

$$\mu_{sN}(0)\left[\frac{R_N}{R_{c1}+R_{F1}} + \frac{\exp\left(d/L_{sN}\right)}{\sinh\left(d/L_{sN}\right)}\right] - \mu_{sN}(d)\frac{1}{\sinh\left(d/L_{sN}\right)} = -jR_N\frac{P_{\Sigma 1}R_{c1} + P_{\sigma F1}R_{F1}}{R_{c1}+R_{F1}}. \tag{176}$$

[11] Mind the sign convention: the reference current is always along the positive axis.

The spin current in the spin injector contact is

$$j_{sF1} = j\frac{P_{\Sigma 1}R_{c1} + P_{\sigma F1}R_{F1} + \mu_{sN}(0)/j}{R_{c1} + R_{F1}}. \tag{177}$$

Finally, we define the spin injection efficiency for the spin injector in the nonlocal geometry as

$$P_{j1} = \frac{j_{sN}(0+)}{j}. \tag{178}$$

Only the spin current which gives rise to the spin accumulation at the detector circuit is relevant.

Spin detector: $F2/N$ **junction.** There is no electric current flowing in the spin detector: $j = 0$. The flow of spin is indicated in Fig. 8. Similarly to the $F1/N$ junction, we obtain for the spin currents in the ferromagnet and the contact,

$$j_{sF2}(0) = -\frac{1}{R_{F2}}\mu_{sF2}(0), \tag{179}$$

$$j_{sc2} = \frac{1}{R_{c2}}\left[\mu_{sF2}(0) - \mu_{sN}(d)\right]. \tag{180}$$

The spin current along x is again discontinuous at $x = d$, due to the presence of the spin sink:

$$j_{sN}(d-) = \frac{1}{R_N}\left[-\frac{\mu_{sN}(0)}{\sinh\left(d/L_{sN}\right)} + \mu_{sN}(d)\coth\left(d/L_{sN}\right)\right], \tag{181}$$

$$j_{sN}(d+) = -\frac{1}{R_N}\mu_{sN}(d). \tag{182}$$

$$\tag{183}$$

The continuity for the spin currents at $x = d$ gives

$$j_{sN}(d-) = j_{sN}(d+) + j_{sc2} = j_{sN}(d+) + j_{sF2}(0). \tag{184}$$

Solving the above algebraic system yields

$$\mu_{sN}(0)\frac{1}{\sinh\left(d/L_{sN}\right)} - \mu_{sN}(d)\left[\frac{R_N}{R_{c2}+R_{F2}} + \frac{\exp\left(d/L_{sN}\right)}{\sinh\left(d/L_{sN}\right)}\right] = 0. \tag{185}$$

Let us denote the spin current in the contact as j_{s2}:

$$j_{s2} \equiv j_{sc2} = j_{sF2}(0). \tag{186}$$

We will need to know the value of this current to calculate the emf at the detector circuit. We find that

$$j_{s2} = -\frac{\mu_{sN}(d)}{R_{c2}+R_{F2}}. \tag{187}$$

Spin quasichemical potentials. Equations 176 and 185 form a closed system, allowing us to extract the spin quasichemical potentials at the two contacts:

$$\mu_{sN}(0) = -j\frac{R_N P_{j1}^0}{2}\left[1-\frac{R_N}{2R_2}\left(1+e^{-2d/L_{sN}}\right)\right]\frac{1}{\kappa}, \tag{188}$$

$$\mu_{sN}(d) = -j\frac{R_N P_{j1}^0}{2}\left(\frac{R_{c2}+R_{F2}}{R_2}\right)\frac{e^{-d/L_{sN}}}{\kappa}, \tag{189}$$

where

$$\kappa = \frac{R_N^2}{4R_1R_2}\left[\left(1+2\frac{R_{c1}+R_{F1}}{R_N}\right)\left(1+2\frac{R_{c2}+R_{F2}}{R_N}\right)-e^{-2d/L_{sN}}\right]. \tag{190}$$

Recall that R_1 and R_2 are the total effective resistances of the two junctions; see Eqs. 135. Similarly, P_{j1}^0 and P_{j2}^0 are the spin injection efficiencies of the individual F/N junctions, given by Eq. 87.

> **Problem.** Calculate the spin injection efficiency P_{j1} of the spin injection circuit. What do you get in the limit of $d \gg L_{sN}$? Does the result agree with that for an isolated F/N junction studied earlier?

emf in the detector circuit. Due to the presence of a nonequilibrium spin in the detector circuit, an emf will develop there. We can obtain it as the increase of the quasichemical potential from the far end of the $F2$ to the far right of the N region, as shown in Fig. 7. Since the spin flow in $F2$ is confined to the distance of order L_{sF2} from $z = 0$, the quasichemical potential μ_{sF2} far away from the contact, at $y \to \pm\infty$ (we are mixing the third dimension here!) will be the same as that at the contact itself, $y \approx 0$, but at $z = \infty$:[12]

$$\text{emf} = \mu_N(\infty) - \mu_{F2}(\infty). \tag{192}$$

Since $j = 0$ in the $F2/N$ junction, we can write,

$$\mu_{F2}(0) - \mu_{F2}(\infty) = -P_{\sigma F2}\mu_{sF2}(0), \tag{193}$$

$$\mu_N(d) - \mu_{F2}(0) = -P_{\Sigma 2}\left[\mu_{sN}(d) - \mu_{sF2}(0)\right], \tag{194}$$

$$\mu_N(\infty) - \mu_N(d) = 0. \tag{195}$$

Summing these equations up we get, after substituting for the spin quasichemical potentials Eqs. 179 and 180,

$$\text{emf} = (R_{c2}P_{\Sigma 2} + R_{F2}P_{\sigma 2})j_{s2}. \tag{196}$$

This is just another realization of Silsbee-Johnson spin-charge coupling: An electromotive force develops due to the presence of a spin current in a spin-polarized contact or a ferromagnetic conductor. Due to charge neutrality this emf can be detected as a voltage drop.

[12] Since $j = 0$, we have that $\nabla\mu = -P_\sigma\nabla\mu_s$. Integrating this equation in the (y, z) plane of F_2 we get

$$\mu_{F2}(y,z) = \mu_{F2}(y_0,z_0) - P_{\sigma F2}\left[\mu_{sF2}(y,z) - \mu_{sF2}(y_0,z_0)\right], \tag{191}$$

where (y_0, z_0) is a reference point. Choosing y_0 from the contact region and letting $z_0 \to \infty$, we get that $\mu_{F2}(\infty, z) \approx \mu_{F2}(y_0, \infty)$, since the spin accumulation vanishes both at $y \to \infty$ and $z_0 \to \infty$. The far ends of the $F2$ electrodes ($y_0 \to \infty$) are thus equipotential with the $z = \infty$ points in the contact region.

Substituting for j_{s2} using Eq. 187 and using Eq. 189 for $\mu_{sN}(d)$, the emf can be readily obtained:

$$\mathrm{emf} = j\frac{R_N}{2}P_{j1}^0 P_{j2}^0 \frac{e^{-d/L_{sN}}}{\kappa}. \tag{197}$$

The emf is in general positive for parallel and negative for antiparallel magnetization orientations.
Often what is detected is the *nonlocal resistance*,

$$\mathcal{R}_{\mathrm{nl}} = \frac{\mathrm{emf}}{j} = \frac{R_N}{2}P_{j1}^0 P_{j2}^0 \frac{e^{-d/L_{sN}}}{\kappa}, \tag{198}$$

or the corresponding difference in the nonlocal resistance for parallel and antiparallel orientations of the magnetizations of $F1$ and $F2$:

$$\Delta\mathcal{R}_{\mathrm{nl}} = \mathcal{R}_{\mathrm{nl}}^{\uparrow\uparrow} - \mathcal{R}_{\mathrm{nl}}^{\uparrow\uparrow} = 2|\mathcal{R}_{\mathrm{nl}}|. \tag{199}$$

Tunnel contacts. For tunnel contacts we find $\kappa \approx 1$ and

$$\mathrm{emf} = j\frac{R_N}{2}P_{\Sigma 1}P_{\Sigma 2}e^{-d/L_{sN}}, \tag{200}$$

as one would expect. The factor of "1/2" appears due the geometry of the spin injector: only half of the injected spin current in the $F1/N$ junction flows towards the $F2/N$ junction. The other half flows towards $x \to -\infty$.

Transparent contacts. The most general expression for transparent contacts is the same as Eq. 197, with $R_{c1} = R_{c2} = 0$. In the conductivity mismatch regime, for $R_N \gg R_{F1}, R_{F2}$, the emf simplifies to

$$\mathrm{emf} = 2jR_N P_{\sigma F1}P_{\sigma F2}\left(\frac{R_{F1}R_{F2}}{R_N^2}\right)\frac{e^{-d/L_{sN}}}{1-e^{-2d/L_{sN}}}. \tag{201}$$

The conductivity mismatch limits the spin injection/detection in the nonlocal geometry.

> **Problem.** *Tunnel/transparent contacts.* Calculate emf for the mixed case of tunnel and transparent contacts of the nonlocal spin injection geometry.

References

[1] Aronov A G 1976 *Zh. Eksp. Teor. Fiz. Pisma Red.* **24** 37–39 [JETP Lett. **24**, 32-34 (1976)]

[2] Johnson M and Silsbee R H 1987 *Phys. Rev. B* **35** 4959–4972

[3] Johnson M and Silsbee R H 1988 *Phys. Rev. B* **37** 5312–5325

[4] Žutić I, Fabian J and Das Sarma S 2004 *Rev. Mod. Phys.* **76** 323–410

[5] Fabian J, Matos-Abiague A, Ertler C, Stano P and Žutić I 2007 *Acta Phys. Slov.* **57** 565–907

[6] van Son P C, van Kempen H and Wyder P 1987 *Phys. Rev. Lett.* **58** 2271–2273

[7] Valet T and Fert A 1993 *Phys. Rev. B* **48** 7099–7113

[8] Fert A and Jaffres H 2001 *Phys. Rev. B* **64** 184420

[9] Hershfield S and Zhao H L 1997 *Phys. Rev. B* **56** 3296–3305

[10] Schmidt G, Ferrand D, Molenkamp L W, Filip A T and van Wees B J 2000 *Phys. Rev. B* **62** R4790–R4793

[11] Fabian J, Žutić I and Das Sarma S 2002 *Phys. Rev. B* **66** 165301

[12] Žutić I, Fabian J and Das Sarma S 2002 *Phys. Rev. Lett.* **88** 066603

[13] Jedema F J, Nijboer M S, Filip A T and van Wees B J 2003 *Phys. Rev. B* **67** 085319

[14] Rashba E I 2000 *Phys. Rev. B* **62** R16267–R16270

[15] Rashba E I 2002 *Eur. Phys. J. B* **29** 513–527

[16] Vignale G and D'Amico I 2003 *Solid State Commun.* **127** 829

[17] Jonker B T, Erwin S C, Petrou A and Petukhov A G 2003 *MRS Bull.* **28** 740–748

[18] Takahashi S and Maekawa S 2003 *Phys. Rev. B* **67** 052409

[19] Fert A, George J M, Jaffres H and Mattana R 2007 *IEEE Trans. Electronic Devices* **54** 921–932

[20] Žutić I, Fabian J and Erwin S C 2006 *Phys. Rev. Lett.* **97** 026602

C 2 Anomalous Hall Effect

Yuriy Mokrousov
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Theories of the AHE **5**

3 Semiclassical approach to transport **8**

4 Intrinsic anomalous Hall conductivity **10**

5 Theory meets experiment **12**

6 Boltzmann equation and semiclassical theory of side jump **15**

7 AHE within the semiclassical Boltzmann equation **19**

8 Separating different contributions to the AHE **25**

9 Acknowledgments **29**

1 Introduction

The anomalous Hall effect (AHE) is one of the oldest and most famous transport phenomena in magnetic materials. Long ago it was discovered by Edwin Hall [1, 2] that in ferromagnets the resistivity of the sample in the direction perpendicular to external electric **E** and magnetic **H** field acquires an additional contribution due to the magnetization of the sample [3, 4]:

$$\rho_{xy} = \rho_{OH} + \rho_{AH} = R_0 H + 4\pi R_S M, \tag{1}$$

where R_0 is the usual Hall coefficient, R_S is the anomalous (spontaneous) Hall coefficient, ρ_{OH} is the ordinary Hall resistivity and ρ_{AH} is the anomalous Hall resistivity. (In the course of this manuscript we will be equivalently using ρ_{AH} and ρ_{xy} for the anomalous Hall resistivity). Graphically the anomalous contribution to ρ_{xy} is very easy to see: upon increasing the magnetic field H the magnetization of the sample eventually reaches the saturation value M_S and the transverse resistivity becomes linear in the magnetic field. The term $4\pi R_S M_S$ is then easy to determine as an offset along the ρ_{xy} axis in the $\rho_{xy}(H)$ dependence. In Fig. 1 it is illustrated how in such a way the anomalous Hall coefficient can be determined as a function of temperature, sample thickness, sample's direction of magnetization etc.

The revival of interest in the anomalous Hall effect during past years can be attributed to several reasons. From the practical point of view, utilizing the AHE for the purposes of modern spintronics would enable us with a unique tool for generating and controlling spin-polarized currents in complex magnetic nanoscale systems. On the other hand, deeper understanding of the mechanisms behind the AHE is necessary for further advances in young and promising field of spin Hall effect related phenomena, which are based on propagation of pure spin currents without accompanying charge transport. Moreover, recent progress in understanding the AHE from the point of view of Berry phases and topological structure of the crystal electronic bands led to the fact that many abstract quantities and objects in this part of quantum physics became measurable and feasible to tackle experimentally.

The mysterious beauty of the anomalous Hall effect lies in the fact that even without an applied magnetic field a non-vanishing resistivity ρ_{xy} can be measured. Indeed, everywhere further on in this lecture we will assume that the only external field present in the system is the electric field **E** along the x-axis of the system, which generates a longitudinal flow of electrons J_x and a transverse current J_y. Expression (1) implicitly assumes that the anomalous contribution is proportional to magnetization M, and it is used as an experimental tool to measure magnetization as a function of temperature in e.g. studies on ferromagnetic semiconductors with dilute magnetic impurities, which are promising materials for applications in spintronics. Only recently it was demonstrated experimentally and theoretically that ρ_{xy} can be highly non-linear as a function of external parameters, including the magnetization M.

The finite anomalous Hall resistivity, arising due to the spin-orbit coupling (SOC), can be observed only in materials with broken time reversal symmetry, of which the widest class constitute ferromagnets. When the time reversal symmetry is present in the system, the anomalous contribution to the conductivity cancels out when integrated over degenerate bands. In ferromagnets, the time reversal symmetry is broken due to the presence of a finite intra-atomic exchange field, which causes splitting between the bands of different spin.

Upon moving in the longitudinal electric field E_x electrons experience scattering of different physical origin responsible for "anomalous" contribution to the transverse velocity. In a simple picture, due to the spin-orbit coupling, this scattering depends on the spin of an incoming electron which results in the anomalous velocity of opposite sign for electrons of opposite spin

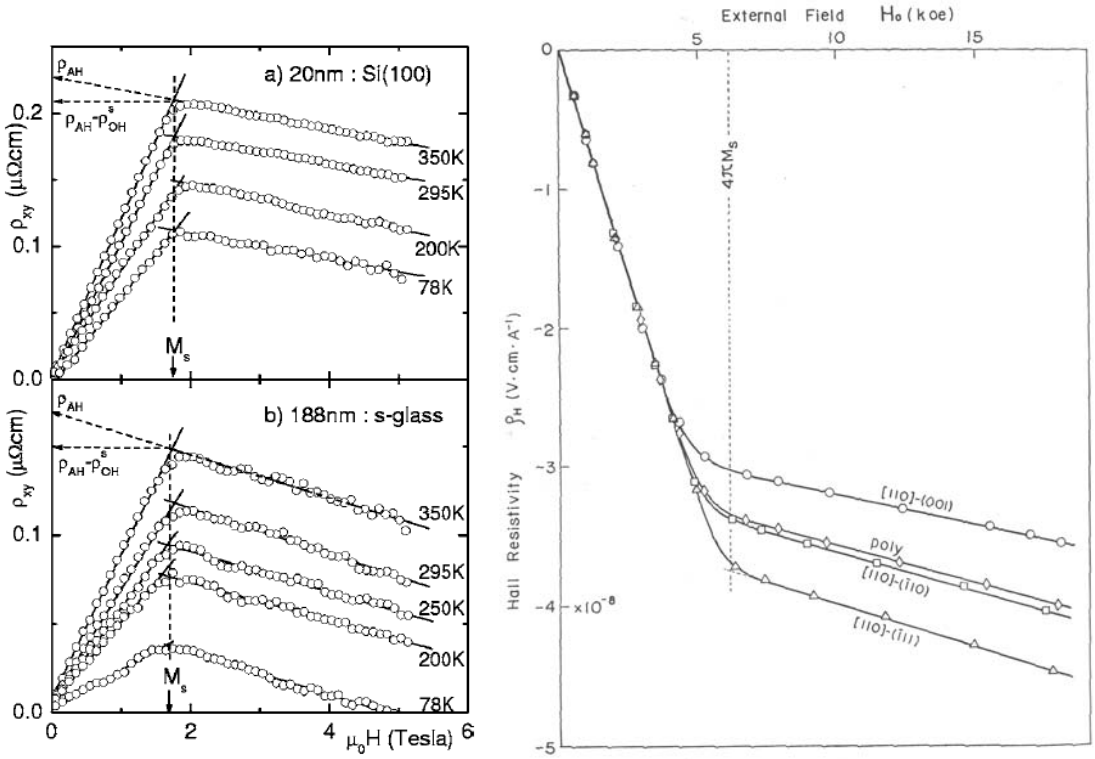


Fig. 1: *Left: isothermal Hall resistivity of a 20 nm and 188 nm Co films in magnetic fields applied perpendicularly to the film plane. The determination of the anomalous Hall resistivity ρ_{AH} and the ordinary Hall resistivity at the field of magnetization saturation ρ_{OH} is indicated. The data are taken from [5]. Right: the Hall resistivity plotted against the external field at room temperature for different directions of the magnetization in single crystals of Ni, as well as in polycrystal samples. The data are taken from [6].*

(Fig. 2). In a non-magnetic material with an equal number of spin-up and spin-down electrons this results in a zero transverse anomalous (charge) conductivity, but leads to the so-called spin Hall effect [7, 8]. In the spin Hall effect, spin-up transverse current propagates to the "left", while the spin-down transverse current propagates to the "right", which leads to a non-zero spin accumulation of opposite sign at the opposite edges of the sample, which can be measured experimentally.

In a ferromagnet, due to a finite exchange splitting of the electronic bands the number of spin-up electrons is different from the number of spin-down electrons. Integrated over the occupied states, the anomalous contribution to the transverse electron velocity is not cancelled and the transverse current in the AHE is not zero. Following this line of thought is seems logical to assume that the anomalous Hall current would be directly proportional to the magnetization M of the sample, or, equivalently, to the difference between the number of spin-up and spin-down carreers. Indeed, almost all conventional theories for the AHE which we will review in the next section derive equation (1), as they are based on perturbative expansion of the spin-orbit coupling λ, which is responsible for asymmetric scattering, i.e. $R_S \sim \lambda$. It is not always the case, however, as it was shown recently that certain points in the band structure can lead to significant singularity-like contributions to the anomalous Hall conductivity (AHC), when the Fermi energy is varied in their vicinity [9].

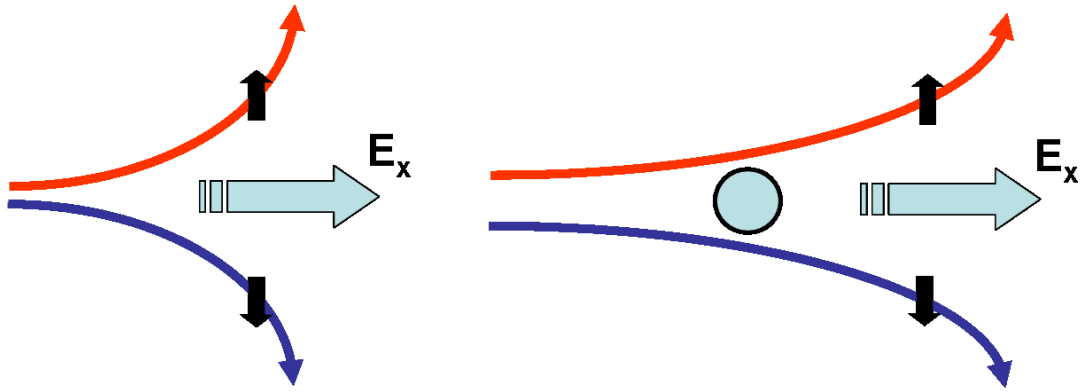


Fig. 2: *Spin-asymmetric scattering picture of the anomalous and spin Hall effects. Upon moving in the longitudinal electric field E_x electrons experience "intrinsic" (left) or impurity (right) scattering which leads to the "anomalous" contribution to the transverse velocity opposite for electrons of opposite spin.*

It has been a long time since the first attempts to explain the AHE were made. The history of theoretical understanding of the AHE is particularly messy, with different models and ways of solving them providing totally different results, correcting, disregarding and contradicting each other. The complete and consistent picture of the AHE emerges quite slowly with years and debates on certain issues are still going on. Nevertheless, in this manuscript we would like to provide as little confusion as possible by considering the AHE in terms of, perhaps, the most appealing language of semiclassical dynamics of wavepackets and semiclassical Boltzmann equation. Semiclassical philosophy and formalism of the anomalous transport in solids has recently reached a degree of self-consistency and consistency with other more rigorous quantum mechanically approaches [10] which makes it possible to use it as a common foundation for considering most of the mechanisms of the AHE.

Among various novel issues in the area of the AHE which we do not consider in this manuscript we would like to mention particularly the many-body effects in the AHE and Berry curvature contribution to them [11], as well as optically induced AHE [12]. Recently, it was demonstrated that the AHE emerges even in paramagnetic materials in applied external magnetic field which is able to polarize the electron spins [13]. In this case the semiclassical theory of the AHE should be modified accordingly, as it is difficult to make a clear separation between the anomalous and ordinary Hall contributions. While in the manuscript we restrict our discussion of the AHE to the case of ferromagnets, another interesting and important aspect of current investigations in the field is the appearance and behavior of AHE in systems with non-collinear spins. In a situation where a spin hops along the loop connecting three non-collinear spins $\mathbf{S}_i$, $\mathbf{S}_j$ and $\mathbf{S}_k$ the total phase acquired by the electron along this loop will be equal to the half of the solid angle Ω spanned by these three spins on a unit sphere. For a small Ω, this phase will be proportional to the scalar spin chirality $\xi_{ijk} = \mathbf{S}_i \cdot (\mathbf{S}_j \times \mathbf{S}_k)$. Appearance of this phase can act as a gauge, i.e. fictitious magnetic field, and it is expected to produce an effect when its uniform component is finite [14]. The issue of the AHE for a periodic non-coplanar spin configuration on the Kagome lattice was studied by Ohgushi *et al.* [15], who pointed out that the spin chirality produces a gauge field in the reciprocal space leading to the AHE. Recently, the relationship between the spin chiralities in real and reciprocal spaces was established by Onoda *et al.* [16].

2 Theories of the AHE

The first attempt to relate the AHE and the spin-orbit interaction was made by Karplus and Luttinger (KL) in 1954 [17]. They argued that due to relativistic corrections the effective Hamiltonian of an electron in a solid acquires an additional term due to spin-orbit interaction:

$$\hat{H}_{so} = -\frac{1}{4m^2c^2}\hat{\sigma}\cdot(\mathbf{p}\times\nabla V), \tag{2}$$

where $V(\mathbf{r})$ is the one-electron effective lattice potential. This additional part of the Hamiltonian modifies the periodic parts $u_n(\mathbf{k})$ of the Bloch wavefunctions $\psi_n(\mathbf{k})$ in a characteristic way. Applying an external electric field results in an additional contribution to the Hamiltonian $\Delta\hat{H} = e\mathbf{E}\cdot\mathbf{r}$, which couples states of different bands $u_n(\mathbf{k})$ and $u_m(\mathbf{k})$, $n \neq m$. This mixing eventually leads to the linear in the electric field "anomalous" contribution to the transverse velocity of electrons. Upon treating the electric field as a perturbation Karplus and Luttinger correctly predicted that the ***intrinsic***, i.e., impurity-free band structure contribution to the anomalous Hall resistivity is proportional to the diagonal part of the resistivity tensor squared ρ_{xx}^2, which agreed with many experiments available at the time. By arguing that the average value of the Hamiltonian (2) for a magnetic electron would be proportional to the magnetization M, they stated the proportionality of the Hall current to the sample magnetization.

Theory of the AHE suggested by Karplus and Luttinger was a clear success at the time. However, this theory was still very far from the "ultimate" theory: the influence of the impurities was not included and the expressions for the current were not *gauge invariant*, i.e. dependent on unphysical phase of the wavefunctions. These were serious obstacles on the way to practical applications of the KL theory to real materials. Smit was the first to try to develop a gauge invariant theory of the AHE [18, 19]. He approached the problem from a different side via solving a Boltzmann-like equation for the wavepackets (for a definition of a wavepacket see section 3). In this way he was able to treat semiclassically the influence of the impurities on the dynamics of the wavepackets in the phase space. By doing so he discovered the so-called ***skew scattering*** mechanism of the AHE. This mechanism can be attributed to the asymmetry in scattering to the "right" and to the "left", or, in other words, $\mathbf{k}\to\mathbf{k}'$ and $\mathbf{k}'\to\mathbf{k}$ scattering. It arises already in the second order of the Born expansion of the scattering rate if the impurity Hamiltonian includes a $\mathbf{k}'\leftrightarrow\mathbf{k}$ asymmetric part, in particular, spin-orbit coupling. In contrast to the KL predictions, Smit's anomalous Hall resistivity is proportional to ρ_{xx}. His estimated magnitude of the AHE appeared to be four orders of magnitude smaller than the experimentally observed, which was, as realized later, due to neglecting the intrinsic spin-orbit coupling in the crystal. Interestingly, investigating the dynamics of a wavepacket in an external electric field Smit also found an intrinsic anomalous velocity of electrons which he wrongly argued is exactly canceled out.

Investigating the dynamics of wavepackets Smit discovered that the center of mass of a propagating wavepacket $\mathbf{r}_c$ is changed according to the following law (compare to equation (54) in section 6):

$$\mathbf{r}_c(\mathbf{k},t) = \frac{\partial\varepsilon(\mathbf{k})}{\partial\mathbf{k}} + \mathbf{A}(\mathbf{k}), \tag{3}$$

where $\mathbf{A}(\mathbf{k})$ is the so-called Berry connection:

$$\mathbf{A}(\mathbf{k}) = \left\langle u(\mathbf{k}) \middle| i\frac{\partial}{\partial\mathbf{k}}u(\mathbf{k}) \right\rangle. \tag{4}$$

According to Smit, when the electric field accelerates the wavepacket, vector $\mathbf{k}$ changes, which in turn changes the Berry connection $\mathbf{A}(\mathbf{k})$, the wavepacket becomes deformed and an additional charge transport in the transverse to the electric field direction emerges. Further on, Smit pointed out that the wavepackets cannot be constantly accelerated by the electric field and it is scattering on impurities which stabilizes the transport and produces on average an exactly opposite change in $\mathbf{k}$ in the steady state. Smit suggested that this is realized via coordinate shifts at scatterings. These coordinate shifts at scatterings, first introduced by Smit [19] play an important role in the modern theory of the AHE (see sections 6 and 7). They were named ***side jumps*** by Berger, who studied them in more detail [20, 21, 22].

Cancellation arguments of Smit fail for two reasons. The Berry connection term $\mathbf{A}(\mathbf{k})$ entering the wavepacket dynamics is not gauge invariant, therefore, applying classical balance arguments to it is not justified. Secondly, the change in the kinetic energy of the particles during side jumps leads to an additional, so-called anomalous correction to the distribution function (see section 7) which can contribute to the Hall current. Nevertheless, Smit's works were at the beginning of the modern semiclassical approach. He made a first attempt to understand the anomalous Hall effect in classical terms, such as corrections to velocity, asymmetric scattering, jumps at impurities – ideas, which are currently incorporated in the modern theory.

In 1958 Luttinger provided a rigorous quantum mechanical description of the AHE [23] which was further generalized to the regime of an ac external field by Lyo and Holstein [24]. Luttinger considered the equation for the evolution of the density matrix $\hat{\rho}$:

$$\frac{\partial \hat{\rho}}{\partial t} = i[\hat{\rho}, \hat{H}], \tag{5}$$

where Hamiltonian $\hat{H}$ includes both electric field and impurity potential:

$$\hat{H} = \hat{H}_0 + V(\mathbf{r}) + eE_x\hat{x} \tag{6}$$

In the stationary state

$$\frac{\partial \hat{\rho}}{\partial t} = 0. \tag{7}$$

For the density matrix which satisfies this equation, the transverse electric current can be found via

$$J_y = -e\mathrm{Tr}[\hat{v}_y\hat{\rho}]. \tag{8}$$

The velocity operator $\hat{v}_y$ is diagonal in $\mathbf{k}$-space, therefore, to calculate the current, only $\mathbf{k}$-diagonal part of $\hat{\rho}$ is in principle needed. However, the velocity operator can have off-diagonal elements in the band index due to the presence of the electric field. From Eqs. (5)-(8) Luttinger derived an analog of the Boltzmann equation, the quantum Boltzmann equation, which operates in terms of the $\mathbf{k}$-diagonal part of the density matrix. Using this equation Luttinger managed to decompose the part of the density matrix linear in electric field $\delta\hat{\rho}$ into series with respect to the strength of the impurity potential V:

$$\delta\hat{\rho} = V^{-2}\delta\hat{\rho}^{(-2)} + V^{-1}\delta\hat{\rho}^{(-1)} + V^{0}\delta\hat{\rho}^{(0)} + ... \tag{9}$$

Luttinger found that the $\delta\hat{\rho}^{(-2)}$ term is diagonal in band index and does not contribute to the Hall current but makes a major contribution to the longitudinal current. The next term, $\delta\hat{\rho}^{(-1)}$, is still diagonal in band index and contains an antisymmetric contribution is the transverse direction

and contributes to the transverse conductivity which behaves as $\frac{1}{n}$ as a function of the impurity concentration n. This term leads to a conventional skew scattering contribution. The zeroth order term can be distinctively separated into four terms:

$$\delta\hat{\rho}^{(0)} = \delta\hat{\rho}_{int} + \delta\hat{\rho}_{sj} + \delta\hat{\rho}_{ad} + \delta\hat{\rho}_{sk}, \tag{10}$$

of which first two are purely off-diagonal and the other two are diagonal in band index, and all four contribute to the Hall current. Since the terms are of zeroth power in V and the velocity operator does not depend on V, the resulting conductivity due to these contributions for a Gaussian correlated disorder becomes independent on the impurity concentration. These four terms lead to intrinsic, side jump, anomalous distribution and intrinsic skew scattering channels of transverse conductivity, respectively, and we will analyze them later.

The approach developed by Luttinger is rigorous quantum-mechanically, but it remains cumbersome and lacks transparency which can be achieved only within the semiclassical picture. In 1959 Adams and Blount tried to develop a semiclassical theory of the AHE which was based on the concept of non-commuting coordinates [25]. Adams and Blount noticed that as far as the electric field couples the states of different bands n and m a part of a wavepacket starts fast oscillations with the frequency $\varepsilon_n(\mathbf{k}) - \varepsilon_m(\mathbf{k})$, and technically a wavepacket looses its classical meaning. The remedy lies in constructing a new set of Bloch orbitals $u'_n(\mathbf{k})$ (to linear order in electric field E_x):

$$|u'_n(\mathbf{k})\rangle = |u_n(\mathbf{k})\rangle + ieE_x \sum_{m\neq n} \frac{\left\langle u_m(\mathbf{k}) \left| \frac{\partial u_n(\mathbf{k})}{\partial k_x} \right\rangle \right.}{\varepsilon_n(\mathbf{k}) - \varepsilon_m(\mathbf{k})} |u_m(\mathbf{k})\rangle. \tag{11}$$

Using this new set of orbitals one can prepare a wavepacket which does not split into parts of different frequencies under applied electric field. It is straightforward then to calculate the transverse velocity, which reads:

$$v_y = \frac{\partial\varepsilon_n(\mathbf{k})}{\partial k_y} - eE_x\Omega_{n,z}(\mathbf{k}), \tag{12}$$

where $\Omega_{n,z}(\mathbf{k})$ is the z-component of the so-called Berry curvature $\mathbf{\Omega}_n(\mathbf{k})$ of the n'th band given by expression:

$$\Omega_{n,z}(\mathbf{k}) = \mathrm{Im}\left(\left\langle \frac{\partial u_n(\mathbf{k})}{\partial k_y} \middle| \frac{\partial u_n(\mathbf{k})}{\partial k_x} \right\rangle - \left\langle \frac{\partial u_n(\mathbf{k})}{\partial k_x} \middle| \frac{\partial u_n(\mathbf{k})}{\partial k_y} \right\rangle \right). \tag{13}$$

(A more general semiclassical derivation of the expression for the Berry curvature-driven anomalous velocity will be given in section 3). Generally speaking, it appears that instead of working with the mixing of different bands due to the electric field, one can assume that the Hamiltonian is the original diagonal Hamiltonian, only the coordinate operators should be changed to:

$$\hat{\mathbf{r}} \longrightarrow i\frac{\partial}{\partial\mathbf{k}} + \mathbf{A}(\mathbf{k}), \tag{14}$$

where $\mathbf{A}(\mathbf{k})$ is the Berry connection given by (4), $\mathbf{\Omega}(\mathbf{k}) = \nabla \times \mathbf{A}(\mathbf{k})$. The anomalous velocity is due to the fact that these operators are not commuting. The comprehensive theory of non-commuting coordinates in Bloch bands is presented in [26]. Adams and Blount also pointed out that their approach can be applied to the problem of impurity scattering given that a moving wavepacket experiences only a small gradient. Their theory was applied by Chazalviel [27],

Noziéres and Lewiner [28], Berger [20, 21, 22] and by Lyo and Holstein [24] to the AHE in III-V n-type semiconductors.

3 Semiclassical approach to transport

Quantitative treatment of transport phenomena in real materials with standard Green's function techniques is normally quite difficult and at times lacks transparency. Therefore, the theories of the AHE normally employ simple Hamiltonians and normally do not go beyond one-electron mean-field picture. The complexity of the task is amplified by the fact the the anomalous Hall conductivity is normally much smaller than the diagonal part of the conductivity tensor. Within the perturbative analysis in the weak disorder limit the AHE-relevant terms appear only in subdominant terms of higher powers of small parameters. Many techniques fail in those orders and it becomes difficult to fish out and group the terms of the same magnitude. On the other hand, using such rigorous techniques as Kubo or Keldysh normally results in operating with non-gauge invariant objects until the very end of the calculation. This in turn complicates the analysis of the physics of the relevant terms and microscopic origins of the AHE.
An alternative and very powerful approach to electron transport properties presents the classical Boltzmann equation. In sufficiently clean materials the motion of electrons can indeed be considered as motion of localized in real and reciprocal space wavepackets rather then delocalized in real space Bloch functions. In the dilute disorder limit such a wavepacket can travel in a solid for a long time as a conventional classical particle. Moreover, it is possible to describe the motion of a wavepacket in external fields from this classical point of view and describe the electron dynamics in clear and concise terms.
A general Hamiltonian of a slowly perturbed crystal reads:

$$H(\hat{\mathbf{r}}, \hat{\mathbf{p}}; \beta_1(\hat{\mathbf{r}}, t), ..., \beta_r(\hat{\mathbf{r}}, t)), \tag{15}$$

where $\{\beta_i(\hat{\mathbf{r}}, t)\}$ are the modulation functions characterizing the perturbations. They can represent deformations, gauge potentials, electromagnetic fields or slowly varying impurity potentials.
Let us take now a wavepacket centered at $\mathbf{r}_c$ in real space with its spread small compared to the length scale of the perturbations. In this case the approximate Hamiltonian which the wavepacket feels can be obtained by linearizing the Hamiltonian around $\mathbf{r}_c$:

$$\hat{H} = \hat{H}_c + \sum (\hat{\mathbf{r}} - \hat{\mathbf{r}}_c) \nabla_{\mathbf{r}_c} \beta_i(\hat{\mathbf{r}}_c, t), \tag{16}$$

where $\hat{H}_c = H(\hat{\mathbf{r}}, \hat{\mathbf{p}}; \{\beta_i(\hat{\mathbf{r}}_c, t)\})$ is the so-called "local" Hamiltonian. The terms in the right hand side of Eq. (16) following the local Hamiltonian are normally small and are usually treated perturbatively. The local Hamiltonian has the periodicity of the lattice and has an energy spectrum of bands with Bloch eigenstates satisfying:

$$\hat{H}_c(\hat{\mathbf{r}}_c, t) \, |\psi_{\mathbf{k}}(\mathbf{r}_c, t)\rangle = \varepsilon_c(\mathbf{r}_c, \mathbf{k}, t) |\psi_{\mathbf{k}}(\mathbf{r}_c, t)\rangle, \tag{17}$$

where we omit the band index for simplicity. Both the wavepacket center $\mathbf{r}_c$ and time t enter the Bloch states and energies parametrically which results in Berry phase effects in the equation of motion.
The Bloch eigenstates form a convenient basis to construct a wavepacket Ψ:

$$|\Psi\rangle = \int a(\mathbf{k},t)|\psi_{\mathbf{k}}(\mathbf{r}_c,t)\rangle\, d\mathbf{k}, \tag{18}$$

where $a(\mathbf{k},t)$ is the amplitude with the normalization condition

$$\int |a(\mathbf{k},t)|^2\, d\mathbf{k} = \langle\Psi|\Psi\rangle = 1. \tag{19}$$

In order that Ψ is really a "packet", it is required that $|a(\mathbf{k},t)|^2$ is peaked in the Brillouin zone around a certain vector $\mathbf{k}_c$:

$$\mathbf{k}_c = \int \mathbf{k}\cdot|a(\mathbf{k},t)|^2\, d\mathbf{k}. \tag{20}$$

Then packet's center of mass in real space lies at $\mathbf{r}_c$:

$$\mathbf{r}_c = \langle\Psi|\hat{\mathbf{r}}|\Psi\rangle. \tag{21}$$

The equations which describe how the center of the wavepacket $(\mathbf{r}_c,\mathbf{k}_c)$ moves in $(\mathbf{r},\mathbf{k})$-space when the perturbations β are present describe the electron dynamics in the system. In the simplest case of free electrons in the external electric field $\mathbf{E}$ such equations are easy to derive. In this case the Hamiltonian of the system is $\hat{H} = \frac{\hat{\mathbf{k}}^2}{2} + e\mathbf{E}\cdot\hat{\mathbf{r}}$, so the local Hamiltonian is the Hamiltonian of free electrons $\hat{H}_c = \frac{\hat{\mathbf{k}}^2}{2}$ and the basis used for the expansion of wavepackets is the plane-wave basis $\{e^{i\mathbf{k}\cdot\mathbf{r}-i\frac{k^2}{2m}t}\}$. It is a textbook exercise then to derive the semiclassical equations of motion for the wavepackets:

$$\dot{\mathbf{r}}_c = \frac{\mathbf{k}_c}{m} \quad \text{and} \quad \dot{\mathbf{k}}_c = -e\mathbf{E}. \tag{22}$$

In the general case, equations for the motion of the mean position $\mathbf{r}_c$ and crystal momentum $\mathbf{k}_c$ can be derived from the Schrödinger equation for the wavepacket, which is most conveniently obtained using a time-dependent variational principle with the Lagrangian given by [29]:

$$L = \left\langle \Psi \left| i\frac{d}{dt} - \hat{H} \right| \Psi \right\rangle, \tag{23}$$

where $\frac{d}{dt}$ means the derivative with respect to the time dependence of the wave function explicitly or implicitly through $\mathbf{r}_c$ and $\mathbf{k}_c$. Taking into account the locality of the wavepacket, it can be shown that $L \approx L(\mathbf{r}_c,\dot{\mathbf{r}}_c,\mathbf{k}_c,\dot{\mathbf{k}}_c,t)$ reads:

$$L = -\mathcal{E} + \mathbf{k}_c\cdot\dot{\mathbf{r}}_c + \dot{\mathbf{k}}_c\cdot\left\langle u\left| i\frac{\partial u}{\partial \mathbf{k}_c}\right\rangle + \dot{\mathbf{r}}_c\cdot\left\langle u\left| i\frac{\partial u}{\partial \mathbf{r}_c}\right\rangle + \left\langle u\left| i\frac{\partial u}{\partial t}\right\rangle\right.\right.\right., \tag{24}$$

where energy $\mathcal{E}$ is the $\varepsilon_c(\mathbf{r}_c,\mathbf{k}_c,t)$ energy corrected by a gradient correction term $\Delta\mathcal{E}$ [29]. In case of a magnetic field $\mathbf{B}$ present in the system the energy correction $\Delta\mathcal{E}$ acquires a form of the magnetization energy $\frac{e}{2m}\mathbf{B}(\mathbf{r}_c)\cdot\mathbf{L}(\mathbf{k}_c)$, where $\mathbf{L}$ is the angular momentum of the wavepacket around its center of mass. The first two terms in this Lagrangian are the same as those for a classical particle with a classical Hamiltonian corresponding to energy $\mathcal{E}$. The last three terms can be grouped into a single term $\langle u|i\frac{du}{dt}\rangle$, which turns out to be the net rate of change of Berry phase of a wavepacket motion within the band. This term is geometric, meaning that its contribution to the action depends on the trajectory in the phase space but not on the rate of the

motion along this trajectory. It is truly remarkable that, as we will see, the intrinsic anomalous Hall conductivity stems from the latter term and is a purely geometrical property.
The Lagrangian formulation provides a fully gauge invariant approach to the study of the wavepacket dynamics. When external magnetic and electric fields $\mathbf{B}$ and $\mathbf{E}$ are present in the system, the wavepacket equations of motion can be derived variationally from the Lagrangian (24):

$$\dot{\mathbf{r}}_c = \frac{\partial \mathcal{E}}{\partial \mathbf{k}_c} - \dot{\mathbf{k}}_c \times \mathbf{\Omega} \quad \text{and} \quad \dot{\mathbf{k}}_c = -e\mathbf{E} - e\mathbf{r}_c \times \mathbf{B}, \tag{25}$$

and with electric field only they reduce to

$$\dot{\mathbf{r}}_c = \frac{\partial \varepsilon}{\partial \mathbf{k}_c} - \dot{\mathbf{k}}_c \times \mathbf{\Omega} \quad \text{and} \quad \dot{\mathbf{k}}_c = -e\mathbf{E}, \tag{26}$$

where $\mathbf{\Omega}$ is the Berry curvature of the Bloch state defined in section 2:

$$\mathbf{\Omega}(\mathbf{k}) = -\mathrm{Im} \left\langle \frac{\partial u}{\partial \mathbf{k}} \right| \times \left| \frac{\partial u}{\partial \mathbf{k}} \right\rangle . \tag{27}$$

Comparing equations (22) and (26) we observe that the difference is the $\dot{\mathbf{k}}_c \times \mathbf{\Omega}$ term to the velocity, also known as the *anomalous velocity* . We will see that the anomalous velocity gives rise to the intrinsic anomalous Hall conductivity. Wavepacket equations (26) show that the Berry curvature $\mathbf{\Omega}$ acts as an unusual magnetic field acting in the momentum space. Unlike the "real" magnetic field of classical electrodynamics, however, the Berry curvature originates from a point source, or, so-called magnetic monopole. Its analog in Bloch bands are the points of band crossings in the band structure [9, 26, 30].

4 Intrinsic anomalous Hall conductivity

The semiclassical dynamics coupled with the Boltzmann equation provides a comprehensive theory of transport phenomena in solid state physics [31]. In this theory the Bloch electrons are considered as classical wavepackets whose distribution in the $(\mathbf{r}, \mathbf{k})$ space is described by a *distribution function* $f(\mathbf{r}, \mathbf{k}, t)$. The normalization of the distribution function is fixed by the requirement that $f(\mathbf{r}, \mathbf{k}, t)\, d\mathbf{r}\, d\mathbf{k}/2\pi$ is the number of electrons in the phase space volume $d\mathbf{r}\, d\mathbf{k}$ around the point $(\mathbf{r}, \mathbf{k})$.
There are two sources that make f change in time. Firstly, it is changing of $\mathbf{r}$ and $\mathbf{k}$ according to semiclassical equations of motion from the previous section, and, secondly, abrupt jumps in $\mathbf{k}$ due to collision of wavepacket with impurities, phonons and other electrons. It can be seen already from (22) that under the action of electric field the wavepacket will accelerate indefinitely. This never happens in metals because of the second scattering mechanism. Microscopically, it is impossible to trace trajectories of all wavepackets and the natural language to describe such a system is provided by the semiclassical Boltzmann equation. In the relaxation time approximation [31] the Boltzmann equation for the distribution in the steady state reads:

$$\dot{\mathbf{r}} \cdot \frac{\partial f}{\partial \mathbf{r}} + \dot{\mathbf{k}} \cdot \frac{\partial f}{\partial \mathbf{k}} = \frac{f_0 - f}{\tau(\mathbf{k})}, \tag{28}$$

where $\tau(\mathbf{k})$ is the transport relaxation time of the Bloch electron and $f_0 = 1/(e^{(\varepsilon(\mathbf{k})-\mu)/k_B T} + 1)$ is the equilibrium distribution function. The problem of calculating the relaxation time τ has to

be considered separately and we will come to this issue later, but for the purposes of this section it is enough to assume it to be constant. The solution of the Boltzmann equation depends on τ only for those states that are near the Fermi surface within an energy width of $k_B T$.
In the electric field $\mathbf{E}$ we can substitute expressions (26) for the $\dot{\mathbf{r}}$ and $\dot{\mathbf{k}}$ into Eq. (28) and solve it to the first order in the electric field, which yields:

$$f = f_0 + \tau e \cdot \mathbf{E} \cdot \frac{\partial \varepsilon}{\partial \mathbf{k}} \frac{\partial f_0}{\partial \varepsilon}, \tag{29}$$

where we have assumed that the chemical potential and temperature are constant in space. The $\frac{\partial f_0}{\partial \varepsilon}$ factor enables that the electric field modifies the distribution only around the Fermi surface. Then the electron current density can be decomposed in the following way:

$$\mathbf{J} = -\frac{e}{(2\pi)^3} \int \dot{\mathbf{r}} f \, d\mathbf{k} = \mathbf{J}^{\Omega} + \mathbf{J}^{\tau}, \tag{30}$$

where the diffusion current due to disorder scattering is given by

$$\mathbf{J}^{\tau} = -\frac{e^2 \tau}{(2\pi)^3} \int \frac{\partial \varepsilon}{\partial \mathbf{k}} \left(\mathbf{E} \cdot \frac{\partial \varepsilon}{\partial \mathbf{k}} \right) \frac{\partial f_0}{\partial \varepsilon} \, d\mathbf{k}, \tag{31}$$

while $\mathbf{J}^{\Omega}$ is the anomalous Hall current:

$$\mathbf{J}^{\Omega} = -\frac{e^2}{(2\pi)^3} \, \mathbf{E} \times \int f_0(\varepsilon(\mathbf{k})) \, \mathbf{\Omega}(\mathbf{k}) \; d\mathbf{k}. \tag{32}$$

The $\mathbf{J}^{\tau}$ current gives a symmetric conductivity tensor and depends only on the properties of the system near the Fermi energy. The $\mathbf{J}^{\Omega}$ current comes from the anomalous velocity due to the Berry curvature and gives a Hall conductivity tensor which is antisymmetric. For systems with time-reversal symmetry the Berry curvature is antisymmetric in $\mathbf{k}$, while f_0 is $\mathbf{k}$-symmetric. This means that for materials with time-reversal symmetry, e.g., non-magnetic, the anomalous Hall current averages to zero. In magnetic materials in which the time-reversal symmetry is broken, e.g., ferromagnets, the anomalous Hall current is non-zero. In this case the spin-orbit coupling has to be taken into account, as it leads to the modifications in the spatial wave functions due to time-reversal asymmetry in spin [26].
Expression (32) allows us to write down the antisymmetric part of the conductivity (already in the multi-band case):

$$\sigma_{xy} = -\frac{e^2}{(2\pi)^3} \sum_n \int_{BZ} f_0^n(\mathbf{k}) \, \Omega_{n,z}(\mathbf{k}) \, d\mathbf{k}, \tag{33}$$

where $\Omega_{n,z}(\mathbf{k})$ is the z-component of the Berry curvature for n'th band given by Eq. (13) from section 2 and $f_0^n(\mathbf{k})$ is $f_0(\varepsilon_n(\mathbf{k}))$. Analogous expression for the transverse conductivity in case of insulators was derived already in 1982 in the milestone paper by Thouless, Kohmoto, Nightingale and den Nijs [32]. Their derivation was based on the Kubo formula for the conductivity [33]:

$$\sigma_{xy}(\omega) = i \sum_{\mathbf{k},n,m} \frac{\langle m\mathbf{k}|J_x|n\mathbf{k}\rangle\langle n\mathbf{k}|J_y|m\mathbf{k}\rangle}{\varepsilon_n(\mathbf{k}) - \varepsilon_m(\mathbf{k})} \cdot \frac{f_0^n(\mathbf{k}) - f_0^m(\mathbf{k})}{\omega + i\delta + \varepsilon_m(\mathbf{k}) - \varepsilon_n(\mathbf{k})}, \tag{34}$$

where $i\delta$ enables the analyticity of the response function. In the static limit of $\omega = 0$ and setting $i\delta$ to zero, we arrive at the following expression:

$$\sigma_{xy} = 2\mathrm{Im} \sum_{\mathbf{k},n\neq m} \frac{\langle m\mathbf{k}|J_x|n\mathbf{k}\rangle\langle n\mathbf{k}|J_y|m\mathbf{k}\rangle}{(\varepsilon_n(\mathbf{k}) - \varepsilon_m(\mathbf{k}))^2} f_0^n(\mathbf{k}), \tag{35}$$

where the current operators are defined as $J_\mu = \sum_{\mathbf{k}} c_{\mathbf{k}}^+ \frac{\partial h(\mathbf{k})}{\partial \mathbf{k}_\mu} c_{\mathbf{k}}$ with $c_{\mathbf{k}}$ as creation operators and Hamiltonian as $\hat{H} = \sum_{\mathbf{k}} c_{\mathbf{k}}^+ h(\mathbf{k}) c_{\mathbf{k}}$. The equivalent expression for the Berry curvature therefore reads:

$$\Omega_{n,z}(\mathbf{k}) = -2\mathrm{Im} \sum_{m\neq n} \frac{\langle m\mathbf{k}|J_x|n\mathbf{k}\rangle\langle n\mathbf{k}|J_y|m\mathbf{k}\rangle}{(\varepsilon_n(\mathbf{k}) - \varepsilon_m(\mathbf{k}))^2} \tag{36}$$

Now taking into account derived from the Feynman theorem [34] expression

$$\langle n\mathbf{k}|J_\mu|m\mathbf{k}\rangle = (\varepsilon_m(\mathbf{k}) - \varepsilon_n(\mathbf{k})) \left\langle n\mathbf{k} \left| \frac{\partial}{\partial \mathbf{k}_\mu} \right| m\mathbf{k} \right\rangle \tag{37}$$

we arrive at the expression (33). Note that assuming that $i\delta = 0$ is justified only in the case of a band insulator such as for example an integer quantum Hall system [32]. In this case the conductivity is $e^2\times$ integer, and this integer is a topological number called the *Chern number* [35].

5 Theory meets experiment

Until recently, it was unclear when we will be able to calculate and predict the anomalous Hall effect in real materials. Attempts to estimate the magnitude of the anomalous Hall conductivity stemming from extrinsic contributions at least qualitatively fail quite often, as the real impurity potentials in considered materials are normally unknown. Depending on the impurity potential taken for the calculation computed conductivity values can differ by orders of magnitude. One of the first successes in computing the AHE in a very good agreement to experiments came when Jungwirth *et al.* [36, 37] applied Sundaram-Niu derived expression for the intrinsic AHC (Eq. (33)) to study the AHE in (III,Mn)V ferromagnetic semiconductors, and found a very good agreement with experiments.

The first truly *ab initio* successful prediction of an intrinsic AHC in elementary bcc Fe was made by Yao *et al.* [38] in 2004. The reported calculated value of σ_{xy} of 751 $(\Omega\cdot\mathrm{cm})^{-1}$ was only 25% off the experimental value of 1032 $(\Omega\cdot\mathrm{cm})^{-1}$, extracted from data on iron whiskers at room temperature [39]. In this calculation, millions of **k**-points were used in order to obtain reliable values of the AHC. Such slow convergence is caused by the appearance of large contributions of both signs to $\Omega_z(\mathbf{k})$ which come from very narrow regions in **k**-space. Generally speaking, the effect of the spin-orbit coupling is very small in Fe, except for the case when states are mixed which are degenerate or nearly degenerate. A completely occupied pair of such states contributes very little to the Berry curvature and the largest contribution appears when the Fermi energy lies in the middle of a spin-orbit induced gap. For example, as shown in Fig. 3 the large spike near H(1,0,0) point in the direction of P($\frac{1}{2},\frac{1}{2},\frac{1}{2}$) is due to a pair of spin-orbit coupled bands, one occupied and one unoccupied in a small **k**-interval. The small energy gap gives rise to a small energy denominator in Eq. (36), making the contribution to the Berry curvature very large in this small interval. The largest peaks and valleys in the distribution of the Berry curvature are, however, located off the **k**-space symmetry lines. For example, as can be seen from Fig. 3, the Berry curvature shows sharp peaks and valleys of several orders of magnitude in height and depth at general **k**-points in the (010)-plane.

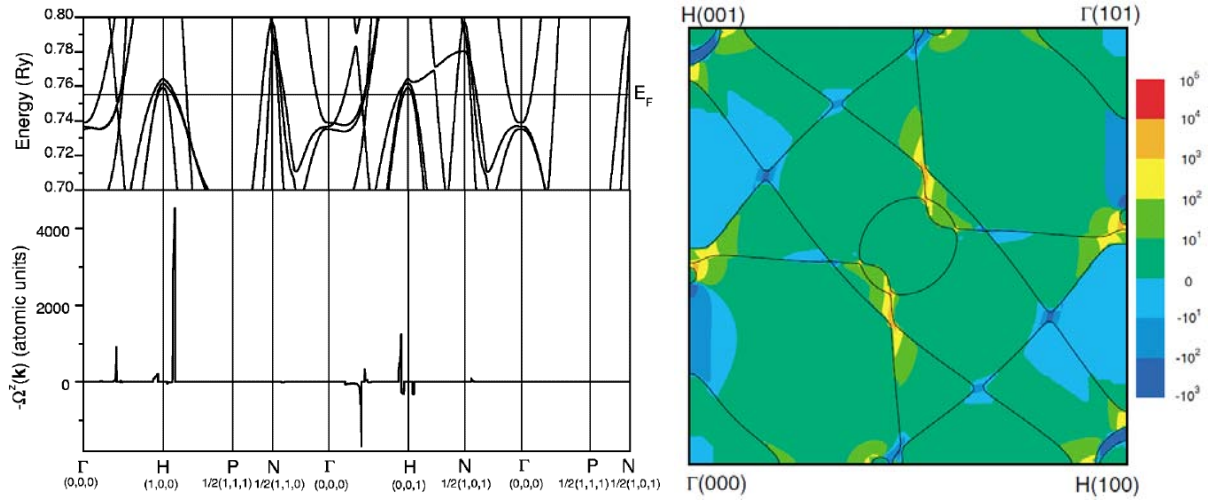


Fig. 3: *Left: Band structure of bcc Fe near the Fermi energy (upper panel) and corresponding summed over occupied bands Berry curvature $-\Omega_z(\mathbf{k})$. Right: Fermi surface of bcc Fe in (010) plane (solid lines) and the color map of the Berry curvature $-\Omega_z(\mathbf{k})$ in atomic units.*

Fang *et al.* considered the influence of the spin-orbit induced band degeneracies on the anomalous Hall conductivity in more detail on an example of $SrRuO_3$ [9]. They (as well as Onoda and Nagaosa in two dimensions [40]) considered a general case of a two-band Hamiltonian:

$$\hat{H}(\mathbf{k}) = a_0(\mathbf{k}) + \sum_{\mu=1,2,3} f_\mu(\mathbf{k})\hat{\sigma}_\mu, \tag{38}$$

where $\hat{\sigma}_\mu$ are the Pauli matrices and $\{a_0, f_\mu\}$ are certain functions of a Bloch vector $\mathbf{k}$. For a particular example of spin-up (t_{2g}) d_{xz}, d_{yz} orbitals with SOC in a cubic perovskite $SrRuO_3$ these functions are given by $a_0(\mathbf{k}) = -2t_1\cos k_z - t_1(\cos k_z + \cos k_y)$, $f_1(\mathbf{k}) = 2t_2 \sin k_x \sin k_y$, and $f_2(\mathbf{k}) = -\lambda M$, $f_3(\mathbf{k}) = -t_1(\cos k_x - \cos k_y)$, where t_1 and t_2 are effective inter- and intra-orbital hopping elements, λ is the SOC constant and M is the magnetization.
Let us establish a correspondence between a $\mathbf{k}$-vector and a vector $\mathbf{f}(\mathbf{k}) = (f_1(\mathbf{k}), f_2(\mathbf{k}), f_3(\mathbf{k}))$. The eigenvalues of Hamiltonian (38) are $\varepsilon_\pm(\mathbf{k}) = f_0(\mathbf{k}) \pm |\mathbf{f}(\mathbf{k})|$ and the band degeneracy occurs when $\mathbf{f} = 0$. Substituting the corresponding eigenvectors into Eq. (33) we get the contribution to the σ_{xy}-AHC from the two bands:

$$\sigma_{xy} = \frac{e^2}{8\pi h} \int (f_0(\varepsilon_+(\mathbf{k})) - f_0(\varepsilon_-(\mathbf{k}))) \left(\frac{\partial\varphi}{\partial k_x}\frac{\partial\theta}{\partial k_y} - \frac{\partial\varphi}{\partial k_y}\frac{\partial\theta}{\partial k_x} \right) \sin\theta \, d\mathbf{k}, \tag{39}$$

where f_0 is the Fermi distribution function and φ and θ are the polar coordinates of vector $\mathbf{f}(\mathbf{k}) = (f\sin\theta\cos\varphi, f\sin\theta\sin\varphi, f\cos\theta)$. The expression for the AHC can be rewritten as:

$$\sigma_{xy} = \frac{e^2}{8\pi h} \int (f_0(\varepsilon_+(\mathbf{k})) - f_0(\varepsilon_-(\mathbf{k}))) \, dk_z \, d\Omega_f, \tag{40}$$

where $d\Omega_f = [\partial(\theta,\phi)/\partial(k_x,k_y)] \sin\theta \, dk_x \, dk_y = \sin\theta \, d\varphi \, d\theta$ is the solid angle in $\mathbf{f}$-space corresponding to an infinitesimal area $dk_x \, dk_y$ in $\mathbf{k}$-space (see Fig. 4(A)). It can be easily shown that $d\Omega_f$ is the integral of the gauge field

$$\boldsymbol{\Omega}(\mathbf{f}) = \pm\frac{\mathbf{f}}{|\mathbf{f}|^3} \tag{41}$$

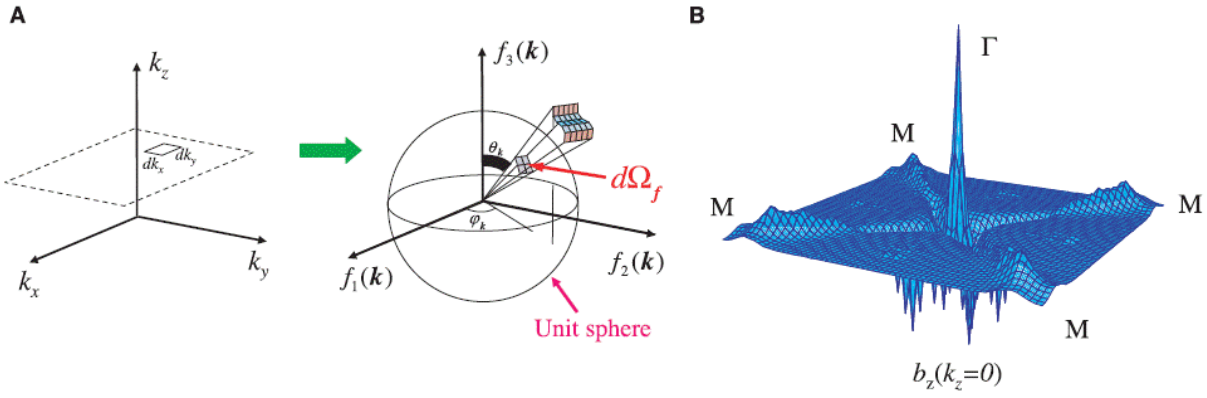


Fig. 4: *A: geometrical meaning of the contribution to the AHC when two bands are nearly degenerate, B: calculated Berry curvature* $\Omega_z(\mathbf{k})$ *for* t_{2g} *bands as a function of* (k_x, k_y) *with* $k_z = 0$ *for cubic SrRuO*$_3$*. The sharp peak around* $k_x = k_y = 0$ *and the ridges along* $k_x = \pm k_y$ *are due to the near degeneracy of* d_{xz} *and* d_{yz} *bands.*

due to a "magnetic monopole" at $\mathbf{f} = 0$ (i.e. the point of degeneracy) over the infinitesimal surface in $\mathbf{f}$-space corresponding to $dk_x\, dk_y$. This means that when the Fermi energy is near the monopole point of degeneracy, the electrons are subject to a strong gauge field and the contribution to the conductivity is large. This observation establishes a close relation between the anomalous Hall conductivity and the magnetic monopole in reciprocal space, existence of which was postulated in real space by Dirac in 1931 [41] and which was found as a solution of enormous energy ($\approx 10^{16}$ GeV) to the equation of the non-Abelian gauge theory for grand unification [42, 43]. From expression (40) it is also easy to see that the integrated contribution from fully occupied couple of bands is zero due to f_0 distribution function.

In case of $SrRuO_3$ the parameters of the Hamiltonian satisfy $t_1 >> t_2 >> \lambda M$, which results in near degeneracy of the bands along the lines $k_x = k_y = 0$ and $k_x = \pm k_y$ due to symmetry [9]. According to first-principles calculations, this results in a large peak of the Berry curvature $\Omega_z(\mathbf{k})$ at the Γ point and its enhancement along the $k_x = \pm k_y$ lines in the Brillouin zone, as can be seen in Fig. 4(B).

Experimental measurements as well as *ab initio* calculations of Fang *et al.* show that in $SrRuO_3$ the anomalous Hall resistivity changes non-monotonously with temperature and even includes a sign change. Such behavior is far beyond the expectation based on the conventional expression in Eq. (1). In Fig. 5 the AHC σ_{xy} is shown as a function of the Fermi level position with respect to the real converged value of E_F. When the Fermi level is shifted not only the absolute value of σ_{xy} is changed drastically, but also its sign. The sharp and spiky structure in this plot is a natural consequence of the singular behavior of the magnetic monopoles in reciprocal space. For the converged solution without any shift in the Fermi energy the calculated value of AHC of $-60\ (\Omega \cdot \mathrm{cm})^{-1}$ is very close to that obtained in experiment of $-100\ (\Omega \cdot \mathrm{cm})^{-1}$. In general, the calculated AHC follows the same trend and matches very well with experiment as a function of the sample magnetization M, shown in Fig. 5. In this graph the change in the magnetization is taken into account by the rigid splitting of spin-up and spin-down electronic bands. The non-monotonous behavior of σ_{xy} as a function of temperature T can be attributed to the non-monotonous dependency of σ_{xy} on the magnetization M (M itself is a monotonous function of T), which arises due to integration over the occupied bands in $\mathbf{k}$-space.

To summarize, evaluating the anomalous Hall conductivity according to Eq. (33) becomes a

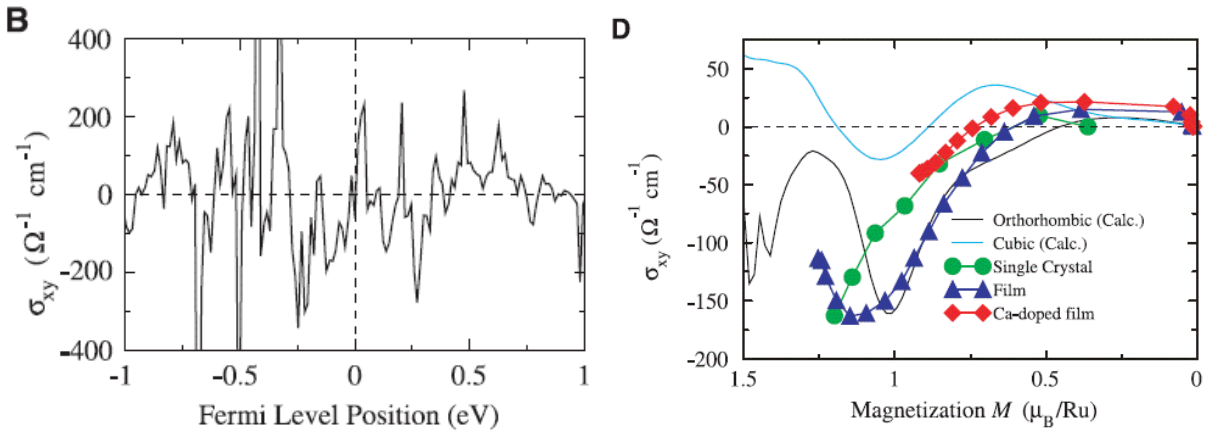


Fig. 5: *B: AHC σ_{xy} as a function of the Fermi level position for the orthorhombic structure of single-crystal of $SrRuO_3$. The Fermi level is shifted rigidly relative to the converged solution (zero point), D: AHC σ_{xy} as a function of the magnetization M together with the results of first-principles calculations for cubic and orthorhombic structures.*

common technique in electronic structure community [45, 46]. Due to the simplicity of this intrinsic mechanism formulation it became possible to work out the magnitudes of the quantities observed in experiment directly from the electronic structure of considered materials, which can be computed with any electronic structure code nowadays. The ways to optimize the integration of the Berry curvature over the occupied bands in the Brillouin zone, which can be quite tricky due to Berry curvature's nontrivial structure, were proposed [44], and even reformulation of Eq. (33) as a Fermi surface integral was suggested [45]. The whole arsenal of methods routinely applied in electronic structure calculations, such as LDA+U, magnetic non-collinearity etc., can be now used to study the anomalous Hall effect in, e.g., non-collinear magnets or strongly correlated materials directly from their electronic structure without referring to unknown impurity potentials. In some sense, the validity of equation (33) marks the beginning of the "first-principles" era in the practical application of the anomalous Hall effect.

6 Boltzmann equation and semiclassical theory of side jump

The near equilibrium dynamics of a uniform ($\frac{\partial f}{\partial \mathbf{r}} = 0$) system of classical charged particles (wavepackets) in a weak electric field $\mathbf{E}$ with elastic impurities is described by Boltzmann equation for the distribution function f ($l = (n, \mathbf{k})$):

$$\frac{\partial f_l}{\partial t} + e\mathbf{E}\frac{\partial f_l}{\partial \mathbf{k}} = -\sum_{l'} \omega_{ll'}\,(f_l - f_{l'}), \tag{42}$$

where $\omega_{ll'}$ is the scattering rate (probability) between volumes in phase space centered around l and l'. The scattering rate depends on details of the scattering potential and should be found separately. The semiclassical equation (42) is very powerful since it automatically takes care of the summation of various infinite series of Feynman diagrams that appear in quantum linear response theory and it keeps the physical meaning of all terms transparent. It operates only with gauge invariant quantities, such as the side jump, the scattering rate, anomalous and usual velocities, distribution function.

In spite of the classical form of Eq. (42) the scattering rate $\omega_{ll'}$ has to be calculated purely quantum mechanically and it is given by the golden rule. The golden rule connects the classical and quantum descriptions of a scattering event and can be expressed in terms of the scattering matrix T:

$$\omega_{ll'} = 2\pi |T_{ll'}|^2 \delta(\varepsilon_{l'} - \varepsilon_l), \tag{43}$$

where the elements of the scattering matrix T are defined as

$$T_{ll'} = \langle l|\hat{V}|\psi_{l'}\rangle, \tag{44}$$

where $\hat{V}$ is the impurity potential and ψ_l is the eigenstate of the full Hamiltonian $\hat{H} = \hat{H}_0 + \hat{V}$ satisfying the Lippmann-Schwinger equation:

$$|\psi_l\rangle = |l\rangle + \frac{\hat{V}}{\varepsilon_l - \hat{H}_0 + i\eta}|\psi_l\rangle. \tag{45}$$

The scattering matrix can be written as a Born series in powers of disorder strength. At weak disorder one can approximate the scattering state $|\psi\rangle$ by a truncated series in powers of $\langle l'|\hat{V}|l\rangle$:

$$|\psi_l\rangle \approx |l\rangle + \sum_{l'} \frac{V_{l'l}}{\varepsilon_l - \varepsilon_{l'} + i\eta}|l'\rangle, \tag{46}$$

which leads to the following expression for the T matrix up to the second order in V:

$$T_{ll'} \approx V_{ll'} + \sum_{l''} \frac{V_{l'l''}V_{l''l}}{\varepsilon_l - \varepsilon_{l''} + i\eta}. \tag{47}$$

In general, the semiclassical description which relies on the golden rule only will not be complete. This becomes clear after noticing that the scattering rates $\omega_{ll'}$ pick up only the real-part information encoded in the scattering matrix elements $T_{ll'}$. This is the reason why in general Eq. (42) fails to describe the AHE properly. While the skew scattering can be described within the framework of Eq. (42) completely by going to the second order in expansion of the T-matrix with respect to the impurity strength given by (47), it does not contain the information about the side jump at a scattering event, thus, it has to be taken into account when constructing a kinetic equation for the semiclassical distribution function which describes the AHE.

The side jump is the coordinate shift $\delta\mathbf{r}_{ll'}$ acquired by a particle during the scattering event. Recently, Sinitsyn *et al.* [48, 49] showed that the equations for the wavepacket dynamics can be used to derive a gauge invariant expression for the side jump scattering on an impurity with the potential which can be treated within the Born approximation. Lets assume for the moment that $l \equiv \mathbf{k}$. Consider a wavefunction $\psi_{\mathbf{k}}(\mathbf{r}, t)$ which moves under an influence of a weak impurity. The solution of the Schrödinger equation in this case can be written in terms of an unperturbed Hamiltonian:

$$\psi^{out}_{\mathbf{k}}(\mathbf{r}, t) = \int C(\mathbf{k}', t)\, \psi_{\mathbf{k}'}(\mathbf{r}, t)\, d\mathbf{k}' \tag{48}$$

To the lowest order in the strength of the potential perturbation theory leads to the following expression for time-dependent coefficients [51]:

$$C(\mathbf{k}', t) = -i\, V_{\mathbf{k}'\mathbf{k}} \int_{-\infty}^{t} e^{(\varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k}))t'}\, dt' + \delta(\mathbf{k}' - \mathbf{k}), \tag{49}$$

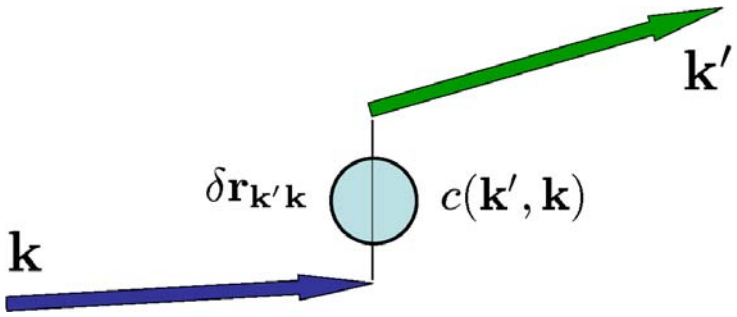


Fig. 6: *Scattering of a wavepacket with the reciprocal vector* $\mathbf{k}$ *on an impurity, during which the incoming wavepacket acquires a coordinate shift* $\delta\mathbf{r}_{\mathbf{k}'\mathbf{k}}$.

where $V_{\mathbf{k}'\mathbf{k}} = \langle\psi_{\mathbf{k}'}(\mathbf{r})|\hat{V}|\psi_{\mathbf{k}}(\mathbf{r})\rangle$ is the matrix element of the disorder potential between two eigenstates of the unperturbed Hamiltonian. Higher order terms can be incorporated into the above formula by substituting the T-matrix instead of the disorder potential [51]:

$$C(\mathbf{k}',t) = -i\,T_{\mathbf{k}'\mathbf{k}}\int_{-\infty}^{t} e^{(\varepsilon(\mathbf{k}')-\varepsilon(\mathbf{k}))t'}\,dt' + \delta(\mathbf{k}'-\mathbf{k}). \tag{50}$$

The time integral in Eq. (50) is divergent, but after performing a regularizing procedure at large positive time we find [51]:

$$C(\mathbf{k}',+\infty) = c(\mathbf{k}',\mathbf{k}) + \delta(\mathbf{k}'-\mathbf{k}), \tag{51}$$

where

$$c(\mathbf{k}',\mathbf{k}) = -2\pi i\,T_{\mathbf{k}'\mathbf{k}}\,\delta(\varepsilon(\mathbf{k}')-\varepsilon(\mathbf{k})). \tag{52}$$

For $\mathbf{k}' \neq \mathbf{k}$ the square of the amplitude $|c(\mathbf{k}',\mathbf{k})|^2$ is the scattering probability from the state with momentum $\mathbf{k}$ into the state with momentum $\mathbf{k}'$, from which the golden rule follows.
Consider now a wavepacket centered around $\mathbf{k}_0$ in reciprocal space, which approaches the impurity from infinity:

$$\Psi_{\mathbf{k}_0}(\mathbf{r},t) = \int a(\mathbf{k}-\mathbf{k}_0)\psi_{\mathbf{k}}(\mathbf{r},t)\,d\mathbf{k}. \tag{53}$$

As usual for a wavepacket (see also section 3), when multiplied with a smooth function of momentum the envelope function $a(\mathbf{k}-\mathbf{k}_0)$ can be treated as a δ-function, however, when multiplied with a true δ-function, it is considered smooth, reflecting the finite width of the wavepacket in reciprocal space. In real space, the spread of the wavepacket is much larger than the lattice constant, but much smaller than any other length scales. Using these properties of the envelope function and Eq. (21) it is possible to write down the law for the center of mass of the wavepacket on infinity [49]:

$$\mathbf{r}_c(\mathbf{k}_0,t)_{t\to-\infty} = \mathbf{v}_{\mathbf{k}_0}t + \delta\mathbf{r}_{-\infty} = \frac{\partial\varepsilon(\mathbf{k}_0)}{\partial\mathbf{k}_0}t + \left\langle u_{\mathbf{k}_0}\middle| i\frac{\partial}{\partial\mathbf{k}_0}u_{\mathbf{k}_0}\right\rangle. \tag{54}$$

On the other hand, after interaction with an impurity potential, we can formally construct an outgoing wavepacket:

$$\Psi^{out}(\mathbf{r},t) = \int a(\mathbf{k}-\mathbf{k}_0)\psi^{out}_{\mathbf{k}}(\mathbf{r},t)\, d\mathbf{k}, \tag{55}$$

where $\psi^{out}_{\mathbf{k}}(\mathbf{r},t)$ is given by (48). Using the perturbation expansion (47) for the T-matrix:

$$T_{\mathbf{k'k}} = V_{\mathbf{k'k}} + \sum_{\mathbf{k''}} \frac{V_{\mathbf{k'k''}}\, V_{\mathbf{k''k}}}{\varepsilon(\mathbf{k''}) - \varepsilon(\mathbf{k}) + i\eta} + ..., \tag{56}$$

and the representation for the potential matrix $V_{\mathbf{k'k}} = |V_{\mathbf{k'k}}| \cdot \exp(i \arg(V_{\mathbf{k'k}}))$ the following expression for the motion of the center of mass of the outgoing wavepacket on infinity (which contributes to the Hall current) in the second order in $V_{\mathbf{k'k}}$ can be derived making use of (50), (51) and (52) ([49]):

$$\mathbf{r}_c(t)_{t\to+\infty} = \int |C(\mathbf{k'},+\infty)|^2 \left(\mathbf{v}_{\mathbf{k'}} t + \left\langle u_{\mathbf{k'}} \right| i\frac{\partial}{\partial \mathbf{k'}} u_{\mathbf{k'}} \right\rangle - \hat{\mathbf{D}}_{\mathbf{k'k}_0} \arg(V_{\mathbf{k'k}_0}) \right) d\mathbf{k'}, \tag{57}$$

where

$$\hat{\mathbf{D}}_{\mathbf{k'k_0}} = \frac{\partial}{\partial \mathbf{k'}} + \frac{\partial}{\partial \mathbf{k}_0}. \tag{58}$$

The coefficient $|C(\mathbf{k'},+\infty)|^2$ can be interpreted as the scattering probability into state $\mathbf{k'}$ from initial state $\mathbf{k}_0$ (see Fig. 6). Equation (57) has a semiclassical meaning that the average final coordinate is the sum over probabilities of final states multiplied with corresponding coordinate shifts. Now presenting the center of mass motion of the $\mathbf{k}$-wavepacket at $t \to -\infty$ scattering into $\mathbf{k'}$-wavepacket at $t \to +\infty$ as $\mathbf{r}_c(\mathbf{k},t)_{t\to-\infty} = \mathbf{v}_{\mathbf{k}} t + \delta\mathbf{r}_{-\infty}$ and $\mathbf{r}_c(\mathbf{k'},t)_{t\to+\infty} = \mathbf{v}_{\mathbf{k'}} t + \delta\mathbf{r}_{+\infty}$ we define the scattering-induced side jump as

$$\delta\mathbf{r}_{\mathbf{k'k}} = \delta\mathbf{r}_{+\infty} - \delta\mathbf{r}_{-\infty}. \tag{59}$$

Making use of equations (54) and (57) in the multiband case we can formulate a gauge-invariant expression for side jump:

$$\delta\mathbf{r}_{l'l} = \left\langle u_{l'} \right| i\frac{\partial}{\partial \mathbf{k'}} u_{l'} \left\rangle - \left\langle u_l \right| i\frac{\partial}{\partial \mathbf{k}} u_l \right\rangle - \hat{\mathbf{D}}_{\mathbf{k'k}} \arg(V_{l'l}). \tag{60}$$

Such type of expression was found already in 1982, when Belinicher *et al.* [52] applied it to the studies of the photovoltaic effect. They showed that when electrons absorb polarized light they shift according to Eq. (60) with V corresponding to the electron-photon interaction. Unlike the golden rule which in the lowest Born approximation depends on the absolute value of the scattering potential, the coordinate shift expression (60) depends on its phase but does not depend on its absolute value, so, in some sense, it can be considered as complimentary to the golden rule.

In case when the radial impurity potential is spin-independent, interestingly, the side jump does not depend on the impurity potential at all:

$$\delta\mathbf{r}_{l'l} = \left\langle u_{l'} \right| i\frac{\partial}{\partial \mathbf{k'}} u_{l'} \left\rangle - \left\langle u_l \right| i\frac{\partial}{\partial \mathbf{k}} u_l \right\rangle - \hat{\mathbf{D}}_{\mathbf{k'k}} \arg(\langle u_l | u_{l'} \rangle), \tag{61}$$

moreover, in the case of a very small scattering angle $|\mathbf{k}' - \mathbf{k}| \ll |\mathbf{k}|$ for non-degenerate bands one can disregard interband scattering and make a couple of approximations [49] which, to the first order in $|\mathbf{k}' - \mathbf{k}|$, allow to write down the side jump as:

$$\delta\mathbf{r}_{\mathbf{k}'\mathbf{k}} \approx \mathbf{\Omega}(\mathbf{k}) \times (\mathbf{k}' - \mathbf{k}), \tag{62}$$

where $\mathbf{\Omega}(\mathbf{k})$ is the Berry curvature of the Bloch band.
There are two distinctive effects due to side jump. The first one is the side jump accumulation. After averaging over many scatterings the side jumps do not cancel and lead to the velocity renormalization by a correction:

$$\mathbf{v}_l^{sj} = \sum_{l'} \omega_{l'l}\, \delta\mathbf{r}_{l'l}. \tag{63}$$

The second effect can be attributed to the fact that when the scattering takes place in an electric field, there is a change in the potential energy after scattering given by

$$\Delta U_{l'l} = e\mathbf{E} \cdot \delta\mathbf{r}_{l'l}. \tag{64}$$

This change of energy ultimately influences the Hall conductivity and has to be accounted for in the Boltzmann equation. We will discuss this further in the next section.
In a simple model case, it can be shown that the side jump contribution to the conductivity depends only on the properties in the vicinity of the Fermi surface [50]. This is in contrast to the intrinsic contribution which depends on the integrated over the occupied bands in the whole Brillouin zone quantities. This is probably why in many cases the side jump contribution to the AHC, which does not depend on the impurity concentration as well as the intrinsic contribution which disregards the impurities all together, is much smaller than the intrinsic one – in order for side jump to dominate certain resonances have to occur around the Fermi surface, while the integrated over the BZ quantities must be small. An example when the side jump contribution is exactly zero presents the so-called quantum spin Hall effect. In 2D Dirac bands, e.g., in graphene, the Fermi level is placed inside the bulk spectrum gap. In this case the gapless excitations are forbidden (except near the sample edges), so the side jump and skew scattering do not contribute to the conductivity, but if the band has a non-zero Berry curvature, then there is a non-zero quantized intrinsic contribution [53, 54] (see also discussion at the end of next section).

7 AHE within the semiclassical Boltzmann equation

The semiclassical Boltzmann equation for the distribution function f_l which takes into account both the change of the momentum and the coordinate shift during scattering in a homogeneous crystal in the presence of an external electric field $\mathbf{E}$ can be written down in the following way up to the linear order in the field [49, 50]:

$$\frac{\partial f_l}{\partial t} + e\mathbf{E} \cdot \mathbf{v}_l^0 \cdot \frac{\partial f_0(\varepsilon_l)}{\partial \varepsilon_l} = -\sum_{l'} \omega_{ll'} \left(f_l - f_{l'} - e\mathbf{E} \cdot \delta\mathbf{r}_{ll'} \cdot \frac{\partial f_0(\varepsilon_l)}{\partial \varepsilon_l} \right), \tag{65}$$

where $\mathbf{v}_l^0 = \frac{\partial \varepsilon_l}{\partial \mathbf{k}}$ is the normal electron velocity. The total distribution function in the steady state $\frac{\partial f_l}{\partial t} = 0$ can be decomposed into the sum of equilibrium f_0 and two non-equilibrium distributions:

$$f_l = f_0(\varepsilon_l) + g_l + g_l^{ad}. \tag{66}$$

The g_l and g_l^{ad} independently solve self-consistent time-independent equations (compare to (42)):

$$e\mathbf{E}\cdot\mathbf{v}_l^0\cdot\frac{\partial f_0(\varepsilon_l)}{\partial\varepsilon_l} = -\sum_{l'}\omega_{ll'}(g_l - g_{l'}), \tag{67}$$

and

$$\sum_{l'}\omega_{ll'}\left(g_l^{ad} - g_{l'}^{ad} - \frac{\partial f_0(\varepsilon_{l'})}{\partial\varepsilon_{l'}}\,e\mathbf{E}\cdot\delta\mathbf{r}_{ll'}\right) = 0. \tag{68}$$

The appearance of equation (68) in addition to standard equation (67) and an anomalous distribution function g_l^{ad} happens due to following reasons. In the presence of an external field the particle upon side jump changes its potential energy $\Delta U_{l'l}$ (64). Since the total energy is conserved, the scattered particle acquires additional kinetic energy $\Delta\varepsilon_{l'l} = \varepsilon_{l'} - \varepsilon_l = \Delta U_{l'l}$ in order to compensate the change in the potential energy. The equilibrium distribution would then become unstable:

$$\frac{\partial f_l}{\partial t} = -\sum_{l'}\omega_{ll'}(f_0(\varepsilon_l) - f_0(\varepsilon_{l'})) = \sum_{l'}\omega_{ll'}\frac{\partial f_0(\varepsilon_{l'})}{\partial\varepsilon_{l'}}\Delta\varepsilon_{l'l} \neq 0, \tag{69}$$

unless compensated by an additional anomalous correction g_l^{ad} to the distribution function. The analog of this equation can be found also in Luttinger's paper [23] and its validity was also confirmed by numerical simulations [48].

To find the current induced by an electric field and the conductivity we need an appropriate expression for the velocity of semiclassical particles in addition to solving the Boltzmann equation for the distribution function. In the anomalous Hall effect, one has to take into account the renormalizations of the conventional band group velocity $\mathbf{v}_l^0$ due to accumulations of the coordinate shifts after many scatterings (side jump) and impurity-independent band mixing due to electric field (the anomalous velocity):

$$\mathbf{v}_l = \mathbf{v}_l^0 + \mathbf{v}_l^a + \mathbf{v}_l^{sj} = \frac{\partial\varepsilon_l}{\partial\mathbf{k}} + \mathbf{\Omega}_l\times e\mathbf{E} + \sum_{l'}\omega_{ll'}\delta\mathbf{r}_{ll'}. \tag{70}$$

While the second term in this equation is the previously discussed here in detail Berry curvature induced contribution, the last term is due to discussed in the previous section accumulations of the side jumps on impurities. The total current in the system is given by:

$$\mathbf{J} = e\sum_l f_l\cdot\mathbf{v}_l = e\sum_l(f_0(\varepsilon_l) + g_l + g_l^{ad})\cdot(\mathbf{v}_l^0 + \mathbf{v}_l^a + \mathbf{v}_l^{sj}), \tag{71}$$

of which we will consider only contributions proportional to electric field. The first term in this product gives the expression for the intrinsic conductivity

$$\sigma_{xy}^{int} = -e^2\sum_l f_0(\varepsilon_l)\cdot\Omega_{l,z} \tag{72}$$

discussed in detail in section 4. Below we will discuss other contributions to the total conductivity which come out of equation (71).

Substituting expansion for the scattering matrix (56) into the golden rule (43) we can make use of the decomposition of the scattering rates in the following way:

$$\omega_{ll'} = \omega_{ll'}^{(2)} + \omega_{ll'}^{(3)} + \omega_{ll'}^{(4)} + ..., \tag{73}$$

where

$$\omega_{ll'}^{(2)} = 2\pi \langle |V_{ll'}|^2 \rangle_{dis}\, \delta(\varepsilon_l - \varepsilon_{l'}), \tag{74}$$

and

$$\omega_{ll'}^{(3)} = 2\pi \left(\sum_{l''} \frac{\langle V_{ll'} V_{l'l''} V_{l''l} \rangle_{dis}}{\varepsilon_l - \varepsilon_{l''} - i\eta} + c.c. \right) \delta(\varepsilon_l - \varepsilon_{l'}) \tag{75}$$

and so on. The superscript of the scattering rate means the order of the impurity potential it corresponds to: $\omega_{ll'}^{(2)} \sim V^2$, $\omega_{ll'}^{(3)} \sim V^3$ and $\omega_{ll'}^{(4)} \sim V^4$. The skew scattering comes from the antisymmetric part of the scattering rate [18]:

$$\omega_{ll'}^{(a)} = \frac{\omega_{ll'} - \omega_{l'l}}{2}. \tag{76}$$

While $\omega_{ll'}^{(2)}$ is symmetric, $\omega_{ll'}^{(3)}$ and $\omega_{ll'}^{(4)}$ have antisymmetric as well as symmetric parts. Their symmetric parts are not essential for the further discussion and they only renormalize the $\omega_{ll'}^{(2)}$ contribution. At order V^3, for example, the antisymmetric part of the scattering rate $\omega_{ll'}^{(3)}$ can be deduced [23, 50, 56]:

$$\omega_{ll'}^{(3a)} = -(2\pi)^2 \sum_{l''} \delta(\varepsilon_l - \varepsilon_{l''}) \, \langle V_{ll'} V_{l'l''} V_{l''l} \rangle_{dis}\, \delta(\varepsilon_l - \varepsilon_{l'}). \tag{77}$$

Usually the properties of the skew scattering were inferred only from the lowest, third, order of antisymmetric part of $\omega_{ll'}$, thus it was normally assumed that $\omega_{ll'}^{(a)}$ is proportional to the impurity concentration n. In the next, fourth order, the antisymmetric scattering is proportional to the product of four disorder vertexes. For a Gaussian correlated potential $\langle V \cdot V \cdot V \cdot V \rangle_{dis} \sim \langle V \cdot V \rangle_{dis} \cdot \langle V \cdot V \rangle_{dis} \sim n^2$, with n as impurity concentration. Thus the higher order antisymmetric contribution behaves as $\omega_{ll'}^{(4a)} \sim n^2$. This fourth order term is different parametrically from the third order term and has a distinct microscopic origin, therefore, should not be neglected. Interestingly, as we will see, the contribution to the conductivity coming from $\omega_{ll'}^{(4a)}$ is similar to the side jump related contribution. Overall, for an appropriate analysis of the AHE the scattering rate has to be decomposed in the following way:

$$\omega_{ll'} = \omega_{ll'}^{(2)} + \omega_{ll'}^{(3a)} + \omega_{ll'}^{(4a)}. \tag{78}$$

Correspondingly, the g_l distribution correction from (66) should be decomposed into the sum of symmetric and antisymmetric in the transverse direction contributions (g^{ad} is antisymmetric):

$$g_l = g_l^{(s)} + g_l^{(3a)} + g_l^{(4a)}, \tag{79}$$

which solve a set of self-consistent time-independent equations corresponding to equation (67) [10, 49, 50]:

$$e\mathbf{E}\cdot\mathbf{v}_l^0\cdot\frac{\partial f_0(\varepsilon_l)}{\partial\varepsilon_l} = -\sum_{l'}\omega_{ll'}^{(2)}(g_l^{(s)}-g_{l'}^{(s)}), \tag{80}$$

$$\sum_{l'}\omega_{ll'}^{(3a)}(g_l^{(s)}-g_{l'}^{(s)}) + \sum_{l'}\omega_{ll'}^{(2)}(g_l^{(3a)}-g_{l'}^{(3a)}) = 0, \tag{81}$$

$$\sum_{l'}\omega_{ll'}^{(4a)}(g_l^{(s)}-g_{l'}^{(s)}) + \sum_{l'}\omega_{ll'}^{(2)}(g_l^{(4a)}-g_{l'}^{(4a)}) = 0. \tag{82}$$

Because $\omega_{ll'}^{(2)}\sim n$, we conclude that $g_l^{(s)}\sim n^{-1}$ from (80). In turn $\omega_{ll'}^{(3a)}\sim n$ and from (81) we see that $g_l^{(3a)}\sim n^{-1}$, while from the fact that $\omega_{ll'}^{(4a)}\sim n^2$ we conclude that $g_l^{(4a)}\sim n^0$ based on (82). The anomalous distribution $g_l^{ad}\sim n^0$ from (68).
Besides an intrinsic conductivity (72) (which depends as n^0 on impurity concentration), there are four other terms coming out of equation (71) which contribute to transverse current and transverse conductivity. The anomalous distribution conductivity comes from the g_l^{ad} correction to the distribution function:

$$\sigma_{xy}^{ad} = -e\sum_l\frac{g_l^{ad}}{E_x}\cdot(\mathbf{v}_l^0)_y\sim n^0, \tag{83}$$

while the direct side jump conductivity corresponding to the side jump part of the transverse velocity comes from the product with the symmetric part $g_l^{(s)}$:

$$\sigma_{xy}^{sj} = -e\sum_l\frac{g_l^{(s)}}{E_x}\cdot(\mathbf{v}_l^{sj})_y\sim n^0. \tag{84}$$

The skew scattering conductivity has two terms, which correspond to $g_l^{(3a)}$ and $g_l^{(4a)}$ parts of g_l:

$$\sigma_{xy}^{sk1} = -e\sum_l\frac{g_l^{(3a)}}{E_x}\cdot(\mathbf{v}_l^0)_y\sim\frac{1}{n}, \tag{85}$$

and

$$\sigma_{xy}^{sk2} = -e\sum_l\frac{g_l^{(4a)}}{E_x}\cdot(\mathbf{v}_l^0)_y\sim n^0. \tag{86}$$

The first skew scattering conductivity (85) is the conventional skew scattering conductivity, which was intensively discussed in the past (see section 2 and e.g. [23, 56]). The second one (86) is generally disregarded intrinsic skew scattering conductivity which is parametrically the same as the side jump and intrinsic conductivity [50]. Overall, there are five separate gauge invariant contributions, summarized in Table 1, each having distinct origins:

$$\sigma_{xy} = \sigma_{xy}^{int}+\sigma_{xy}^{ad}+\sigma_{sj}^{int}+\sigma_{xy}^{sk1}+\sigma_{xy}^{sk2}. \tag{87}$$

It is possible to regroup them into three: besides intrinsic, the side jump and anomalous contributions both originate from coordinate shifts at scattering events. Similarly, the conventional and the intrinsic skew scattering both appear from the asymmetry in the collision term kernel in the semiclassical Boltzmann equation. However, it is very insightful to remember the differences between all five. The intrinsic skew scattering conductivity is independent of the impurity

Intrinsic	Side jump	Anomalous distribution	Conventional skew scattering	Intrinsic skew scattering
$\mathrm{O}(n^0)$	$\mathrm{O}(n^0)$	$\mathrm{O}(n^0)$	$\mathrm{O}(n^{-1})$	$\mathrm{O}(n^0)$
$-e\,\mathrm{Tr}(f_0 \cdot v^a)$	$-e\,\mathrm{Tr}(g^{(s)} \cdot v^{sj})$	$-e\,\mathrm{Tr}(g^{ad} \cdot v^0)$	$-e\,\mathrm{Tr}(g^{(3a)} \cdot v^0)$ $\omega_{ll'}^{(3a)} \sim n$	$-e\,\mathrm{Tr}(g^{(4a)} \cdot v^0)$ $\omega_{ll'}^{(4a)} \sim n^2$

Table 1: *Mechanisms of the anomalous Hall effect within the semiclassical Boltzmann equation [10].*

concentration and is parametrically very similar to the side jump related contributions, in contrast to a common belief that the skew scattering conductivity is inversely proportional to the impurity concentration n. On the other hand, the side jump contribution is a direct consequence of coordinate shifts, while the anomalous distribution contribution requires more complicated derivation within the semiclassical theory.

Finally, it has to be noted, that sometimes such a decomposition of the AHC into five terms the way described above may not be very transparent. For example, for a particular case of a two-dimensional electron gas ferromagnet a new extrinsic regime of the AHE was predicted – the so-called hybrid skew scattering is inversely proportional to the impurity concentration (like conventional skew scattering), but does not depend on the impurity strength (like side-jump). Moreover, in the discussion above we did not consider a contribution to the AHC which involves the product of g^{ad} with v^{sj}, which disappears in case of an isotropic scattering but can give rise to a nonzero contribution given an appropriate anisotropic scattering and anisotropic Fermi distribution.

Let us now consider a concrete example of a massive two-dimensional Dirac model Hamiltonian with randomly placed weak δ-function-like spin-independent impurities [50, 53]. The impurity-free 2D Dirac Hamiltonian which breaks the time-reversal symmetry can be written down as:

$$\hat{H}_0 = v(k_x\sigma_x + k_y\sigma_y) + \Delta\sigma_z, \tag{88}$$

where σ_i are Pauli matrices and Δ characterizes the strength of the spin-orbit coupling. The non-zero mass Δ opens a gap in the spectrum and splits the Dirac band into subbands above and below the gap with dispersions $\varepsilon_{\mathbf{k}}^{\pm} = \pm\sqrt{\Delta^2 + (vk)^2}$, where $k = |\mathbf{k}|$, while $+$ and $-$ distinguish bands with positive and negative energies. In the following we will assume that the Fermi energy ε_F is positive and lies above the gap (see Fig. 7). We denote as V_0^2 and V_1^3 the terms $\langle V^2\rangle_{dis}$ and $\langle V^3\rangle_{dis}$, respectively. In case of this model Hamiltonian, the Berry curvature can be easily found for both $+$ and $-$ bands:

$$\Omega_z^{\pm}(k) = \mp\frac{\Delta v^2}{2(\Delta^2 + (vk)^2)^{3/2}}, \tag{89}$$

and the corresponding intrinsic conductivity looks like:

$$\sigma_{xy}^{int} = -\frac{e^2}{(2\pi)^2}\int_0^{\infty}\Omega_z^{-}(k)\,d\mathbf{k} - \frac{e^2}{(2\pi)^2}\int_{k_F}^{\infty}\Omega_z^{+}(k)\,d\mathbf{k} = -\frac{e^2\Delta}{4\pi\sqrt{\Delta^2 + (vk_F)^2}}. \tag{90}$$

Concerning the side jump, for 2D Dirac Hamiltonian it can be shown that $\sigma_{xy}^{ad} = \sigma_{xy}^{sj}$, thus the total side jump conductivity is just twice the one of the two contributions, and reads:

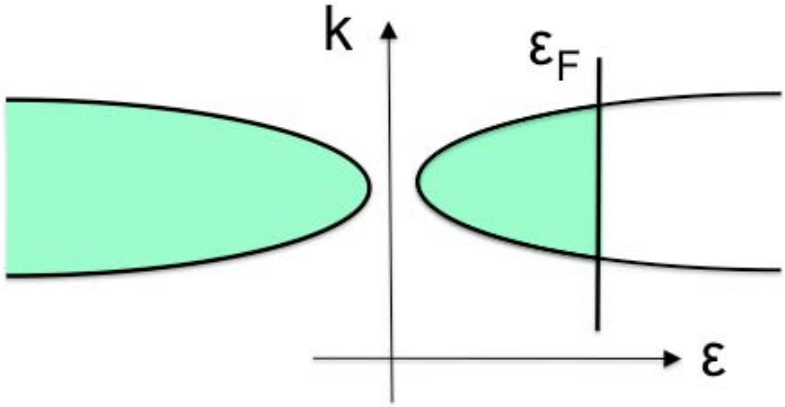


Fig. 7: *Schematic energy dispersion for the massive 2D Dirac Hamiltonian. The Fermi level is assumed to be above the gap.*

$$\sigma_{xy}^{ad} + \sigma_{xy}^{sj} = -\frac{e^2\Delta}{\pi\sqrt{\Delta^2 + (vk_F)^2}} \frac{(vk_F)^2}{(vk_F)^2 + 4\Delta^2}. \tag{91}$$

In turn, the skew scattering contribution to the Hall conductivity is:

$$\sigma_{xy}^{sk} = -\frac{e^2V_1^3}{2\pi n V_0^4} \frac{\Delta(vk_F)^4}{(4\Delta^2 + (vk_F)^2)^2} - \frac{3e^2}{4\pi\sqrt{\Delta^2 + (vk_F)^2}} \frac{\Delta(vk_F)^4}{(4\Delta^2 + (vk_F)^2)^2}, \tag{92}$$

where the first term is the conventional skew scattering conductivity σ_{xy}^{sk1}, inversely proportional to the impurity concentration. The second term, independent on the impurity concentration, is the σ_{xy}^{sk2} conductivity, subdominant in case of weak disorder. Interestingly, when the chemical potential lies in the gap (Fermi vector $k_F = 0$), all contributions to the conductivity except for the intrinsic one vanish. In this case the conductivity equals

$$\sigma_{xy} = -\frac{e^2}{(2\pi)^2} \int_0^\infty \Omega_z^-(k)\, d\mathbf{k} = -\frac{e^2}{4\pi}. \tag{93}$$

This quantization of the Hall conductivity in closely related to the *quantum spin Hall effect* [54, 57, 58] . In two dimensions, there exist expressions for the correction to the distribution function g_l [49, 50, 55]:

$$g_l = -\frac{\partial f_0(\varepsilon_l)}{\partial \varepsilon_l} \cdot eE_x \cdot (A_l|\mathbf{v}_l^0| \cos\varphi_l + B_l|\mathbf{v}_l^0| \sin\varphi_l), \tag{94}$$

where φ_l is the angle of $\mathbf{v}_l^0$ with the x-axis, and $A_l = \tau_l^{\parallel}$ and $B_l = (\tau_l^{\parallel}/\tau_l^{\perp})^2$ in the assumption that the transverse conductivity is much smaller than the longitudinal one, while the relaxation times are:

$$\frac{1}{\tau_l^{\parallel}} = \sum_{l'} \int \omega_{ll'} \left(1 - \frac{|\mathbf{v}_{l'}^0|}{|\mathbf{v}_l^0|} \cos(\varphi - \varphi')\right) \frac{d\mathbf{k}'}{2\pi}, \tag{95}$$

$$\frac{1}{\tau_l^{\perp}} = \sum_{l'} \int \omega_{ll'} \frac{|\mathbf{v}_{l'}^0|}{|\mathbf{v}_l^0|} \sin(\varphi - \varphi') \frac{d\mathbf{k}'}{2\pi}. \tag{96}$$

From (94) we see that g_l is a Fermi surface property. From the arguments around equation (69) which serve as a derivation of the anomalous distribution contribution to the distribution function it is rather clear that g_l^{ad} constitutes a Fermi surface property as well. This means that of all parts listed in Table 1 the only conductivity to which the whole Fermi sea contributes is the intrinsic conductivity σ_{xy}^{int}. This is probably why in many cases the calculated from first principles intrinsic conductivity corresponds to experimental values so well, while the only other conductivity which can compete with σ_{xy}^{int} is the conventional skew scattering due to its $1/n$ dependence on the impurity concentration, which can be quite large in the clean limit of small n. In the next section we consider two examples of how these two conductivities can be separated from each other.

Research on diluted magnetic semiconductors stimulated investigations of the AHE using different from semiclassical approaches. Due to relative simplicity of several important models, such as, e.g., 2D Rashba model, it was possible to perform rigorous quantum mechanical calculations using Kubo and Kubo-Streda formulas [50, 53, 59, 60, 61, 62] and via quantum Boltzmann and Keldysh techniques [63, 64, 65, 66]. Sinitsyn *et al.* [10, 50] demonstrated the one to one correspondence between semiclassical contributions to the AHE and the summation of relevant subseries of Feynman diagrams in the Kubo-Streda formulas and similar agreement was established with Luttinger's theory.

8 Separating different contributions to the AHE

It is useful to write down the current flowing in a 2D system under applied external electric field along the x axis (supposing that the sample is spatially uniform):

$$\begin{pmatrix} J_x \\ J_y \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & -\sigma_{xy} \\ \sigma_{xy} & \sigma_{xx} \end{pmatrix} \cdot \begin{pmatrix} E_x \\ 0 \end{pmatrix} = \begin{pmatrix} \sigma_{xx} E_x \\ \sigma_{xy} E_x \end{pmatrix}, \tag{97}$$

where $\hat{\sigma}$ is the conductivity tensor and σ_{xy} component contains the anomalous contribution to the Hall conductivity. Normally, $J_y \ll J_x$, and hence $\sigma_{xy} \ll \sigma_{xx}$. The resistivity tensor $\hat{\rho}$ is inverse to the conductivity tensor $\hat{\rho} = \hat{\sigma}^{-1}$, and its elements can be expressed in the following way:

$$\rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} \approx \frac{1}{\sigma_{xx}} \tag{98}$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \approx \frac{\sigma_{xy}}{\sigma_{xx}^2} \approx \sigma_{xy} \cdot \rho_{xx}^2 \tag{99}$$

The anomalous part of the transverse conductivity σ_{xy} can be written down in the following way:

$$\sigma_{xy} = \left(b + \frac{a}{\rho_{xx}} \right) = \sigma^{IAH} + \sigma^{sk}, \tag{100}$$

where we dropped the xy-subscript and denoted by σ^{sk} the conventional skew scattering part of the conductivity, while σ^{IAH} encodes intrinsic, side jump, intrinsic skew scattering and anomalous distribution part of the transverse conductivity. Finally, the anomalous Hall part of the transverse resistivity reads:

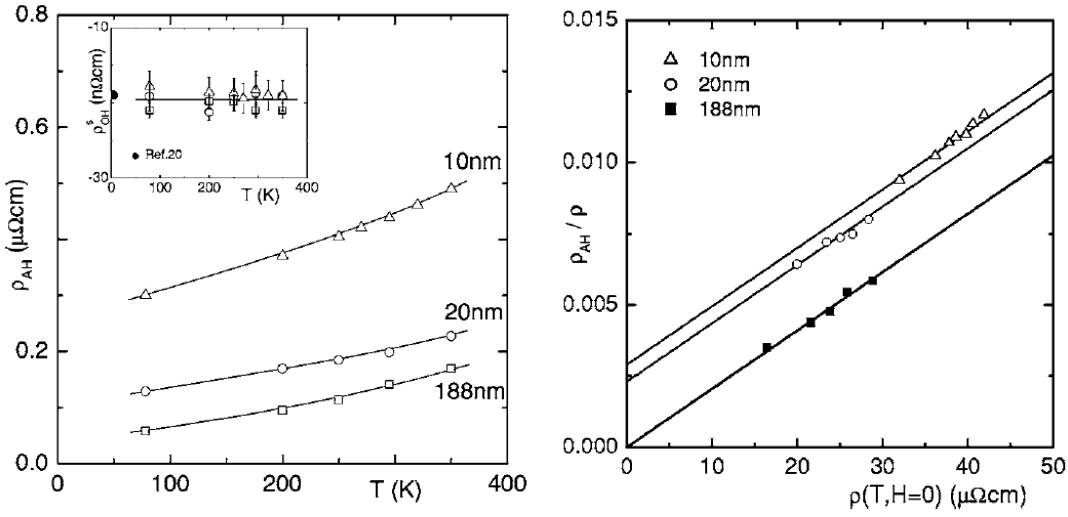


Fig. 8: *Left: Temperature dependence of the anomalous and ordinary (at $H = M_s$, inset) Hall resistivities; the solid lines are guides to the eye. Right: Analysis of the anomalous Hall resistivity in terms of the longitudinal resistivity $\rho(T)$. The straight lines are fits to Eq. (102) which allow to separate contributions to ρ_{AH} which are linear and quadratic in $\rho(T)$. In this graph $\rho \equiv \rho_{xx}$. Taken from [5].*

$$\rho_{xy} = \rho_{AH} = a\rho_{xx} + b\rho_{xx}^2 = a\rho_{xx} + \sigma^{IAH}\rho_{xx}^2 \tag{101}$$

This relation serves as a main playground for analyzing experimental data. Experimentally, it is a real challenge to separate the intrinsic contribution σ^{IAH} to the AHE from the skew scattering contribution, while distinguishing various parts of σ^{IAH} among each other for a given material is close to impossible at the current stage of development in the field. Moreover, on the experimental side data on the AHE were frequently fitted in terms of a single power law ρ_{xx}^n, were n was ranging from 1 up to 4 [18, 19, 68, 69]. In this section we would like to discuss two examples where a clear separation between the two contributions to the AHC could be successfully performed.

Kötzler and Gil [5] report on a systematic study of AHE in hcp cobalt. They performed measurements of the ρ_{OH}, ρ_{AH} and ρ_{xx} resistivities for $10-200$ nm thin polycrystalline Co films in the temperature range of $78-350$ K and found that the anomalous resistance of all films is dominated by a common, intrinsic term, while the extrinsic contribution due to skew scattering turns out to be small and related with the structural disorder in the films. For two films, the field dependence of several isothermal Hall resistivities is shown in Fig. 1. The anomalous and ordinary Hall resistivities $\rho_{AH}(T)$ and $\rho_{OH}(T)$ were determined as the saturation value of the anomalous Hall resistivity and the ordinary Hall resistivity at the saturation field $\rho_{OH}(H = M_s, T)$, respectively, as illustrated in this figure.

Investigating the temperature dependence of the AHE in this system, Kötzler and Gil found that both ρ_{AH} and ρ_{xx} increase with increasing temperature and the ρ_{AH} resistivity also decreases with increasing thickness of the films (see Fig. 8). In order to see how relevant decomposition (101) is, they plotted the ratio $\rho_{AH}(T)/\rho_{xx}(T)$ as a function of $\rho_{xx}(T)$:

$$\frac{\rho_{AH}}{\rho_{xx}} = a + \sigma^{IAH}\rho_{xx} \tag{102}$$

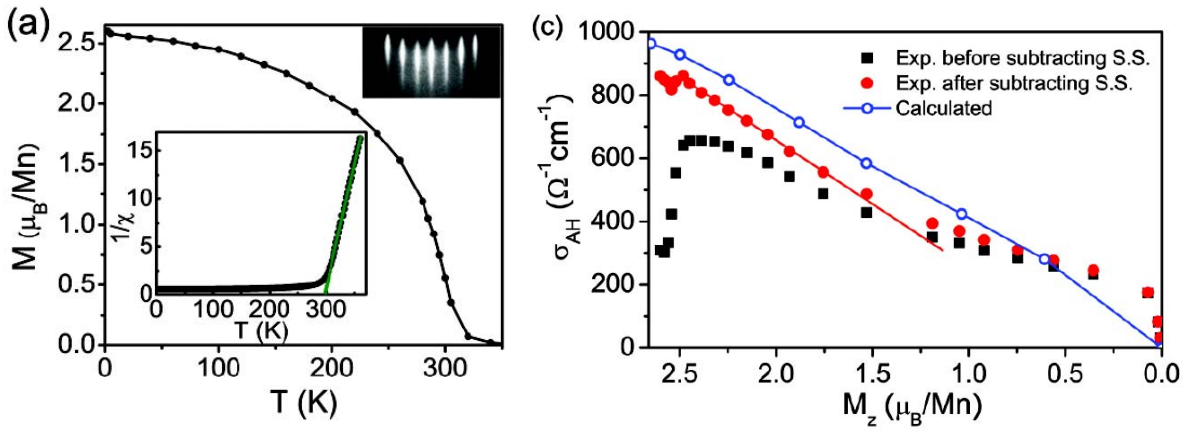


Fig. 9: *Left: Temperature-dependent spontaneous magnetization M for Mn_5Ge_3. Upper inset shows the RHEED pattern of a Mn_5Ge_3 thin film grown on Ge/GaAs(111). The lower inset shows the temperature dependence of the inverse magnetic susceptibility, measured at $T = 0.7$ T. The green line is a linear fit which extrapolates the T_c of 298 K. Right: Experimental AHC as a function of M before and after subtracting the skew scattering (S.S.) contribution. The solid line across the solid dots is a linear fit. Calculated σ_{IAH} values versus the z component of the magnetization M_z ($M_z = M = M_0 \cos\theta$, see text) are also shown. Taken from [46].*

As can be seen from Fig. 8, relation (102) holds quite nicely for a given film thickness and temperature range, which allows for a determination of constants a and σ^{IAH}. In principle, parameters a and σ^{IAH} via different mechanisms could depend explicitly on temperature T or magnetization of the sample M, which would destroy perfect linear dependence of $\rho_{AH}(T)/\rho_{xx}(T)$ on $\rho_{xx}(T)$. This is not the case for Co films, however: the estimated value of σ^{IAH} constitutes 205 $(\Omega \cdot \mathrm{cm})^{-1}$ and does not depend on the film thickness, temperature or details of the impurity scattering present in the system. This leads to the conclusion that in polycrystalline films of hcp Co the part of the AHC, proportional to the n^0, is mainly determined by the band structure Berry curvature related part σ_{xy}^{int}. The non-zero σ_{xy}^{sk1} part of the AHC can be calculated via extracting the non-zero offset a from Fig. 8.

A more complicated situation was reported and analyzed in Ref. [46] for ferromagnetic (0001) thin films of Mn_5Ge_3 grown on Ge(111). For this material, the magnetization M varies significantly in the range of considered temperatures up to 300 K (the spontaneous magnetization M as a function of temperature is presented in Fig. 9 (left)), so that it makes sense to assume that the coefficients a and σ^{IAH} of Eq. (102) depend on magnetization explicitly:

$$\frac{\rho_{AH}}{\rho_{xx}} = a(M) + \sigma^{IAH}(M)\rho_{xx}. \tag{103}$$

Without making further assumptions on the functional dependence of the quantities in Eq. (103) on the magnetization M it would be impossible to separate the intrinsic and skew scattering contributions from temperature or field-dependent measurements on a single sample because temperature changes M, $a(M)$, $\sigma^{IAH}(M)$ and ρ_{xx} simultaneously. Zeng *et al.* employed the result of Noziéres and Lewiner [28] that the conventional skew scattering coefficient a is normally linear in M, $a = \gamma M$, while coefficient γ can be obtained by plotting $\rho_{\mathrm{AH}}/M\rho_{xx}$ versus ρ_{xx}, as shown in Fig. 10 (right):

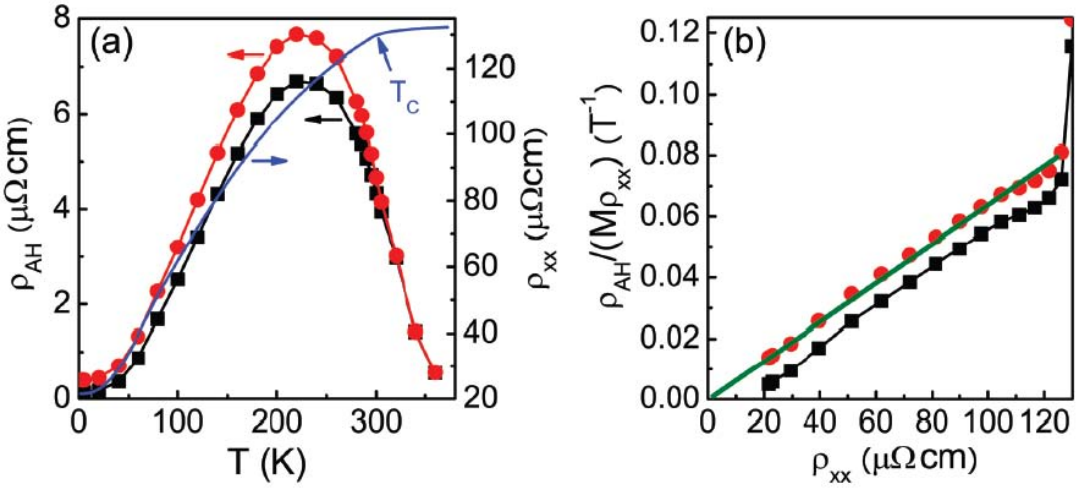


Fig. 10: *Left: ρ_{AH} as a function of temperature before (black squares) and after (red circles) subtracting the skew scattering contribution. ρ_{xx} is also shown with a blue line. Right: ratio $\rho_{AH}/M\rho_{xx}$ as a function of ρ_{xx} before (black squares) and after (red circles) subtracting the skew scattering term. The solid red line is a fit crossing the origin. Taken from [46].*

$$\frac{\rho_{AH}}{M\rho_{xx}} = \gamma + \frac{\sigma^{IAH}(M)}{M}\rho_{xx}. \tag{104}$$

This plot is linear below $\rho_{xx} \approx 110$ $\mu\Omega$·cm, or, equivalently, $T \approx 220$ K, which suggests that the intrinsic anomalous Hall conductivity in this material is linear in magnetization M in a wide range of temperatures:

$$\sigma^{IAH}(M) \sim M. \tag{105}$$

The parameter γ can be found as an offset by extrapolating to $\rho_{xx} = 0$, which, in turn, knowing the dependence of $M(T)$, $\rho_{AH}(T)$ and $\rho_{xx}(T)$ (Figs. 9 (left) and 10 (left), respectively) allows to reconstruct the dependence of σ_{xy} on M, presented in Fig. 9 (right, black squares):

$$\sigma_{xy}(M) = \frac{\rho_{AH}}{\rho_{xx}^2} = \frac{\gamma M}{\rho_{xx}} + \sigma^{IAH}(M), \tag{106}$$

from which, in turn, by subtracting the $\frac{\gamma M}{\rho_{xx}}$ term the intrinsic conductivity $\sigma^{IAH}(M)$ can be obtained (Fig. 9, right, red circles). We can see that indeed, while the total conductivity σ_{xy} is highly nonlinear in M, the intrinsic conductivity σ^{IAH} is linear in M in a wide interval of the magnetization and extrapolates to 860 $(\Omega \cdot \text{cm})^{-1}$ at zero temperature. These experimental findings are in a very good agreement with first principles calculations for the Berry curvature part of σ^{IAH} which give a value of 964 $(\Omega \cdot \text{cm})^{-1}$ at 0 K.

The linear dependence on the magnetization of $\sigma^{IAH}(M)$ can be qualitatively accounted for by the long-wavelength, low-frequency fluctuations of the spin orientation at low temperatures. According to experimental data the linear dependence is observed as far as the magnetization falls off quadratically with T (up to 240 K) [46]. In this range the magnitude of the magnetization M_0 stays roughly constant, but the local magnetization is deviating from the easy axis.

Making an assumption that the local conductivity is determined by the angle θ of the local magnetization M_0 with the easy axis one can determine the local intrinsic conductivity $\sigma^{IAH}(\theta, \varphi)$ by performing a calculation for an infinite ferromagnetic crystal with the magnetization at polar (θ, φ) angles with the easy axis. By averaging over the in-plane φ angle one can obtain a value for $\sigma^{IAH}(\theta) = \sigma^{IAH}(M = M_0 \cos\theta)$, which is presented in Fig. 9 (right, blue circles) and can be directly compared to experimental values (Fig. 9, right, red circles). This comparison appears to be surprisingly good until the vicinity of the T_C, where the interacting spin waves are responsible for a different fall off of the magnetization. In general, however, looking beyond Mn_5Ge_3, such "local approximation" which employs the rigid picture of the magnetization fluctuations [70, 71] serves as only one of the possible scenarios of the temperature or magnetization dependence of the anomalous Hall conductivity.

9 Acknowledgments

I would like to thank Ivo Souza, Eric Roman and David Vanderbilt for numerous discussions on the anomalous Hall effect which helped me to shape this manuscript greatly.

References

[1] E.H. Hall, *On a New Action of the Magnet on Electric Currents*, Amer. Jour. Math. **2**, 287 (1879)

[2] E.H. Hall, *On the new action of magnetism on a permanent electric current*, Philos. Mag. **10**, 301 (1880)

[3] C.M. Hurd, *The Hall Effect in Metals and Alloys*, Plenum, New York (1972)

[4] C.L. Chien and C.R. Westgate, *The Hall Effect and Its Applications*, Plenum, New York (1980)

[5] J. Kötzler and W. Gil, *Anomalous Hall resistivity of cobalt film: Evidence for the intrinsic spin-orbit effect*, Phys. Rev. B **72**, 060412 (2005)

[6] T. Hiraoka, *Hall Effect and the Pressure Effect in Single Crystals of Nickel*, J. Sci. Hiroshima **32**, 153 (1968)

[7] J.E. Hirsch, *Spin Hall Effect*, Phys. Rev. Lett. **83**, 1834 (1999)

[8] J. Sinova, D. Culcer, Q. Niu, N.A. Sinitsyn, T. Jungwirth and A.H. MacDonald, *Universal Intrinsic Spin Hall Effect*, Phys. Rev. Lett. **92**, 126603 (2004)

[9] Zh. Fang *et al*, *The Anomalous Hall Effect and Magnetic Monopoles in Momentum Space*, Science **302**, 92 (2003)

[10] N.A. Sinitsyn, *Semiclassical theories of the anomalous Hall effect*, J. Phys.: Condens. Matter **20**, 023201 (2008)

[11] F.D.M. Haldane, *Berry Curvature on the Fermi Surface: Anomalous Hall Effect as a Topological Fermi-Liquid property*, Phys. Rev. Lett. **93**, 206602 (2004)

[12] W. Yao, A.H. MacDonald and Q. Niu, *Optical Control of Topological Quantum Transport in Semiconductors*, Phys. Rev. Lett. **99**, 047401 (2007)

[13] J. Cumings, L.S. Moore, H.T. Chou, K.C. Ku, G. Xiang, S.A. Crooker, N. Samarth and D. Goldhaber-Gordon, *Tunable Anomalous Hall Effect in a Nonferromagnetic System*, Phys. Rev. Lett. **96**, 196404 (2006)

[14] J. Ye, Y.B. Kim, A.J. Millis, B.I. Shraiman, P. Majumdar and Z. Tesanovich, *Berry Phase Theory of the Anomalous Hall Effect: Application to Colossal Magnetoresistance Manganites*, Phys. Rev. Lett. **83**, 3737 (1999)

[15] K. Ohgushi, S. Murakami and N. Nagaosa, *Spin anisotropy and quantum Hall effect in the kagomé lattice: chiral spin state based on a ferromagnet*, Phys. Rev. B **62**, R6065 (2000)

[16] M. Onoda, G. Tatara and N. Nagaosa, *Anomalous Hall Effect and Skyrmion Number in Real and Momentum Spaces*, J. Phys. Soc. Jpn. **73**, 2624 (2004)

[17] R. Karplus and J.M. Luttinger, *Hall Effect in Ferromagnetics*, Phys. Rev. **95**, 1154 (1954)

[18] J. Smit, *The spontaneous Hall effect in ferromagnetic substances I*, Physica **21**, 877 (1955)

[19] J. Smit, *The spontaneous Hall effect in ferromagnetics II*, Physica **24**, 29 (1958)

[20] L. Berger, *Side-Jump Mechanism for the Hall Effect of Ferromagnets*, Phys. Rev. B **2**, 4559 (1970)

[21] L. Berger, *Comment on Side-Jump and Side-Slide Mechanisms for Ferromagnetic Hall effect: A Reply*, Phys. Rev. B **8**, 2351 (1973)

[22] L. Berger, *Account of scattering-independent contributions to the Hall conductivity in ferromagnets: A reply*, Phys. Rev. B **17**, 1453 (1978)

[23] J.M. Luttinger, *Theory of the Hall Effect in Ferromagnetic Substances*, Phys. Rev. **112**, 739 (1958)

[24] S.K. Lyo and T. Holstein, *Side-Jump Mechanism for Ferromagnetic Hall Effect*, Phys. Rev. Lett. **29**, 423 (1972)

[25] E.N. Adams and E.I. Blount, *Energy bands in the presence of an external force field - II: Anomalous velocities*, J. Phys. Chem. Solids. **10**, 286 (1959)

[26] E.I. Blount, *Solid State Physics*, vol. 13 (New York: Academic), p. 305

[27] J.N. Chazalviel, *Spin-dependent Hall effect in semiconductors*, Phys. Rev. B **11**, 3918 (1975)

[28] P. Noziéres and C. Lewiner, *A simple theory of the anomalous Hall effect in semiconductors*, J. Phys. France **34**, 901 (1973)

[29] G. Sundaram and Q. Niu, *Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and Berry-phase effects*, Phys. Rev. B **59**, 14915 (1999)

[30] N. Nagaosa, *Anomalous Hall Effect - A New Perspective* -, J. Phys. Soc. Japan **75**, 042001 (2006)

[31] N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (W.B. Saunders, Philadelphia, 1976)

[32] D.J. Thouless, M. Kohmoto, M.P. Nightingale and M. den Nijs, *Quantized Hall Conductance in a Two-Dimensional Periodic Potential*, Phys. Rev. Lett. **49**, 405 (1982)

[33] G.D. Mahan, *Many-Particle Physics* (Plenum, 1990)

[34] N. Nagaosa, *Spin Currents in Semiconductors, Metals and Insulators*, J. Phys. Soc. Japan **77**, 031010 (2008)

[35] J.E. Avron, D. Osadchy and R. Seiler, *A Topological Look at the Quantum Hall Effect*, Physics Today **38** (Aug. 2003)

[36] T. Jungwirth, Q. Niu and A.H. MacDonald, *Anomalous Hall Effect in Ferromagnetic Semiconductors*, Phys. Rev. Lett. **88**, 207208 (2002)

[37] T. Jungwirth, J. Sinova, K.Y. Wang, K.W. Edmonds, R.P. Campion, B.L. Gallagher and C.T. Foxon, *Dc-transport properties of ferromagnetic (Ga,Mn)As semiconductors*, Appl. Phys. Lett. **83**, 320 (2003)

[38] Y. Yao, L. Kleinman, A.H. MacDonald, J. Sinova, T. Jungwirth, D.-S. Wang, E. Wang and Q. Niu, *First Principles Calculation of Anomalous Hall Conductivity in Ferromagnetic bcc Fe*, Phys. Rev. Lett. **92**, 037204 (2004)

[39] P.N. Dheer, *Galvanometric Effect in Iron Whiskers*, Phys. Rev. **156**, 637 (1967)

[40] M. Onoda and N. Nagaosa, *Topological Nature of Anomalous Hall Effect in Ferromagnets*, J. Phys. Soc. Japan **71**, 19 (2002)

[41] P.A.M. Dirac, *Quantized Singularities in the Electromagnetic Field*, Proc. R. Soc. London **133**, 60 (1931)

[42] G. 't Hooft, Nucl. Phys. **B79**, 276 (1974)

[43] A.M. Polyakov, JETP Lett. **20**, 194 (1974)

[44] X. Wang, J.R. Yates, I. Souza and D. Vanderbilt, *Ab initio calculation of the anomalous Hall conductivity by Wannier interpolation*, Phys. Rev. B **74**, 195118 (2006)

[45] X. Wang, D. Vanderbilt, J.R. Yates and Ivo Souza, *Fermi-surface calculation of the anomalous Hall conductivity*, Phys. Rev. B **76**, 195109 (2007)

[46] C. Zeng, Y. Yao, Q. Niu and H.H. Weiterling, *Linear Magnetization Dependence of the Intrinsic Anomalous Hall Effect*, Phys. Rev. Lett. **96**, 037204 (2006)

[47] A.M. Perelomov and Ya. B. Zel'dovich, *Quantum Mechanics: Selected Topics*, (Singapore, World Scientific, 1999)

[48] N.A. Sinitsyn, Q. Niu, J. Sinova and K. Nomura, *Disorder effects in the anomalous Hall effects induced by Berry curvature*, Phys. Rev. B **72**, 045346 (2005)

[49] N.A. Sinitsyn, Q. Niu and A.H. MacDonald, *Coordinate shift in the semiclassical Boltzmann equation and the anomalous Hall effect*, Phys. Rev. B **73**, 075318 (2006)

[50] N.A. Sinitsyn, A.H. MacDonald, T. Jungwirth, V.K. Dugaev and J. Sinova, *Anomalous Hall effect in a two-dimensional Dirac band: The link between the Kubo-Streda formula and the semiclassical Boltzmann equation approach*, Phys. Rev. B **75**, 045315 (2007)

[51] E. Merzbacher, *Quantum Mechanics*, (Wiley, New York, 1970)

[52] V.I. Belinicher, E.L. Ivchenko and B.I. Sturman, Zh. Eksp. Teor. Fiz. **83**, 649 (1982)

[53] N.A. Sinitsyn, J.E. Hill, H. Min, J. Sinova and A.H. MacDonald, *Charge and Spin Hall Conductivity in Metallic Graphene*, Phys. Rev. Lett. **97**, 106804 (2007)

[54] C.L. Kane and E.J. Mele, Z_2 *Topological Order and the Quantum Spin Hall Effect*, Phys. Rev. Lett. **95**, 146802 (2005)

[55] J. Schliemann and D. Loss, *Anisotropic transport in a two-dimensional electron gas in the presence of spin-orbit coupling*, Phys. Rev. B **68**, 165311 (2003)

[56] P. Leroux-Hugon and A. Ghazali, *Contribution to the theory of the anomalous Hall effect: Influence of the band structure on the skew scattering*, J. Phys. C: Solid State Phys. **5**, 1072 (1972)

[57] C.L. Kane and E.J. Mele, *Quantum Spin Hall Effect in Graphene*, Phys. Rev. Lett. **95**, 226801 (2005)

[58] L. Sheng, D.N. Sheng, C.S. Ting and F.D.M. Haldane, *Nondissipative Spin Hall Effect via Quantized Edge Transport*, Phys. Rev. Lett. **95**, 136602 (2005)

[59] V. Dugaev, P. Bruno, M. Taillefumier, B. Canals and C. Lacroix, *Anomalous Hall effect in a two-dimensional electron gas with spin-orbit interaction*, Phys. Rev. B **71**, 224423 (2005)

[60] J.I. Inoue, T. Kato, Y. Ishikawa, and H. Itoh, G.E.W. Bauer and L.W. Molenkamp, *Vertex Corrections to the Anomalous Hall Effect in Spin-Polarized Two-Dimensional Electron Gases with a Rashba Spin-Orbit Interaction*, Phys. Rev. Lett. **97**, 046604 (2005)

[61] V.K. Dugaev, A. Crépieux and P. Bruno, *Localization corrections to the anomalous Hall effect in a ferromagnet*, Phys. Rev. B **64**, 104411 (2001)

[62] A. Crépieux and P. Bruno, *Theory of the anomalous Hall effect from the Kubo formula and the Dirac equation*, Phys. Rev. B **64**, 014416 (2001)

[63] S.Y. Liu and X.L. Lei, *Kinetic-equation approach to the anomalous Hall effect in a diffusive Rashba two-dimensional electron system with magnetization*, Phys. Rev. B **72**, 195329 (2005)

[64] S.Y. Liu, N.J.M. Horing and X.L. Lei, *Anomalous Hall effect in Rashba two-dimensional electron systems based on narrow-band semiconductors: Side-jump and skew scattering mechanisms*, Phys. Rev. B **74**, 165316 (2006)

[65] S. Onoda, N. Sugimoto and N. Nagaosa, *Theory of Non-Equilibirum States Driven by Constant Electromagnetic Fields Non-Commutative Quantum Mechanics in the Keldysh Formalism* , Prog. Theor. Phys. **116**, 61 (2006)

[66] D. Culcer and Q. Niu, *Geometrical phase effects on the Wigner distribution of Bloch electrons*, Phys. Rev. B **74**, 035209 (2006)

[67] A.A. Kovalev, K. Výborný and J. Sinova, *Hybrid skew scattering regime of the anomalous Hall effect in Rashba systems: Unifying Keldysh, Boltzmann, and Kubo formalisms*, Phys. Rev. B **78**, 041305 (2008)

[68] P. Xiong, G. Xiao, J.Q. Wang, J.Q. Xiao, J.S. Jiang and C.L. Chien, *Extraodinary Hall effect and giant magnetoresistance in the granular Co-Ag system*, Phys. Rev. Lett. **69**, 3220 (1992)

[69] A. Gerber, A. Milner, A. Finkler, M. Karpovski, L. Goldsmith, J. Tuaillon-Combes, O. Boisron, P. Mélinon and A. Perez, *Correlation between the extraodinary Hall effect and resistivity*, Phys. Rev. B **69**, 224403 (2004)

[70] C. Zener, *Classical Theory of the Temperature Dependence of Magnetic Anisotropy Energy*, Phys. Rev. **96**, 1335 (1954)

[71] H.B. Callen and E. Callen, *The present status of the temperature dependence of magnetocrystalline anisotropy, and the* $l(l+1)/2$ *law*, J. Phys. Chem. Solids **27**, 1271 (1966)

C 4 Quantum Spin Hall Effect

H. Buhmann
Physikalisches Institut, EP3
Universität Würzburg

Contents

1 Introduction **2**

2 Quantum Spin Hall Effect **2**

3 HgTe Quantum Wells **4**

4 Measurement of the QSH effect **6**

5 Non-locality of the QSH effect **8**

6 Stability of the QSH effect **9**

7 Applications **9**

8 Conclusion **11**

1 Introduction

Spin polarization, manipulation and detection are currently the most challenging tasks in semiconductor science. Spin injection in metals already revolutionized the data storage on hard disk drives [1, 2]. However, spin injection into semiconductors has turned out to be much more difficult. This is basically due to the facts that on the one hand injection of spin polarized electron from a ferro-magnetic metal into a semiconductor is very inefficient due to the enormous mismatch of the density of states and on the other hand that only few ferro-magnetic semiconductor materials exist which exhibit Currie temperatures below 100 K. Thus, the search for spin polarizing effects in semiconducting materials guides a lot of research activities.
So far, spin injection and spin valve behaviour have been realized in magnetic semiconductors and ferro-magnetic semiconductor tunnel devices [3]. Another effect, which has recently been discusses as a candidate for spin polarization in semiconductors, is the spin-orbit (SO) interaction. In crystal structures the lack of inversion symmetry can lift the spin degeneracy. Two prominent effects are the bulk inversion asymmety (BIA) or Dresselhaus-effect and the structural inversion asymmetry (SIA). The latter is a dominant effect in quantum well (QW) structures and know as Rashba-effect [4]. The symmetry of the QW potential is controllable to a large extend via external gate-voltages which makes this effect one possible candidate to spin-controlled transport in semiconductors. The most famous device concept in this context is the spin field effect transistors or Datta-Das transistor [5]. But again for a realization a controlled spin polarization injection and detection in semiconductor systems is required. Concepts like the spin Hall effect (SHE) [6, 7, 8, 9] and birefringent electron optics [10] are explored with promising results.
Another effect, which is known for quite a long time now and which is related to spin polarized charge transport, is the quantum Hall effect (QHE) [11]. For odd filling factors (ν), and especially for $\nu = 1$, the current is highly spin polarized. The only drawback is the fact that in order to create QHE states the device has to be subjected to a high external magnetic field which disqualifies this effect from possible application purposes. More desirable would be an effect which provides well defined stable transport channels for spin polarized carriers similar to the QHE states.
In this presentation I will discuss the so-called quantum spin-Hall effect (QSHE) which is observed in the bulk insulating regime in materials with a very high SO coupling. Due to band structure effects counter propagating edge channels are formed which are oppositely spin polarized [12]. This effect has been proposed to be observable in narrow graphene structure [13] and which has been demonstrated experimentally in HgTe QW structures. For an understanding of the fundamental details of the QSHE I will first introduce the concept of the QSHE on the basis of the QHE and discuss the special properties of HgTe QW structures. Then I will present the experimental signatures of the QSHE and discuss limits of stability and possible appliplications.

2 Quantum Spin Hall Effect

The most successful picture for the explanation of the experimental observation in the quantum Hall (QH) regime is the formation of one dimensional edge channels. Figure 1 sketches the situation of an electron moving in a magnetic field perpendicular to the plane of motion. Classically, the orbits are closed and localized in the center of the sample while at the edge of the sample skipping orbits contribute to an effective charge transport. Quantum mechanically

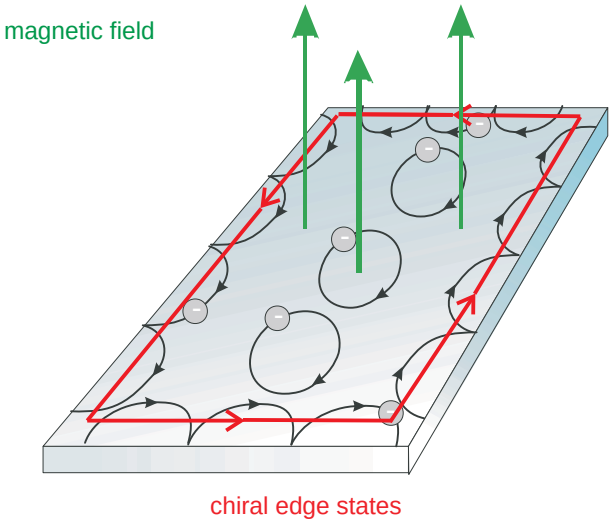


Fig. 1: *Schematic view of the electron motion in a magnetic field. Classically, closed cyclotron and skipping orbit explain the transport phenomena. Quantum mechanically, the skipping orbits are represented by an one-dimensional (chiral) edge channel (red line).*

the skipping orbits are represented by one-dimensional edge channels. In the QH regime the number of such edge channels is given by the number of occupied Landau levels (LL). The quantized Hall conductance as well as the vanishing longitudinal resistance can be explained within the Landauer-Büttiker formalism [14].

In this picture the QHE can be view as a first example for a non-trivial insulator. The stability of the quantization effect is given by the fact that left and right moving channel are confined to opposite sample edges and backscattering is prohibited as long as the edge channel wave function do not overlap. Even the presence of scattering potentials located at the sample edges do not influence the current transport. The edge channel will move around this scatterer on an equipotential line (cf. Fig. 2).

Using this model picture he QSH state can now be thought as two copies of the QH state, one for each spin component i.e., the appropriate magnetic field direction has to be opposite for each copy. The situation of two separate copies is shown in Fig. 3 a). A combination of both states in a single sample will result in two counter propagating edge channels with opposite spin [cf. Fig. 3 b)]. This new state does not break time reversal symmetry and exists even without magnetic field [12]. It should be noted that even though counter propagating channels are now

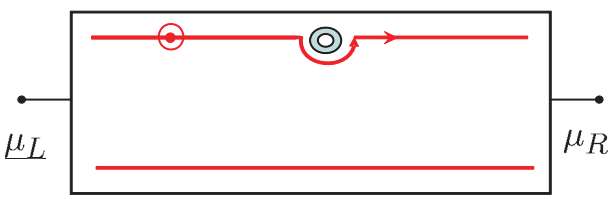


Fig. 2: *Schematic view of the electron transport experiment in the QHE regime. Current transport takes place only in the upper channel. Small imposed potentials do not result in backscattering.*

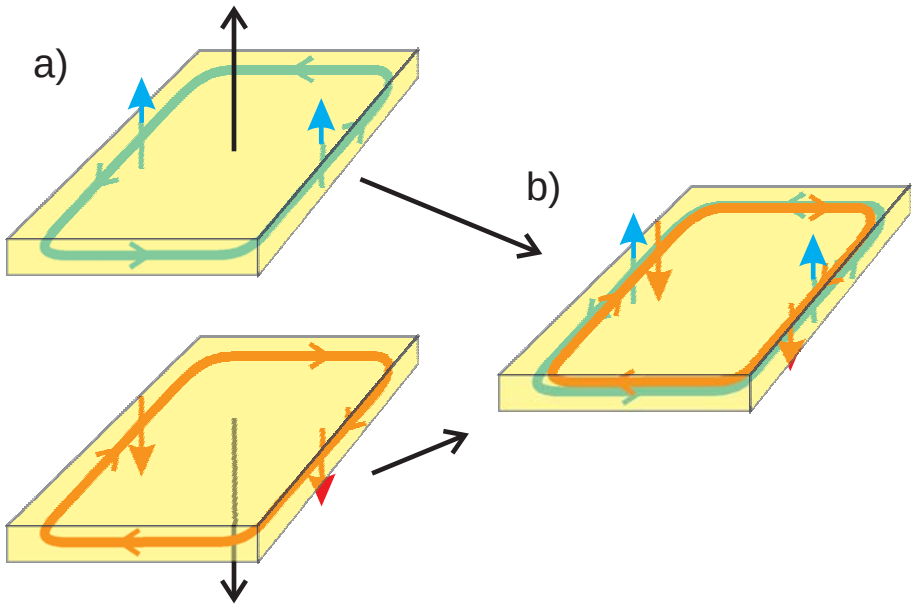


Fig. 3: *a) Two copies of a $\nu = 1$ chiral QHE edges state for inverted magnetic field. Black arrow indicates the direction of the magnetic field and blue and red arrows the spin polarization directions. b) A combination of two $\nu = 1$ QHE states leads to a quantum spin Hall state without magnetic field.*

present at each edge, backscattering is not possible because these states are so-called Kramer's doublets, $\Psi_{k,s^+}, \Psi_{-k,s^-}$, protect by time reversal symmetry. In order to introduce backscattering more than two Kramer's pairs have to be present at the same edge [12], as for example in the metallic conducting regimes.
The question which rises is now: In what kind of materials does the quantum spin Hall insulator state exist? The first candidate which was discussed in literature was graphene [13]. For certain edge configurations the existence of helical edge states has been predicted for narrow channels. But due to the very small SO energy in graphene, the resulting insulating state is far too small for QSH edge channel transport to be observable with current experimental techniques. In December 2005 Bernevig *et al.* [15] presented calculations which showed that the QSH insulator state should be observable in HgTe QWs with an inverted subband structure ordering.

3 HgTe Quantum Wells

Bulk HgTe is a semimetal with a degenerate heavy hole (HH) and light hole state (LH) of the Γ_8 symmetry band at the Γ point (cf. Fig. 4). Due to the strong SO interaction in this material the p-orbital like Γ_8 band is energetically shifted above the s-orbital like Γ_6 band. The related energy difference is of about 300 meV. From this, HgTe can also be viewed as a semiconductor with a negative band gap ($E_g := E_{\Gamma_6} - E_{\Gamma_8} = -300$ meV). If a thin layer of HgTe is embedded between $Hg_{1-x}Cd_xTe$ (with $x > 10\%$) the degeneracy between LH and HH state is lifted due to the quantum confinement and a gaped semiconductor is formed.
$Hg_{0.32}Cd_{0.68}Te$ is a normal gap semiconductor as shown in Fig. 4. The formed HgTe/HgCdTe-quantum well (QW) is a so-called type-III structure where the barrier conduction band edge is

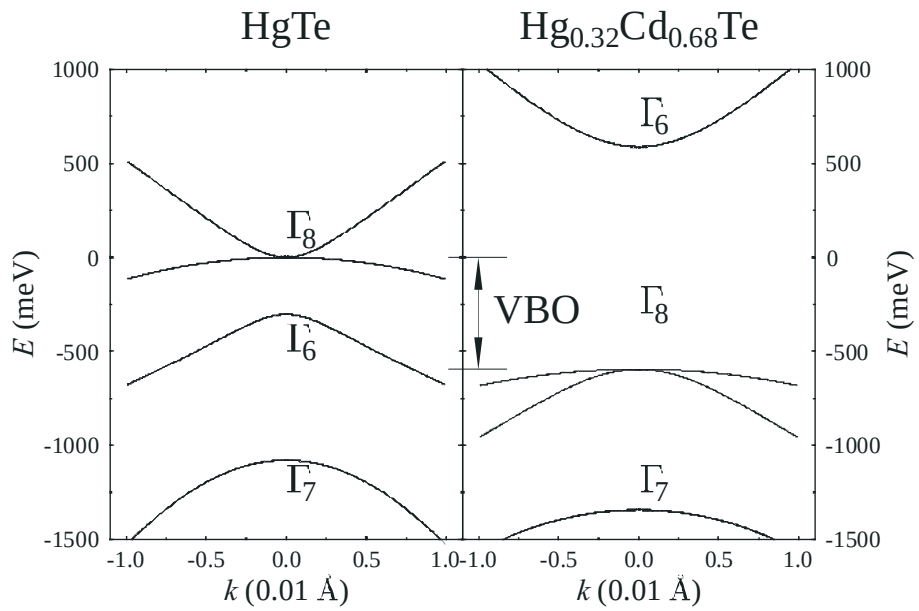


Fig. 4: *Energy dispersion for conduction and valence band of HgTe and HgCdTe near the* Γ *point. Indicated by arrows is the valence band offset (VBO* $\approx 570meV$*).*

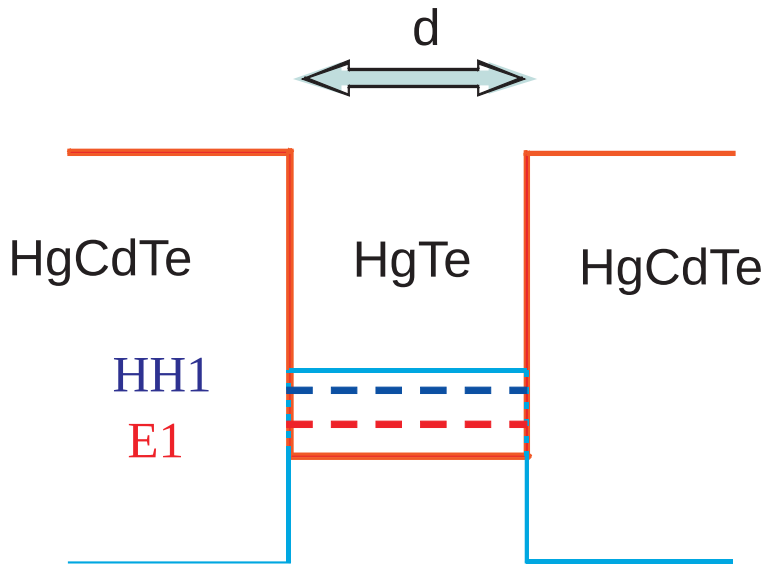


Fig. 5: *Type-III HgTe QW structure with inverted subband structure ordering: HH1 conduction and E1 valence band.*

below the valence band edge within the QW (cf. Fig. 5). For wide wells the confined Γ_8-like levels are above the Γ_6-like. Such a situation is schematically shown in Fig. 5, which depicts a type-III QW with an inverted subband structure ordering. For a larger confinement the subband structure ordering becomes again normal i.e., $E_{E1} > E_{HH1}$. However, for the observation of the quantum spin Hall-effect the inverted subband structure ordering is essential which is the case for QW widths, d, larger than 6.3 nm [16]. In this case, the QW exhibits a HH1-like conduction band and an E1-like valance band with a finite gap. As band structure calculations in a tight binding nearest neighbor hoping model showed two bands of two counter propagation anti-

parallel spin-polarized electronic one-dimensional states exist at the the sample edges within the bulk energy gap. These states exhibit the expected properties for the formation of the quantum spin Hall (QSH) effect [15].

4 Measurement of the QSH effect

The existence of the QSHE edge states can be demonstrated when the Fermi energy of the QW structure is located within the bulk energy gap. Then, electronic transport is expected to be carried by two edge channels, which are oppositely polarized and propagating at opposite sample edges. For a sample in Hall-bar geometry the situation is visualized schematically in Fig. 6. Each color symbolizes one spin polarized state. A current that enters from the left into the sample will propagate with one spin polarization along the top (blue) edge channel and with the opposite spin state along the bottom edge (red channel). The conductance of a ballistic one-dimensional channel is quantized and corresponds to e^2/h.

For the measurement we used 7 nm wide HgTe-QW structures with and n-type doping. The carrier concentration was $n_s \approx 3 \times 10^{11}$ cm^{-2} and the carrier mobility was determined to be of the order of 150000 cm$^2/(Vs)$ which yields an elastic mean free path of $1 - 2$ μm. In order to assure that elastic scattering does not influence the carrier transport samples have to be fabricated with dimensions of approximately the same order.

Within the ballistic limit, edge channel transport is independent of the sample length, L, and width, W. Thus, various samples were fabricated of various dimensions. Note, that all samples were equipped with a top gate electrode which allows for a continuous variation of the carrier density. The gate electrodes are only slightly larger than the devices. Thus, the contact regions remain n-type conducting which assures the possibility of transport measurements. Starting from an n-type conductance at zero gate-voltage the Fermi energy passes the bulk insulating regime for increasing negative gate-voltages and finally reaches a p-type conducting regime for voltages smaller than -2 to -4 V. The latter is a due to a rather narrow band gap of about 30 to

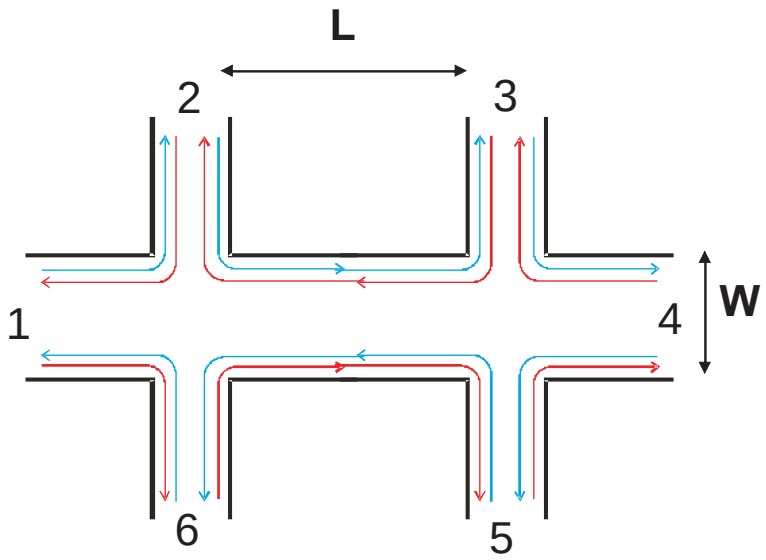


Fig. 6: *Multi-terminal device sturcture of width W and length L. Indicated are the spin polarized edge channels in the QSH regime by colors (red and blue) as well as the ohmic contact by number ($1 \ldots 6$).*

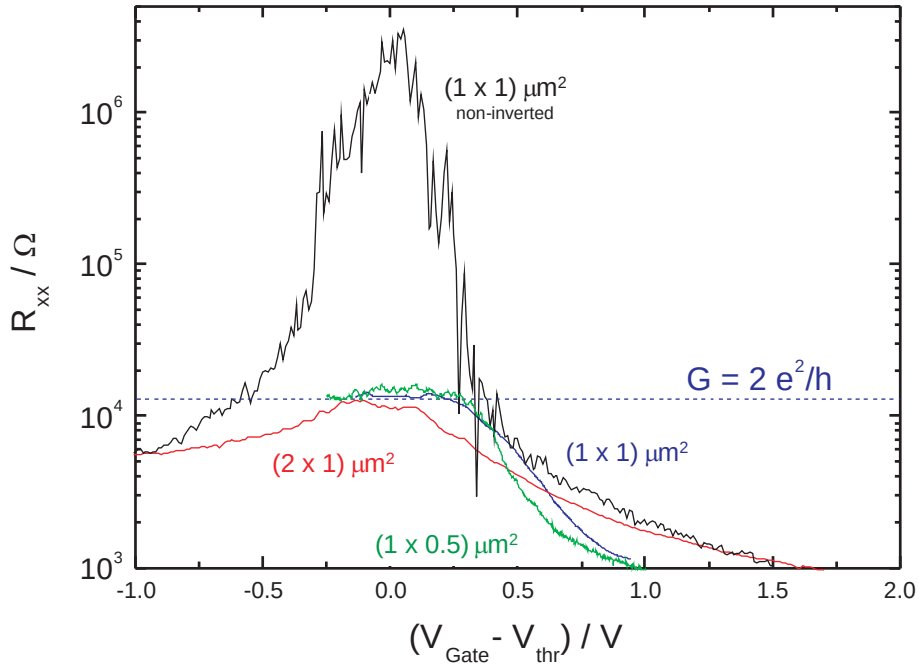


Fig. 7: *QSHE measurement for inverted QW structures (red, blue and green) and normal QW (black). The device dimensions $L \times W$ are given in the figure.*

50 meV for the current device structures.
In Fig. 7 the results are shown for four-terminal measurements on different devices. For all inverted QW devices the resistance $R_{14,23}$ reaches a value of $h/2e^2$ ($\approx$ 12.9 kΩ) in the bulk insulating regime while normal QW structures ($d < 6$ nm), which have been measuresed for comparison, display a real insulating behaviour. The measured resistance value suggests indeed a quantized conductance exists which does not depend on the device dimensions.
A value of $R = h/2e^2$ would have been expected for a two-terminal device in the QSH regime. In order to estimate the resistance value of a multi-terminal device in four-probe geometry the Landauer-Büttiker formalism for one-dimensional ballistic conductors has to be employed [14]. Considering the two counter propagating edge channels and the fact that within the ohmic contacts an equilibration of the two spin channels is possible, the resistance can be calculated by using the following transmission matrix:

$$T = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 & 1 & -2 \end{bmatrix} . \tag{1}$$

Solving the equation

$$I_i = e/h \sum_{j \neq i} T_{ij} \left(\mu_i - \mu_j\right) \tag{2}$$

yields $R_{14,23} = h/2e^2$, a value which has been observed for all measurement using this contact

configuration.

5 Non-locality of the QSH effect

Following the Landauer-Büttiker formalism it is possible to verify the ballistic edge channel transport concept of the QSH state by investigating different contact configurations. From Eqs. (1) and (2) one infers that, for example, the two-terminal resistance $R_{14,14}$ should be $3h/2e^2 \approx 38.7\ \mathrm{k}\Omega$. The agreement between model and experiment is shown in Fig. 8. More different contact configurations which confirm the non-locality of the QSH effect can be found in Ref. [17].

It should be noted that despite the similarity between quantum Hall effect and QSH effect the conductance quantization in the latter case will strongly depend on the quality of the samples. While in the QH regime edge channel transport will not be influenced by potentials fluctuations (cf. Fig. 2) such fluctuation destroy the insulating state in the QSH regime by introducing locally n- or p-conducting metallic regions. Such metallic regions at the edge of the sample act as additional metallic contacts and equilibrate the chemical potentials of left and right moving states i.e., introducing backscattering [17]. The probability of finding such metallic region at the sample edge increases with increasing device size and with decreasing sample quality. Thus, the device sizes for a quantized conductance in the QSH regime is limited mainly by the elastic mean free path, in contrast to the QHE.

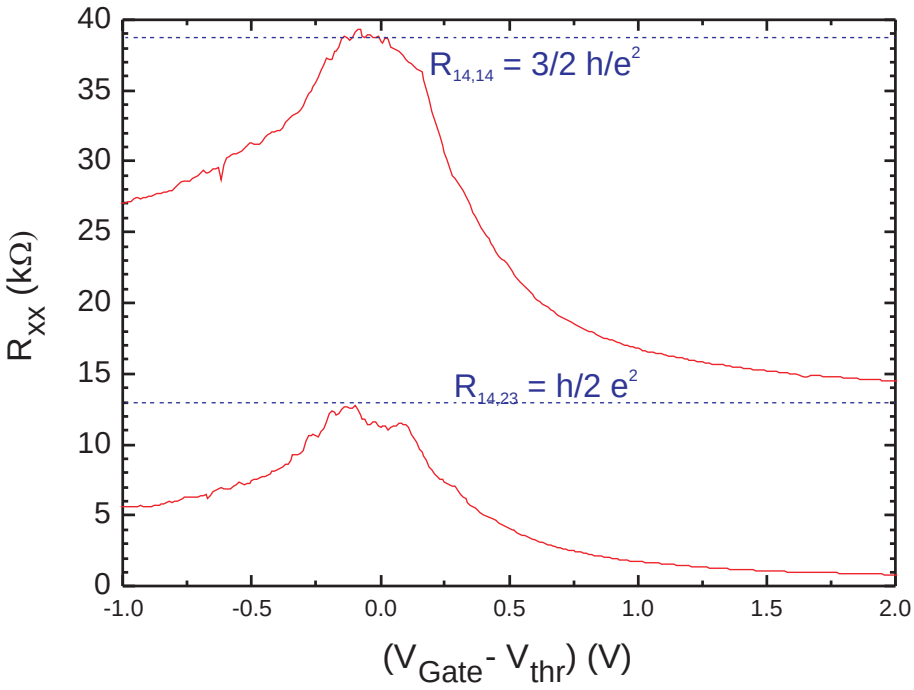


Fig. 8: *Quantized resistance for a two- ($R_{14,14}$) and four-terminal ($R_{14,23}$) measurement configuration in the QSH regime.*

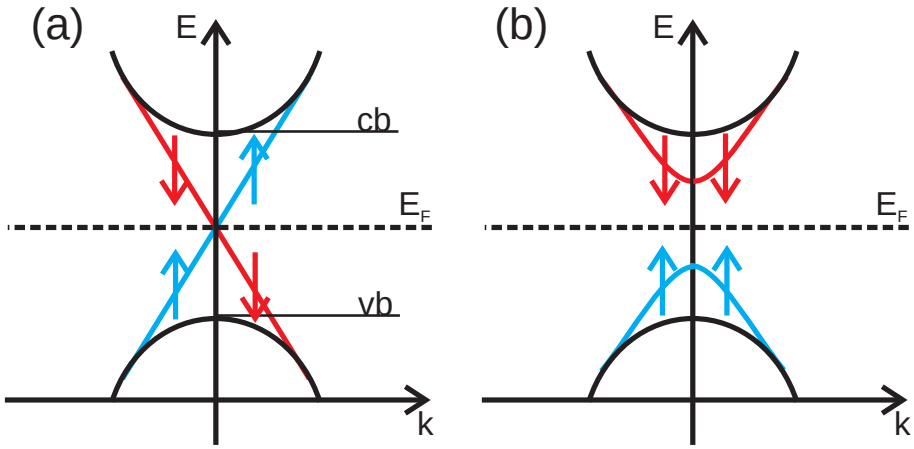


Fig. 9: *Schematic energy dispersion for QSH edge channels without (a) and with (b) external magnetic field.*

6 Stability of the QSH effect

The influence of potential fluctuations on the quantized conductance in the QSH regime will mainly depend on the actual energy gap and especially on the position of the Fermi energy with respect to the conduction and valence band edge. Fig. 9 (a) gives a simple schematic illustration of the energy dispersion at the sample edge. If the Fermi energy is located at mid-gap position the stability is largest. In case of HgTe QWs the energy gap is of the order of 30 to 50 meV which is quite large and assures that the QSH effect is observable for temperatures up to at least 4 K [18].

Edge channel transport is also limited by the extension of the edge channel wave function into the bulk. If the sample width decreases below a certain limit the wave functions of opposite edges overlap and backscattering can take place [19]. An estimation shows that this might happen for device sizes lower than 200 . . . 250 nm, which has been confirmed by measurements on narrow device structures [9].

Another effect which influences the stability of the QSH state will be introduced by an external magnetic field. The magnetic field destroys the time reversal symmetry and opens a gap in the linear edge channel dispersion [schematically shown in Fig. 9 (b)]. The strength of the gap depends on the direction of the magnetic field relativ to the plane of the QW [16, 20]. Fig. 10 shows the experimental dependence of the QSH effect on the direction of magnetic field. The influence is strongest if the magnetic field is applied perpendicular to the QW plane. The quantized conductance is destroy at a field strength of about 25 mT. In parallel configuration the critical magnetic field has to be at least ten times higher to result in the same effect [20].

7 Applications

The remaining question is whether or not it is possible to use the QSH effect for spin injection, detection, and control purposes. In order to explore this possibility systems or regions that exhibits an QSH state have to be connected to systems with different defined spin transport properties.

A QSH system can be used to separate spin states in multi terminal devices (cf. 6). Thus, it

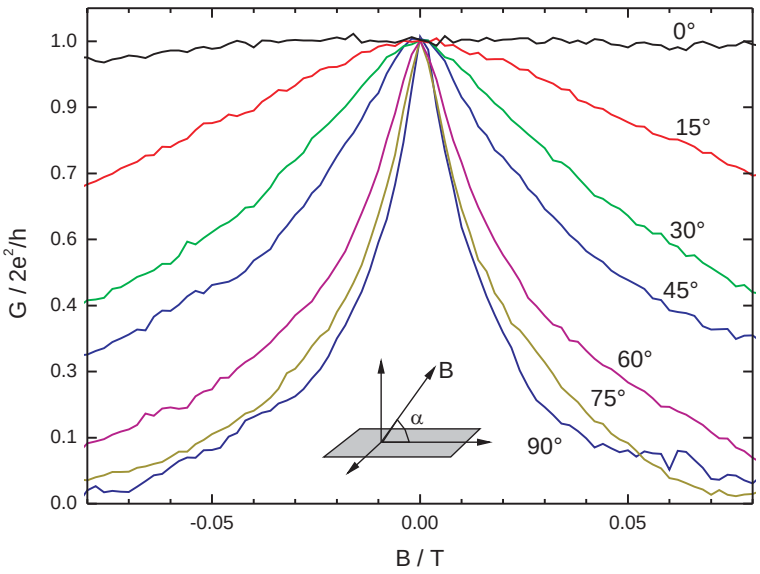


Fig. 10: *Angular magnetic field dependence of the four-termial resistance ($R_{14,23}$) The angle between the magnetic field direction and the QW plane is indicated in the figure.*

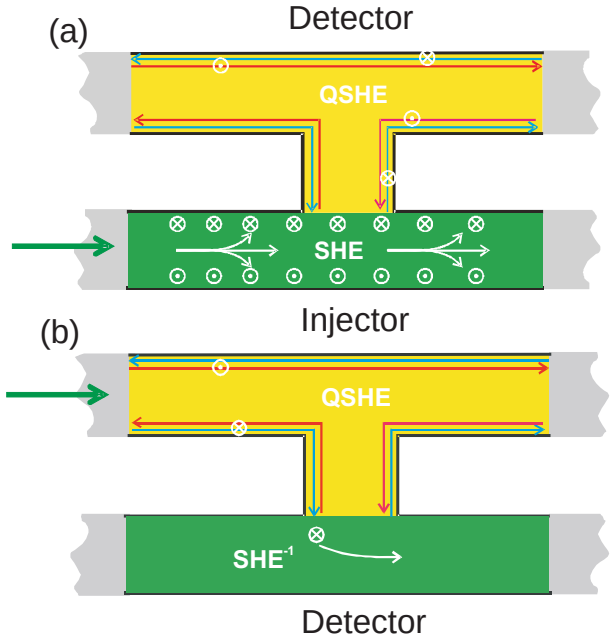


Fig. 11: *Schematic measurement configuration for the demonstration of the spin injection and detection properties of the QSH effect. Yellow: device region gate into the QSH regime, green: metallic, either n- or p-type semiconductor.*

can easily be used to transport and detect spin polarized currents or to inject spin polarized currents into a metallic regime. However, it will not be an easy task to fabricate planar devices of different semiconductor materials. Thus, for first application the spin properties of metallic HgTe QW have to be used. Due to the narrow gap usual optical spin detection and injection techniques cannot be employed. Hence, electronic spin dependent transport mechanisms have to be employed.
HgTe QWs exhibit one of the larges Rashba SO splitting for semiconductors. Rashba energies of up to 30 meV have been observed [21]. In systems with a strong Rashba SO splitting the spin Hall effect (SHE) can be used to create spin polarizations at the sample edges [7, 8, 22]. Indeed, an intrinsic SHE signal has been observable for metallic HgTe QW structures in an all-electric measurements [9]. However, it turned out that only for the p-type conducting regime the spin related voltage signal was detectable. For the n-type conducting regime the spin signal remained unresolved.
The idea is now to combine a QSH system with a SHE system using split-gate technologies. Thus, these two regimes, which exist in HgTe QW for different gate-voltages, can be realized in close vicinity. Schematic device structures are shown in Fig. 11. Both regions are connected by a single contact. This device can now be used in two configurations: First, spin polarization is created using the SHE. The spin accumulation at the sample edge is then detected by the QSH edge channels [Fig. 11 a)]. Second, a spin polarized current is injected from the QSH regime into the metallic n- or p-type system. The inverse SHE [6, 9] creates a voltage signal due to the injected spin current [Fig. 11 b)]. Measurements of all configurations confirm these application concept and moreover, reveal a very high spin sensitivity because even a SHE signal in the n-type conducting regime became detectable [23], which has been unresolved in an all-electric SHE measurement configuration [9].

8 Conclusion

The quantum spin Hall effect (QSHE) is a new non-trivial insulator state which exhibits extraordinary spin transport properties with a high potential for spintronics applications. The big advantages of this state are that neither external magnetic field or magnetic materials nor polarized light sources are required for the creation or detection of spin polarized carriers, and that spin states are transported dissipationless in ballistic one-dimensional edge channels.
Up to now, experimentally, this effect has only been demonstrated in HgTe quantum well systems, but the search for further materials with these extraordinary properties is on and its realizations are awaited.

Acknowledgements

The experimental work on HgTe QW structures is done at the Physikalisches Institut, Lehrstuhl für Experimentelle Physik (Prof. L.W. Molenkamp), Universität Würzburg, by the Quantum Transport Group, (Prof. H. Buhmann, C. Brüne, M. König, M. Leberecht, R. Rommel, A. Roth, S. Wiedmann, in collaboration with the following theoreticians: E.M. Hankiewicz , E.G. Novik (Univ. Würzburg), X.-L. Qi, T. Hughes, and S.-C. Zhang (Standford University).

References

[1] M.N. Baibich, J.M. Broto, A. Fert, F. Nguyen Van Dau, and F. Petroff, Phys. Rev. Lett. **61**, 2472 (1988).

[2] G. Binasch, P. Grnberg, F. Saurenbach, and W. Zinn, W. Phys. Rev. B **39**, 4828 (1989).

[3] see for example: D.D. Awschalom and M.E. Flatte, Nature Physics **3**, 153 (2007).

[4] Yu.A. Bychkov and E.I Rashba, Pisma Zh. Eksp. Teor. Fiz. **39**, 66 (1984) [JETP Lett. **39**, 78 (1984)].

[5] S. Datta and B. Das, Appl. Phys. Lett. **56**, 665 (1990).

[6] J.E. Hirsch, Phys. Rev. Lett. **83**, 1834 (1999).

[7] Y.K. Kato, R.C. Myers, A.C. Gossard, and D.D. Awschalom, Science **306**, 1910 (2004).

[8] J. Wunderlich, B. Kaestner, J. Sinova, and T. Jungwirth, Phys. Rev. Lett. **94**, 047204 (2005).

[9] C. Brüne, A. Roth, E.G. Novik, M. König, H. Buhmann, E.M. Hankiewicz, W. Hanke, J. Sinova, and L.W. Molenkamp, submitted to Natur Physics (2008); cond-mat/0812.3768.

[10] M. Khodas, A. Shekhter, and A.M. Finkelstein, Phys. Rev. Lett. **92**, 086602 (2004).

[11] K.v. Klitzing, G. Dorda und M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

[12] C.L. Kane, E.J. Mele, Phys. Rev. Lett. **95**, 146802 (2005); C. Wu, B.A. Bernevig, S.C. Zhang, Phys. Rev. Lett. **96**, 106401 (2006); C. Xu, J. Moore, Phys. Rev. B **73**, 045322 (2006).

[13] C.L. Kane and E.J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).

[14] M. Büttiker, IBM J. Res. Dev. **32** 317 (1988).

[15] B.A. Bernevig, T.L. Hughes, S.C. Zhang, Science **314**, 1757 (2006).

[16] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L.W. Molenkamp, X.-L. Qi, S.-C. Zhang, Science **318**, 766 (2007).

[17] A. Roth, C. Brüne, H. Buhmann, L.W. Molenkamp, J. Maciejko, X.-L. Qi, and S.-C. Zhang, submitted to Science (2009).

[18] Currently, measurements of the QSH effect have only been conducted in the temperature range between 10 mK and 4 K. The actual temperature limits for HgTe QW structures have yet to be explored experimentally.

[19] B. Zhou, H.-Z. Lu, R.-L. Chu, S.-Q. Shen, and Q. Niu (2008), cond-mat/0806.4810.

[20] M. König, H. Buhmann, L.W. Molenkamp, T. Hughes, C.-X. Liu, X.-L. Qi, and S.-C. Zhang, Journ. Phys. Soc. Japan **77**, 031007 (2008).

[21] Y.S. Gui, C.R. Becker, N. Dai, J. Liu, Z.J. Qiu, E.G. Novik, M. Schfer, X.Z. Shu, J.H. Chu, H. Buhmann, and L. W. Molenkamp, Phys. Rev. **70**, 115328 (2004).

[22] J. Sinova, D. Culcer, Q. Niu, N.A. Sinitsyn, T. Jungwirth, and A.H. MacDonald, Phys. Rev. Lett. **92**, 126603 (2004).

[23] Results are in preparation for publication.

C 5 Electron spins in carbon based materials

M. Morgenstern
II. Institute of Physics B and Jülich-Aachen Research Alliance: Fundamentals of Future Information Technology (JARA-FIT)
RWTH Aachen University, D-52074 Aachen

Contents

1 Introduction **2**

2 General properties of carbon materials **3**

3 Spin transport in carbon nanotubes **8**

4 Spin transport in graphene **12**

5 Coherent manipulation of spin centers in diamond **13**

6 Quantum dots and spin centers in graphene and carbon nanotubes **21**

7 Summary **24**

8 Acknowledgement **24**

1 Introduction

Carbon is one of the most abundant elements of the universe (3.5 %) mainly existing as ^{12}C (98.9 % ^{12}C, 1.1% ^{13}C). The electronic configuration of carbon is similar to silicon being $1s^2 2s^2 2p^2$, i.e., carbon exhibits a half-filled sp-shell like Si and Ge. While Si has been overwhelmingly used for electronics taking over after an initial success of Ge, carbon has only recently been discussed as an important alternative.

The neglect of carbon so far[1] is due to the fact that the two bulk phases of carbon are either semimetallic (graphite), i.e., they lack a band gap, or it has been rather difficult to obtain n-type doping (diamond). Only recently, n-type doping using either P, N [2] or a so-called surface transfer doping [3] has been achieved for diamond. Consequently, first diamond field-effect transistors with gate length of 1.5 μm have been built [4], which are extremely attractive for high-temperature applications.

The recent interest into carbon by the transistor community is due to the fact that further nanoscopic allotropes have been found, most notably carbon nanotubes [5] and graphene [6]. These materials exhibit extremely high mobilities even at room temperature. Thus, both nanoscopic structures have already been used within efficient field-effect transistors [7, 8] and carbon nanotubes are meanwhile part of the alternatives considered within the International Technology Roadmap for Semiconductors [9].

Major challenges in that respect are the efficient separation of metallic and semiconducting nanotubes [10] preferable by a selective growth process and, for graphene, the cost-efficient production of graphene templates, respectively graphene wafer materials [11].

Besides their perspectives with respect to charge-based electronics, carbon materials are also extremely attractive for spin-based electronics. In addition to the high mobility of graphene and carbon nanotubes, two specific reasons favor these materials for spin applications. Firstly, carbon is lighter than any other element with a half-filled sp-shell and, thus, exhibits a small spin-orbit coupling, which to first order scales with the charge of the nucleus Z. On the other hand, it is well known that spin-orbit coupling is the main source for spin-relaxation in solids. Thus, one expects an extremely stable spin degree of freedom in carbon materials. Indeed, a spin relaxation length of 50 μm (2 μm) has already been deduced from spin-dependent transport experiments for carbon nanotubes (graphene) as a lower bound still restricted by preparational conditions [12, 13]. Secondly, the nucleus of ^{12}C has no nuclear spin and, using laser purification, an isotope purity of, at least, 99.983 % ^{12}C is possible [14]. Since it is known that spin coherence is limited by the hyperfine interaction between electron and nuclear spins in III-V materials [15], the absence of nuclear spin is extremely favorable for applications in spin-based quantum information. Indeed, using negatively charged nitrogen-vacancy complexes in diamond as spin centers, a coherence time of 350 μs has been reported at room temperature [16, 17]. This is significantly larger than the coherence time of 1 μs in GaAs quantum dots at 135 mK [15]. Even without spin-echo techniques, the room-temperature dephasing time of the diamond centers is about 2 μs compared with a dephasing time of 10 ns in GaAs quantum dots at 135 mK. The spare distribution of nuclear spins in diamond, moreover, allows to transfer the coherent information to the nuclei providing a coherence time of more than 20 ms [18].

These properties make carbon materials a very promising option for spintronics. The current

[1] This article does not deal with organic materials which, of course, also contain carbon as an ingredient and which are important, in particular, with respect to low-cost electronic applications. However, the spin transport properties of organic materials, although prospective, have only partly been studied so far [1], which leads to a currently incomplete framework.

article summarizes the recent results and is organized as follows. Chapter 2 describes general properties of graphite, diamond, graphene and carbon nanotubes. Chapter 3 and 4 deal with the spin transport properties of carbon nanotubes and graphene. Chapter 5 will describe the results of coherent manipulation of spins using nitrogen-vacancy centers in diamond and chapter 5 will discuss the success and the perspectives for using spin centers in graphene or carbon nanotubes.

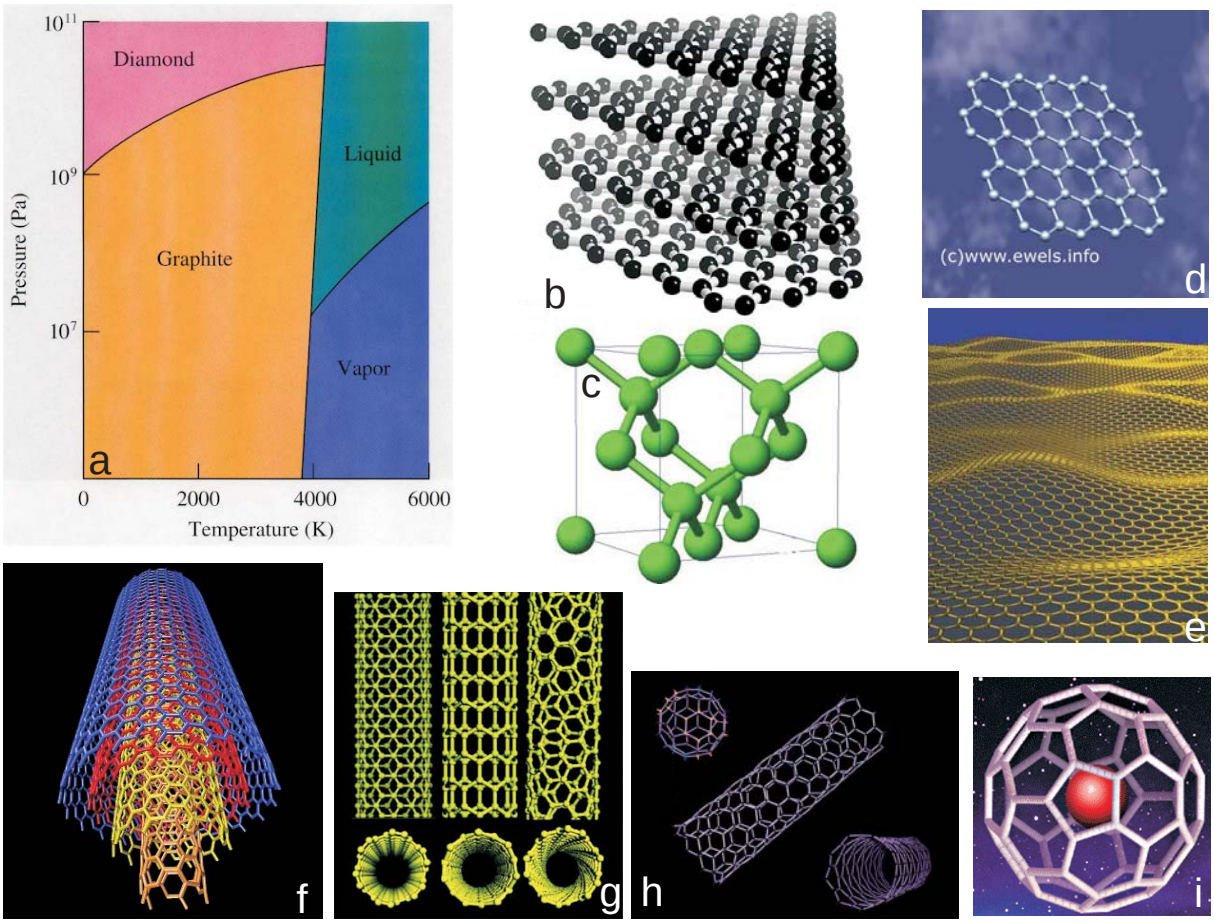


Fig. 1: *(a) Schematic phase diagram of carbon; (b) crystal structure of graphite; (c) crystal structure of diamond; (d) idealized structure of graphene; (e) corrugated structure of graphene; (f) structure of multiwall carbon nanotube; (g) structure of carbon nanotubes with different chirality: left: armchair, middle: zigzag; (h) structure of C_{60} fullerene and two zigzag carbon nanotubes; (i) structure of C_{60} containing a dopant atom in the center.*

2 General properties of carbon materials

Figure 1 (a) shows the phase diagram of carbon bulk materials. The most common phase energetically favored at atmospheric pressure is graphite with a crystal structure shown in Fig. 1 (b). It consists of planes of graphene (Fig. 1 (d)), where the carbon atoms exhibit a sp_2+p_z configuration leading to strong binding via the sp_2 hybrid orbitals and a weaker π-bonding via the p_z orbitals. The latter orbitals are responsible for the electrically metallic properties of

graphite. The graphene sheets are stacked, i.e., they are only weakly bound by van der Waals forces between the π orbitals. Consequently, it is rather easy to separate stacks of graphene sheets from each other, e.g., by ripping it using scotch tape.

The high pressure phase of carbon is diamond being metastable at atmospheric pressure with the crystal structure shown in Fig. 1 (c). Here the carbon atoms exhibit a sp_3 configuration of the atomic orbitals. This results in the so-called diamond structure shown in Fig. 1 (c) which consists of two fcc lattices shifted by (1/4, 1/4, 1/4) along the diagonal of the cube. The sp_3 bonds of diamond are much stronger than the van der Waals forces of graphite making diamond an extremely hard material. Since the π-bond is missing, diamond is not a metal, but a large gap insulator.

The easy separation between planes in graphite can be exploited to make single layers of graphene. Therefore, one multiply uses scotch tape in order to thin down the graphene stacks until stacks of only a few layers exist on the scotch tape [6]. Pressing these layers to a Si substrate with a 300 nm thick SiO_2 allows to produce single layers and to distinguish these monolayers from multilayers by their optical appearance due to an interference effect [19]. Additionally, Raman spectroscopy can be used to identify single layers [20]. The graphene sheets have the tendency to undulate, i.e., they look more like Fig. 1(e) than like Fig. 1(d) exhibiting a wave length of about 10 nm and a wave amplitude of about 1 nm [21, 22]. Graphene exhibits a very interesting electronic structure due to the two inequivalent atoms within the unit cell. This leads to cones at the edges of the Brillouin zone crossing the Fermi level, i.e., to a linear dispersion $E(k) = \hbar \cdot v_F \cdot k$ similar to light with a Fermi velocity of $v_F \simeq 10^6$ m/s representing the effective velocity of light of the quasiparticles in graphene. The two inequivalent atoms are an additional degree of freedom which can be described by a pseudospin leading to a Dirac-like Hamiltonian $H = \hbar v_F \cdot \sigma \cdot \underline{k}$ where σ represents the 2D Pauli matrix [6]. The energetic similarity between sp_3- and sp_2-orbitals favors the existence of additional allotropes and indeed nanostructures like nanotubes (Fig. 1 (f), (g), (h)) or fullerenes (Fig. 1 (h), (i)) can be produced, e.g., by simple discharge techniques using carbon electrodes and various catalysts [23]. Single-wall nanotubes are basically rolled up graphene sheets classified by their circumference vector as shown in Fig. 2 (a). If the circumference vector runs parallel (perpendicular) to a lattice vector, the tube is called zigzag (armchair) nanotube. The electronic properties of nanotubes depend on the circumference vector leading to the easy rule: If (a_1, a_2) is the circumference vector in units of the lattice vectors of graphene, the nanotube is metallic for $a_1 - a_2 = 3n$ ($n \in Z$) and semiconducting otherwise. This can be easily understood looking at the band structure of graphene in Fig. 2 (c). The bonding and antibonding π and $\pi^\star$ bands of graphene cross the Fermi level exactly at the corners of the Brillouin zone due to the two degenerate atoms within the unit cell. Rolling-up the graphene sheet leads to an additional quantization condition for the wave vector $\underline{k}$, i.e., the wave vector along the circumference $k_\parallel$ multiplied by the radius of the nanotube r must be an integer ($k_\parallel \cdot r = n$ mit $n \in Z$). Consequently, only discrete $k_\parallel$-values are allowed leading to equidistant $\underline{k}$-lines perpendicular to the circumference direction as depicted in Fig. 2 (d) and (e) together with the Brillouin zone of graphene. Only if these lines cut the corners of the Brillouin zone, the nanotube provides states at the Fermi level leading to metallic conductivity. Figure 2 (f) and (g) show the resulting band structure along the tube. Only in the case of metallic tubes, there are two branches with linear dispersion crossing the Fermi level and leading to a constant density of states ($DOS(E) \propto$ const). All other branches in metallic and semiconducting tubes show a nearly parabolic energy dispersion at the onset energy E_i leading to $DOS(E) \propto 1/\sqrt{E - E_i}$ which results in the characteristic van-Hove singularities at the onset energies. The width of the semiconducting gap is bassically inversely proportional to

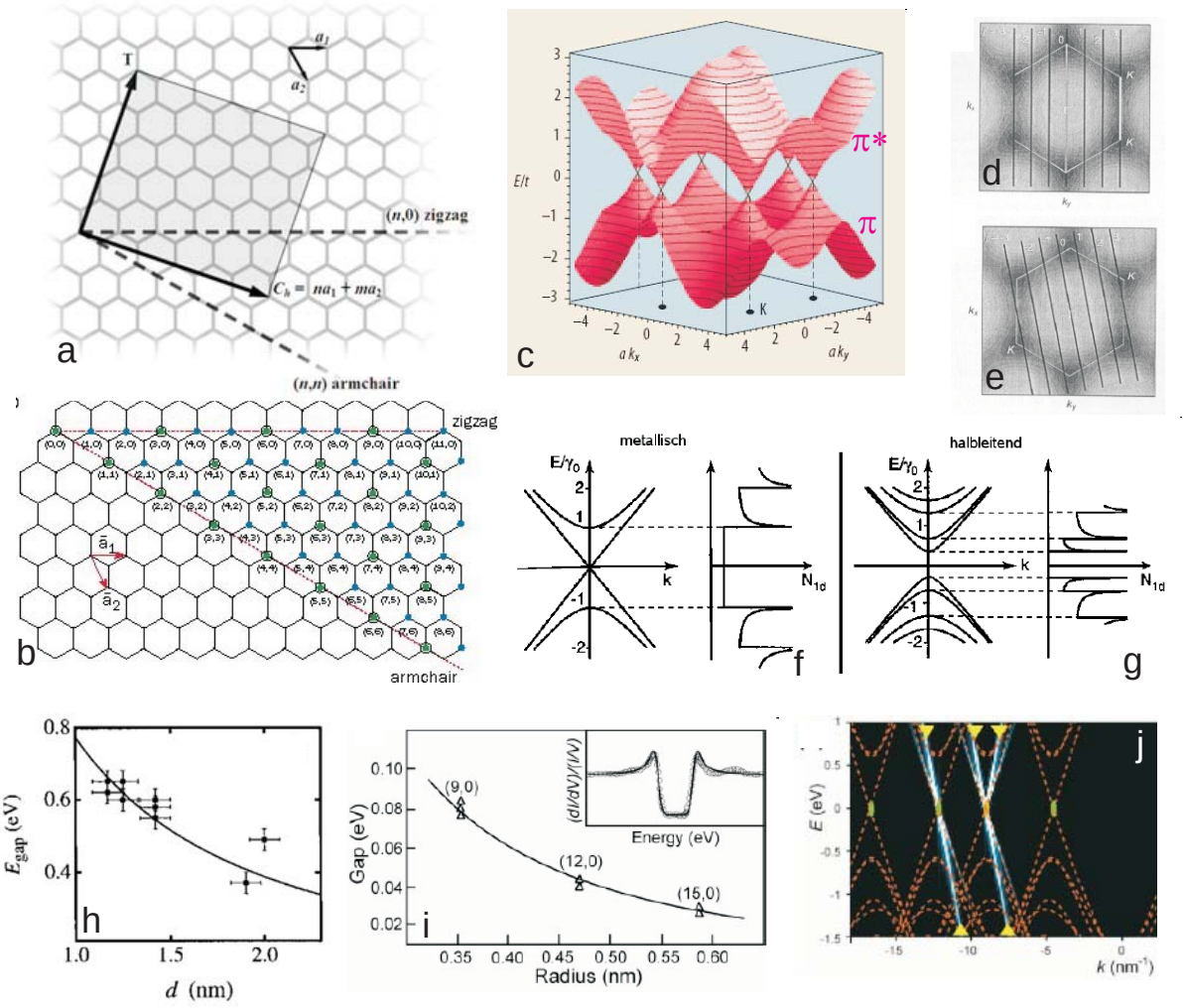


Fig. 2: *(a) Graphene sheet with lattice vectors a_1, a_2 and circumference vector C_k indicated; the vector T points along the tube axis and the dashed lines mark the circumference directions for armchair and zigzag nanotubes; (b) atomic structure of graphene with marked circumference vectors; small (blue) dots are semiconducting, larger (green encircled) dots metallic nanotubes; (c) band structure of graphene; (d), (e): Brillouin zone of graphene (white hexagon) with measured Fermi surface as a color plot and marked allowed $\underline{k}$ values of two nanotubes with different circumference vector; (f),(g) band structure and resulting density of states of a metallic (f) and a semiconducting (g) nanotube; (h) primary band gap $E_{\rm gap}$ of semiconducting tubes as a function of the tube diameter d as measured by scanning tunneling spectroscopy (STS) [25]; (i) secondary band gap of metallic nanotubes as a function of radius r as measured by STS; inset shows the measured secondary gap for a (9,0)-nanotube [26]; (j) calculated dispersion of a carbon nanotube as a Luttinger liquid; dashed orange lines mark the one-particle dispersion of the nanotube, while blue lines mark the dispersion of charge (steeper) and spin (less steep) density wave modes [28].*

the tube diameter being about 0.8 eV at a diameter of 1 nm as shown in Fig. 2 (h) [24, 25].

Of course, the chemical bonds in nanotubes are not perfect sp_2-bonds and curvature leads indeed to an opening of a band gap even for nominally metallic tubes. This is shown in the inset of Fig. 2(i) for a tube with a circumference vector (9,0) [26]. This so-called secondary gap scales

with $E_{\mathrm{Gap}} \propto 1/r^2$ and is more than one order of magnitude smaller than the semiconducting primary gap described above [24]. Only for armchair tubes the secondary gap is absent and, thus, only armchair tubes are real metals even at low temperature.

It should be mentioned that metallic nanotubes are one-dimensional systems and, as such, exhibit Luttinger properties [27, 28], i.e., the principal excitations are not single particle excitations but collective ones. In particular, charge and spin density waves are the excitations close to the Fermi level. They exhibit a different dispersion as shown in Fig. 2 (j). The different velocities of spin density wave and charge density wave lead to spin-charge separation, i.e., the charge moves faster than the spin, which might be additionally advantageous for spin transport properties.

Finally, it should be kept in mind that the electronic properties of nanotubes can be modified intentionally by either filling the nanotube with molecules as C_{60} (Fig. 3(a)) or by exploiting natural defects within the nanotube (Fig. 3 (b)) and at their end (Fig. 3(c)). In addition, multiwall nanotubes offer the possibility to provide electrical conductivity even, if some of the tubes are semiconducting.

A very attractive property of graphene and nanotubes is the long mean free path which is caused by the fact, that extended defects can not lead to backscattering. The formal reason is the pseudospin related to the two inequivalent cones of the band structure of graphene (Fig. 2 (c)) which leads to a cancelation of backscattering by a Berry phase [33]. In simplified terms, the Fourrier components of the extended disorder have a too low wave number to couple cones at the opposite corners of the Brillouin zone as required for backscattering. Also low-energy and, thus, low wave length phonons, which are present at low temperature, provide not enough $\underline{k}$-vector to couple the opposite edges of the Brillouin zone. A minor correction to this statement comes from the fact that the band structure close to the Fermi level is not a perfect cone, but exhibits an outwards warping. However, the major sources of backscattering remain atomically sharp defect potentials and optical phonons. In graphene, the edges can be an additional source of backscattering.

The transport properties of single wall carbon nanotubes are indeed exceptional [34], i.e., they do not provide weak localization at low temperature. Instead, interference effects of the electronic waves due to backscattering at the contacts have been observed up to length of 1 μm showing a long mean free path, at least, at 1.5 K [35]. For multiwall carbon nanotubes, the elastic mean free path is also rather long (e.g. 250 nm), but diffusive transport [37] as well as weak localization has been clearly observed [38], although quantized conductance values indicating ballistic transport properties have also been found [39]. Interestingly Aharanov-Bohm measurements performed with the setup shown in Fig. 3 (e) indicate that the current in multi-wall nanotubes flows primarily through the outermost nanowire [36].

Graphene exhibits an exceptional mobility of 2 m^2/Vs at room temperature if deposited on SiO_2, which translates to a mean free path of 0.3 μm [6]. Suspending the graphene and additionally annealing it by an extremely high current density of $J = 2 \times 10^{12}$ A/m^2 increases the mobility up to 23 m^2/Vs at 5 K [40]. Weak localization has firstly not been observed [41], which was ascribed to the phase-breaking influence of the rippling as visible in Fig. 3(h) [21, 22], but later weak localization has been found in other samples and has been ascribed primarily to edge scattering [42].

It is obvious that the very long transport length combined with the low mass of carbon leading to low spin-orbit coupling and the nearly absent nuclear spin leading to low hyperfine interaction make graphene and carbon nanotubes extremely attractive for spintronics.

Since fullerenes have barely been used for spin transport so far, they will be described only

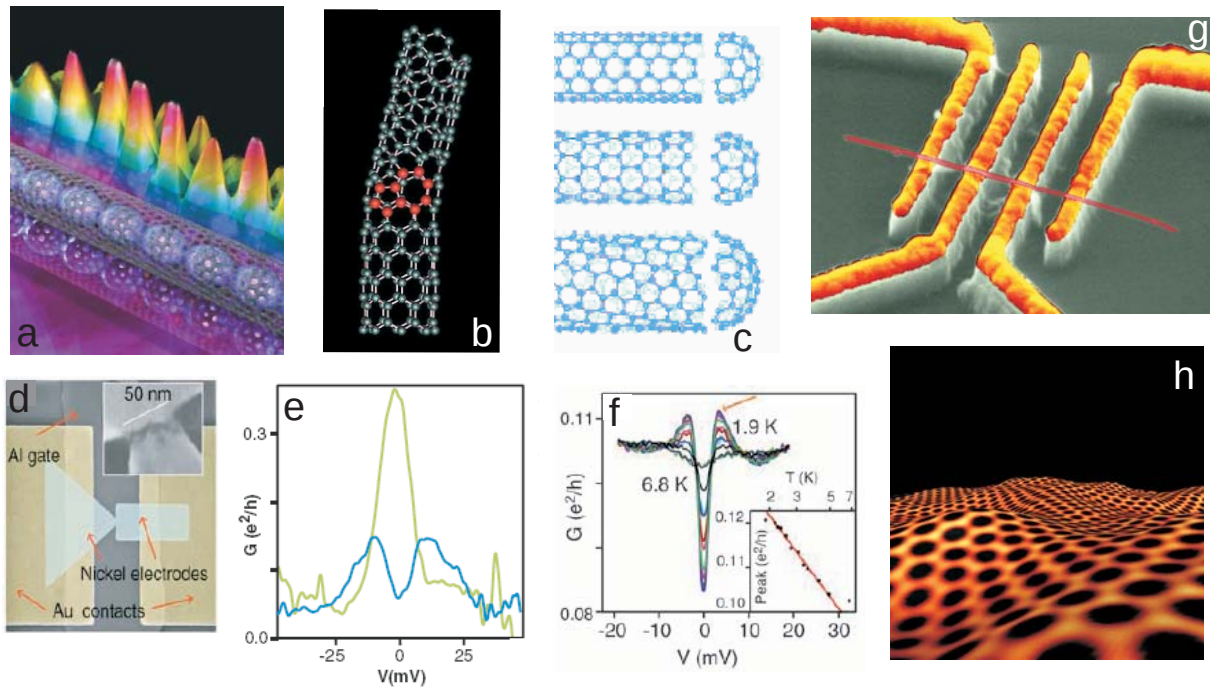


Fig. 3: *(a) Carbon nanotube filled with fullerenes; the resulting conduction band DOS as measured by STS is shown in the background [32]; (b) defect structures (red) combining two nanotube areas which results in pentagons and heptagons of carbon; (c) end area of a nanotube; note the similarity to fullerenes; (d) setup for magnetotransport measurements of C_{60} with ferromagnetic Ni contacts; the inset shows the contact region with C_{60} molecules present but barely visible; (e) conductance curves of the setup in (d) as a function of source-drain voltage between the Ni contacts at $T = 1.5$ K; the blue (green) curve is measured at 250 mT (15 mT) with Ni contact magnetization oriented parallel (antiparallel) with respect to each other; (f) temperature dependence of the conductance curve of (e) with parallel magnetization of the contacts; inset shows the logarithmic temperature dependence of the peak values [43]; (g) Multiwall carbon nanotube deposited on four gold electrodes for a four-point measurement [36]; (h) STM image of graphene as deposited on SiO_2; the color scale displays the atomic resolution while the long range corrugation is displayed in 3D [22].*

briefly. As shown in Fig. 1(i), they consist of hexagons and pentagons of carbon atoms. In particular, the pentagons deviate strongly from the sp_2 hybridization. C_{60} is the most stable configuration and consists of 20 hexagons and 12 pentagons. Also all other fullerenes contain 12 pentagons, but a different number of hexagons. Similar configurations are found at the end of nanotubes. The electronic structure of fullerenes is molecular-like exhibiting discrete levels. Most importantly, they have a lowest unoccupied (LUMO) and a highest occupied (HOMO) energy level, both derived from π- and $\pi^\star$-like elementary binding orbitals. Also fullerenes can be filled and, thereby, doped leading to interesting properties as, e.g., superconductivity [29]. Ferromagnetic properties of polymerized C_{60} [30] as well as of other carbon materials after specific treatment, e.g., ion bombardment with hydrogen, [31] have been reported, but these results are still under debate. More interestingly, C_{60} molecules have been placed between Ni contacts by electromigration (Fig. 3(d)) and exhibit an exchange-split Kondo resonance. The exchange splitting of the Kondo peak can be turned on and off by aligning the contact magnetization parallel or antiparallel to each other as shown in Fig. 3 (e). The logarithmic

temperature dependence of the peaks shown in Fig. 3(f) identifies them clearly as Kondo peaks [43].

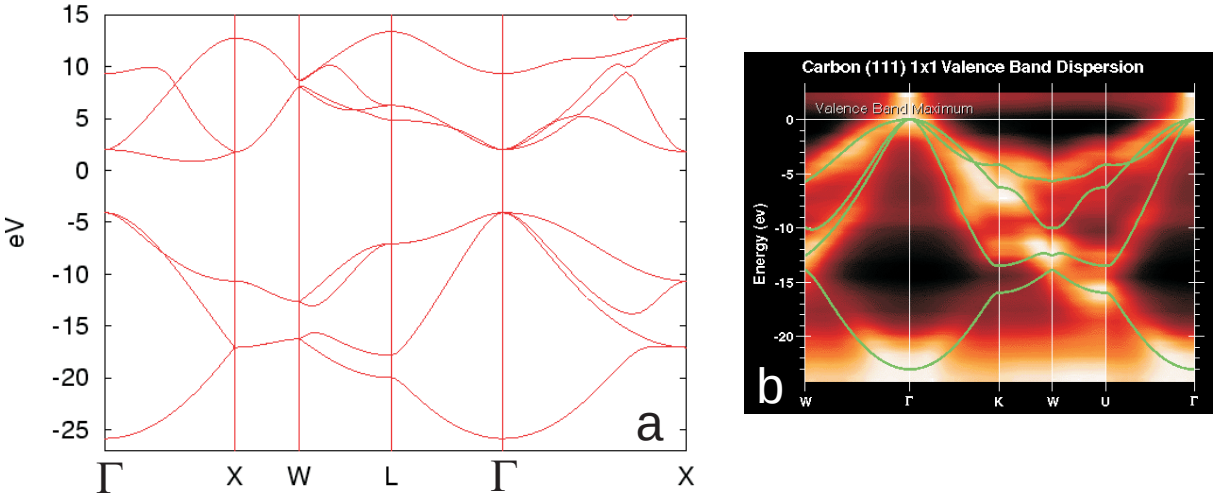


Fig. 4: *(a) Calculated band structure of diamond plotted in the high symmetry directions; (b) photoemission measurement of the valence band of diamond (color plot) with superimposed calculated band structure*

Finally, Fig. 4 shows the band structure of diamond, which exhibits the typical sp_3-appearance consisting of three p-like valence bands and one s-like conduction band similar to Si, GaAs, etc.. The low spin-orbit interaction is apparent by the fact that the valence bands do not exhibit a spin-orbit gap. More detailed spectroscopic measurements of acceptors indicate a spin-orbit gap of 6 meV [44], which is a factor of 60 lower than in GaAs. The band gap is large with 5.48 eV and the effective masses are 0.2 and 0.25 for electrons and holes, respectively.

3 Spin transport in carbon nanotubes

The first successful spin transport measurements in carbon based materials have been performed using carbon nanotubes [45]. Therefore, a multiwall tube (MWNT) with a diameter of 10-40 nm has been placed below two electrodes made out of cobalt. An electron microscope image of the device is shown in Fig. 5 (a). The Co electrodes are about 300 nm apart just connected by the MWNT. The resistance of the device is dominated by the intrinsic resistance of the nanotube with a value of about 250 kΩ.

Measuring this resistance differentially by using lock-in technique as a function of a magnetic field applied within the surface plane (Fig. 5(b)), the resistance changes twice (back and forth) by about 8 % and the switching fields depend on the sweep direction of the field. This is a behaviour well known for many magnetoresistance effects, which is attributed to the different coercive fields of the two electrodes. Within the black curve of Fig. 5(b), where the B-field increases with time, the first electrode switches at about 5 mT leading to an antiparallel orientation of the magnetization of the two electrodes. In this configuration the resistance is high, since the majority of electrons injected from one electrode, e.g., the spin-up electrons $N_{1\uparrow}$ find only few empty states within the second electrode, where the spin-up electrons are minority states. At higher field of about 30 mT also the second electrode switches. Thus, the two electrodes

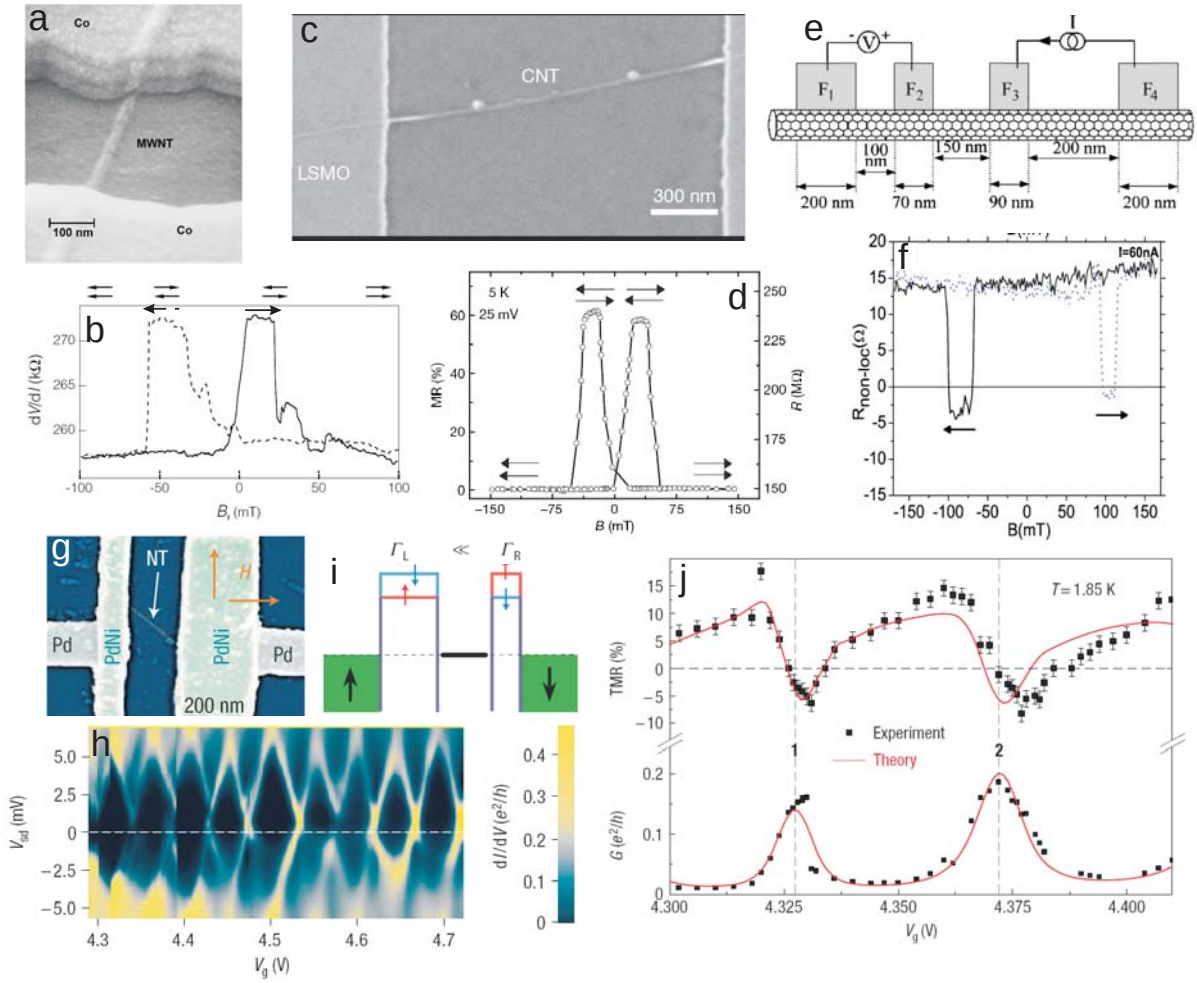


Fig. 5: *(a) Electron microscopy image of a multiwall nanotube (MWNT) in between two Co electrodes [45]; (b) magnetoresistance curve of the sample shown in (a); the arrows within the image mark the field sweep direction, while the arrows above the image mark the relative orientation of the two electrodes; (c) electron microscopy image of MWNT bridging two electrodes of $La_{0.7}Sr_{0.3}MnO_3$ (LSMO) [12]; (d) magnetoresistance curve of the sample shown in (c) with arrows marking the relative orientation of the electrodes; (e) sketch of a four probe measurement of magnetoresistance using four Co electrodes on top of a single-wall nanotube [49];(f) measurement of the non-local resistance dV/dI of the sample sketched in (e) with arrows marking the direction of field sweep; (g) nanotube between two PdNi contacts [50]; (h) differential conductivity of the sample shown in (g) as a function of V_{sd} and V_g; the color code is visible at the scale bar on the right; (i) model of the assumed tunnel barriers between the electrodes (green) and the tube level (black line); (j) lower curve: conductivity of the nanotube exhibiting two Coulomb-peaks; upper curve: measured magnetoresistance (TMR); the line is the result of a model calculation (see text)*

exhibit parallel orientation of magnetization, which reduces the resistance. Going back in magnetic field, the first electrode switches at about - 15 mT and the second one at about - 60 mT. The fact, that there are several additional switching effects of smaller magnitude, is attributed to the presence of magnetic domains at the edges of the electrodes.

A simplified evaluation of the expected resistances is $\sigma_{\|} \propto N_{maj}^2 + N_{Min}^2$ for parallel orientation

and $\sigma_{\uparrow\downarrow} \propto N_{maj} \cdot N_{Min} + N_{Maj} \cdot N_{Min}$ for the antiparrallel orientation of the electrodes. Thereby, $N_{maj/min}$ is the density of states at the Fermi level of the cobalt for the two different spin orientations. Bulk Co exhibits a spin polarization $P_F := (N_{Maj} - N_{Min})/(N_{Min} + N_{Maj})$ of 34 % leading to an expected change in resistance by 21 %. This is significantly larger than the measured value of 8 %.
There are three possible reasons:

1. Firstly, the spins might be switched within the carbon nanotube due to scattering leading to a spin polarization of the injected electrons which decays with distance from the injection electrode. Assuming an exponential decay $P(l) = P_F \cdot \exp(-l/l_0)$, one obtains a decay length called spin relaxation length of $l_0 \simeq 130$ nm.

2. Secondly, scattering centers at the interface between Co and the MWNT might reduce the spin polarization of the injected electrons.

3. Thirdly, the resistance mismatch between the electrodes and the MWNT leads to a reduced spin polarization in the MWNT, if the transport is diffusive [46]. This is due to the fact that the electrochemical potentials basically drop within the MWNT independent of the spin. This problem can be solved either by using ballistic conductors between the spin polarized electrodes or by introducing a tunnel barrier between the ferromagnetic electrodes and the MWNT [47]. Alternatively, the spin polarization P_F could be increased close to 100 %.

The later two solutions are combined in an experiment using MWNT and $La_{0.7}Sr_{0.3}MnO_3$ (LSMO) as an electrode material [12]. LSMO is ferromagnetic with a Curie temperature of 365 K and exhibits, at least, theoretically a spin polarization $P_F \simeq 100$ %. Indeed, the experiment, which uses a longer MWNT bridging a gap of about 2 μm (Fig. 5 (c)) shows a better magnetoresistance of 60 % (Fig. 5 (d)) than it was found within the experiments using Co electrodes. Note that the resistance of this device is three orders of magnitude larger than the resistance of the device with Co electrodes indicating tunneling barriers between the LSMO and the MWNT. A model description of the results based on density functional calculations deduces a spin relaxation length of $l_0 \simeq 5$ μm.
Single-Wall nanotubes have also been probed using Co contacts initially finding a magnetoresistance of 2 % and estimating a spin relaxation length of 1.4 μm [48]. Here, also four-probe measurements have been performed [49]. Therefore, the spin current is driven between two Co contacts and the resulting voltage is measured across two other Co contacts as shown in Fig. 5 (e). The general idea is that spin polarized electrons are injected at contact F3. They drift towards F4 due to the applied electric field, but they also diffuse towards contact F2. Thus, F2 measures an electrochemical potential, which is proportional to the spin polarization directly below the contact, but within the nanotube. Since the spin density within the nanotube decreases with distance from F3, the corresponding potential is less at F1 giving rise to a measurable voltage between F1 and F2. If the injected spin polarization changes sign, the voltage also has to change sign. This is indeed observed in the magnetoresistance curves of Fig. 5 (f), where $R_{non-local}$ is the measured voltage between F1 and F2 divided by the current flowing between F3 and F4. The importance of this non-local method is that it is not affected by spurious effects (change of B-field can influence the current within the tube, anisotropic magnetoresistance within the electrodes, etc.), but only detects the injected spins. A detailed evaluation of the non-local data indicates that the two-terminal resistance in this sample overestimates the

spin injection efficiency by a factor of four, i.e., the magnetoresistance caused by spin injection is not 6 % but only 1.5 %, the other 4.5 % originating from spurious effects.
The most intriguing experiment used a two-terminal measurement, but an additional gate voltage, which was able to tune the strength and even the sign of magnetoresistance [50]. Thus, the first step towards a spin-based field effect transistor, tunability of spin transport by an electric field, has been achieved. Currently, there is no gain of this transistor and it works only at 1.5 K, i.e., there are remaining challenges.
Figure 5 (g) shows the setup using a PdNi alloy for the elctrodes. The different widths of the two electrodes guarantees that they are switching at different B-field. The important difference of this device with respect to previous ones is that it shows clear Coulomb blockade oscillations and the corresponding Coulomb diamonds (Fig. 5(h)), i.e., the electrons are injected one by one into the nanotube. The nanotubes are deposited on conducting, degenerately doped Si with a 400 nm SiO_2 layer on top. A gate voltage V_g applied to the Si couples capacitively to the nanotube and, thus, changes the charge on the nanotube one by one. Each time a new electron is added, the corresponding energy level is at E_F and a current can flow through the nanotube, even if the source-drain voltage V_{sd} between the PdNi electrodes is low. This leads to nearly equidistant peaks in conductivity as a function of gate voltage as shown in the lower part of Fig. 5(j). If V_{sd} is increased , current can also flow, if the energy level within the nanotube is away from E_F, which leads to the characteristic diamond-shaped regions of suppressed conductance in the V_g-V_{sd}-plane visible in Fig. 5 (h). If one now measures the magnetoresistance (TMR), i.e., the difference in resistance between the parallel and the antiparallel orientation of the PdNi electrode magnetization divided by the sum, one gets the upper curve in Fig. 5 (j). The TMR oscillates each time, when a new electron enters the nanotube region. At the resonance, i.e., when an energy level is aligned with E_F, TMR is negative, while it is positive elsewhere.
The authors propose the following explanation partly based on an earlier theoretical study [51]. As depicted in Fig. 5 (i), the nanotube is coupled asymmetrically to the two electrodes, i.e., the tunneling rate Γ_L from the left electrode to the nanotube is much lower than Γ_R. The transmission T of electrons across the nanotube is described as a Breit-Wigner resonance, i.e.,

$$T(E_F) = \frac{\Gamma_L \cdot \Gamma_R}{(E_F - E_0)^2 + (\Gamma_L + \Gamma_R)^2/4} \tag{1}$$

with E_0 being the energy of the level within the nanotube. Importantly the transmission rate of each tunnel barrier depends additionally on the spin polarization P_F of the electrode via $\Gamma^{\uparrow} \propto 1 + P_F$ and $\Gamma^{\downarrow} \propto 1 - P_F$. If $\Gamma_R >> \Gamma_L$, one can approximate $T \propto \Gamma_L/\Gamma_R$, if $E_0 \simeq E_F$, i.e., at resonance, while one gets $T \propto \Gamma_L \cdot \Gamma_R$ if $|E_F - E_0| >> (\Gamma_L + \Gamma_R)/2$, i.e., far away from the resonance. Assuming that P_F is the same for both electrodes, one can calculate the magnetoresistance

$$TMR = \frac{T_{\uparrow\uparrow} + T_{\downarrow\downarrow} - T_{\uparrow\downarrow} - T_{\downarrow\uparrow}}{T_{\uparrow\uparrow} + T_{\downarrow\downarrow} + T_{\uparrow\downarrow} + T_{\downarrow\uparrow}} \tag{2}$$

Inserting $T \propto \Gamma_L/\Gamma_R$ and $\Gamma^{\uparrow} \propto 1+P_F$ and $\Gamma^{\downarrow} \propto 1-P_F$ at resonance, one gets: $TMR \propto -P_F^2$ Consequently, the TMR is negative at resonance as observed experimentally.
Doing the same calculation far away from the resonance ($T \propto \Gamma_R\Gamma_L$), on gets $TMR \propto +P_F^2$. Thus, the sign change of TMR requires different tunneling electrodes at both ends of the nanotube. Refining this model by additionally considering the spin splitting within the dot due to the effective magnetic field of the electrodes, the TMR data points in Fig. 5 (j) can be reproduced quantitatively by the solid line using a spin polarization of the electrodes of $P_F = 20$ %, a

spin splitting energy of 0.26 meV, if the electrodes are parallel, and a ratio of tunneling rates $\Gamma_R/\Gamma_L \simeq 35$ with $\Gamma_R \simeq 10^{12}$/sec.
The experiment has also been performed with MWNTs and works as well, although less easy to be interpreted. Finally an oscillation of magnetoresistance with gate voltage without sign change and with an amplitude of about 4 % has been observed for nanotubes with lower resistance, and is interpreted as due to standing electron waves between the electrodes which have an alternating spin orientation as a function of energy for parallel electrodes [53]. Consequently, carbon nanotubes can act as building blocks of spin transistors, although the spin-orbit coupling is weak and the standard Datta-Das transistor [52] is not working.

4 Spin transport in graphene

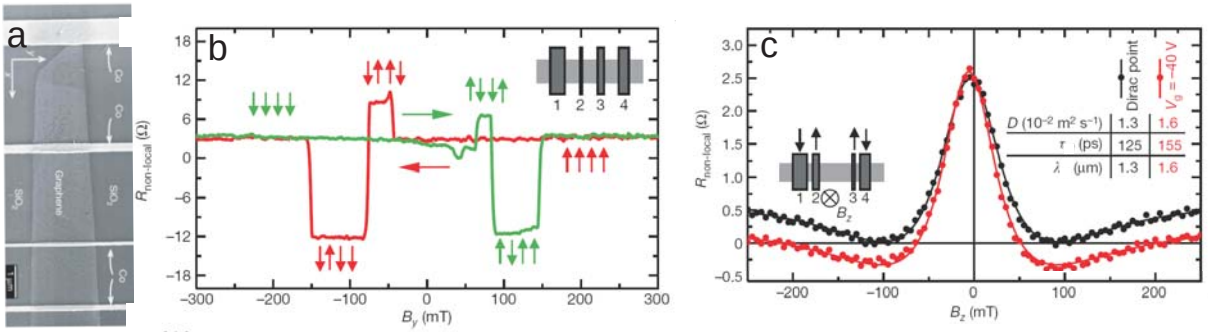


Fig. 6: *(a) Electron microscopy image of a monolayer graphene on Si/SiO$_2$(300 nm) with four Co electrodes separated by tunneling barriers from graphene for a nonlocal measurement of magnetoresistance [13]; (b) non-local magnetoresistance curve of the sample in (a) with B-field applied parallel to the graphene plane; arrows indicate the magnetization direction of the four electrodes; (c) non-local resistance in the parallel configuration of electrode 2 and 3 as a function of B-field applied perpendicular to the graphene plane as sketched in the inset; the two curves are recorded at different gate voltages (red:-40 V, black: 20 V = Dirac point); the lines are fits of the experimental data using diffusion constant D, spin dephasing time τ and spin dephasing length λ as marked in the inset table.*

Shortly after the discovery of graphene [6], it was realized that graphene has the same advantages as carbon nanotubes (low spin-orbit coupling, low hyperfine coupling). The first report of a spin-valve effect using NiFe electrodes was still ambigous [54], but shortly later convincing experiments using a four-probe measurement (Fig. 6(a)) with Co electrodes separated by an Al_2O_3 tunnel barrier from the graphene have been published [13]. Figure 6 (b) shows the non-local resistance (V_{12}/I_{34}) as a function of magnetic field applied in the plane of the graphene. The arrows mark the magnetization directions of the four electrodes. A sign change depending on the relative orientation of the two inner electrodes is clearly observed. Using samples with different distances between the inner two electrodes, the authors deduce a spin-flip length of about $\lambda_{sf} = 1.5 - 2$ μm and a polarization of the injected electrons of $P_F = 10$ %. Besides flipping from spin-up to spin-down, the spins can also precess, i.e., they can change the phase of the spinor. A magnetic field applied perpendicular to the direction of the injected spins leads to a precessional motion around the magnetic field with a precession frequency $\omega = g\mu_B B/\hbar$ ($g \simeq 2$: g-factor of graphene, μ_B: Bohr magneton). Within a one-dimensional ballistic system,

this would lead to a cosine-oscillation of the non-local resistance. However, the diffusive motion of the electrons within the 2D graphene from the injection electrode 3 towards the detection electrode 2 leads to a mixture of different spin orientations at electrode 2, i.e., to a damped oscillation, which is indeed visible in Fig. 6(c). The appearance of a mixture of spin orientations due to different paths of the electrons is called spin dephasing and the corresponding length scale, on which originally parallel spins decay exponentially, is called spin dephasing length λ_{sd}. It can be deduced from the curves in Fig. 6(c), which are called Hanle-curves, and turns out to be $\lambda_{sd} = 1.3 - 1.6$ μm slightly depending on the electron density of the graphene, which is tuned by a gate voltage. The fact that $\lambda_{sd} \simeq \lambda_{sf}$ indicates that spin relaxation might govern the apparent mixture of spins below electrode 2. Further experiments revealed that the spin relaxation is slightly anisotropic in graphene being 20 % lower if the spins are injected perpendicular to the graphene plane [55] and that the injected spins can be drifted effectively by an electric field [56].
The gate dependence of the magnetoresistance has been studied by four-probe measurements using permalloy contacts without tunnel barriers and five sign changes have been observed between gate voltages of 70 V and -70 V [57]. They are interpreted in terms of Fabry-Perot resonances similar to the results found with SWNTs [53], but the still very irregular gate voltage dependence calls for further studies. A sign change of magnetoresistance has also been observed for multilayer graphene in two-probe measurements [58].
Although the spin relaxation length appears to be still lower than in carbon nanotubes, which might be improved by optimizing the preparation process, graphene offers advantages with respect to nanotubes, e.g.,

- Graphene is always conducting and the electron density is easily tunable by a gate voltage, while carbon nanotubes are only conducting for armchair tubes.

- Graphene can be structured by oxidation using standard lithography, which allows to make quantum dots, double quantum dots, etc. opening a large variety of possible experiments.

Thus, graphene might be a future material for spintronics as well.

5 Coherent manipulation of spin centers in diamond

The favorable properties for spintronics are equally valid for diamond and most promising results concerning the realization of qubits have been achieved using so-called nitrogen-vacancy (N-V) centers in diamond. These N-V centers can be induced either by electron bombardment with energies in the MeV range using the natural nitrogen impurities within diamond [70] or more directly by nitrogen bombardment and subsequent annealing [60]. Both methods provide the general possibility to control the position of the defects either by focused ion beam or by using an electron microscope [60]. The advantage of the ion beam technique is that ultrapure diamond without nitrogen impurities can be used as a template, which is important since nitrogen impurities are the major source for decoherence of the spins in the N-V center [61, 62]. However, the success yield of N-V center production by N^+ ion bombardment is only 5 % so far. In order to probe the created N-V centers optically, nanocrystalline diamond is typically used, since the nanoparticles exhibit a higher transmission at the interface to vacuum due to near-field effects [65].

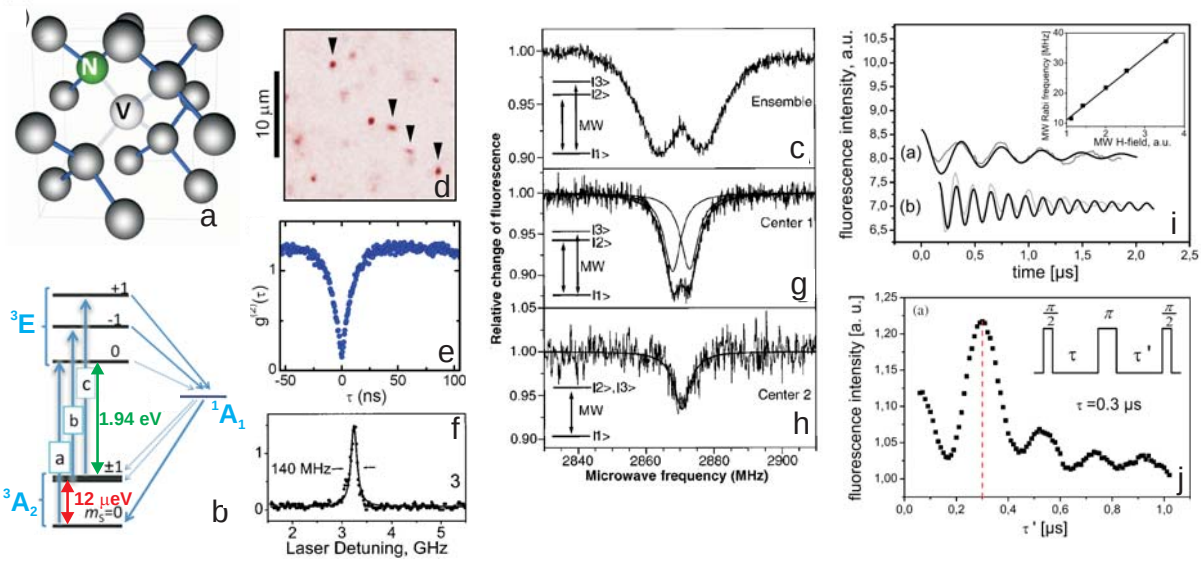


Fig. 7: *(a) Atomic arrangement of the Nitrogen(N)-Vacancy(V) center in diamond; (b) electronic states of the N-V center (not to scale);* $^3A_2$ *is the ground state triplet spin split by dipolar interactions,* $^1A_1$ *is the excited singlet and* ^3E *is the excited triplet spin split by dipolar and spin-orbit interactions; a,b,c mark the dipole allowed optical transitions, the full (dashed) arrows towards and from* $^1A_1$ *mark spin-orbit driven strong (weak) internal spin conversions; (c) fluorescence of an ensemble of N-V centers at excitation wavelength* $\lambda = 637$ *nm as a function of the frequency of an additional microwave field [70]; (d) confocal optical microscope image showing fluorescence of individual defects [62]; (e) time correlation function of single photon detection from an individual spot of (d); (f) fluorescence line of a single defect [65]; (g),(h) same as (c), but for single defects; (i) fluorescence intensity of a single defect measured after the application of a microwave pulse with pulse lengths displayed on the x-axis; the two curves are recorded with different microwave amplitudes and the resulting oscillation frequency as a function of amplitude is shown in the inset [73] ; (j) optically detected spin-echo experiment using the sequence displayed in the inset [73].*

The structure of the N-V center consists of a substitutional nitrogen atom and a neighboring vacancy and is shown in Fig. 7 (a). It is known that the N-V complex is mostly negatively charged, i.e., it consists of six lone pair electrons, in particular, one excess electron from nitrogen, four lone-pair electrons from the broken bonds surrounding the vacancy and the extra electron from charging. These six electrons occupy the four dangling bonds of the vacancy [66], which exhibit a T_d-like symmetry ignoring the difference between the neighboring nitrogen and carbon atoms. The T_d symmetry leads to two completely filled molecular states with the highest A_1 symmetry and six molecular bonds with the lower T_2 symmetry being only 4-fold occupied. Basically these are the p_{xy}, p_{xz} and p_{yz} states. Correspondingly, two holes remain in the T_2-shell which can form either a singlet (total spin $S = 0$) or a triplet ($S = 1$). The fact, that one of the dangling bonds of the vacancy is adjacent to nitrogen reduces the symmetry from T_d towards C_{3V}, which results in a splitting of the sixfold degenerate T_2 levels into two energetically lower A_1 levels and four energetically higher E levels. It turns out that the ground state of the system leaving two empty states in the E shell is a triplet with A_2 symmetry, while the singlet with two empty states in the E shell is the excited state having A_1 symmetry. Thus, the ground state of the defect provides an intrinsic spin polarization, which is not limited by life-time effects. The

energy difference between triplet and singlet is 1.92 eV. As shown in Fig. 7(b), only $\Delta E = 20$ meV above the singlet state is the next triplet state consisting of one empty state in the E level and one empty state in the A_1 level. This pair exhibits the antisymmetric E symmetry.
The ground state triplet exhibits a dipolar spin-spin splitting between the ($m_s = 0$)- and the ($m_s = +1/-1$)-states of about 12 μeV (2.9 GHz) as displayed in Fig. 7 (b) and an additional hyperfine splitting (interaction between the electron spins and the nuclear spin of ^{14}N) of about 20 neV (5 MHz). The excited triplet also exhibits a m_s-splitting due to spin-spin interaction and spin-orbit interaction leading to three different level energies, but there is no consensus on the exact arrangement and the corresponding energies [66, 62].
Since optical excitations do not couple strongly to the spin degree of freedom, the optically preferred excited state is the triplet with E symmetry corresponding to a wave length of the exciting light of $\lambda = 637$ nm. The excitation does not change the spin projection m_s and since the wave functions are barely changed by spin-spin and spin-orbit splitting, the optical matrix elements are similar for the three possible transitions a, b and c in Fig. 7(b) leading to decay rates of 8-12 ns [71] . The important point is that the spin-orbit coupling leads to a mixing of the A_1 singlet to the E and the A_2 triplets, which results in a slow phonon-induced internal conversion depending on m_s. The preferred transitions for internal conversion are $^3E(m_s = \pm 1) \rightarrow {}^1A_1(m_s = 0) \rightarrow {}^3A_2(m_s = 0)$. The transition rates can even be tuned by an external electric field [72].
The time constant of the first process is a few microseconds, while the second step lasts only several 100 ns [67, 16]. Thus, optical excitation leads to a preferred occupation of the $^3A_2(m_s = 0)$ ground state on the time scale of a few microseconds, even if the temperature is much higher than the energy difference between $^3A_2(m_s = 0)$ and $^3A_2(m_s = \pm 1)$ of 12 μeV. The only driving force to reestablish thermal equilibrium within the $^3A_2$ state is the spin-lattice relaxation, which has an activation barrier of 62 meV leading to a time constant of 1.2 ms at 300 K and up to 380 s at low temperature [67]. This leads, e.g., to more than 90 % occupation of the ($m_s = 0$)-level during optical excitation at 300 K [68, 17].
The optically induced spin polarization of the $^3A_2$ level can be probed using an additional microwave, which is tuned to the $(m_s = 0) \rightarrow (m_s = \pm 1)$ transition [69]. Since the excitation to ($m_s = \pm 1$) leads to an increased population of the long-living $^1A_1$ state, the fluorescence gets darker, if the $^3A_2$ spin is flipped by the microwave. The resulting fluorescence intensity as a function of microwave frequency is shown in Fig. 7 (c) exhibiting a 10 % reduction of intensity at $f_{\mathrm{Microwave}} = 2.88$ GHz (12 μeV) with an additional splitting of 14 MHz (58 neV) close to the value of the known hyperfine splitting of the ($m_s = \pm 1$)-states with the ^{14}N [70].
Using a confocal microscope, i.e., two microscopes, one focussing the exciting light within the (x, y)-plane and one collecting the fluorescence from a focus within the (x, z)-plane, also single spins can be addressed. At low N-V defect density, where individual defects are several μm apart, the fluorescence appears only at certain spots (Fig. 7(d)) and the result of a time-correlation experiment of the collected photons from a single spot (Hanbury-Brown-Twiss experiment) proves that the spots origin from individual defects [65, 62], since they never emit two photons at the same time (Fig. 7 (e)) [2]. The linewidth of the fluorescence is only 140 MHz (Fig. 7 (f)) corresponding to a lifetime of 2 ns [65] and, thus, much smaller that the m_s splitting, but larger than the hyperfine splitting. The linewidth can be further reduced by using diamond samples with extremely low nitrogen concentration going down to $\Delta f = 13$ MHz and, thus, exhibiting a quality factor of $f/\Delta f \simeq 10^7$ [72]. Adding the microwave during the confocal ob-

[2] The N-V center can, thus, be used as a single photon source [63] and "which path" interference experiments on the single photon level have been realized [64].

servation of a single defect reveals that the hyperfine splitting at 2,88 GHz is different for each defect partly going down to zero (Fig. 7(g) and (h)) [70]. The reason is a different environment of ^{13}C and substitutional ^{14}N nuclear spins for each defect leading to different hyperfine fields. Since the spectral width of the fluorescence line is much smaller than the splitting of the different (m_s)-levels in the ground state $^3A_2$ and the excited state E, individual spin configurations of a single N-V-center can be addressed optically [65].

Consequently, coherent single spin manipulation as required for quantum algorithms can be performed in the ground state of the defect. Moreover, the spin configuration is readable due to the m_s-dependent internal conversion to the singlet A_1. The coherent manipulation is proven by using a pulsed microwave field slightly detuned from the m_s-transition of 2.88 GHZ. The resulting beating between the exciting microwave frequency and the transition frequency leads to oscillations in the occupancy of the excited and the ground state called Rabi oscillations, which quantitatively also depend on the amplitude of the exciting field (inset of Fig. 7 (i)). The main part of Fig. 7(i) shows the resulting oscillations in fluorescence for two different detunings [73]. A damped oscillation is visible with a decay constant of 1.5 to 2.0 μs. The decay is a measure of the spin dephasing and, it has been shown that it strongly depends on the density of nitrogen impurities within the diamond lattice. The substitutional nitrogen impurities without vacancy exhibit $S = 1/2$ and, thus, can exchange m_s with the N-V center. At low N concentration, dephasing times of $T_2 = 350$ μs have been achieved [17]. Notice that the read-out via fluorescence destroys the spin phase. Thus, the experiment has to prepare the ($m_s = 0$)-state by optical spin alignement using a pulsed red laser, afterwards to switch off the light for the manipulation of the spin by a microwave pulse and, finally, to read-out the spin by fluorescence with the red laser turned on again. This experiment is then repeated several times for each pulse length of the microwave resulting in one of the curves in Fig. 7(i).

If the dephasing is caused by a slow process, which is not relevant on the time scale of a single experiment, but relevant on the time scale of the repeating with different pulse length, spin-echo experiments are used to suppress this type of dephasing. Therefore, the spin is first prepared in a superposition of the ground state and the excited state by a so-called $\pi/2$-pulse, then it is reversed by a so-called π-pulse after a certain time τ of the evolution and, finally, an additional $\pi/2$-pulse is applied after the same time of evolution in order to return the spin to the ground state. This effectively suppresses the influence of residual magnetic fields due to neighboring spins, which might fluctuate in different measurement cycles of the repeating but not on the time scale $2 \cdot \tau$ in a single experiment. The reason is that the precession of the reversed spin superposition is decelerated by the additional field exactly as much as the precession of the original spin superposition is accelerated by the same field. Fig. 7 (j) shows the fluorescence as a function of delay time τ' after the π-pulse, if the delay time before the π-pulse is fixed to $\tau = 0.3$ μs. Indeed, the largest signal is observed, if the two delay times are equal. But there are additional maxima within Fig. 7 (j), which are caused by the hyperfine splitting of the excited state reflecting that $m_s = 1$ and $m_s = -1$ exhibit different frequencies due to the nuclear field of the ^{14}N nuclear spin. Changing $\tau = \tau'$, i.e., repeating the spin echo experiment for different time intervals τ, the remaining dephasing called decoherence can be measured as shown, e.g., in Fig. 8 (e). In principle, the nuclear ^{14}N spin can be entangled with the electron spin in order to realize a coherent coupling of qubits. However, the small interaction between the ^{14}N nuclear spin and the electron spin mainly located at the vacancy makes this difficult. A larger hyperfine splitting of 130 MHz (540 neV) is caused by the abundant ^{13}C nuclei neighboring the vacancy as shown in Fig. 8 (a). Typically, 3 % of the N-V centers have one direct ^{13}C neighbor which has a nuclear spin $I = 1/2$ and corresponding $m_I = 1/2, -1/2$.

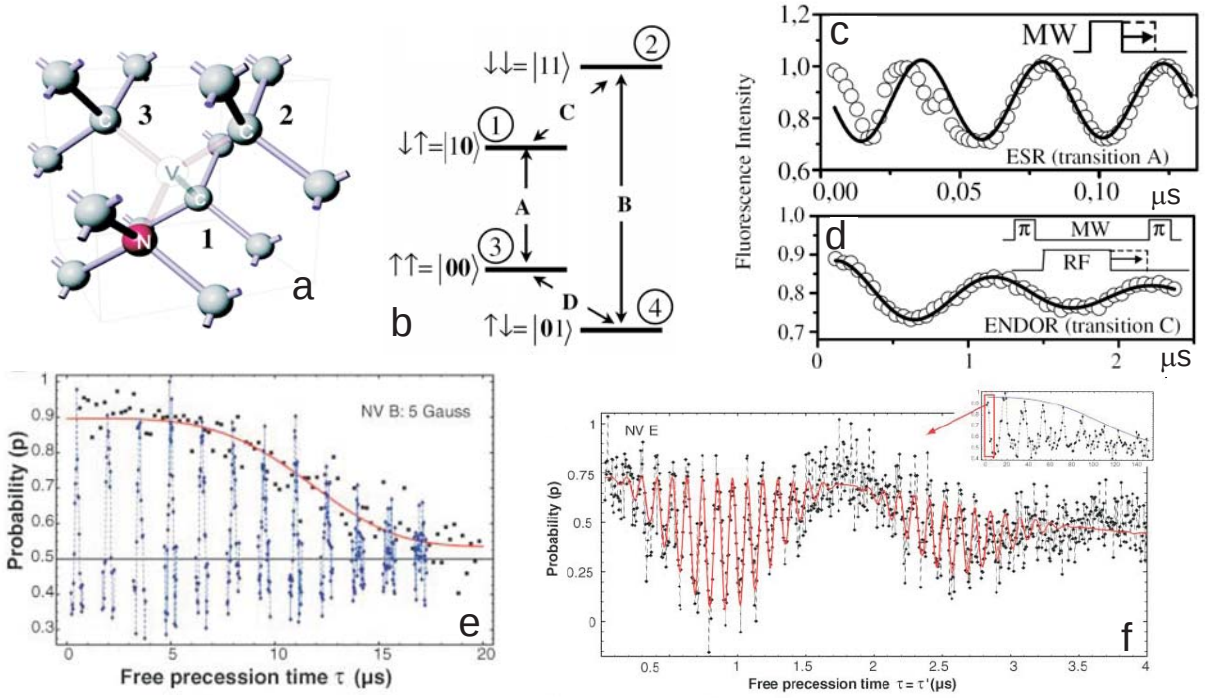


Fig. 8: *(a) atomic arrangement of the Nitrogen(N)-Vacancy(V) center in diamond with neighboring ^{13}C nuclear spins; (b) electronic states of the $^3A_2$ level coupled to one neighboring ^{13}C nuclear spin (see text); (c) fluorescence modified by a microwave pulse of different length tuned to transition A; (d) fluorescence modified by the pulse sequence displayed in the inset of two π-pulses tuned to the transition A in (b) and an intermediate radio frequency pulse tuned to the transition C of varying pulse length [73]; (e) spin-echo experiment of a single N-V center measured at $B = 0.5$ mT; each curve represents the τ'-dependent fluorescence signal for a given τ exhibiting a maximum at $\tau = \tau'$; the envelope function is a fit of the maxima according to $p(\tau) \propto \exp(-\tau/\tau_c)^4$ as expecetd for the interaction of the electron spin with a bath of nuclei exhibiting different Lamor frequencies due to hyperfine interaction; (f) the inset shows the revival of the spin echo signal after each period of the Lamor frequency of the bare ^{13}C spins at $B = 4.2$ mT; the main part shows the more complicated oscillating pattern present within the first peak of the revival curve in the inset. It is caused by the modified Lamor frequencies of the nuclei close to the N-V center, which are strongly influenced by the hyperfine interaction of the ^{13}C spins with the electron spin [16].*

This leads to an energy spectrum of the $^3A_2$ triplet as shown in Fig. 8(b), where the states are labeled according to $||m_I \pm 1/2||m_s| >$. States 1 and 2 are split by 130 MHz depending on the relative orientation of nuclear and electronic spin. State 3 and 4 are also slightly split by 2-10 MHz depending on the position of the ^{13}C due to the nuclear interaction between ^{14}N and ^{13}C [68]. Microwave fields can induce transitions between the m_s states (transition A, B), while radio frequency fields induce transitions between the m_I states (transition C, D). To initialize a certain ground state (3 or 4), the microwave field is first tuned to the transition of the other state, which leads to the population of the ^1A singlet by an additional laser field, only if this state is populated leading to a darkening of the fluorescence. If, e.g., state 3 is initialized, a microwave signal tuned to transition A leads to Rabi oscillations between state 3 and 1 as shown in Fig. 8 (c), while a microwave tuned to transition B does not. An additional radio frequency pulse

tuned to transition C and applied after a π pulse tuned to transition A now leads to an oscillating occupation of the nuclear spin levels, in turn leading to an oscillating detuning of the microwave π-pulse applied afterwards with the relevant electron spin transition. This leads to oscillating fluorescence as visible in Fig. 8 (d), which indicates the coherent rotation of the nuclear spin [68]. The dephasing time of the nuclear spin, as determined by spin-echo experiments of the nuclear spin is, at least, 20 ms.

Importantly, this scheme realizes a conditional rotation of the nuclear spin. The nuclear spin can only be rotated by a 130 Mhz π-pulse, if the electron spin is excited. Such a conditional coherent rotation (CROT) is very close to a conditional not gate (CNOT), which is a central building block of a quantum computer. The fidelity of the CROT-operation has been determined to be about 90 % using complex pulse sequences in order to determine the density matrix of the combined state [68]. Since the time scale of a single rotation is only about 0.1 μs, several 1000 coherent operations appear possible even at 300 K.

A more reliable production scheme of coupled spins with a success rate of 10 % is achieved by using N_2^+ bombardment of nanocrystalline diamond which leads to a $S = 1/2$ electron at a substitutional N-atom close to a N-V center [17]. The electron spin of the nitrogen can be polarized by resonance with the ($m_s = -1$)-state of the N-V center, which has to be induced by an appropriate external magnetic field. Coherent coupling has been proven by a beating pattern of the spin echo signal of the NV-center similar to the experiment in Fig. 7 (j) showing a coupling strength of 14 MHz and a dephasing rate of the $S = 1/2$ level of 3 ms. The spin echo beating revealed an additional coherent coupling of the electron spin ensemble to the two nuclear ^{14}N spins. The nuclear spin of the bare N-atom could even be polarized by flip-flop with its electron spin which is also probed at resonance between the two electron spins via fluorescence of the N-V center [74]. This provides, at least, an optically controllable three-qubit system, which is, however, limited by the relatively short coherence time of the electron spins.

A more promising approach with a much longer coherence time would be the addressing of individual ^{13}C nuclear spins. This approach uses the Lamor frequency of individual nuclear spins in magnetic field, which is significantly modified by the interaction with the electron spin [16, 18]. The inset in Fig. 8(f) shows that the decoherence of the electron spins in magnetic field is driven by the Lamor precessions of several nuclear spins which periodically change the effective magnetic field of the electron spin. If the spin echo time equals the time period of the Lamor precession of the nuclear spins, the reversed electron spin experiences exactly the same changing effective field as the original one and, thus, does not dephase effectively. Indeed the spin echo signal revives after each cycle of the bare Lamor precession of the nuclear spins. This revival is superimposed by a decay with $\tau_c = 240$ μs probably due to spin flip-flop within the nuclear bath. However, the nuclear spins close to the electron spin exhibit a different Lamor precession due to hyperfine interaction with the electron spin, which leads to the complex spin echo signal at short time scales τ shown in the main part of Fig. 8 (f). This pattern consists of a fast oscillation which depends on the individual defect, but not on the external magnetic field and a slow envelope oscillation, which depends linearly on magnetic field with a slope being different for each defect. The pattern indicates coherent coupling of the electron spin and a single nearby nuclear spin. The fast oscillation is caused by the dipolar interaction of the nuclear spin with the ($m_s = 1$)-component of the electron spin superposition[3] leading to an enhanced precession frequency of the nuclear spin independent of magnetic field. The slow

[3]The π-pulse in this experiment is tuned to the transition ($(m_s = 0) \rightarrow (m_s = 1)$), which is distinct from the ($(m_s = 0) \rightarrow (m_s = -1)$)-transition due to the applied field.

oscillation is caused by the terms of the interaction matrix between electron spin and nuclear spin not coupling to the z-component of the electron spin. These elements arise primarily due to a misalignement between the preferred zero-field m_s-axis and the external magnetic field, which mixes the (m_s)-states. Hence, the strength of this term depends on the B-field value and direction. The misalignment can be described as a modified effective g-factor of the nuclear spins, which, of course, is anisotropic as found experimentally [16]. Other defects, which are coupled to several nuclear spins, show more complex behaviour, which can be explained quantitatively by their different Lamor precessing caused by the electron spin.

Note that the additional effective magnetic field provided by the electron spin decouples the particular nuclear spin close to the electron spin effectively from all other nuclear spins prohibiting spin flip-flop. Moreover, it makes it addressable by radio frequency pulses due to its characteristic coupling strength to the electron spin [18]. Fig. 9 (a) shows the energy levels of an electron spin coupled to a nuclear spin in magnetic field. The hyperfine coupling allows to address a conditional transition, e.g., the transition ($m_s = 0 \rightarrow m_s = 1$) only if the nuclear spin is $\downarrow$. This process allows to map a superposition of a nuclear spin onto a superposition of an electronic spin, which is readable by fluorescence. Therefore, one excites a specific transition by a π-pulse, which leads to the excited state, only if the nuclear spin is down and fastly stops the Lamor precession of the nuclear spin due to the strong hyperfine field. If the electron spin was not excited, i.e., the nuclear spin is $\uparrow$, one simply waits for half a Lamor period, which transfers the nuclear spin to $\downarrow$. This corresponds to the following mapping $|0> \bigotimes(\alpha|\uparrow> + \beta|\downarrow>) \rightarrow |\downarrow> \bigotimes(\alpha|0> + \beta|1>)$. The reverse process can be done by starting with the finite position of the previous process, then waiting for half a period of the nuclear Lamor frequency and finally applying a π-pulse, which conditionally excites the $|0>$-state into the $|\downarrow> |1>$ state. This protocol has been used to map a prepared electron spin on the nuclear spin, let evolve the nuclear spin for a certain time and finally remap the nuclear spin back to the electron spin. The coherence of the evolution has then be probed by the remaining amplitude of the Rabi oscillations of the electron spin for different initial electron spin superpositions [18]. The long term precession of the nuclear spin measured by this protocol is shown in Fig. 9 (b) showing oscillations longer than 500 μs. The additional beating is caused by interactions with neighboring ^{13}C nuclear spins. This dephasing can again be controlled by spin echo, where π-pulses for the nuclear spin are created via excitation of ($m_s = 1$) for a controlled time. This leads to a nuclear spin decoherence time of, at least, 20 ms. Importantly, the decay rate of the nuclear spin is barely influenced by the fluorescence light used to prepare the electron spin. At $B = 2$ mT only one out of 1000 photons destroys coherence of the nuclear spin, which is traced back to a changed contact interaction in the excited state, which itself should become irrelevant at lower B-field.

The approach using nuclear spins can be extended to two nuclear spins around a N-V defect. Fig. 9 (c) shows the optically detected electron spin resonance of a single defect which shows four lines due to hyperfine splitting with two neighboring nuclear spins [75]. Fig. 9 (d) shows the corresponding energy level diagram. The different levels of the ($m_s = -1$)-state can be entangled by firstly preparing the ($m_s = 0$)-state optically, secondly applying a microwave π-pulse tuned to the $|00>$ transition in order to initialize the $|00>$-state, thirdly applying a $\pi/2$ radio frequency pulse tuned to the $|00> \rightarrow |01>$ transition in order to get $1/\sqrt{2} \cdot (|00> + |01>)$ and finally applying a π-pulse tuned to the $|01> \rightarrow |11>$ transition resulting in the Bell state $\Phi^{\pm} = 1/\sqrt{2} \cdot (|00> \pm |11>)$. The adequate preparation can be probed by so-called tomography, which uses certain pulse sequences for the read-out of different superpositions. The presence of off-diagonal elements in Fig. 9 (e) clearly proves the entanglement of the

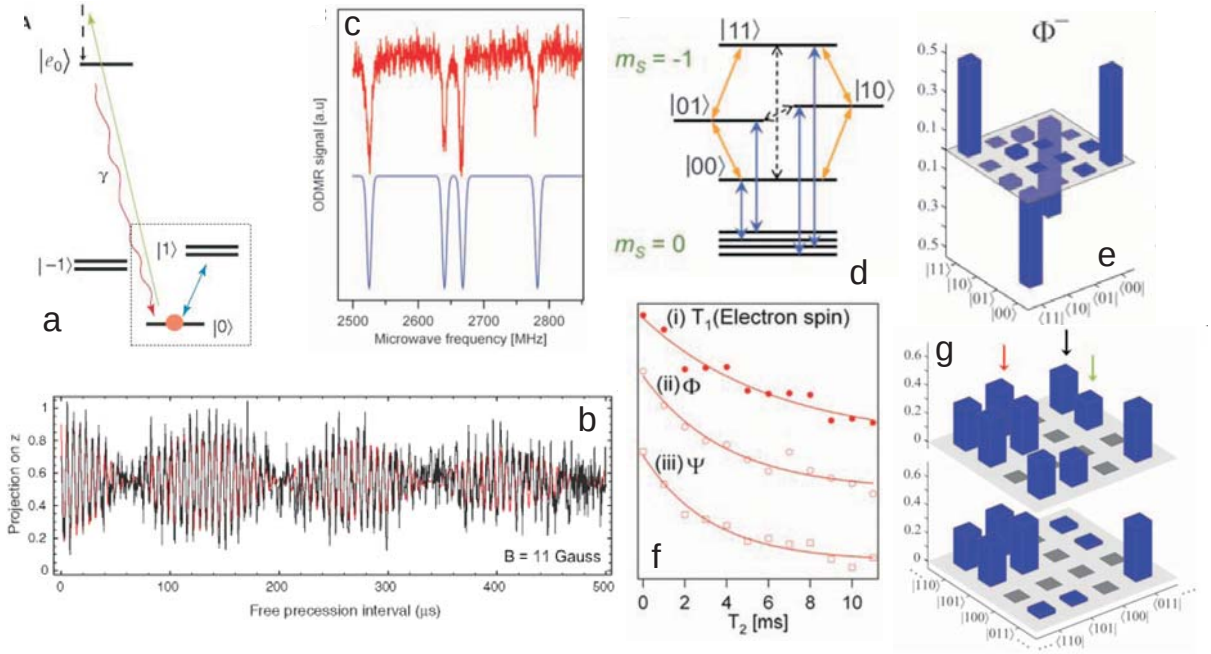


Fig. 9: *(a) Energy levels of the electron spin of the N-V center in magnetic field including the hyperfine splitting of the $m_s = 1$ states in the ${}^3A_2$ level; (b) Lamor precession of a nuclear spin close to the electron spin measured by mapping the electron spin configuration to the nuclear spin and remapping it to the electron spin for read-out [18]; (c) Optically detected magnetic resonance of a single electron spin with fourfold hyperfine splitting indicating two ${}^{13}C$ nuclei in the neighborhood of the electron spin (top: experiment, bottom: simulation) [75]; (d) level scheme corresponding to (c); (e) measured density matrix of the state entangling two ${}^{13}C$ nuclear spins $\Phi^- = 1/\sqrt{2} \cdot (|00> - |11>)$; (f) time dependence of the strength of the off-diagonal elements of the density matrix for $\Phi^\pm$ and the state $\Psi^\pm = 1/\sqrt{2} \cdot (|01> \pm |10>)$ in comparison with the spin relaxation time T_1 of the electron spin in the N-V center; (g) density matrix of the state $W = 1/\sqrt{3} \cdot (|110> + e^{i\phi} \cdot |101> + e^{i\psi} \cdot |011>)$ measured directly after preparation (top) and $\tau = 4.4$ μs later (bottom) [75].*

Φ^-- state. The coherent evolution of the state can be probed using the coherent evolution of the density matrix probed by microwave pulses applied after a certain delay to the state preparation. The result is shown in Fig. 9 (f) demonstrating that the entanglement decoheres on the time scale of milliseconds probably limited by the spin relaxation of the electron, which changes the hyperfine field of the nuclear spins dramatically. Using more complex pulse sequences, also the electron spin can be entangled with the two nuclear spins leading, e.g., to a state $W = 1/\sqrt{3} \cdot (|110> + e^{i\phi} \cdot |101> + e^{i\psi} \cdot |011>)$, where the first entry of the ket marks the electron spin using $m_s = 1 \rightarrow |1>$ and $m_s = 0 \rightarrow |0>$. Thereby, ψ and ϕ are arbitrary phases [75]. Fig. 9 (g) shows the resulting tomography pattern measured directly after preparation and 4.4 μs later. Interestingly, the off-diagonal elements describing superpositions with different electron spins have disappeared due to the fast electron spin decoherence, while the off-diagonal elements with the same electron spin survived due to the longer decoherence time of the nuclear spins.

Although the approach of entangling nuclear spins via the electron spin of the N-V center appears to be restricted to only a few nuclei around one electron spin, distant coupling of electron

spins by photons has been proposed to enhance this number using the spin dependent transition frequencies of the electron spin according to transition a,b, and c in Fig. 7 (a) [71, 76].

6 Quantum dots and spin centers in graphene and carbon nanotubes

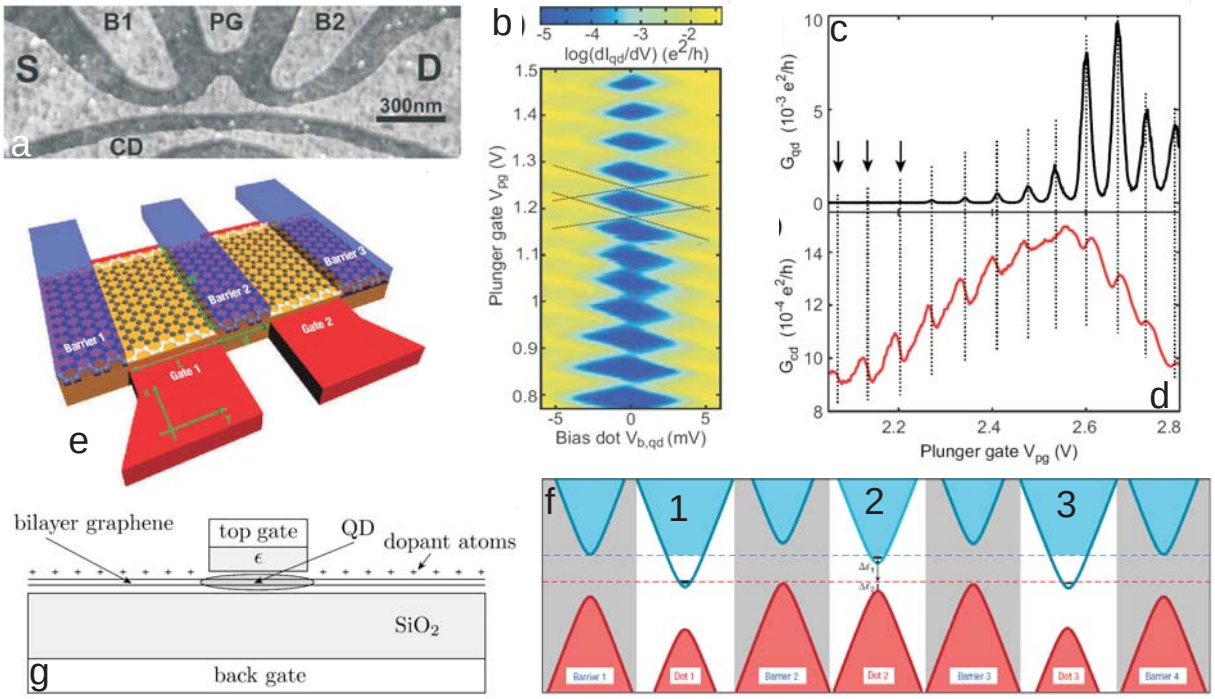


Fig. 10: *(a) Atomic force microscope image of a quantum dot with nearby gates B1, PG, B2 and a quantum point contact CD, all made out of graphene [81]; (b) color plot of the differential conductance of the dot shown in (a) as a function of source-drain voltage $V_{b,qd}$ and plunger gate voltage V_{pg}; (c) conductance $G_{qd} = dI_{qd}/dV$ of the dot as a function of V_{pg} at $V_{b,qd} = 0.5$ mV; (d) conductance of the quantum point contact obtained simultaneously with (c) at $V_{CD} = 0.5$ mV; (e) sketch of two quantum dots for qubits using armchair edges of a graphene ribbon of certain width in order to provide a gap; gates are used to confine electrons within this gap in lateral direction [83]; (f) sketch of the band structure of three coupled quantum dots; tuning of the valence band of the area between dot 1 and dot 3 close to the resonant states within these two dots (red dashed line) leads to efficient coupling between the resonant states in dot 1 and dot 3; (g) sketch of a circular quantum dot induced by a gate relying on the natural gap of bilayer graphene under electric field provided by the dopant atoms [88].*

The only disadvantage of the N-V centers in diamond is that they are addressed optically, which might limit the scalability of the approach, although electrical control by the Stark effect has been proposed [72]. Electrically addressable qubits have been implemented using either superconducting flux qubits [77] or quantum dots in GaAs [15, 78]. The later approach is limited by the spin dephasing due to hyperfine interaction [79], which is strong, since both nuclei, Ga and As, provide a nuclear spin. Thus, graphene quantum dots probably isotopically purified might be a valuable alternative.

Along the road for implementing qubits in graphene quantum dots, the first breakthroughs have been realized, i.e., quantum dots have been formed by lateral confinement using oxygen etching combined with standard lithography and signatures of Coulomb blockade and confinement energies have been found [80]. Moreover, the minimal-invasive detection of the charge within the quantum dot by a close-by quantum point contact has been realized [81]. Figure 10 (a) shows an atomic force microscope image of a quantum dot structure. The tunneling barriers from the reservoirs to the dot can be tuned by the two finger gates B1 and B2, while the electrostatic potential and, thus, the filling of the dot can be tuned by the plunger gate PG. Applying a voltage between S and D leads to transport through the dot, if a level of the dot is close to the Fermi level. Figure 10 (b) shows the Coulomb diamonds of the differential conductance of the dot as a function of source-drain and gate voltage revealing a charging energy of about 4 meV. Clearly, the number of charges on the dot can be controlled. Qubit proposals require a detection of the charge, e.g., in quantum dot molecules [82], without using this charge itself. An elegant way is a quantum -point contact, which has a conductance of $N \cdot e^2/h$ with N being an integer and giving the number of subbands contributing to the transport. Importantly, the conductance changes rapidly by e^2/h, if an additional subband is shifted across the Fermi level by an electrostatic field. Tuning the field to such a transition, the quantum point contact is an excellent detector of additional fluctuations in electric field and can detect additional single charges in distances up to several 100 nm. Thus, the quantum point contact can be used as a charge detector CD and is implemented in the device shown in Fig. 10(a). The signal of the charge detector as a function of plunger gate is shown in Fig. 10 (d) in comparison with the conductance through the quantum dot in Fig. 10 (c). Each time an additional electron is brought to the dot, which results in a peak in the quantum dot conductance, the quantum point contact exhibits a kink in its conductance. Thus, changes in charge of the quantum dot can easily be detected quasi non-invasively.

In order to use the graphene quantum dots as qubits, spin states have to be prepared selectively. Besides the confinement of the electrons, it is necessary to break the so-called valley symmetry of the quantum dots, i.e., the existence of inequivalent states at the opposite edges of the Brillouin zone. The following two proposals have been made. The first is displayed in Fig. 10 (e) [83]. It uses a ribbon of graphene with armchair edges, which has similar quantization conditions as a zigzag nanotube, i.e., if the number of rows in the ribbon is not a multiple of three, the ribbon exhibits a gap and is semiconducting. For a width of the ribbon $W = 30$ nm, the gap size is $\Delta E = h \cdot v_F/(3W) = 60$ meV with $v_F = 10^6$ m/s being the Fermi velocity of graphene. As usual, the gap can be used to confine electrons, if the gap of adjacent areas is tuned to, e.g., the conduction band area of the dot in between. This is correct for any semiconducting ribbon, but not sufficient for qubit operation. The additional requirement is, that the valley degeneracy of the confined states is broken. If the valley degeneracy persists, there will be no energy difference between a triplet and a singlet of an electron pair in the quantum dot, and, thus, there will be no exchange coupling to a neighboring quantum dot, which requires an energy penalty for triplet occupation with respect to singlet occupation. Consequently, only dots without valley degeneracy can be used for conditional qubit operations as CNOT.

Using the armchair edges to induce the gap leads to an additional lifting of the valley degeneracy due to the fact, that both atoms contribute to the edges, which requires a mixing of k states [84]. A favorable property of these type of quantum dots is, that the gap is relatively small, allowing a long distance coupling of quantum dots as displayed in Fig. 10 (f). The states in quantum dot 1 and quantum dot 3 are tuned into resonance and the areas in between the dots are tuned such that their valence bands are very close to the resonant state providing a minimum of a tunnel barrier

and, thus, a very small exponential decay of the resonant states within the barrier. Consequently, a significant density of the state of quantum 1 overlaps with quantum dot 3 leading to an efficient coupling. Coupling energies of several meV have been found for distances between the dots of 300 nm. This long distance coupling appears favorable, e.g., for fault-tolerant qubit operations [85].

Another possibility is to use natural gaps in graphene as they have been reported for bilayer graphene on SiO_2 [86] or for graphene on SiC [87]. These gaps can again be tuned by gate electrodes as shown in Fig. 10 (g) leading to confined states. The degeneracy of the k-valleys can be lifted using circular symmetry of the dots and a magnetic field applied perpendicular to the graphene plane, which basically makes use of the chirality of the graphene band structure, i.e., states with a positive angular momentum correspond to $\underline{k}$, while states with negative angular momentum correspond to $-\underline{k}$ [88]. Typical numbers for valley splitting are 3.5 meV for a dot size of 50 nm and an applied field of 3.5 T. Thus, well-developed proposals for qubits

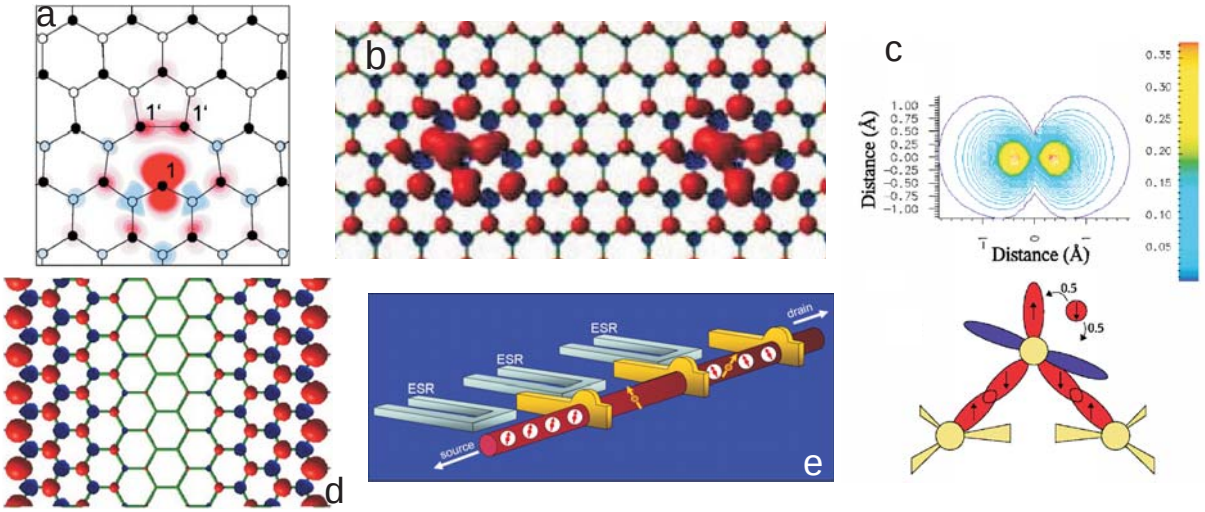


Fig. 11: *(a) Calculated spin density of a vacancy in graphene (color) superimposed on the atomic structure [89]; (b) spin density of two coupled vacancies, each saturated with two hydrogen atoms [91]; (c) spin density of a carbon adatom on graphene (top) and corresponding atomic structure with dangling bonds of the adatom (bottom) [93]; (d) spin density of zigzag edges of graphene nanoribbons [94]; (e) sketch of a proposal for qubits in carbon nanotubes filled with endofullerenes $C_{59}N$.*

in graphene are existing, but experiments have not detected single spins in graphene quantum dots so far. Another opportunity would be the use of interinsic spins in graphene present in the ground state of defects similar to the N-V center in diamond. Such ground state spins have been found theoretically for vacancies [89, 90], for vacancies in combination with hydrogen atoms [91, 92], for adatoms [93] as well as for the zigzag edges of nanoribbons [94, 95]. Fig. 11 (a)-(d) display the corresponding spin densities of the different defects partly having $S \simeq 1/2$ as the adatom and partly having $S \simeq 1$ as the vacancy. Neighboring defects are coupled ferromagnetically if located on the same sublattice and antiferromagnetically, if located on different sublattices and the coupling strength might be tuned by gates in between the defects [96, 95]. The g-factor is most likely $g \simeq 2$ [97]. Nothing is known about spin relaxation and spin coherence of these defects, but they might have similar properties as the spins of the N-V centers in diamond. If it is possible to address these spins electrically or if it is easier to use quantum dots

as building blocks for qubits, is currently unknown and requires further research.
Finally, it should be mentioned that also carbon nanotubes can act as building blocks for qubits having similar properties as the ribbons displayed in Fig. 10 (e) , i.e., a similar distant coupling as shown in Fig. 10 (f) can also be achieved for nanotubes [98]. A more advanced concept could even use the filling properties of carbon nanotubes in order to deposit spins into the tube, e.g., as C_{59}N-molecules exhibiting $S = 1/2$. This device, which is displayed in Fig. 11 (e), could store the qubits within the molecules, which are initialized and read out by there interaction with the nanotube current modified by spin carrying quantum dots and manipulated by side gates providing microwave fields [99]. The major idea is to further decouple the qubit spins from the environement as, e.g., from the interaction with nanotube phonons very similar as it happens during the transfer of electron spin to nuclear spins in diamond. However, so far only first steps towards quantum dot control in carbon nanotubes have been realized [100].

7 Summary

Within this review, I have summarized the current development concerning the use of spin information within carbon materials. The big advantages of spins in carbon, which are the small spin-orbit interaction and the small hyperfine interaction, makes this field of research rapidly growing. Major achievements so far are the realization of a gate-tunable spin valve using carbon nanotubes [50], the observation of the spin valve effect in graphene [13], and the realization of entangled three-partite states using the N-V-center in diamond and two adjacent ^{13}C nuclear spins [75]. But there are many more proposals, which will probably be realized in the near future. Important breakthroughs like, e.g., the realization of tunable quantum dots in graphene [80, 81], have already been achieved. Thus, spins in carbon materials promise to remain an exciting area of research.

8 Acknowledgement

It is pleasure to acknowledge the careful reading of the manuscript by D. Subramaniam.

References

[1] W. J. M. Naber, S. Faez, W. G. van der Wiel, J. Phys. D: Appl. Phys. **40**, R205 (2007); Z. H. Xiong, D. Wu, Z. Valy Vardeny, J. Shi, Nature **427**, 821 (2004); J. R. Petta, S. K. Slater, D. C. Ralph, Phys. Rev. Lett. **93**, 136601 (2004); D. R. McCamey, H. A. Seipel, S. Y. Paik, M. J. Walter, N. J. Borys, J. M. Lupton, C. Boehme, Nature Mat. **7**, 723 (2008).

[2] S. Koizumi, M. Kamo, Y. Sato, H. Ozaki, T. Inuzuka, Appl. Phys. Lett. **71**, 1065 (1997); H. Kato, S. Yamasaki, H. Okushi, Appl. Phys. Lett. **86**, 222111 (2005); S. Bhattacharyya, O. Auciello, J. Birrel, J. A. Carlisle, L. A. Curtiss, A. N. Goyette, D. N. Gruen, A. R. Krauss, J. Schlueter, A. Sumant, P. Zapol, Appl. Phys. Lett. **79**, 1441 (2001); S. Bhattacharyya, Phys. Rev. B **70**, 125412 (2004).

[3] P. Strobel, M. Riedel, J. Ristein, L. Ley, Nature **430**, 439 (2004).

[4] K. Hirama, S. Myamoto, H. Matsudaira, K. Yamada, H. Kawarada, T. Chikyo, H. Koinuma, K. Hasegawa, H. Umezawa, Appl. Phys. Lett. **88**, 112117 (2006).

[5] D. S. Bethune, C. H. Klang, M. S. de Vries, G. Gorman, R. Savoy, J. Vazquez, R. Beyers, Nature **363**, 605 (1993).

[6] A. K. Geim, K. S. Novoselov, Nature Mat. **6**, 183 (2007).

[7] S. J. Tans, A. R. M. Verschueren, C. Dekker, Nature **393**, 49 (1998); A. Bachtold, P. Hadley, T. Nakanishi, C. Dekker, Science **294**, 1317 (2001); A. Javey, J. Guo, Q. Wang, M. Lundstron, H. Dai, Nature **424**, 654 (2003); S. J. Wind, J. Appenzeller, P. Avouris, Phys. Rev. Lett. **91**, 058301 (2003).

[8] M. C. Lemme, T. J. Echtermeyer, M. Baus, H. Kurz, IEEE Electron Device Lett. **28**, 282 (2007); B. Özyilmaz, P. Jarillo-Herrero, D. Efetov, P. Kim, Appl. Phys. Lett. **91**, 192107 (2007).

[9] http://www.itrs.net.

[10] M. S. Arnold, A.A. Green, J. F. Hulvat, S. I. Stupp, M. C. Hersam, Nature Nanotechnol. **1**, 60 (2006).

[11] C. Berger, Z. M. Song, X. B. Li, X. S. Wu, N. Brown, C. Naud, D. Mayou, T. B. Li, J. Hass, A. N. Marchenkov, E. H. Conrad , P. N. First, W. A. de Heer, Science **312**, 1191 (2006).

[12] L. E. Hueso, J. M. Pruneda, V. Ferrari, G. Burnell, J. P. Valdes-Herrera, B. D. Simons, P. B. Littlewood, E. Artacho, A. Fert, N. D. Mathur, Nature **445**, 410 (2007).

[13] N. Tombros, C. Josza, M. Popinciuc, H. T. Jonkman, B. van Wees, Nature **448**, 571 (2007).

[14] J. Göthel, M. Ivanenko, P. Hering, W. Fuss, K. L. Kompa, Appl. Phys. B-Lasers and Optics **62**, 329 (1996).

[15] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, A. C. Gossard, Science **309**, 2180 (2005).

[16] L. Childress, M. V. G. Dutt, J. M. Taylor, A. S. Zibrov, F. Jelezko, J. Wrachtrup, P. R. Hemmer, M. D. Lukin, Science **314**, 281 (2006).

[17] T. Gaebel, M. Domhan, I. Popa, C. Wittmann, P. Neumann, F. Jelezko, J. R. Rabeau, N. Stavrias, A. D. Greentree, S. Prawer, J. Meijer, J. Twamley. P. R. Hemmer, J. Wrachtrup, Nature Phys. **2**, 408 (2006).

[18] M. V. G. Dutt, L. Childress, L. Jiang, E. Togan, J. Maze, F. Jelezko, A. S. Zibrov, P. R. Hemmer, M. D. Lukin, Science **316** 1312 (2007).

[19] P. Blake, E. W. Hill, A. H. Castro Neto, K. S. Novoselov, D. Jiang, R. Yang, T. J. Booth, A. K. Geim, Appl. Phys. Lett. **91**, 063124 (2007); D. S. L. Abergel, A. Russell, V. I. Falko, Appl. Phys. Lett. **91**, 063125 (2007).

[20] A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, A. K. Geim, Phys. Rev. Lett. **97**, 187401 (2006).

[21] J. C. Meyer, A. K. Geim, M. I. Katsnelson, K. S. Novoselov, T. J. Booth, S. Roth, Nature **446**, 60 (2007); A. Fasolino, J. H. Los, M. I. Katsnelson, Nature Mat. **6**, 858 (2007);

[22] V. Geringer, M. Liebmann, T. Echtermeyer, S. Runte, M. Schmidt, R. Rückamp, M. Lemme, M. Morgenstern, cond-mat 0806.1028, Phys. Rev. Lett., in press.

[23] S. Krummacher, *29. Ferienkurs des Instituts für Festkörperforschung: Physik der Nanostrukturen*, D1 (FZ Jülich, Jülich, 1998); B. Kessler, *29. Ferienkurs des Instituts für Festkörperforschung: Physik der Nanostrukturen*, D4 (FZ Jülich, Jülich, 1998).

[24] C. L. Kane, E. J. Mele, Phys. Rev. Lett. **78**, 1932 (1997).

[25] J. W. G. Wildoer, L. C. Venema, A. G. Rinzler, R. E. Smalley, C. Dekker, Nature **391**, 59 (1998).

[26] M. Ouyang, J. L. Huang, C. L. Cheung, C. M. Lieber, Science **292**, 702 (2001).

[27] M. Bockrath, D. H. Cobden, J. Lu, A. G. Rinzler, R. E. Smalley, T. Balents, P. L. McEuen, Nature **397**, 598 (1999).

[28] J. Lee, S. Eggert, H. Kim, S. J. Kahng, H. Shinohara, Y. Kuk, Phys. Rev. Lett. **93**, 166403 (2004).

[29] A. F. Hebard, M. J. Rosseinsky, R. C. Haddon, D. W. Murphy, S. H. Glarum, T. T. M. Palstra, A. P. Ramirez, A. R. Kortan, Nature **350**, 600 (1991).

[30] T. L. Makarova, B. Sundqvist, R. Höhne, P. Esquinazi, Y. Kopelevich, P. Scharff, V. A. Davydov, L. S. Kashevarova, A. V. Rakhmanina, Nature **413**, 716 (2001); S. Mathew, B. Satpati, B. Joseph, B. N. Dev, R. Nirmala, S. K. Malik, R. Kesavamoorthy, Phys. Rev. B **75**, 075426 (2007).

[31] P. Esquinazi, D. Spemann, R. Höhne, A. Setzer, K. H. Han, T. Butz, Phys. Rev. Lett. **91**, 227201 (2003); S. Talapatra, P. G. Ganesan, T. Kim, R. Vajtai, M. Huang, M. Shima, G. Ramanath, D. Srivastava, S. C. Deevi, P. M. Ajayan, Phys. Rev. Lett. **95**, 097201 (2005).

[32] D. J. Hornbaker, S. J. Kahng, S. Misra, B. W. Smith, A. T. Johnson, E. J. Mele, D. E. Luzzi, A. Yazdani, Science **295**, 828 (2002).

[33] T. Ando, H. Suzuura, J. Phys. Soc. Jpn. **71**, 2753 (2002); T. Ando, T. Nakanishi, R. Saito, J. Phys. Soc. Jpn. **67**, 2857 (1998); T. Ando, T. Nakanishi, J. Phys. Soc. Jpn. **67**, 1704 (1998).

[34] S. J. Tans, M. H Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs, C. Dekker, Nature **386**, 474 (1997).

[35] J. Kong, E. Yenilmez, T. W. Tombler, W. Kim, H. Dai, R. B. Laughlin, L. Liu, C. S. Jayanthi, S. Y. Wu, Phys. Rev. Lett. **87**, 106801 (2001).

[36] A. Bachthold, C. Strunk, J. P. Salvetat, J. M. Bonard, L. Forro, T. Nussbaumer, C. Schönenberger, Nature **397**, 673 (1999).

[37] A. Bachthold, M. S. Fuhrer, S. Plyasunov, M. Forereo, E. H. Anderson, A. Zettl, P. L. McEuen, Phys. Rev. Lett. **84**, 6082 (2000).

[38] K. Liu, Ph. Avouris, R. Martel, W. K. Hsu, Phys. Rev. B **63**, 161404 (2001).

[39] S. Frank, P. Poncharal, Z. L. Wang, W. A. de Heer, Science **280**, 1744 (1998).

[40] K.I. Bolotin, K. J. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, H. L. Störmer, Sol. St. Com. **146**, 351 (2008).

[41] S. V. Morozov, K. S. Novoselov, M. I. Katsnelson, F. Schedin, L. A. Ponomarenko, D. Jiang, A. K. Geim, Phys. Rev. Lett. **97**, 016801 (2006).

[42] F. V. Tikhonenko, D. W. Horsell, R. V. Gorbachev, A. K. Savchenko, Phys. Rev. Lett. **100**, 056802 (2008).

[43] A. N. Pasupathy, R. C. Bialczak, J. Martinek, J. E. Grose, L. A. K. Donev, P. L. McEuen, D. C. Ralph, Science **306**, 86 (2004).

[44] M. I. Eremets, Semicond. Sci. Technol. **6**, 439 (1991).

[45] K. Tsukagoshi, B. W. Alphenaar, H. Ago, Nature **401**, 572 (1999).

[46] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip, B. J. van Wees. Phys. Rev. B **62**, R4790 (2000).

[47] E. I. Rashba, Phys. Rev. B **62**, R16267 (2000); A. Fert, H. Jaffres, Phys. Rev. B **64**, 184420 (2001).

[48] J. R. Kim, H. M. So, J. J. Kim, J. Kim, Phys. Rev. B **66**, 233401 (2002).

[49] N. Tombros, S. J. van der Molen, B. J. van Wees, Phys. Rev. B **73**, 233403 (2006).

[50] S. Sahoo, T. Kontos, J. Furer, C. Hoffmann, M. Gräber, A. Cottet, C. Schönenberger, Nature Phys. **1**, 99 (2005).

[51] T. Schäpers, J. Nitta, H. B. Heersche, H. Takayanagi, Phys. Rev. B **64**, 125314 (2001).

[52] S. Datta, B. Das, Appl. Phys. Lett. **56**, 665 (1990).

[53] H. T. Man, I. J. W. Wever, A. F. Morpurgo, Phys. Rev. B **73**, 241401 (2006).

[54] E. W. Hill, A. K. Geim, K. Novoselov, F. Schedin, P. Blake, IEEE Trans. Magn. **42**(10), 2694 (2006).

[55] N. Tombros, S. Tanabe, A. Veligura, C. Josza, M. Popinciuc, H. T. Jonkman, B. J. van Wees, Phys. Rev. Lett. **101**, 046601 (2008).

[56] C. Josza, M. Popinciuc, N. Tombros, H. T. Jonkman, B. J. van Wees, Phys. Rev. Lett. **100**, 236603 (2008).

[57] S. Cho, Y. F. Chen, M. S. Fuhrer, Appl. Phys. Lett. **91**, 123105 (2007).

[58] M. Nishioka, A. M . Goldman, Appl. Phys. Lett. **90**, 252505 (2007).

[59] A. Gruber, A. Dräbenstedt, C. Tietz, L. Fleury, J. Wrachtrup, C. von Borczyskowski, Science **276**, 2012 (1997).

[60] J. R. Rabeau, P. Reichart, G. Tamanyan, D. N. Jamieson, S. Prawer, F. Jelezko, T. Gaebel, I. Popa, M. Domhan, J. Wrachtrup, Appl. Phys. Lett. **88**, 023113 (2006).

[61] T. A. Kennedy, J. S. Colton, J. E. Butler, R. C. Linares, P. J. Doering, Appl. Phys. Lett. **83**, 4190 (2003).

[62] R. J. Epstein, F. M. Mendoza, Y. K. Kato, D. D. Awschalom, Nature Phys. **1**, 94 (2005).

[63] A. Beveratos, R. Brouri, T. Gacoin, A. Villing, J. P. Poizat, P. Grangier, Phys. Rev. Lett. **89**, 187901 (2002).

[64] V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, J. F. Roch, Science **315**, 966 (2007).

[65] F. Jelezko, I. Popa, A. Gruber, C. Tietz, J. Wrachtrup, A. Nizovtsev, S. Kilin, Appl. Phys. Lett. **81**, 2160 (2002).

[66] N. B. Manson, J. B. Harrison, M. J. Sellars, Phys. Rev. B **74**, 104303 (2006).

[67] D. A. Redman, S. Brown, R. H. Sands, S. C. Rand, Phys. Rev. Lett. **67**, 3420 (1991); J. Harrison, M. J. Sellars, N. B. Manson, Diamond. Rel. Mater. **15**, 586 (2006).

[68] F. Jelezko, T. Gaebel, I. Popa, M. Domhan, A. Gruber, J. Wrachtrup, Phys. Rev. Lett. **93**, 130501 (2004).

[69] E. van Oort, P. Stroomer, M. Glasbeek, Phys. Rev. B **42**, 8605 (1990).

[70] A. Gruber, A. Dräbenstedt, C. Tietz, L. Fleury, J. Wrachtrup, C. von Borczyskowski, Science **276**, 2012 (1997).

[71] A. Batalov, C. Zierl, T. Gaebel, P. Neumann, I. Y. Chan, G. Balasubramanian, P. R. Hemmer, F. Jelezko, J. Wrachtrup, Phys. Rev. Lett. **100**, 077401 (2008).

[72] P. Tamarat, T. Gaebel, J. R. Rabeau, M. Khan, A. D. Greentree, H. Wilson, L. C. L. Hollenberg, S. Prawer, P. Hemmer, F. Jelezko, J. Wrachtrup, Phys. Rev. Lett. **97**, 083002 (2006).

[73] F. Jelezko, T. Gaebel, I. Popa, A. Gruber, J. Wrachtrup, Phys. Rev. Lett. **92**, 076401 (2004).

[74] R. Hanson, F. M. Mendoza, R. J. Epstein, D. D. Awschalom, Phys. Rev. Lett. **97**, 087601 (2006).

[75] P. Neumann, N. Mizuochi, F. Rempp, P. Hemmer, H. Watanabe, S. Yamasaki, V. Jacques, T. Gaebel, F. Jelezko, J. Wrachtrup, Science **320**, 1326 (2008).

[76] L. M. Duan, R. Raussendorf, Phys. Rev. Lett. **95**, 080503 (2005).

[77] T, Hime, P. A. Reichardt, B. L. T. Plourde, T. L. Robertson, C. E. Wu, A. V. Ustinov, J. Clarke, Science **314**, 1427 (2006); M. Steffen, M. Ansmann, R. C. Bialczak, N. Katz, E. Lucero, R. McDermott, M. Neeley, E. M. Weig, A. N. Cleland, J. M. Martinis, Science **313**, 1423 (2006); A. O. Niskanen, K. Harrabi, F. Yoshihara, Y. Nakamura, S. Lloyd, J. S. Tsai, Science **316**, 723 (2007); J. H. Plantenberg, P. C. de Groot, C. J. P. M. Harmans, J. E. Mooij, Nature **447**, 836 (2007); P. J. Leek, J. M. Fink, A. Blais, R. Bianchetti, M. Göppl, J. M. Gambetta, D. I. Schuster, L. Frunzio, R. J. Schoelkopf, A. Wallraff, Science **318**, 1889 (2007).

[78] F. H. L. Koppens, C. Buizert, K. J. Tielrooij, I. T. Vink, K. C. Nowack, T. Meunier, L. P. Kouwenhoven, L. M. K. Vandersypen, Nature **442**, 766 (2006).

[79] A. C. Johnson, J. R. Petta, J. M. Taylor, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, A. C. Gossard, Nature **435**, 925 (2005).

[80] L. A. Ponomarenko, F. Schedin, M. I. Katsnelson, R. Yang, E. W. Hill, K. S. Novoselov, A. K. Geim, Science **320**, 356 (2008); C. Stampfer, J. Güttinger, F. Molitor, D. Graf, T. Ihn, K. Ensslin, Appl. Phys. Lett. **92**, 012102 (2008).

[81] J. Güttinger, C. Stampfer, S. Hellmüller, F. Molitor, T. Ihn, K. Ensslin, Appl. Phys. Lett. **93**, 212102 (2008).

[82] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, L. M. K. Vandersypen, Rev. Mod. Phys. **79**, 1217 (2007).

[83] B. Trauzettel, D. V. Bulaev, D. Loss, G. Burkard, Nature Phys. **3**, 192 (2007).

[84] L. Brey, H. A. Fertig, Phys. Rev. B **73**, 235411 (2006).

[85] K. M. Swore, B. M. Terhal, D. P. DiVicenzo, Phys. Rev. A **72**, 022317 (2005)

[86] E. V. Castro, K. S. Novoselov, S. V. Morozov, N. M. R. Peres, J. M. B. Lopes dos Santos, J. Nilsson, F. Guinea, A. K. Geim, A. H. Castro Neto, Phys. Rev. Lett. **99**, 216802 (2007); J. B. Oostinga, H. B. Heersche, X. Liu, A. F. Morpurgo, L. M. K. Vandersypen, Nature Mat. **7**, 151 (2008).

[87] S. Y. Zhou, G. H. Gweon, A. V. Federov, P. N. First, W. A. de Heer, D. H. Lee, F. Guinea, A. H. Castro Neto, A. Lanzara, Nature Mat. **6**, 770 (2007); T. Ohta, A. Bostwick, T. Seyller, K. Horn, E. Rotenberg, Science **313**, 951 (2006).

[88] P. Recher, J. Nilsson, G. Burkard, B. Trauzettel, cond-mat 0810.0419 (2008).

[89] O. V. Yazyev, L. Helm, Phys. Rev. B, **75**, 125408 (2007).

[90] P. O. Lehtinen, A. S. Foster, Y. Ma, A. V. Krasheninnikov, R. M. Nieminen, Phys. Rev. Lett. **93**, 187202 (2004).

[91] L. Pisani, B. Montanari, N. M. Harrison, New J. Phys., **10**, 033002 (2008).

[92] E. J. Duplock, M. Scheffler, P. J. D Lindan, Phys. Rev. Lett. **92**, 225502 (2004).

[93] P. O. Lehtinen, A. S. Foster, A. Ayuela, A. Krasheninnikov, K. Nordlund, R. M. Nieminen, Phys. Rev. Lett. **91**, 017202 (2003).

[94] L. Pisani, J. A. Chan, B. Montanari, N. M. Harrison, Phys. Rev. B **75**, 064418 (2007).

[95] Y. W. Son, M. L. Cohen, S. G. Louie, Nature **444**, 347 (2006); H. Lee, Y. W. Son, N. Park, S. Han, J. Yu, Phys. Rev. B **72**, 174431 (2005).

[96] V. K. Dugaev, V. I. Litvinov, J. Barnas, Phys. Rev. B **74**, 224438 (2006).

[97] K. W. Lee, C. E. Lee, Phys. Rev. Lett. **97**, 137206 (2006).

[98] D. V. Bulaev, B. Trauzettel, D. Loss, Phys. Rev. B **77**, 235301 (2008).

[99] C. Meyer, private communication

[100] H. Ingerslev Jorgensen, K. Grove-Rasmussen, K.-Y. Wang, A. M. Blackburn, K. Flensberg, P. E. Lindelof, D. A. Williams; Nature Phys. **4**, 536 (2008); N. Mason, M. J. Biercuk, C. M. Marcus, Science **303**, 655 (2004); S. Sapmaz, C. Meyer, P. Beliczynski, P. Jarillo-Herrero, L. P. Kouwenhoven, Nano Lett. **6**, 1350 (2006).

C 6 Spin and Transport through Quantum Dots

M. R. Wegewijs
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Single electron charging effects on tunneling **3**

3 Single quantum dots **4**
3.1 Linear transport 6
3.2 Non-linear transport 7
3.3 Discrete states and excitations 9
3.4 Transport theory - occupations and rates 12

4 Single electron spin-valve **14**
4.1 Spin dynamics due to transport 15
4.2 Charge dynamics and non-linear transport 18
4.3 Outlook: mapping the exchange field with higher order tunneling 20

1 Introduction

Quantum dots are small devices in which few mobile electrons are confined and can be connected to macroscopic electrodes. The properties of such systems are dominated by the quantization of electronic charge- and possibly other degrees of freedom (orbit, *spin*). This quantization makes quantum dots very similar to atoms (coupled dots similar to molecules). Therefore one might think that no fundamental new physics is involved. However, their tunability and the possibility of wiring them up in an external electric circuit enables new regimes to be accessed. In particular in *transport*, on can study in detail the interplay of strong interactions, quantized motion and non-equilibrium boundary conditions imposed by the external electrodes. A fascinating aspect is that in these solid state devices *single electrons* can experimentally be manipulated to a high degree despite the coupling to huge numbers of other electrons and other degrees of freedom around (phonons, nuclear spins, etc). It is the dominating role of the Coulomb interaction in these confined systems which allows for this degree of control: once a single electron has jumped to some new position in such a device, other electrons are prevented from following it since the change in Coulomb interacting energy is larger than thermal and quantum fluctuations. This suppression of transport is called *Coulomb blockade*. Coulomb interaction is also responsible for many other effects such as exchange interactions, spin blockade, the exchange-field induced by ferromagnetic electrodes, and also the Kondo effect. The degree of control, however, comes at a price: the understanding of the operation of such devices is complicated due to the inherent many-body nature of the transport processes.

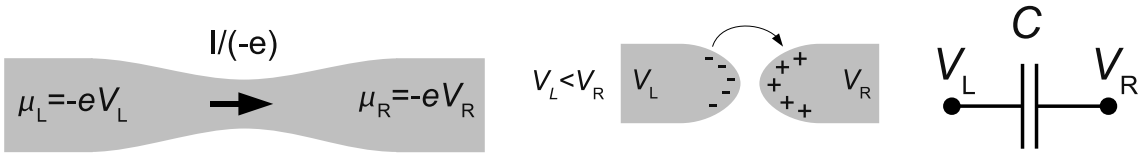
However, much progress has been made. Although single electron experiments date back to Millikan's determination of the electron charge [1] $q_e = -e$ where $e \approx 1.602\ 10^{-19}$C, only much later solid-state experiments involving single electron effects were done. Important developments came from transport through metallic layers which break up into small islands due to surface tension effects [2] and the progress in the field of superconducting tunnel-junctions. This stimulated the development of basic theoretical models and approaches still in use today [3]. After the advent of the first *tunable* single electron devices [4] the field of quantum dot physics boomed, see the key reviews [5, 6, 7, 8, 9, 10, 11]. In the last two decades quantum dots realized in nano-structured *solid-state* devices have reached a high level of sophistication, ranging for "on-chip" fundamental physics experiments and hybrid devices to device applications. Nowadays nano-gap fabrication combined with chemical functionalization even allows quantum dot devices to be fabricated in single *real* molecules. Thus, in a way the field of quantum dots has now come full circle: the techniques which have developed for controlling "artificial" atoms and molecules can now be applied to the real thing: the complicated problem of transport through single molecules greatly benefits from this wealth of experience and techniques, both in experiment and theory. Clearly, quantum dot systems are encountered in many areas of physics, electronics, chemistry and biology and will continue to be crucial in the full range from basic physics to nano-scale device applications.

The importance of interactions in the transport processes of charge and spin makes understanding and operating quantum dot devices a complicated task. It is the aim of this lecture to introduce some of the basic notions of single quantum dot devices underlying current research efforts. Basics of Coulomb blockade and orbital quantization effects will be introduced in Sections 2-3 based on the above cited key reviews where more details and references can be found. In the last decade spin effects have been at the center of attention, including Hund's-rule / shell-filling effects, Zeeman splitting, exchange interaction in double dots, spin-polarized electrodes. In the remaining Section 4) a specific quantum dot devices will be discussed where the elec-

tron spin plays a crucial role in the transport. Specifics of different realizations of quantum dots systems are avoided if possible to focus on common principles and models. Other interesting avenues where spin effects are important but which are not considered here are Pauli spin-blockade in double quantum dots [11], higher-order tunneling effects ("cotunneling", level renormalization, broadening) [12, 13], Kondo effect [14, 15], time-dependent transport [7, 16, 17], metallic dots (without orbital quantization effects) [8], many-electron quantum dots and their statistical properties [7, 18], superconducting quantum dots and/or electrodes [19] and device application aspects of single electron circuits [8]. Specifics on transport through quantum dots realized in semiconductor hetero-structures [9], wires [20], carbon nano-tubes [21, 22] and nanometer-size molecular quantum dots [23] can be found in the indicated references (reviews and some key publications).

2 Single electron charging effects on tunneling

To understand how Coulomb interactions allow single electrons to be controlled, we consider the simplest structure, a single tunnel junction [5, 7, 8, 9] between two electrodes at potentials V_r, where the left and right electrode are denoted by $r = L, R$. We first fix some conventions. The electrochemical potentials of the electrodes is $\mu_r = \epsilon_F - eV_r$, where $e > 0$ denotes the elementary charge *magnitude* (i.e. the electron charge is $= -e$). Therefore a chemical potential or energy bias $\mu_L - \mu_R = -e(V_L - V_R) > 0$ will drive an electron *particle*-current in the direction $L \to R$ i.e. L acts as the source and R as the drain electrode. This requires a negative bias voltage $V = V_L - V_R < 0$ to be applied, and the *electric* current by definition flows in the opposite in the opposite direction, $I < 0$. We will always discuss the electron particle $I/(-e)$ current versus the applied energy bias $-eV$ (or drain-source voltage) to avoid the inconvenient minus signs, as is common practice. [1]



(a) Quantum point contact (b) Tunnel junction (c) Circuit

One can think of a tunnel junction as a pinched-off ballistic conduction channel. For such a system the two-terminal conductance is due to contact resistance and is quantized in units of the fundamental quantum of conductance $G_Q = e^2/h$ [24]. If the conduction channel is gradually pinched off, the conductance is reduced in steps of magnitude $2G_Q$ (factor 2 for spin) until finally the current drops exponentially to zero and

$$dI/dV \ll e^2/h. \tag{1}$$

This is referred to as *charge quantization*, which is a semi-classical effect. The electrons are classically confined to either electrode (i.e. with well-defined charge numbers). To see some effect of this in transport one needs to have quantum mechanical tunneling through the insulating barrier: the junction has to be considered as a "leaky" capacitor. When an electron

[1] Note that in the differential conductance, this sign cancels out when converted to energy units and particle currents: $dI/dV = ed(I/(-e))/d(\mu_L - \mu_R)$.

tunnels across the junction the polarization charge changes abruptly by the elementary quantum of charge e. As a result potential across the junctions jumps up or down by $\Delta V = e/C$. The associated change in the energy of the junction, $e\Delta V = e^2/C$, is called the *charging energy*. If the applied bias voltage energy eVor the thermal fluctuations kT are not able to overcome this energy barrier, the transport is blocked. Thus the tunneling of a *single electron* can have a dominating effect on the macroscopically measurable transport current. Since the capacitance C scales linearly with the junction dimensions, *single electron tunneling* (SET) effects become important for sufficiently small junctions and / or sufficiently small temperatures:

$$kT \leq e^2/C. \tag{2}$$

Generally, the suppression of tunneling due to the charging energy is termed *Coulomb blockade*.[2] A tunnel junction is very different from an ordinary resistor where an incompressible electron liquid flows in a quasi-continuous way, responding instantaneously to electric fields. In contrast, the leakage of junction charge is a slow process i.e. only once in a while an electron tunnels, the duration of this quantum mechanical process being much shorter than the average time between such events [6]. The time for the charge in the junction to rearrange itself after a tunnel process is also very short (dielectric relaxation time related to plasmon dynamics [25, 6]). Surprisingly, despite the clear arguments above single electron charging effects are *not* very important in single tunnel junctions [26, 27] and were first observed in double tunnel junctions where a quantum dot is formed to be discussed in Sect. 3. The point is that above we have tacitly assumed that the polarization charge on the junction is a classically well-defined variable. However, this depends strongly on the nature of the electromagnetic environment created by the external circuit, which is can be described by a frequency dependent impedance $Z(\omega)$ in series with the junction. Only for a high impedance environment $Z(0) \gg 1/G_Q$ the junction charge is classically well-defined. However, this is almost never the case unless exceptional precautions are taken [28]. Therefore, quantum fluctuations of the junction charge and voltage are strong and can nearly completely lift the Coulomb blockade of transport: during such a fluctuation charge moves into the circuit leads and the Coulomb barrier is lowered, allowing the electron to tunnel. One may say that an electron tunnels directly into the leads, rather than staying localized onto the capacitor plate. See for a clear exposition and examples [25]. A precise understanding of the electromagnetic environment requires a full quantum description of the circuit to calculate the transport, see the reviews [26, 29, 27] and key papers [30, 31, 32] for further details.

3 Single quantum dots

Clearly, when introducing a second tunnel junction, thus creating an island or *dot* between two electrodes, the above problem with the low impedance of the external circuit is solved: when viewed from one junction, the other junction shunts the low environmental impedance [26]. Therefore charging effects are always important for two or more tunnel junctions in series under the conditions of charge quantization (1) and of low temperature (2), even if the environmental

[2] Another way of seeing how the quantum conductance enters into the condition for observing charging effects at low temperature is based on the time-energy uncertainty relation, which states that in the classical limit $\Delta E \Delta t \gg \hbar$. The energy uncertainty is set by the charging $\Delta E \sim e^2/C$ and the time for discharging the capacitor is governed by the smallest RC-time $\Delta t \sim RC$ where R is the largest resistance dominating the current. Plugging this in the capacitance cancels out and one obtains the condition for classical charging effects to be $R \gg h/e^2 = 25.8k\Omega$ i.e. Eq. 1.

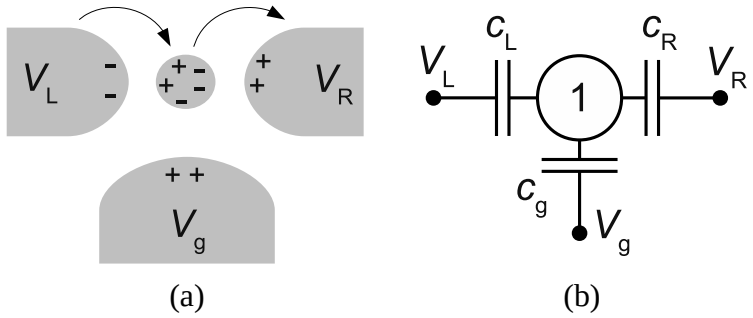


Fig. 1: *(a) Quantum dot connected to electrodes and a gate terminal. (b) Double junction electrostatic circuit. Tunneling through the "leaky capacitors" is not indicated.*

impedance is low compared to $1/G_Q$. Such systems are considered from hereon and the issue of the environment can be ignored.

A *quantum dot* denotes a confined electron system where both charge *and* orbital quantization effects are important, see the key reviews [7, 8, 9, 11]. We first discuss charge quantization by extending the above capacitance model to a double junction. The success of the capacitance model derives from the smallness of screening length compared to the electron wavelength [7].[3] The small capacitances needed to see charging effects are limited from below by *self-capacitance* which for a sphere equals $4\pi\epsilon L$ and for a disk $8\epsilon L$ (height $\ll L$), i.e. roughly $10\epsilon L$ where L denotes the radius. The condition $e^2/C = kT$ for room-temperature $T = 300$ K is satisfied for $C \sim 2$ aF, whereas for low temperature $T = 1$ K this becomes $C \sim 6$ pF. In vacuum, this translates into typical dimensions of $L \sim e^2/(kT10\epsilon_0) = 70$ nm and $L \sim 20$ μm for $T = 300$ K and $T = 1$ K, respectively. The effective capacitance is actually larger than this estimate due to the fact that the tunnel-barrier consists of materials with $\epsilon > \epsilon_0$, the *mutual capacitances* to the nearby conducting electrodes [34, 36] and the *quantum-capacitance* due to non-locality of charge [37]. Therefore one needs smaller dimensions that the above rough estimates, see [7].

Now consider a dot (labeled by 1) coupled by two tunnel junctions to electrodes $r = L, R$, as sketched in Fig. 1. A third, capacitively coupled, gate electrode g, allows the potential on the dot to be varied in a *continuous* and linear way by changing the voltage V_g[4]. This originates from a continuous shift of the incompressible conduction electron liquid relative to the ionic cores in the material [25]. The dot is thus considered as a conducting body with a small capacitance: when *discrete* charges are moved from the electrodes to the dot the potential of the dot V_1 strongly changes and may block the transfer if the corresponding energy change is larger than kT. The calculation of the energy of a set of conductors is a textbook [38, 39] problem. However, the issue is that we need the energy as function of the electrode potentials and dot charge. We include these voltage sources in the network of conducting elements[5] as

[3] In semiconductors this condition is less well satisfied than in metals and there is a weak voltage dependence of the capacitances [33, 34]. In molecular quantum dots this dependence is even stronger [35, 36]. In such cases one can often work with differential capacitances i.e. linearized with respect to voltage.

[4] The gate electrode can also model constant stable background charges, which are always present and give rise to a constant gate voltage offset. Experimentally, one can simply compensate for offset charges by changing V_g, so we leave them out here. However, for applications background charges present a major problem. We note that instable background charges less trivial effect and can give rise to a dynamic "resetting" of the gate voltage scale. These have to be incorporated into the transport model, resulting in a double dot model. This is especially important in nanometer sized molecular quantum dots.

[5] There are in fact two equivalent ways of approaching this [40], both used in the literature. The approach taken

conductors with large capacitance to ground and large charge i.e. their potential is insensitive to their charge. The opposite is true for the dot: its potential is highly sensitive to changes in the charge Q_1 due to the small total capacitance $C_1 = \sum_{r=L,R} c_r + c_g$. This is seen by expressing the charge in the capacitances in Fig. 1 and the voltage differences across the 3 different junctions, $Q_1 = \sum_{r=L,R} c_r(V_1 - V_r) + c_g(V_1 - V_g)$. Solving for the dot potential V_1 we find:

$$V_1 = (C_1)^{-1} \left(Q_1 + \sum_r c_r V_r + c_g V_g \right). \tag{3}$$

To find out if tunneling is possible, we need to calculate the change in the total energy of the whole system due to a single electron tunneling event. Introducing the electron number $N_1 = Q_1/(-e)$ and inserting it into the electrostatic energy $E_{N_1} = (V_1)^2/2C_1$ we obtain

$$E_{N_1} = U_1 \left(N_1 - \left(\sum_r c_r V_r + c_g V_g \right) /e \right)^2 \tag{4}$$

where U_1 is the *charging energy*:

$$U_1 = \frac{e^2}{2C_1}. \tag{5}$$

As function of the externally induced continuous charge $\sum_r c_r V_r + c_g V_g$ each charge state defines a potential energy parabola with a minimum at $eN_1 = \ldots, -e, 0, e, \ldots$. Clearly, at zero applied voltages $V_{L,R,g} = 0$ the state with $N_1 = 0$ electrons has the lowest energy. The *energy change* associated with adding a single electron, the *electrochemical potential* of the dot, takes on discrete values:[6]

$$\mu_{N_1} = E(N_1) - E(N_1 - 1) = U_1 \left(N_1 + \frac{1}{2} \right) - \sum_{i=L,R,g} \alpha_i e V_i \tag{6}$$

Here we introduced the capacitance ratios $\alpha_i = c_i/C_1 < 1$ for $i = L, R, g$. We now use the conservation of total energy of the dot and the reservoirs for a single electron tunnel process. For *weak tunnel coupling* this is sufficient, see below Sect. 3.4.[7]. While in the electrodes with a continuum of electron energies is are available, only those electrons with energy equal to one of the discrete values μ_N can enter the dot.

3.1 Linear transport

The charge quantization becomes evident for small applied voltages and low temperature. As one varies the gate electrode to more positive values, one attracts electrons to the dot by the

here is simpler to extend to multiple dot systems [10] and to quantum-mechanical systems, see [41] (Appendix). Alternatively, one can consider the dot as an open system, treating the voltage sources as external. If one then considers a tunnel event, one has to calculate the change in free energy plus the *work done by the voltage sources*. In this picture, tunneling only occurs when the free energy decreases ("downhill" process), the energy difference being dissipated in the external sources [40].

[6] One should be aware that often these values of the electrochemical potential are referred to as "levels" of the dot, even though they are energy changes (related to electron affinity, ionization potential). Energy level diagrams where electrons are drawn in as occupying such a level often lead to confusion at first.

[7] For strong tunnel coupling this breaks down, since higher order processes shift the dot energy levels.

positive *continuous* induced charge. As a result, all μ_N shift down uniformly and one achieves subsequent resonances $\mu_N = \mu_r$ for $N = 1, 2, \ldots$. At such a *charge-degeneracy point* an electron can enter / leave the dot without changing its energy. The small bias $\mu_L - \mu_R \ll kT$ then results in a directed transport current and a finite linear conductance $(dI/dV)_{V=0}$: for $\mu_L > \mu_N > \mu_R$ an electron can be extracted from the left, inserted into the dot and finally extracted on the right, while conserving the total energy, see Fig. 2. Clearly, the probability for the opposite process is smaller. Plotting the conductance as function of V_g one obtains what are called Coulomb blockade oscillations, see Fig. 2.

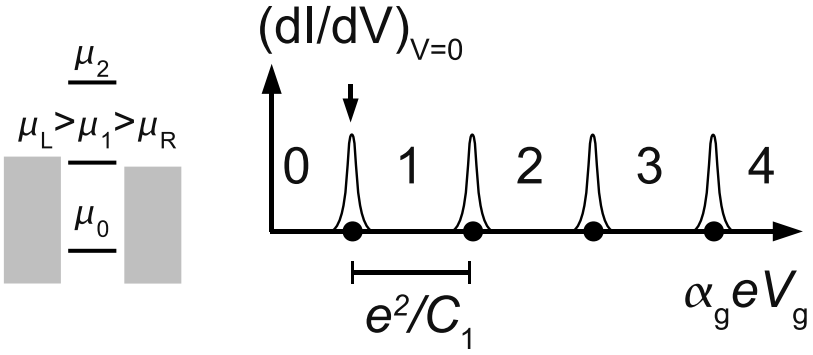


Fig. 2: *Coulomb blockade oscillations of the linear conductance: at the dots on the gate energy axis the resonance condition $\mu_N = \mu_L \approx \mu_R$ is satisfied for some $N = 0, 1, 2, \ldots$. The left inset shows the electrochemical potentials at the degeneracy point marked by the arrow. The energy width of the peak is set by the electron temperature of the electrodes, kT, which smears the sharp Fermi levels. Between the conductance peaks the charge on the dot is fixed to the indicated value due to Coulomb blockade.*

The distance between subsequent degeneracy points $\mu_{N_1} - \mu_{N_1-1}$ is called the N_1th *addition energy*:[8]

$$\mu_{N_1} - \mu_{N_1-1} = U_1 = \frac{e^2}{C_1}. \tag{7}$$

This is the *extra* energy required to add the N_1th electron, given that one already has the energy available to add the $N_1 - 1$th electron. The addition energy is thus an electrochemical potential *difference* i.e. a second difference of the energy E_N with respect to N (curvature of energy versus N). In the electrostatic model considered here, the addition energy is the same for all N and equal to the charging energy U_1. The distance between the degeneracy points allows the addition energy to be read off directly from experimental data, giving the capacitance C_1.[9] To determine α_g independently, one needs to measure the non-linear transport as well.

3.2 Non-linear transport

For finite bias $-eV \gg kT$ each degeneracy point from the linear transport regime widens into a finite region where single electron transport (SET) is possible. The resulting *stability diagram* is

[8] There is no chemical contribution here i.e. the Fermi-levels of the metallic islands are assumed to be identical. Such differences would only cause an offset in the gate voltage, similar to background charged-impurities.

[9] The precise position of the degeneracy points μ_{N_1} are not very useful, since they depend on the Fermi-level of the electrodes, the number of electrons N_1 (not directly accessible from the transport measurement) and uncontrollable offsets due to charged impurities close to the device.

sketched in Fig. 3, the light regions indicating SET, the dark regions Coulomb blockade where the current is suppressed and the charge state on the dot is fixed despite the finite temperature and voltage. By using bias and gate *energies* on the axis, the meaning of the boundary lines at the SET region are easily remembered: the left (right) diagonal line through the Nth degeneracy point corresponds to resonance with the left (right) electrode, $\mu_L = \mu_N$ ($\mu_R = \mu_N$). These basic rules of thumb are illustrated by the left two insets in Fig. 3. Clearly, the boundary lines must cross at the charge degeneracy point where $\mu_L = \mu_N = \mu_R$. The slopes of the

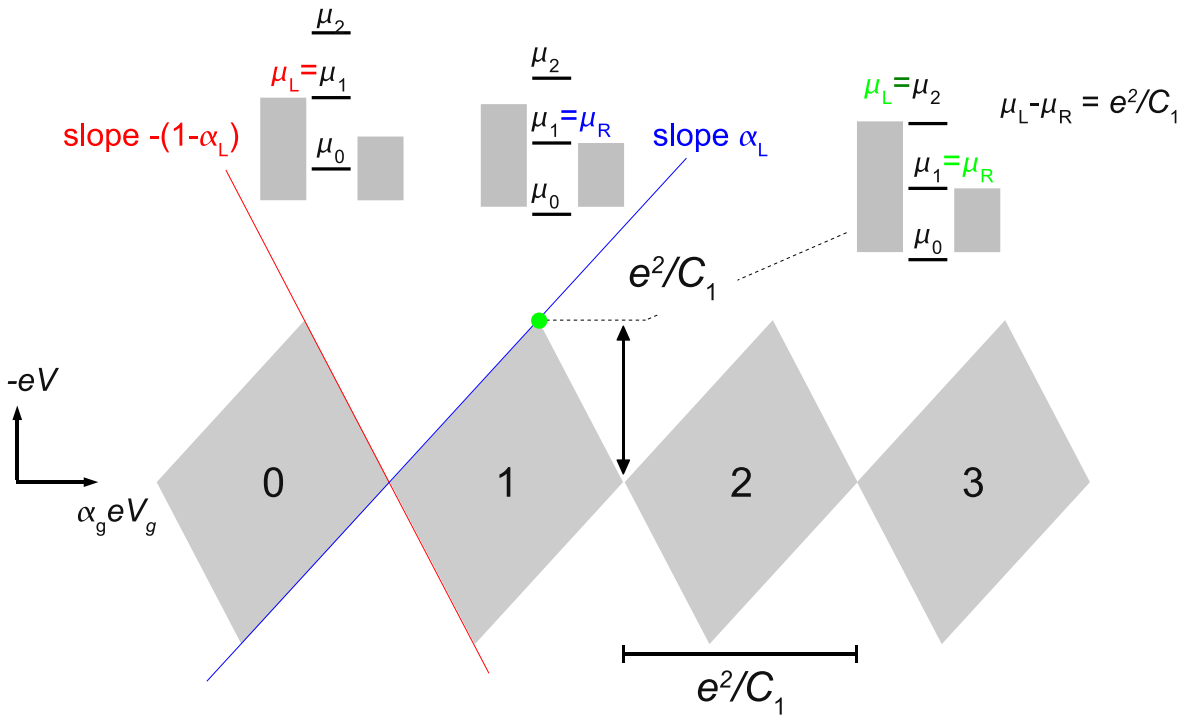


Fig. 3: *Stability diagram in $(\alpha_g eV_g, -eV)$ plane. The vertical and horizontal size of the diamond shaped Coulomb blockade regions is in general a measure of the stability of the charge state (indicated by the number), given here by the charging energy $U_1 = e^2/C_1$. The boundary lines correspond to the electrochemical potential diagrams indicated in the left two insets. Their slopes are given for the experimentally most relevant case of asymmetric bias $V_L = V, V_R = 0$: they follow by equating $\mu_L = -eV = \mu_N$ and $\mu_R = 0 = \mu_N$ and using Eq. (6). The right inset makes clear that when subsequent electrochemical potentials of the dot are simultaneously at resonance, that the energy bias $\mu_L - \mu_R = -eV$ equals the addition energy, which equals the charging energy.*

resonance lines in the stability diagram actually depend on how the bias $-(V_L - V_R)$ is applied. In experiments one usually sets $V_R = 0$ for practical reasons, which has the added advantage that the charging energy can be read off directly as the height of the diamond shaped Coulomb blockade region, see the upper right inset to Fig. 3. In theoretical works one favors a symmetric bias $-V_L = V_R = V/2$ since then for symmetric coupling $C_L = C_R$ (a common simplifying assumption) the electrochemical potentials μ_{N_1} are independent of the bias voltage and are only shifted by V_g, see Eq. (6).

More generally, a stability diagram indicates the boundaries between regimes where the current has a different value. At the boundary dI/dV is nonzero since the current changes drastically

due to a resonance condition being satisfied (energy conservation). Therefore color scale plots of dI/dV as function of the applied gate and bias voltage are the standard starting point for transport spectroscopic studies and provide a compact format for summarizing many theoretical results in a single figure. From such a plot, an important characteristic transport signature of quantum dots without quantized orbital states (usually metallic particles) is immediately evident: the Coulomb diamonds have identical shape and size, see (7). See the review [42] for experimental examples and theoretical modeling.

3.3 Discrete states and excitations

If one makes a quantum dot sufficiently small, *size quantization* effects will become important due to the wave nature of electrons.[10] For such systems the above electrostatic treatment accounting only for charge quantization is clearly insufficient. However, it turns out that by a simple extension one obtains a very successful model. One first neglects Coulomb electron-electron interactions altogether and calculates the single-particle energies ("particle in a box problem"). The total energy is now approximated by the sum of the electrostatic contribution (4) and the single-particle energies. This is the so-called *constant-interaction model*. There is now both a *chemical* (orbital) and electrostatic contribution to the energy change upon charge addition, starting in the $N_1 - 1$ and ending in N_1 electron *ground* state:

$$\mu_{N_1} = \epsilon_{N_1,\text{free}} + E(N_1) - E(N_1 - 1) = \epsilon_{N_1,\text{free}} + U_1 \left(N_1 + \frac{1}{2} \right) - \sum_{i=L,R,g} \alpha_i e V_i. \qquad (8)$$

Here $\epsilon_{N_1,\text{free}}$ denotes the energy of the lowest free orbital to which the N_1th electron is added. Clearly, this term depends on the electron number N_1 and on the details of the spectrum. This implies that in general, the addition energy differs from the charging energy. For example, if the energy spectrum is non-degenerate, starting from a dot where the highest orbital level is half-filled, one only needs to pay the charging energy *extra* to add the next electron to the same level, i.e. the addition energy equals the charging energy, $\mu_{N_1} - \mu_{N_1-1} = U_1$, for N_1 =even. In contrast, when all levels are initially doubly occupied, the N_1th electron is added to the next empty orbital when N_1 =odd. In this case the addition energy is $U + \delta$, where δ is the separation of the highest filled level to the empty level. If the level spacing and the charging energy are approximately constant for several subsequent electron numbers (which is often the case) , then the stability diagram shows an even-odd alternation of the sizes of the Coulomb diamonds, as sketched in Fig. 4. This is common fingerprint of size quantization in quantum dots which can be directly read off from the stability diagram.[11] In contrast, small quantum dots without special symmetries have no orbital shell structure i.e. irregular orbital energy spacings. Irregular addition spectra occurring for low electron numbers are thus another fingerprint of quantization effects.

[10] For a single electron a potential-box of size L the kinetic energy scales as L^{-2} with system size, and will for some value of L equal the charging energy which scales $\propto L^{-1}$.

[11] If the spectrum contains degeneracies, the filling of the degenerate shell will give rise to several addition energies equal to the charging energy U, followed by one addition energy $U + \delta$ when the next shell starts to be filled. For example, carbon-nanotube quantum dots display clear shell-filling over many tens of electron numbers [22]. It is striking that the degeneracy, a fingerprint of the microscopic molecular structure (hexagonal carbon patches), shows up in a macroscopic transport current! In semi-conductor quantum dots with well-defined circular disk shape the shell filling as measured in transport follows accurately a 2-dimensional analogue of the periodic table, see [9] for details.

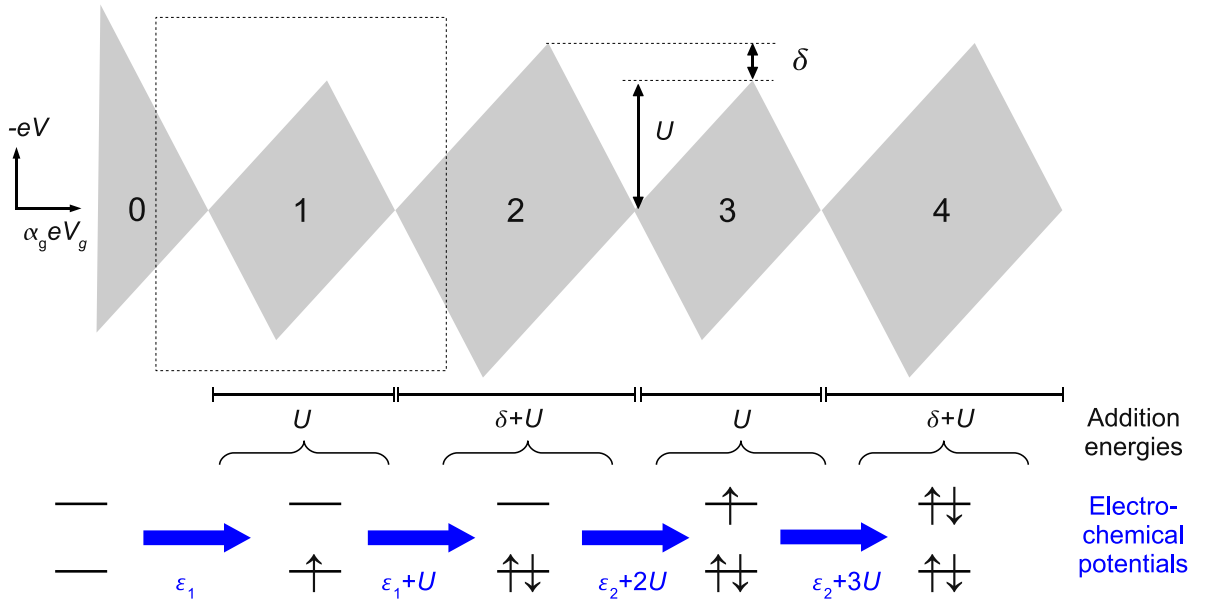


Fig. 4: *Stability diagram in $(\alpha_g eV_g, -eV)$ plane for a quantum dot with multiple quantized orbital states. Bottom: filling of the orbital levels with energies $\varepsilon_1, \varepsilon_2$ in the different charge states. The orbital energy splitting is denoted by $\delta = \varepsilon_2 - \varepsilon_1$. Blue arrows: electrochemical potential i.e.* energy-change *for transitions between these states. The addition energies are obtained by taking the electrochemical potential differences. The Anderson model Hamiltonian gives a good description in the boxed region on the left.*

In a limited gate and bias voltage range the essential transport characteristics are captured by restricting the model to a single level with a charging energy U coupled to two electrodes. The corresponding model Hamiltonian defines the *Anderson model*,

$$H_D = \varepsilon \sum_{\sigma=\uparrow,\downarrow} n_\sigma + U n_\uparrow n_\downarrow \tag{9}$$

where $n_\sigma = d_\sigma^\dagger d_\sigma$ counts the occupation of the orbital with spin projection $\sigma = \uparrow, \downarrow$. This model and its extensions form the basis for most transport calculations on quantum dots. The energy ε here incorporates the bias and gate voltage induced shifts of the electrochemical potential and other contributions independent of the charge number, see Eq. (10).

Electrons may of course be added to other levels than the lowest unoccupied one, resulting in an excited state of the dot. This is possible if the generalized electrochemical potential

$$\mu_{N_1}^{b,a} = E_b - E_a \tag{10}$$

lies in the bias energy window, where b and a denote any pair of ground or excited states with N_1 and $N_1 - 1$ electrons respectively. This plays however no role in the Coulomb blockade regime where no other state than the Coulomb blockaded ground state is occupied. Figure 5 indicates for a simple model, which new resonance lines appear in the SET regimes and which processes they correspond to.

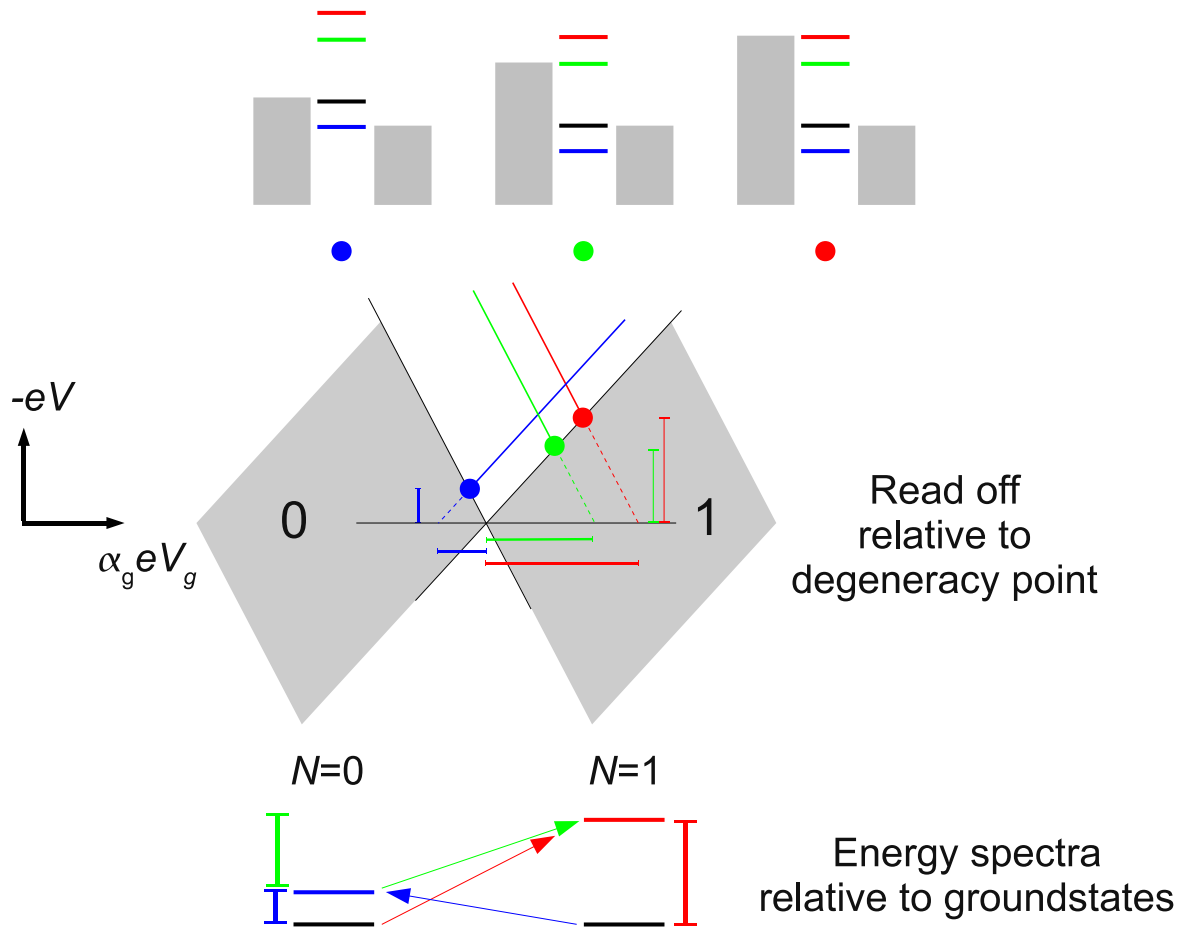


Fig. 5: *Center: Stability diagram in* $(\alpha_g eV_g, -eV)$ *plane for a quantum dot with excited many-electron states. Bottom: energy spectrum for* $N = 0, 1$ *electrons and transition induced by electron addition. Along the full colored lines in the SET region the current changes when the electrochemical potential corresponding to these transitions enters the bias window. The excitation energies, relative to the ground state, can be read off in two ways: (i) either vertically by the position of the colored dots marking the intersection with the SET regime boundary or (ii) horizontally by the position of the lines extrapolated to zero bias voltage (the dashed parts in the* $N = 0$ *or* $N = 1$ *Coulomb blockade regime do not show up in transport). Top: electrochemical potential diagrams for the corresponding colored dots at the SET boundary for method (i). The black "level" marks the electrochemical potential for the ground* $N = 0$ *to ground* $N = 1$*-electron transition (all such transitions for other electron numbers are not shown). The colored lines show the electrochemical potentials for transition involving an* excited state, *which are obtained by shifting the ground-ground level up (red) or down (blue) by each of the excitation energies relative to the ground state, or their difference (green). Note that the green level marks the transition between two excited states, which can only be observed in transport if relaxation mechanisms are weak [43], see [11] for an example of such transitions involving spin.*

3.4 Transport theory - occupations and rates

The aim of this section is twofold. On the one hand, we discuss some of the basic arguments we used above to construct stability diagrams. On the other hand, we indicate the microscopic origin of this picture and the general framework in which quantum transport through quantum dots can be calculated.

The crucial starting point is that the Coulomb interaction effects need to be included non-perturbatively: it cannot be used as an expansion parameter since it is one of the largest energy scales. We thus should treat the tunneling as a small parameter. From the previous sections it is clear that in general transport through a quantum dot is described by a model Hamiltonian containing three terms, $H = H_D + \sum_{r=L,R} H_r + H_T$, which account for the quantum dot, the electrodes and the tunneling between them, respectively. The strong classical electrostatic effects due to capacitive interaction with the electrodes and quantized orbitals we have been discussing so far enter into H_D, as well as further many-electron interactions on the dot (e.g. spin-spin, spin-orbit interactions, etc.), see [41] for a systematic discussion. The metallic electrodes with good screening properties are described by a single particle models H_r, which are fully specified by the density of states ρ_r (which can be considered as constant in nearly all cases of interest). The electrodes remain in equilibrium at their own electrochemical potentials $\mu_r = \epsilon_F - eV_r$ and temperature T. In order to calculate the quantum transport current which can be compared with experimental results, one needs to additionally specify the matrix elements of the microscopic coupling H_T accounting for electron tunneling between the electrodes and the dot.

The calculation of the non-linear current from this well-defined model is in general a non-trivial problem which is still under intense study. Importantly, since the electrodes impose conflicting boundary conditions (different electrochemical potentials) on the microscopic quantum system the occupations of the states are *not* given by a grand-canonical distribution. The density matrix of the dot P is thus in general *not known* and must be calculated from a kinetic equation, which in general takes the form [44]:

$$\frac{dP}{dt}(t) \;=\; -i[H_D, P(t)] + \int_{-\infty}^{t} dt' W(t-t') P(t') \tag{11}$$

The first term on the right hand side derives from the Liouville equation for the density operator P for an isolated quantum dot, $\dot{P}(t) = -i[H_D, P(t)]$ (equivalent to the Schrödinger equation). The second term describes the effect of the electrodes on the quantum dot and includes non-trivial retardation effects through the time integral extending into the past time-evolution. In the long-time limit, $t \to \infty$, where the dot has reached a time-independent, non-equilibrium stationary state $P(t) \to P$ this reduces to

$$\frac{dP}{dt}(\infty) = 0 = -i[H_D, P] + WP \tag{12}$$

where $W = \lim_{\eta\to 0} \int_0^\infty e^{i\eta\tau} W(\tau) d\tau$ includes generalized *transport rate*. This is the zero-frequency component of the memory kernel appearing in the "collision term" on the far right in the kinetic equation (11). From this equation one can find the density matrix P which we need to calculate the expectation value of the transport current.

For comparison with experiments the limit of weak tunnel coupling and non-degenerate quantum dot states is an important starting point. In this limit only the occupations P_a of the many-body quantum dot states labeled by a (diagonal elements of the density matrix P) matter, which

obey a stationary *master equation*:

$$\frac{dP_a}{dt}(\infty) = 0 = -\left(\sum_b W_{ba}\right) P_a + \sum_b W_{ab} P_b \tag{13}$$

Here we sum over quantum dot states b (electron number N_b) with one electron more or less than state a (electron number N_a). In this limit the rates simplify to those obtained by Fermi's golden rule:

$$W_{ba} \;=\; \sum_r \Gamma_r^{ba} \begin{cases} f_r^+(E_b - E_a) & N_b = N_a + 1 \\ f_r^-(E_a - E_b) & N_a = N_b + 1 \end{cases} \tag{14}$$

Here $f_r^{\pm}(x) = 1/(e^{\pm(x-\mu_r)/T} + 1)$ is the Fermi-distribution function for an electron tunneling out / into the electrode r and $f^+ = 1 - f^-$. The rate constant Γ_r^{ba} for electrode $r = L, R$ depends on the density of states: the quantity

$$\hbar\Gamma_r^{ba} = 2\pi \left(T_r^{ba}\right)^2 \rho_r \tag{15}$$

defines characteristic energy scales for quantum charge fluctuations due to tunneling. The matrix elements T_r^{ba} for an electron tunneling onto the dot from electrode r, thereby inducing a transition from state a to b on the dot, completely specify the tunneling part of the model, H_T. The expectation value for the current $I_r(t) = (-e)\langle -\dot{N}_r(t)\rangle$ flowing into the dot from electrode r (electron number N_r) is obtained from the probabilities $P(t)$ through

$$I_r(t) = \mathrm{Tr} \int_{-\infty}^{t} dt' W_{I_r}(t - t') P(t'). \tag{16}$$

Here Tr denotes the trace over the quantum dot eigenstates. In the stationary limit the current follows in a similar fashion from *current rates* W_{I_r}:

$$I_r = \mathrm{Tr} W_{I_r} P. \tag{17}$$

The key insight which Eq. (13) brings is that the stationary occupations follow from a balancing of probability flows. The loss of probability due to transitions from a to all states b (first term) must cancel the gain of probability of state a due to the decay of the latter states (second term). Whenever a positive energy change $E_b - E_a$ falls below the electrochemical potential μ_r due to a change in the bias voltage, the transition rate Eq. (14) increases by an amount Γ_r^{ba} since the tunneling onto the dot becomes energetically allowed. This implies that the probabilities and also the current change: we obtain a non-zero dI/dV and a resonance line in the stability plot. The width of this resonance is set by the thermal energy kT due to the Fermi-function in Eq. (14).[12] One can thus sketch the stability diagrams by considering the available quantum dot states for each electron number and the directions of the processes with non-zero transition rates based on the positions of the chemical potentials (10) relative to the energy bias window

[12] Since the change in the occupations cannot exceed one, the change in the current is on the order of the typical values of $e\Gamma_r^{ba}$ which we loosely denote by $e\Gamma$. The differential conductance is now estimated as $dI/dV = \Delta I/\Delta(\mu_L - \mu_R) \sim e^2\Gamma/(kT)$. From this one easily sees that the charge quantization condition (1), $dI/dV \ll e^2/h$ requires that the life-time broadening due to quantum fluctuations (15) stays much smaller than that due to thermal fluctuations: $\hbar\Gamma \ll kT$. This small ratio is the formal expansion parameter used in perturbation theory, although in special cases higher order effects may complicate matters e.g. due to the Kondo effect.

$\mu_L > \mu_R$. See for instance Fig. 5. Depending on the rate constants one can even in special cases find the qualitative dependence of the occupations and the current on the voltage.
As an explicit example, for the Anderson model the master equation for the probabilities for $N = 0, 1, 2$ electrons reads

$$\frac{d}{dt}\begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix} = \sum_r \Gamma_r \begin{pmatrix} -2f_r^+(\epsilon)P_0 + f_r^-(\epsilon)P_1 \\ 2f_r^+(\epsilon)P_0 - \left(f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)P_1 + 2f_r^-(\epsilon+U)P_2 \\ +f_r^+(\epsilon+U)P_1 - 2f_r^-(\epsilon+U)P_2 \end{pmatrix}. \quad (18)$$

Here $P_1 = \sum_\sigma P_\sigma$ is the occupation of the 1-electron state, irrespective of its spin value $\sigma = \uparrow, \downarrow$ and the probability is conserved, i.e. $P_0 + P_1 + P_2 = 1$. Plugging the solution of these equations into the expression

$$I_r = \sum_r (-e)\Gamma_r \left[2f_r^+(\epsilon)P_0 + \left(-f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)P_1 - 2f_r^-(\epsilon+U)P_2\right] \quad (19)$$

one obtains the current. This produces, for example, the curve labeled $p = 0$ in Fig. 8(a) in the next section.
Finally, we mention that corrections to the above single electron transport picture can be incorporated by calculating non-diagonal density-matrix elements (related to quantum superpositions of quantum dot states), which turn out to be crucial for the single electron spin-valve transistor in Sect. 4. Similarly, higher order tunnel processes can be included, such as for instance "cotunneling" processes involving pairs of electrons using perturbation theory [13] and the Kondo effect [14] using renormalization group methods [15]. Such effects play an important and sometimes even a dominant role in experiments and are key spectroscopic tools for the characterization and control of quantum dots, see Sect. 4.3. However, these issues are beyond the scope of this lecture.

4 Single electron spin-valve

Effects of a spin-polarization of conduction electrons in a tunnel junction have been of interest since the seminal papers of Julliere [45] and Slonczewski [46]. Consider two *ferromagnetic* electrodes connected by a tunnel junction as sketched in Fig. 6(a). The ferromagnets are characterized by a spin-dependent density of states $\rho_{r\sigma}, \sigma = \uparrow, \downarrow$ and a normalized polarization vector $\mathbf{n}_r$ along which the majority spin $\uparrow$ is pointing in electrode r. The *magnetic spin-valve effect* [46] refers to the reduction of the linear tunneling conductance when the polarization vectors $\mathbf{n}_r$ of the two ferromagnetic electrodes $r = L, R$ are at a angle θ defined by $\cos\theta = \mathbf{n}_L \cdot \mathbf{n}_R$:

$$\left.\frac{dI}{dV}\right|_{V=0}(\theta) \propto (1 + p_L p_R \cos\theta) \quad (20)$$

Here the relative spin-polarization of the spin-dependent density of states of each electrode $\rho_{r\sigma}$ is introduced:

$$p_r = \frac{\rho_{r\uparrow} - \rho_{r\downarrow}}{\rho_{r\uparrow} + \rho_{r\downarrow}} \quad (21)$$

i.e. $\rho_{r\sigma} = \frac{1}{2}(1 \pm p_r)\rho_r$ for $\sigma = \uparrow, \downarrow$ where $\rho_r = \sum_{\sigma=\uparrow,\downarrow} \rho_{r\sigma}$ is the total density of states. The angle dependence in Eq. (20) predicted by Slonczewski has been demonstrated experimentally

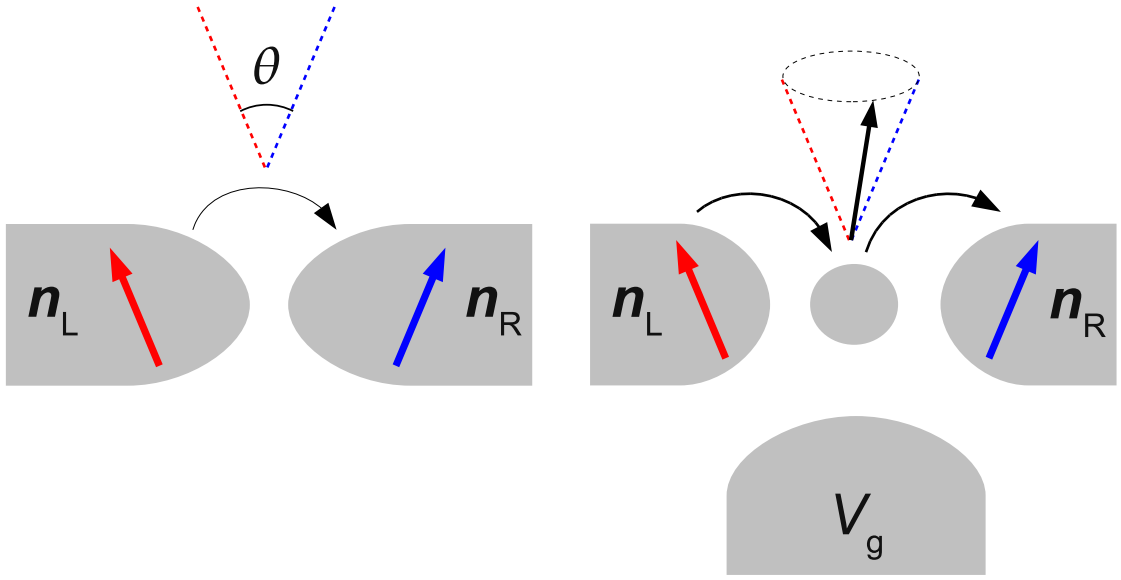


Fig. 6: *(a) Tunnel junction between magnetic electrodes. (b) Quantum dot connected to magnetic electrodes and a normal gate electrode. The electron spin can precess on the dot, even in the absence of an external magnetic field, thereby affecting the transport.*

in detail [47, 48]. It arises from the fact that an electron polarized along e.g. $\mathbf{n}_L$ has spinor component $\cos(\theta/2)$ ($\sin(\theta/2)$) when quantized along $\mathbf{n}_R$ and hence the probability depends on twice the angle i.e. on θ. For perfect ferromagnets, for which $p_r = 1$, the resistance is infinite (finite) for $\theta = \pi$ (0) i.e. we can switch between two resistance states by reversing the magnetization. This *tunnel-magneto resistance* (TMR) effect is based on the quantum mechanical spin degree of freedom and has been rapidly developed into a technologically relevant effect (e.g. MRAM).

A double tunnel junction offers new possibilities due to the role of Coulomb interaction and the tunable level in the quantum dot between the two ferromagnets. This *single-electron spin-valve transistor* is sketched in Fig. 6(b). The basic effects can again be understood from the Anderson model (9) i.e. a single level with Coulomb interaction U, simply generalized to account for a spin-dependent density of states in the electrodes which enter into the tunnel rates through Eq. (15).[13] The *charge* transport is now strongly influenced by the electron *spin* dynamics on the dot, which we now discuss.

4.1 Spin dynamics due to transport

We explain the basic physics of the spin dynamics in a heuristic fashion, arguing that the kinetic equation for the expectation value of the spin of the quantum dot $\mathbf{S}(t) = \langle \mathbf{S}(t) \rangle$ reads

$$\frac{d\mathbf{S}}{dt} = \left(\frac{d\mathbf{S}}{dt}\right)_{\text{acc}} + \left(\frac{d\mathbf{S}}{dt}\right)_{\text{rel}} + \left(\frac{d\mathbf{S}}{dt}\right)_{\text{pre}} \tag{22}$$

This equation can be derived rigorously from the general quantum kinetic equation (11) by accounting for the full density matrix, see the reviews [51, 52] and the original papers [53, 54].

[13] Additional spin effects related to the interface with the ferromagnets can be incorporated as well, see [49, 50].

Equation (22) is in fact a continuity equation for the spin, where the total change in time derives from the injected spin-currents (first 2 terms) and a 3rd precession term. We now discuss the three competing influences in turn.
First, due to transport on average a single electron with a spin $1/2$ will reside on the dot for a finite time. For non-magnetic electrodes the average local spin $\mathbf{S}$ vanishes due to rotational symmetry (contributions with opposite polarization cancel). If at least one electrode is polarized ($p_r \neq 0$) this symmetry is broken and a finite spin value results. This *spin-accumulation* derives from the spin-injection contribution

$$\left(\frac{d\mathbf{S}}{dt}\right)_{\text{acc}} = \sum_r \frac{\hbar}{2}(I_r/e)p_r\mathbf{n}_r \tag{23}$$

Only when both $\mathbf{n}_L = \mathbf{n}_R$ and $p_L = p_R$ does the accumulation term vanish. This follows from Eq. (23) using the current conservation $\sum_r I_r = 0$ in the stationary limit. An extreme example of spin-accumulation occurs for complete ($p_r = 1$) anti-parallel polarizations $\mathbf{n}_L = -\mathbf{n}_R$ where a single electron with spin parallel to the left injecting electrode gets "stuck" on the quantum dot, unable to exit to the right electrode at lower electrochemical potential with opposite polarization.
The transport however, also counteracts the spin accumulation due to the negative *spin-relaxation* term

$$\left(\frac{d\mathbf{S}}{dt}\right)_{\text{rel}} = -\sum_r \Gamma_r/\hbar \left(f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)\left(\mathbf{S} - p_r^2(\mathbf{S}\cdot\mathbf{n}_r)\mathbf{n}_r\right) \tag{24}$$

where $\Gamma_r\left(f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)$ is the Golden Rule rate of tunneling from the one electron state with spin to the zero- and two- electron state with zero spin (singlets). Notably, the relaxation is anisotropic: for a perfect ferromagnet ($p_r = 1$) and an average spin parallel to $\mathbf{n}_r$ one of the relaxation terms vanishes, whereas for opposite average spin this term is maximal.
The third term in Eq. (22) describes a *spin-precession* of the electron spin about a magnetic field as described by the standard Bloch equation[14]

$$\left(\frac{d\mathbf{S}}{dt}\right)_{\text{pre}} = \gamma\mathbf{S}\times\sum_r \mathbf{B}_r. \tag{25}$$

This is surprising since there is no external field applied locally to the quantum dot. In fact, the electron, by virtually tunneling to the magnetic electrodes induces a field on the quantum dot, which acts on its own spin. This local magnetic *exchange field* $\mathbf{B}_r$ induced by the electrode r on the dot reads

$$\gamma\mathbf{B}_r = p_r\mathbf{n}_r\frac{\Gamma_r}{\pi\hbar}P\int_{-\infty}^{\infty} d\omega\left(\frac{f_r^-(\omega)}{\omega-\epsilon} + \frac{f_r^+(\omega)}{\omega-\epsilon-U}\right) \tag{26}$$

where $\gamma = -g\mu_B$ is the negative gyromagnetic ratio determined by the electron g-factor ($-g$) and the Bohr magneton μ_B [15]. The integral is a Cauchy principal value integral i.e.

[14] The Bloch equation follows simply from the Zeeman interaction $H_Z = -\gamma\mathbf{S}\cdot\mathbf{B}$ and the Heisenberg equation of motion: $\dot{\mathbf{S}} = i[H_Z,\mathbf{S}] = \gamma\mathbf{S}\times\mathbf{B}$ using the spin algebra $S_i = i\sum_{jk}\epsilon_{ijk}S_jS_k$.

[15] Note that in many of the original publications the negative gyromagnetic factor γ is absorbed into the exchange field, thereby reversing its direction.

$P\frac{1}{x} = \mathrm{Re}\frac{1}{x+i0}$ with $i0$ denoting an infinitesimal imaginary part. This field is induced by virtual coherent tunnel processes ($\propto \Gamma_r$) to the magnetic electrode r which shift the quantum dot states with opposite spin by a different amount i.e. they split in energy. This splitting, written as a local Zeeman term $g\mu_B \mathbf{S}\cdot\mathbf{B}_r$, defines the exchange field.[16] An important implication is that the exchange field can be very large for strong tunneling Γ_r. Indeed, the exchange field has been detected experimentally in carbon-nanotube [55] and single-molecule quantum dot spin-valves [56], where in the latter case fields on the order of 70 T were reported.

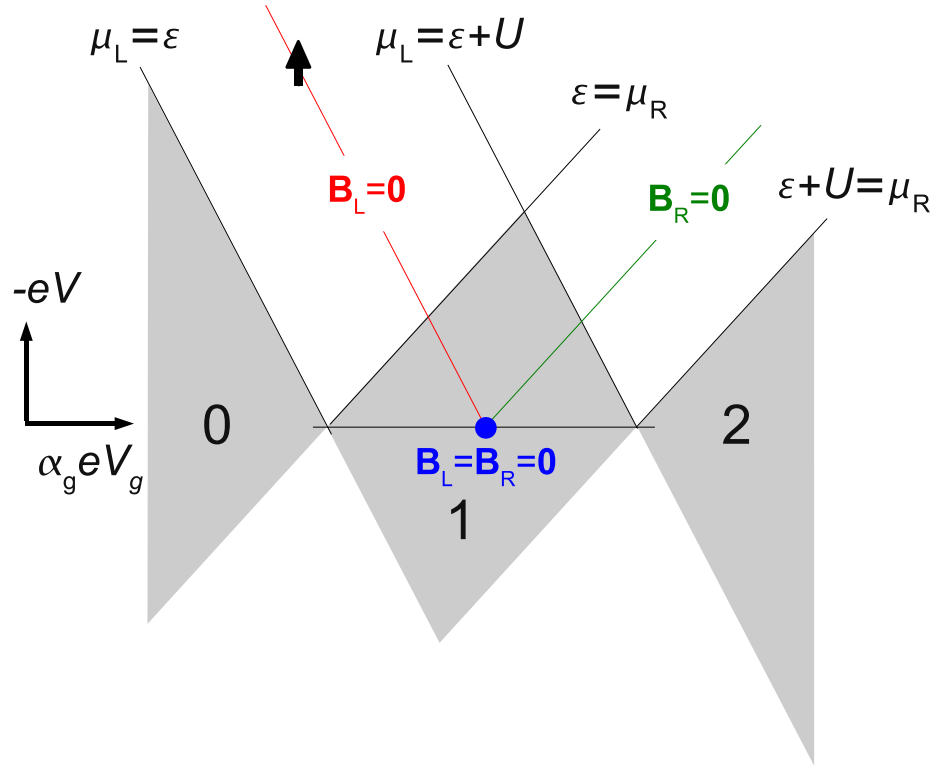


Fig. 7: *Stability diagram for the Anderson model: the boundaries of the finite bias SET regions are defined by the resonance conditions $\epsilon = \mu_r$ and $\epsilon + U = \mu_r$ where $r = L, R$. Midway between these resonances the exchange field induced by the left (right) ferromagnetic electrode vanishes.*

An interesting aspect is that the exchange field is a *Coulomb interaction effect*: it vanishes for $U = 0$ in which case the two terms under the integral cancel exactly (due $f^- + f^+ = 1$ and $P\int d\omega \frac{1}{\omega-\epsilon} = 0$). Here, the first (second) term corresponds to a virtual process where the electron tunnels out of (into) the dot. The finite interaction prevents this destructive interference between these electron and hole processes. The exchange field is also a *many-electron effect* since it involves an integral over the entire spectrum of the electrodes (cut off by the band widths). Furthermore, the exchange field can be *electrically controlled* by the bias (μ_r) and gate (ϵ) voltages, adding a transistor functionality to the device. Due to the interference already mentioned, one of the exchange fields $\mathbf{B}_r$ can even be switched off completely by tuning to

[16] For the simple case where one excludes the $N = 2$ state (i.e. $U = \infty$), the energy shift due to the tunneling in second order perturbation theory reads $E_\sigma - \epsilon = \sum_{k\in\sigma} \frac{t^2}{\epsilon-\omega_k} f_r^-(\omega)$ where the sum extends over the electron states in the band of spin σ. Here $f^- = 1 - f^+$ occurs since the electron fluctuates into unoccupied states in the electrode. Converting the sum into an integral with the density of states $\rho_{r\sigma} = \frac{1}{2}(1 \pm p_r)\rho_r$ and $\Gamma_r = 2\pi\rho_r t^2$, and subtracting the results for spin $\sigma = \uparrow$ and $\downarrow$ one obtains the Zeeman shift leading to the exchange field (26).

the condition $\epsilon - \mu_r = -U/2$ i.e. the electrochemical potential is midway between the two resonances of the Anderson model, $\epsilon = \mu_r$ and $\epsilon + U = \mu_r$. This defines a line in the stability diagram indicated in Fig. 7, along which the two terms in the integral cancel exactly since $\epsilon - \mu_r = -(\epsilon - \mu_r + U)$ and $P\int d\omega f^-(\omega)/(\omega - x) = P\int d\omega f^+(\omega)/(\omega + x)$ for $f^{\pm}(\omega) = (e^{\pm\omega/T} + 1)^{-1}$. The exchange fields of the electrodes can even vanish simultaneously for zero bias $\mu_r = 0$ at the special gate voltage point for which $\epsilon = U/2$ (particle-hole symmetry point). Based on the kinetic equation for the spin (22) we now have the following picture: spin accumulation and relaxation due to transport compete to establish a stationary, non-zero spin value on the dot. The crucial new feature of the spin-valve transistor is that through the exchange fields one can electrically control the *orientation* of this accumulated spin. This one can detect this in the charge transport.

4.2 Charge dynamics and non-linear transport

Having understood the spin dynamics we now turn to the charge dynamics in order to understand the transport current. The occupations of the charge states, $P_N, N = 0, 1, 2$, obey a *generalized* master equation:[17]

$$\frac{d}{dt}\begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix} = \frac{d}{dt}\begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix}_{\text{normal}} + \sum_r p_r\Gamma_r \begin{pmatrix} 2f_r^-(\epsilon) \\ -2f_r^-(\epsilon) + 2f_r^+(\epsilon + U) \\ -2f_r^+(\epsilon + U) \end{pmatrix} \mathbf{n}_r \cdot \mathbf{S}/\hbar. \quad (27)$$

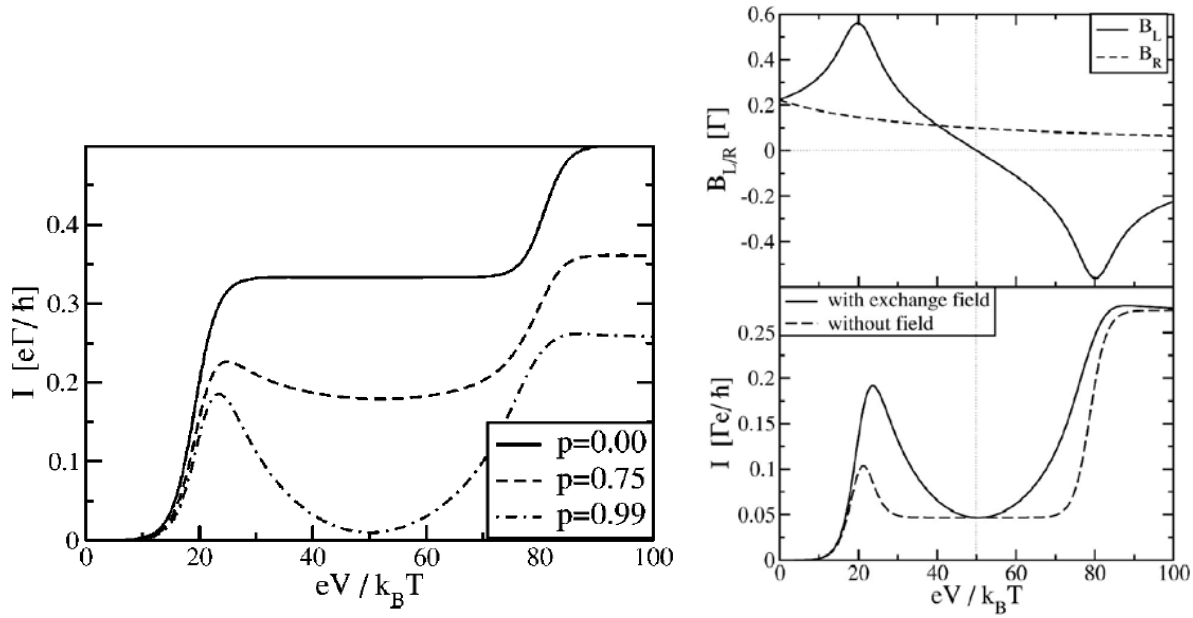
The first term on the right hand side is given by the right hand side of Eq. (18) and does not explicitly depend on the spin. The spin does however, influence it indirectly by changing the occupations. The second term explicitly depends on the spin and arises only for ferromagnetic electrodes: it may cause a gain or loss of probability depending the orientation of spin relative to the polarization vectors. The charge current has a similar structure:

$$I_r/(-e) = (I_r)_{\text{normal}}/(-e) - p_r\Gamma_r\left[f_r^-(\epsilon) + f_r^+(\epsilon + U)\right]\mathbf{n}_r \cdot \mathbf{S}_r/\hbar \quad (28)$$

where $(I_r)_{\text{normal}}$ is the expression for normal electrons on the right hand side of Eq. (19). The second term contains the rate of decay of the one-electron state with non-zero spin.
The exchange field modifies the transport in several ways. For instance, in the linear regime $-eV \ll kT$, it weakens the normal spin-valve effect for a single tunnel junction (see Eq. (20)).[18] Here we focus on the non-linear transport regime where a single electron and therefore a free spin can be present on the dot. In Fig. 8(a) an $I-V$ curve is shown for the case of perpendicular polarizations, i.e. $\theta = \pi/2$. With increasing polarization, the first current plateau is suppressed. In this regime either 0 or 1 electron is present on the dot ($P_2 = 0$) and the current directly is proportional to the occupation P_0. The striking shape of the $I - V$ curve arises due to a kind of *spin blockade* which is counteracted by the precession of the spin around the exchange fields. This is seen by setting $\mathbf{B}_L = \mathbf{B}_R = 0$, in which case the kinetic equation for the spin (22) gives $\mathbf{S} = p\left[(\Gamma_L/\Gamma_R)P_0\mathbf{n}_L - P_1/2\mathbf{n}_R\right]$. One furthermore finds that the dot will be occupied most of the time i.e. $P_1 \approx 1 \gg P_0$. Thus the spin is anti-parallel to the right (drain) electrode,

[17] The second term on the right hand side comes from the non-diagonal elements of the density matrix i.e. Eq. (18) is not a standard master equation which involves only occupancies.

[18] After tunneling through the first junction, the electron spin can precess and thereby adjust itself to the polarization orientation of the drain electrode. This increases the probability of escaping relative to the single junction case.



(a) (b)

Fig. 8: *Results of Braun et. al. [54]. (a) Current-voltage trace for perpendicular polarization directions i.e.* $\theta = \pi/2$*. As the degree of polarization* $p_L = p_R = p$ *is increased there is an overall suppression of the current due to spin blockade, which is counteracted by the precession about the exchange field away from the center of the plateau (*$eV = 50kT$*). The interaction strength is* $U = 30kT \gg \Gamma_L = \Gamma_R = \Gamma/2$ *and* $\epsilon = 10kT$*. (b) Upper panel: magnitude of the exchange fields as function of bias voltage:* $\mathbf{B}_L$ *vanishes at the center of the plateau (*$eV = 50kT$*), while* $\mathbf{B}_R$ *remains finite (which is irrelevant since the spin is parallel to it). Lower panel: comparison of the full result and the result obtain when setting the exchange field to zero "by hand". In the latter case the current uniformly drops to the minimum value due to the exponential dependence the Fermi-functions in the rates, see Eq. (14). In contrast, the much weaker voltage dependence of the exchange field, which lifts the blockade, dominates in the full result. The parameters are the same as in (b) and* $p = 0.95$*.*

$\mathbf{S} \approx -p/2\mathbf{n}_R$. The electron cannot exit due to the low density of minority spin $\downarrow$ in the right (drain) electrode when $p_R \approx 1$. No additional electrons can be transported due to the Coulomb blockade effect and the current is suppressed. This spin blockade would however suppress the current uniformly across the entire plateau as Fig. 8(b) shows. Instead, due to the exchange fields a broad minimum is found. Since the spin blockade causes the spin to be parallel to the exchange field of the right (drain) electrode it is not rotated by it. At the center of the plateau, $\mu_L = V/2 = \epsilon + U/2$, the exchange field of the left (source) electrode vanishes, $\mathbf{B}_L = 0$: only here the spin blockade is thus complete, see Fig. 8(b). However, away from this point the spin will precess around the non-zero $\mathbf{B}_L$ of the source electrode, generating a spin component parallel to the drain polarization $\mathbf{n}_R$ which increases the current.

4.3 Outlook: mapping the exchange field with higher order tunneling

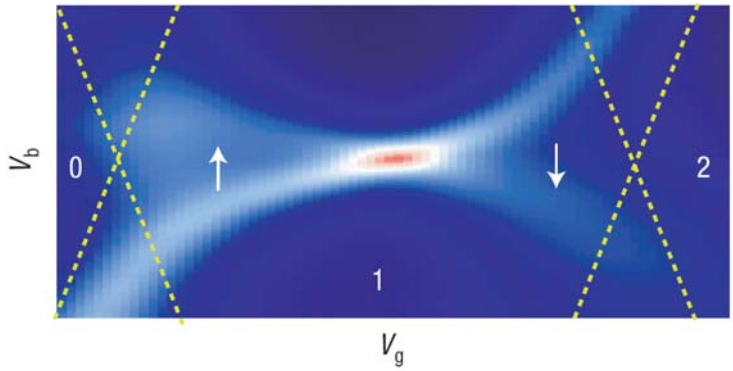


Fig. 9: *Result from Hauptmann et. al. [55]. Experimental stability diagram i.e. differential conductance in color scale as function of applied voltages. The yellow dashed lines indicate the SET regions which are hardly visible here due to the strong tunnel coupling. The white/red color indicates a conductance peak due to higher order tunnel processes. This so-called Kondo peak normally lies at zero bias (i.e. a horizontal line). Due to the exchange field induced by the ferromagnetic electrodes this peak splits and shifts to a finite bias equal to the Zeeman splitting. Clearly, this exchange field goes to zero approximately at the center of the Coulomb diamond, in line with Fig. 7. Away from this point the dependence on the bias follows the same qualitative behavior of $|\mathbf{B}_L|$ as plotted in the upper panel of Fig. 8(b).*

As mentioned above, the exchange field arises due to a splitting induced by virtual tunneling processes. The single electron tunneling discussed so far cannot directly resolve this splitting since thermal fluctuations $\sim kT$ dominate over any effects on the order of the tunnel rates $\hbar\Gamma$, such as the exchange field. However, transport effects involving higher order tunnel processes, such as cotunneling [57] and the Kondo effect [58, 59], are sensitive this splitting. These are beyond the scope of this introduction, see the reviews [51, 52]. As a concluding remark, we merely mention that the exchange field was demonstrated experimentally by measuring the predicted Zeeman shift of the Kondo effect [58, 59] in C_{60} molecular quantum dots [56]. Furthermore, recently [55] this Zeeman splitting was used to directly map out the magnitude of exchange field as function of a level position (gate voltage) in a carbon-nanotube quantum dot contacted with ferromagnetic electrodes, as shown in Fig. 9.

Acknowledgement F. Reckermann, S. Das and M. Baumgärtel are acknowledged for proof reading the manuscript.

References

[1] R. A. Millikan, Phys. Rev. (Series I) **32**(4), 349 (1911).

[2] C. J. Gorter, Physica **15**, 777 (1951).

[3] I. O. Kulik and R. I. Shekhter, Sov. Phys. JETP **4**, 308 (1975).

[4] T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. **59**, 109 (1987).

[5] H. Grabert and M. H. Devoret, eds., *Single Charge Tunneling*, NATO ASI Series (New York, Plenum Press, 1992).

[6] D. K. Ferry and S. M. Goodnick, *Transport in Nanostructures* (Cambridge University Press, 1997), chap. Quantum dots and single electron phenomena.

[7] L. Kouwenhoven, C. Marcus, P. McEuen, S. Tarucha, R. Westervelt, and N. Wingreen, in *Mesoscopic electron transport*, edited by K. Sohn, L. Kouwenhoven, and G. Schön (Kluwer, 1997), chap. Electron Transport in Quantum Dots.

[8] K. K. Likharev, Proc. IEEE **87**, 606 (1999).

[9] L. P. Kouwenhoven, D. G. Austing, and S. Tarucha, Rep. Prog. Phys. **64**, 701 (2001).

[10] W. G. van der Wiel, S. D. Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, and L. P. Kouwenhoven, Rev. Mod. Phys. **75**, 1 (2003).

[11] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. Vandersypen, Rev. Mod. Phys. **79**, 1217 (2007).

[12] D. V. Averin and Y. V. Nazarov, in *Single Charge Tunneling*, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1992).

[13] M. Leijnse and M. R. Wegewijs, Phys. Rev. B **78**, 235424 (2008).

[14] M. Pustilnik and L. Glazman, J. Phys. Condens. Matter **16**, R513 (2004).

[15] H. Schoeller, Eur. Phys. Journ. B (2009), (in press).

[16] T. Fujisawa, D. G. Austing, Y. Tokura, Y. Hirayama, and S. Tarucha, J. Phys. Cond. Mat. **15**, R1395 (2003).

[17] C. Meyer, J. Elzerman, and L. Kouwenhoven, Nano Letters **2**, 295 (2007).

[18] Y. Alhassid, Rev. Mod. Phys. **72**, 895 (2000).

[19] J. A. van Dam, Y. V. Nazarov, E. P. A. M. Bakkers, S. D. Franceschi, and L. P. Kouwenhoven, Nature **442**, 667 (2006).

[20] F. A. Zwanenburg, C. E. W. M. van Rijmenam, Y. Fang, C. M. Lieber, and L. P. Kouwenhoven, *Spin states of the first four holes in a silicon nanowire quantum dot* (2009), arXiv:0811.2914.

[21] S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs, and C. Dekker, Nature **386**, 474 (1997).

[22] M. Biercuk, S. Ilani, C. Marcus, and P. McEuen (2008), vol. 111 of *Topics in Applied Physics*, p. 455.

[23] E. Osorio, T. Bjornholm, J.-M. Lehn, M. Ruben, and H. van der Zant, Jour. Phys. Cond. Mat. **20**, 374121 (2008).

[24] C.W.J. Beenakker and H. van Houten, Solid State Phys. **44**, 1 (1991).

[25] K. Flensberg, S. M. Girvin, M. Jonson, D. R. Penn, and M. D. Stiles, Phys. Script. **T42**, 189 (1992).

[26] H. Grabert, Z. Phys. B **85**, 319 (1991).

[27] G. Schön, *Quantum transport and dissipation* (VCH Verlag, 1997), chap. Single-electron tunneling, p. 105, URL `http: //www.tfp.uni-karlsruhe.de/schoen/setrev4.pdf`.

[28] A. N. C. J. M. Schmidt and J. Clarke, Phys. Rev. Lett. **64**(13), 1565 (1990).

[29] G. Ingold and Y. V. Nazarov, in *Single charge tunneling*, edited by H. Grabert and M. H. Devoret (Plenum Press, New York, 1992), vol. B 294 of *NATO ASI Series*, chap. Charge tunneling rates in ultrasmall junctions.

[30] Yu. V. Nazarov, JETP Lett. **49**, 105 (1989).

[31] M. H. Devoret, D. Esteve, H. Grabert, G.-L. Ingold, H. Pothier, and C. Urbina, Phys. Rev. Lett. **64**, 1824 (1990).

[32] S. M. Girvin, L. I. Glazman, M. Jonson, D. R. Penn, and M. D. Stiles, Phys. Rev. Lett. **64**, 3183 (1990).

[33] S. Bednarek, B. Szafran, and J. Adamowski, Phys. Rev. B **61**, 4461 (2000).

[34] L. D. Hallam, J. Weis, and P. A. Maksym, Phys. Rev. B **53**, 1452 (1996).

[35] E. A. Osorio, K. O'Neill, M. R. Wegewijs, N. Stuhr-Hansen, J. Paaske, T. Bjørnholm, and H. S. van der Zant, Nanolett. **7**, 3336 (2007).

[36] K. Kaasbjerg and K. Flensberg, Nano Lett. **8**, 3809 (2008).

[37] M. Macucci, K. Hess, and G. J. Iafrate, Phys. Rev. B **48**, 17354 (1993).

[38] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1984).

[39] J. D. Jackson, *Classical Electro-dynamics* (Wiley, New York, 1975).

[40] G. Lientschnig, *Simulating Hybrid Circuits of Single-Electron Transistors and Field-Effect Transistors*, Ph.D. thesis, TU Delft (2003), (available from http://www.tudelft.nl).

[41] D. Weinmann, *Quantum transport in nanostructures*, Ph.D. thesis, University of Hamburg (1994), URL `http://www-ipcms.u-strasbg.fr/IMG/pdf/diss.pdf`.

[42] J. von Delft and D. C. Ralph, Phys. Rep. **345**, 61 (2001).

[43] E. A. Osorio, K. O'Neill, N. Stuhr-Hansen, O. F. Nielsen, T. Bjørnholm, and H. S. van der Zant, Adv. Mater. **19**, 281 (2007).

[44] H. Schoeller, *Mesoscopic Electron Transport* (Kluwer, 1997), chap. Transport through interacting quantum dots, p. 291.

[45] M. Julliere, Phys. Lett. A **54**, 225 (1975).

[46] J. C. Slonczewski, Phys. Rev. B **39**, 6995 (1989).

[47] J. S. Moodera and L. R. Kinder, J. Appl. Phys. **79**, 4724 (1996).

[48] H. Jaffrès, D. Lacour, F. Nguyen Van Dau, J. Briatico, F. Petroff, and A. Vaurès, Phys. Rev. B **64**, 064427 (2001).

[49] W. Wetzels, G. E. W. Bauer, and M. Grifoni, Phys. Rev. B **74**, 224406 (2006).

[50] H. Saarikoski, W. Wetzels, and G. E. W. Bauer, Phys. Rev. B **75**, 075313 (2007).

[51] J. König, J. Martinek, J. Barnas, and G. Schön, in *CFN Lectures on Functional Nanostructures*, edited by K. Busch et al. (Springer, 2005), Lecture Notes in Physics, chap. Quantum dots attached to ferromagnetic leads: exchange field, spin precession, and Kondo effect, p. 145.

[52] M. Braun, J. König, and J. Martinek, *Manipulating Single Spins in Quantum Dots Coupled to Ferromagnetic Leads* (Springer, 2006), Lecture Notes in Physics.

[53] J. König and J. Martinek, Phys. Rev. Lett. **90**, 166602 (2003).

[54] M. Braun, J. König, and J. Martinek, PRB **70**, 195345 (2004).

[55] J. R. Hauptmann, J. Paaske, and P. E. Lindelof, Nat. Phys **4**, 373 (2008).

[56] A. N. Pasupathy, R. C. B. and1 J. Martinek, J. E. Grose, L. A. K. Donev, P. L. McEuen, and D. C. Ralph, Science **306**, 86 (2004).

[57] I. Weymann and J. Barnas, Phys. Rev. B **75**, 155308 (2007).

[58] J. Martinek, M. Sindel, L. Borda, J. Barnas, J. Konig, G. Schön, and J. von Delft, Phys. Rev. Lett. **91**, 247202 (2003).

[59] J. Martinek, Y. Utsumi, H. Imamura, J. Barnas, S. Maekawa, J. Konig, and G. Schön, Phys. Rev. Lett. **91**, 127203 (2003).

C 6 Spin and Transport through Quantum Dots

M. R. Wegewijs
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Single electron charging effects on tunneling **3**

3 Single quantum dots **4**
3.1 Linear transport 6
3.2 Non-linear transport 7
3.3 Discrete states and excitations 9
3.4 Transport theory - occupations and rates 12

4 Single electron spin-valve **14**
4.1 Spin dynamics due to transport 15
4.2 Charge dynamics and non-linear transport 18
4.3 Outlook: mapping the exchange field with higher order tunneling 20

1 Introduction

Quantum dots are small devices in which few mobile electrons are confined and can be connected to macroscopic electrodes. The properties of such systems are dominated by the quantization of electronic charge- and possibly other degrees of freedom (orbit, *spin*). This quantization makes quantum dots very similar to atoms (coupled dots similar to molecules). Therefore one might think that no fundamental new physics is involved. However, their tunability and the possibility of wiring them up in an external electric circuit enables new regimes to be accessed. In particular in *transport*, on can study in detail the interplay of strong interactions, quantized motion and non-equilibrium boundary conditions imposed by the external electrodes. A fascinating aspect is that in these solid state devices *single electrons* can experimentally be manipulated to a high degree despite the coupling to huge numbers of other electrons and other degrees of freedom around (phonons, nuclear spins, etc). It is the dominating role of the Coulomb interaction in these confined systems which allows for this degree of control: once a single electron has jumped to some new position in such a device, other electrons are prevented from following it since the change in Coulomb interacting energy is larger than thermal and quantum fluctuations. This suppression of transport is called *Coulomb blockade*. Coulomb interaction is also responsible for many other effects such as exchange interactions, spin blockade, the exchange-field induced by ferromagnetic electrodes, and also the Kondo effect. The degree of control, however, comes at a price: the understanding of the operation of such devices is complicated due to the inherent many-body nature of the transport processes.

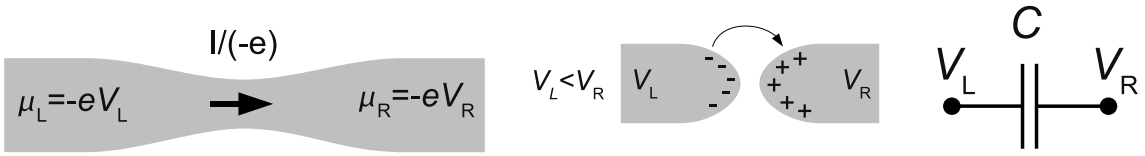
However, much progress has been made. Although single electron experiments date back to Millikan's determination of the electron charge [1] $q_e = -e$ where $e \approx 1.602\ 10^{-19}$C, only much later solid-state experiments involving single electron effects were done. Important developments came from transport through metallic layers which break up into small islands due to surface tension effects [2] and the progress in the field of superconducting tunnel-junctions. This stimulated the development of basic theoretical models and approaches still in use today [3]. After the advent of the first *tunable* single electron devices [4] the field of quantum dot physics boomed, see the key reviews [5, 6, 7, 8, 9, 10, 11]. In the last two decades quantum dots realized in nano-structured *solid-state* devices have reached a high level of sophistication, ranging for "on-chip" fundamental physics experiments and hybrid devices to device applications. Nowadays nano-gap fabrication combined with chemical functionalization even allows quantum dot devices to be fabricated in single *real* molecules. Thus, in a way the field of quantum dots has now come full circle: the techniques which have developed for controlling "artificial" atoms and molecules can now be applied to the real thing: the complicated problem of transport through single molecules greatly benefits from this wealth of experience and techniques, both in experiment and theory. Clearly, quantum dot systems are encountered in many areas of physics, electronics, chemistry and biology and will continue to be crucial in the full range from basic physics to nano-scale device applications.

The importance of interactions in the transport processes of charge and spin makes understanding and operating quantum dot devices a complicated task. It is the aim of this lecture to introduce some of the basic notions of single quantum dot devices underlying current research efforts. Basics of Coulomb blockade and orbital quantization effects will be introduced in Sections 2-3 based on the above cited key reviews where more details and references can be found. In the last decade spin effects have been at the center of attention, including Hund's-rule / shell-filling effects, Zeeman splitting, exchange interaction in double dots, spin-polarized electrodes. In the remaining Section 4) a specific quantum dot devices will be discussed where the elec-

tron spin plays a crucial role in the transport. Specifics of different realizations of quantum dots systems are avoided if possible to focus on common principles and models. Other interesting avenues where spin effects are important but which are not considered here are Pauli spin-blockade in double quantum dots [11], higher-order tunneling effects ("cotunneling", level renormalization, broadening) [12, 13], Kondo effect [14, 15], time-dependent transport [7, 16, 17], metallic dots (without orbital quantization effects) [8], many-electron quantum dots and their statistical properties [7, 18], superconducting quantum dots and/or electrodes [19] and device application aspects of single electron circuits [8]. Specifics on transport through quantum dots realized in semiconductor hetero-structures [9], wires [20], carbon nano-tubes [21, 22] and nanometer-size molecular quantum dots [23] can be found in the indicated references (reviews and some key publications).

2 Single electron charging effects on tunneling

To understand how Coulomb interactions allow single electrons to be controlled, we consider the simplest structure, a single tunnel junction [5, 7, 8, 9] between two electrodes at potentials V_r, where the left and right electrode are denoted by $r = L, R$. We first fix some conventions. The electrochemical potentials of the electrodes is $\mu_r = \epsilon_F - eV_r$, where $e > 0$ denotes the elementary charge *magnitude* (i.e. the electron charge is $= -e$). Therefore a chemical potential or energy bias $\mu_L - \mu_R = -e(V_L - V_R) > 0$ will drive an electron *particle*-current in the direction $L \to R$ i.e. L acts as the source and R as the drain electrode. This requires a negative bias voltage $V = V_L - V_R < 0$ to be applied, and the *electric* current by definition flows in the opposite in the opposite direction, $I < 0$. We will always discuss the electron particle $I/(-e)$ current versus the applied energy bias $-eV$ (or drain-source voltage) to avoid the inconvenient minus signs, as is common practice. [1]



(a) Quantum point contact (b) Tunnel junction (c) Circuit

One can think of a tunnel junction as a pinched-off ballistic conduction channel. For such a system the two-terminal conductance is due to contact resistance and is quantized in units of the fundamental quantum of conductance $G_Q = e^2/h$ [24]. If the conduction channel is gradually pinched off, the conductance is reduced in steps of magnitude $2G_Q$ (factor 2 for spin) until finally the current drops exponentially to zero and

$$dI/dV \ll e^2/h. \tag{1}$$

This is referred to as *charge quantization*, which is a semi-classical effect. The electrons are classically confined to either electrode (i.e. with well-defined charge numbers). To see some effect of this in transport one needs to have quantum mechanical tunneling through the insulating barrier: the junction has to be considered as a "leaky" capacitor. When an electron

[1] Note that in the differential conductance, this sign cancels out when converted to energy units and particle currents: $dI/dV = ed(I/(-e))/d(\mu_L - \mu_R)$.

tunnels across the junction the polarization charge changes abruptly by the elementary quantum of charge e. As a result potential across the junctions jumps up or down by $\Delta V = e/C$. The associated change in the energy of the junction, $e\Delta V = e^2/C$, is called the *charging energy*. If the applied bias voltage energy eVor the thermal fluctuations kT are not able to overcome this energy barrier, the transport is blocked. Thus the tunneling of a *single electron* can have a dominating effect on the macroscopically measurable transport current. Since the capacitance C scales linearly with the junction dimensions, *single electron tunneling* (SET) effects become important for sufficiently small junctions and / or sufficiently small temperatures:

$$kT \leq e^2/C. \tag{2}$$

Generally, the suppression of tunneling due to the charging energy is termed *Coulomb blockade*.[2] A tunnel junction is very different from an ordinary resistor where an incompressible electron liquid flows in a quasi-continuous way, responding instantaneously to electric fields. In contrast, the leakage of junction charge is a slow process i.e. only once in a while an electron tunnels, the duration of this quantum mechanical process being much shorter than the average time between such events [6]. The time for the charge in the junction to rearrange itself after a tunnel process is also very short (dielectric relaxation time related to plasmon dynamics [25, 6]). Surprisingly, despite the clear arguments above single electron charging effects are *not* very important in single tunnel junctions [26, 27] and were first observed in double tunnel junctions where a quantum dot is formed to be discussed in Sect. 3. The point is that above we have tacitly assumed that the polarization charge on the junction is a classically well-defined variable. However, this depends strongly on the nature of the electromagnetic environment created by the external circuit, which is can be described by a frequency dependent impedance $Z(\omega)$ in series with the junction. Only for a high impedance environment $Z(0) \gg 1/G_Q$ the junction charge is classically well-defined. However, this is almost never the case unless exceptional precautions are taken [28]. Therefore, quantum fluctuations of the junction charge and voltage are strong and can nearly completely lift the Coulomb blockade of transport: during such a fluctuation charge moves into the circuit leads and the Coulomb barrier is lowered, allowing the electron to tunnel. One may say that an electron tunnels directly into the leads, rather than staying localized onto the capacitor plate. See for a clear exposition and examples [25]. A precise understanding of the electromagnetic environment requires a full quantum description of the circuit to calculate the transport, see the reviews [26, 29, 27] and key papers [30, 31, 32] for further details.

3 Single quantum dots

Clearly, when introducing a second tunnel junction, thus creating an island or *dot* between two electrodes, the above problem with the low impedance of the external circuit is solved: when viewed from one junction, the other junction shunts the low environmental impedance [26]. Therefore charging effects are always important for two or more tunnel junctions in series under the conditions of charge quantization (1) and of low temperature (2), even if the environmental

[2] Another way of seeing how the quantum conductance enters into the condition for observing charging effects at low temperature is based on the time-energy uncertainty relation, which states that in the classical limit $\Delta E \Delta t \gg \hbar$. The energy uncertainty is set by the charging $\Delta E \sim e^2/C$ and the time for discharging the capacitor is governed by the smallest RC-time $\Delta t \sim RC$ where R is the largest resistance dominating the current. Plugging this in the capacitance cancels out and one obtains the condition for classical charging effects to be $R \gg h/e^2 = 25.8k\Omega$ i.e. Eq. 1.

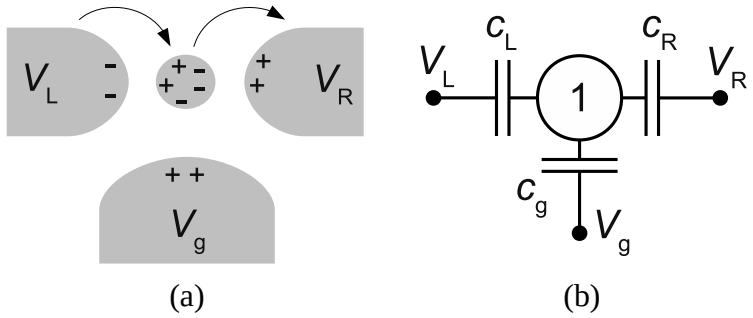


Fig. 1: *(a) Quantum dot connected to electrodes and a gate terminal. (b) Double junction electrostatic circuit. Tunneling through the "leaky capacitors" is not indicated.*

impedance is low compared to $1/G_Q$. Such systems are considered from hereon and the issue of the environment can be ignored.

A *quantum dot* denotes a confined electron system where both charge *and* orbital quantization effects are important, see the key reviews [7, 8, 9, 11]. We first discuss charge quantization by extending the above capacitance model to a double junction. The success of the capacitance model derives from the smallness of screening length compared to the electron wavelength [7].[3] The small capacitances needed to see charging effects are limited from below by *self-capacitance* which for a sphere equals $4\pi\epsilon L$ and for a disk $8\epsilon L$ (height $\ll L$), i.e. roughly $10\epsilon L$ where L denotes the radius. The condition $e^2/C = kT$ for room-temperature $T = 300$ K is satisfied for $C \sim 2$ aF, whereas for low temperature $T = 1$ K this becomes $C \sim 6$ pF. In vacuum, this translates into typical dimensions of $L \sim e^2/(kT10\epsilon_0) = 70$ nm and $L \sim 20$ μm for $T = 300$ K and $T = 1$ K, respectively. The effective capacitance is actually larger than this estimate due to the fact that the tunnel-barrier consists of materials with $\epsilon > \epsilon_0$, the *mutual capacitances* to the nearby conducting electrodes [34, 36] and the *quantum-capacitance* due to non-locality of charge [37]. Therefore one needs smaller dimensions that the above rough estimates, see [7].

Now consider a dot (labeled by 1) coupled by two tunnel junctions to electrodes $r = L, R$, as sketched in Fig. 1. A third, capacitively coupled, gate electrode g, allows the potential on the dot to be varied in a *continuous* and linear way by changing the voltage V_g[4]. This originates from a continuous shift of the incompressible conduction electron liquid relative to the ionic cores in the material [25]. The dot is thus considered as a conducting body with a small capacitance: when *discrete* charges are moved from the electrodes to the dot the potential of the dot V_1 strongly changes and may block the transfer if the corresponding energy change is larger than kT. The calculation of the energy of a set of conductors is a textbook [38, 39] problem. However, the issue is that we need the energy as function of the electrode potentials and dot charge. We include these voltage sources in the network of conducting elements[5] as

[3] In semiconductors this condition is less well satisfied than in metals and there is a weak voltage dependence of the capacitances [33, 34]. In molecular quantum dots this dependence is even stronger [35, 36]. In such cases one can often work with differential capacitances i.e. linearized with respect to voltage.

[4] The gate electrode can also model constant stable background charges, which are always present and give rise to a constant gate voltage offset. Experimentally, one can simply compensate for offset charges by changing V_g, so we leave them out here. However, for applications background charges present a major problem. We note that instable background charges less trivial effect and can give rise to a dynamic "resetting" of the gate voltage scale. These have to be incorporated into the transport model, resulting in a double dot model. This is especially important in nanometer sized molecular quantum dots.

[5] There are in fact two equivalent ways of approaching this [40], both used in the literature. The approach taken

conductors with large capacitance to ground and large charge i.e. their potential is insensitive to their charge. The opposite is true for the dot: its potential is highly sensitive to changes in the charge Q_1 due to the small total capacitance $C_1 = \sum_{r=L,R} c_r + c_g$. This is seen by expressing the charge in the capacitances in Fig. 1 and the voltage differences across the 3 different junctions, $Q_1 = \sum_{r=L,R} c_r(V_1 - V_r) + c_g(V_1 - V_g)$. Solving for the dot potential V_1 we find:

$$V_1 = (C_1)^{-1} \left(Q_1 + \sum_r c_r V_r + c_g V_g \right). \tag{3}$$

To find out if tunneling is possible, we need to calculate the change in the total energy of the whole system due to a single electron tunneling event. Introducing the electron number $N_1 = Q_1/(-e)$ and inserting it into the electrostatic energy $E_{N_1} = (V_1)^2/2C_1$ we obtain

$$E_{N_1} = U_1 \left(N_1 - \left(\sum_r c_r V_r + c_g V_g \right) /e \right)^2 \tag{4}$$

where U_1 is the *charging energy*:

$$U_1 = \frac{e^2}{2C_1}. \tag{5}$$

As function of the externally induced continuous charge $\sum_r c_r V_r + c_g V_g$ each charge state defines a potential energy parabola with a minimum at $eN_1 = \ldots, -e, 0, e, \ldots$. Clearly, at zero applied voltages $V_{L,R,g} = 0$ the state with $N_1 = 0$ electrons has the lowest energy. The *energy change* associated with adding a single electron, the *electrochemical potential* of the dot, takes on discrete values:[6]

$$\mu_{N_1} = E(N_1) - E(N_1 - 1) = U_1 \left(N_1 + \frac{1}{2} \right) - \sum_{i=L,R,g} \alpha_i e V_i \tag{6}$$

Here we introduced the capacitance ratios $\alpha_i = c_i/C_1 < 1$ for $i = L, R, g$. We now use the conservation of total energy of the dot and the reservoirs for a single electron tunnel process. For *weak tunnel coupling* this is sufficient, see below Sect. 3.4.[7]. While in the electrodes with a continuum of electron energies is are available, only those electrons with energy equal to one of the discrete values μ_N can enter the dot.

3.1 Linear transport

The charge quantization becomes evident for small applied voltages and low temperature. As one varies the gate electrode to more positive values, one attracts electrons to the dot by the

here is simpler to extend to multiple dot systems [10] and to quantum-mechanical systems, see [41] (Appendix). Alternatively, one can consider the dot as an open system, treating the voltage sources as external. If one then considers a tunnel event, one has to calculate the change in free energy plus the *work done by the voltage sources*. In this picture, tunneling only occurs when the free energy decreases ("downhill" process), the energy difference being dissipated in the external sources [40].

[6] One should be aware that often these values of the electrochemical potential are referred to as "levels" of the dot, even though they are energy changes (related to electron affinity, ionization potential). Energy level diagrams where electrons are drawn in as occupying such a level often lead to confusion at first.

[7] For strong tunnel coupling this breaks down, since higher order processes shift the dot energy levels.

positive *continuous* induced charge. As a result, all μ_N shift down uniformly and one achieves subsequent resonances $\mu_N = \mu_r$ for $N = 1, 2, \ldots$. At such a *charge-degeneracy point* an electron can enter / leave the dot without changing its energy. The small bias $\mu_L - \mu_R \ll kT$ then results in a directed transport current and a finite linear conductance $(dI/dV)_{V=0}$: for $\mu_L > \mu_N > \mu_R$ an electron can be extracted from the left, inserted into the dot and finally extracted on the right, while conserving the total energy, see Fig. 2. Clearly, the probability for the opposite process is smaller. Plotting the conductance as function of V_g one obtains what are called Coulomb blockade oscillations, see Fig. 2.

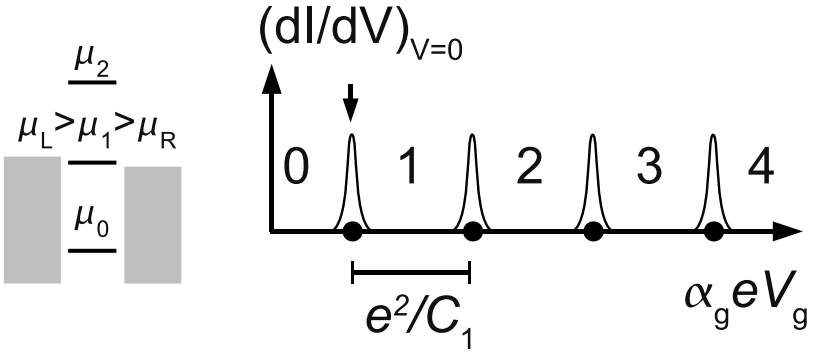


Fig. 2: *Coulomb blockade oscillations of the linear conductance: at the dots on the gate energy axis the resonance condition $\mu_N = \mu_L \approx \mu_R$ is satisfied for some $N = 0, 1, 2, \ldots$. The left inset shows the electrochemical potentials at the degeneracy point marked by the arrow. The energy width of the peak is set by the electron temperature of the electrodes, kT, which smears the sharp Fermi levels. Between the conductance peaks the charge on the dot is fixed to the indicated value due to Coulomb blockade.*

The distance between subsequent degeneracy points $\mu_{N_1} - \mu_{N_1-1}$ is called the N_1th *addition energy*:[8]

$$\mu_{N_1} - \mu_{N_1-1} = U_1 = \frac{e^2}{C_1}. \tag{7}$$

This is the *extra* energy required to add the N_1th electron, given that one already has the energy available to add the $N_1 - 1$th electron. The addition energy is thus an electrochemical potential *difference* i.e. a second difference of the energy E_N with respect to N (curvature of energy versus N). In the electrostatic model considered here, the addition energy is the same for all N and equal to the charging energy U_1. The distance between the degeneracy points allows the addition energy to be read off directly from experimental data, giving the capacitance C_1.[9] To determine α_g independently, one needs to measure the non-linear transport as well.

3.2 Non-linear transport

For finite bias $-eV \gg kT$ each degeneracy point from the linear transport regime widens into a finite region where single electron transport (SET) is possible. The resulting *stability diagram* is

[8] There is no chemical contribution here i.e. the Fermi-levels of the metallic islands are assumed to be identical. Such differences would only cause an offset in the gate voltage, similar to background charged-impurities.

[9] The precise position of the degeneracy points μ_{N_1} are not very useful, since they depend on the Fermi-level of the electrodes, the number of electrons N_1 (not directly accessible from the transport measurement) and uncontrollable offsets due to charged impurities close to the device.

sketched in Fig. 3, the light regions indicating SET, the dark regions Coulomb blockade where the current is suppressed and the charge state on the dot is fixed despite the finite temperature and voltage. By using bias and gate *energies* on the axis, the meaning of the boundary lines at the SET region are easily remembered: the left (right) diagonal line through the Nth degeneracy point corresponds to resonance with the left (right) electrode, $\mu_L = \mu_N$ ($\mu_R = \mu_N$). These basic rules of thumb are illustrated by the left two insets in Fig. 3. Clearly, the boundary lines must cross at the charge degeneracy point where $\mu_L = \mu_N = \mu_R$. The slopes of the

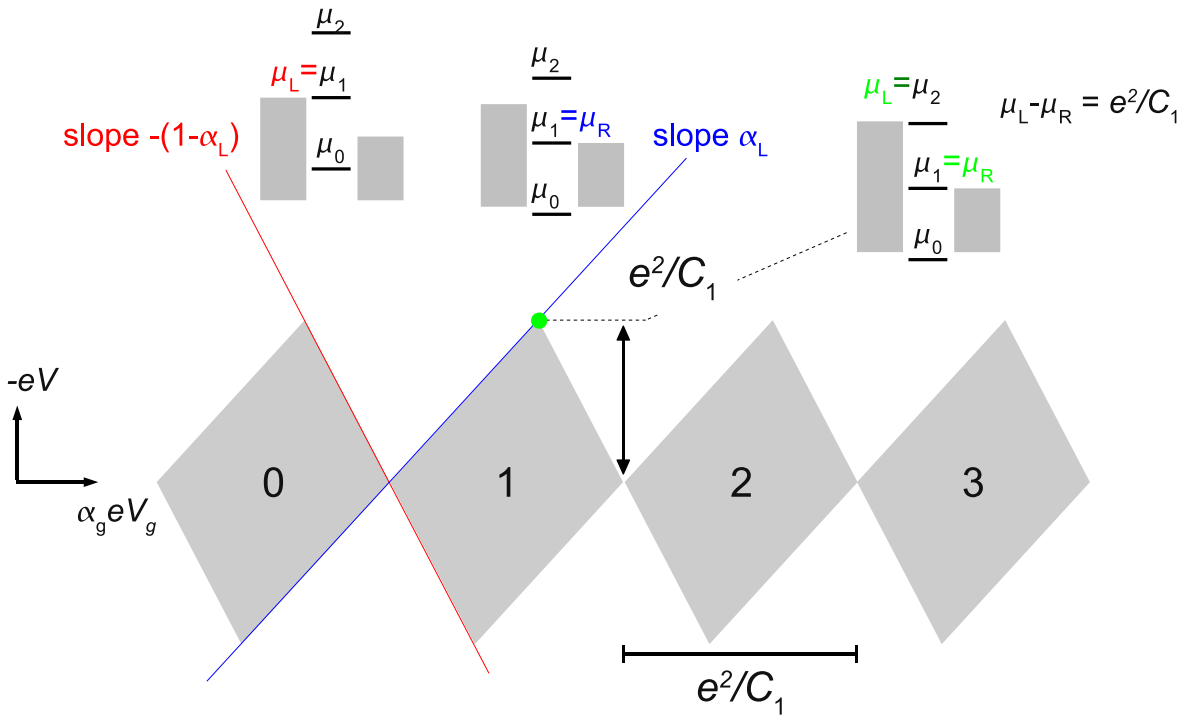


Fig. 3: *Stability diagram in $(\alpha_g eV_g, -eV)$ plane. The vertical and horizontal size of the diamond shaped Coulomb blockade regions is in general a measure of the stability of the charge state (indicated by the number), given here by the charging energy $U_1 = e^2/C_1$. The boundary lines correspond to the electrochemical potential diagrams indicated in the left two insets. Their slopes are given for the experimentally most relevant case of asymmetric bias $V_L = V, V_R = 0$: they follow by equating $\mu_L = -eV = \mu_N$ and $\mu_R = 0 = \mu_N$ and using Eq. (6). The right inset makes clear that when subsequent electrochemical potentials of the dot are simultaneously at resonance, that the energy bias $\mu_L - \mu_R = -eV$ equals the addition energy, which equals the charging energy.*

resonance lines in the stability diagram actually depend on how the bias $-(V_L - V_R)$ is applied. In experiments one usually sets $V_R = 0$ for practical reasons, which has the added advantage that the charging energy can be read off directly as the height of the diamond shaped Coulomb blockade region, see the upper right inset to Fig. 3. In theoretical works one favors a symmetric bias $-V_L = V_R = V/2$ since then for symmetric coupling $C_L = C_R$ (a common simplifying assumption) the electrochemical potentials μ_{N_1} are independent of the bias voltage and are only shifted by V_g, see Eq. (6).

More generally, a stability diagram indicates the boundaries between regimes where the current has a different value. At the boundary dI/dV is nonzero since the current changes drastically

due to a resonance condition being satisfied (energy conservation). Therefore color scale plots of dI/dV as function of the applied gate and bias voltage are the standard starting point for transport spectroscopic studies and provide a compact format for summarizing many theoretical results in a single figure. From such a plot, an important characteristic transport signature of quantum dots without quantized orbital states (usually metallic particles) is immediately evident: the Coulomb diamonds have identical shape and size, see (7). See the review [42] for experimental examples and theoretical modeling.

3.3 Discrete states and excitations

If one makes a quantum dot sufficiently small, *size quantization* effects will become important due to the wave nature of electrons.[10] For such systems the above electrostatic treatment accounting only for charge quantization is clearly insufficient. However, it turns out that by a simple extension one obtains a very successful model. One first neglects Coulomb electron-electron interactions altogether and calculates the single-particle energies ("particle in a box problem"). The total energy is now approximated by the sum of the electrostatic contribution (4) and the single-particle energies. This is the so-called *constant-interaction model*. There is now both a *chemical* (orbital) and electrostatic contribution to the energy change upon charge addition, starting in the $N_1 - 1$ and ending in N_1 electron *ground* state:

$$\mu_{N_1} = \epsilon_{N_1,\text{free}} + E(N_1) - E(N_1 - 1) = \epsilon_{N_1,\text{free}} + U_1 \left(N_1 + \frac{1}{2} \right) - \sum_{i=L,R,g} \alpha_i e V_i. \qquad (8)$$

Here $\epsilon_{N_1,\text{free}}$ denotes the energy of the lowest free orbital to which the N_1th electron is added. Clearly, this term depends on the electron number N_1 and on the details of the spectrum. This implies that in general, the addition energy differs from the charging energy. For example, if the energy spectrum is non-degenerate, starting from a dot where the highest orbital level is half-filled, one only needs to pay the charging energy *extra* to add the next electron to the same level, i.e. the addition energy equals the charging energy, $\mu_{N_1} - \mu_{N_1-1} = U_1$, for N_1 =even. In contrast, when all levels are initially doubly occupied, the N_1th electron is added to the next empty orbital when N_1 =odd. In this case the addition energy is $U + \delta$, where δ is the separation of the highest filled level to the empty level. If the level spacing and the charging energy are approximately constant for several subsequent electron numbers (which is often the case) , then the stability diagram shows an even-odd alternation of the sizes of the Coulomb diamonds, as sketched in Fig. 4. This is common fingerprint of size quantization in quantum dots which can be directly read off from the stability diagram.[11] In contrast, small quantum dots without special symmetries have no orbital shell structure i.e. irregular orbital energy spacings. Irregular addition spectra occurring for low electron numbers are thus another fingerprint of quantization effects.

[10] For a single electron a potential-box of size L the kinetic energy scales as L^{-2} with system size, and will for some value of L equal the charging energy which scales $\propto L^{-1}$.

[11] If the spectrum contains degeneracies, the filling of the degenerate shell will give rise to several addition energies equal to the charging energy U, followed by one addition energy $U + \delta$ when the next shell starts to be filled. For example, carbon-nanotube quantum dots display clear shell-filling over many tens of electron numbers [22]. It is striking that the degeneracy, a fingerprint of the microscopic molecular structure (hexagonal carbon patches), shows up in a macroscopic transport current! In semi-conductor quantum dots with well-defined circular disk shape the shell filling as measured in transport follows accurately a 2-dimensional analogue of the periodic table, see [9] for details.

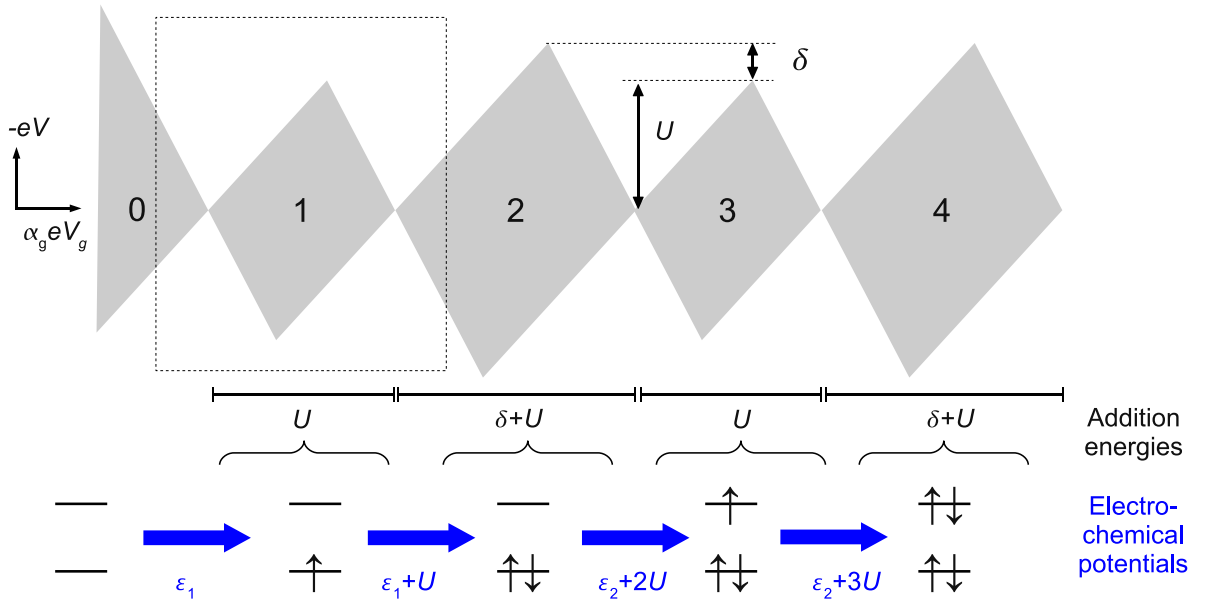


Fig. 4: *Stability diagram in $(\alpha_g eV_g, -eV)$ plane for a quantum dot with multiple quantized orbital states. Bottom: filling of the orbital levels with energies $\varepsilon_1, \varepsilon_2$ in the different charge states. The orbital energy splitting is denoted by $\delta = \varepsilon_2 - \varepsilon_1$. Blue arrows: electrochemical potential i.e.* energy-change *for transitions between these states. The addition energies are obtained by taking the electrochemical potential differences. The Anderson model Hamiltonian gives a good description in the boxed region on the left.*

In a limited gate and bias voltage range the essential transport characteristics are captured by restricting the model to a single level with a charging energy U coupled to two electrodes. The corresponding model Hamiltonian defines the *Anderson model*,

$$H_D = \varepsilon \sum_{\sigma=\uparrow,\downarrow} n_\sigma + U n_\uparrow n_\downarrow \tag{9}$$

where $n_\sigma = d_\sigma^\dagger d_\sigma$ counts the occupation of the orbital with spin projection $\sigma = \uparrow, \downarrow$. This model and its extensions form the basis for most transport calculations on quantum dots. The energy ε here incorporates the bias and gate voltage induced shifts of the electrochemical potential and other contributions independent of the charge number, see Eq. (10).

Electrons may of course be added to other levels than the lowest unoccupied one, resulting in an excited state of the dot. This is possible if the generalized electrochemical potential

$$\mu_{N_1}^{b,a} = E_b - E_a \tag{10}$$

lies in the bias energy window, where b and a denote any pair of ground or excited states with N_1 and $N_1 - 1$ electrons respectively. This plays however no role in the Coulomb blockade regime where no other state than the Coulomb blockaded ground state is occupied. Figure 5 indicates for a simple model, which new resonance lines appear in the SET regimes and which processes they correspond to.

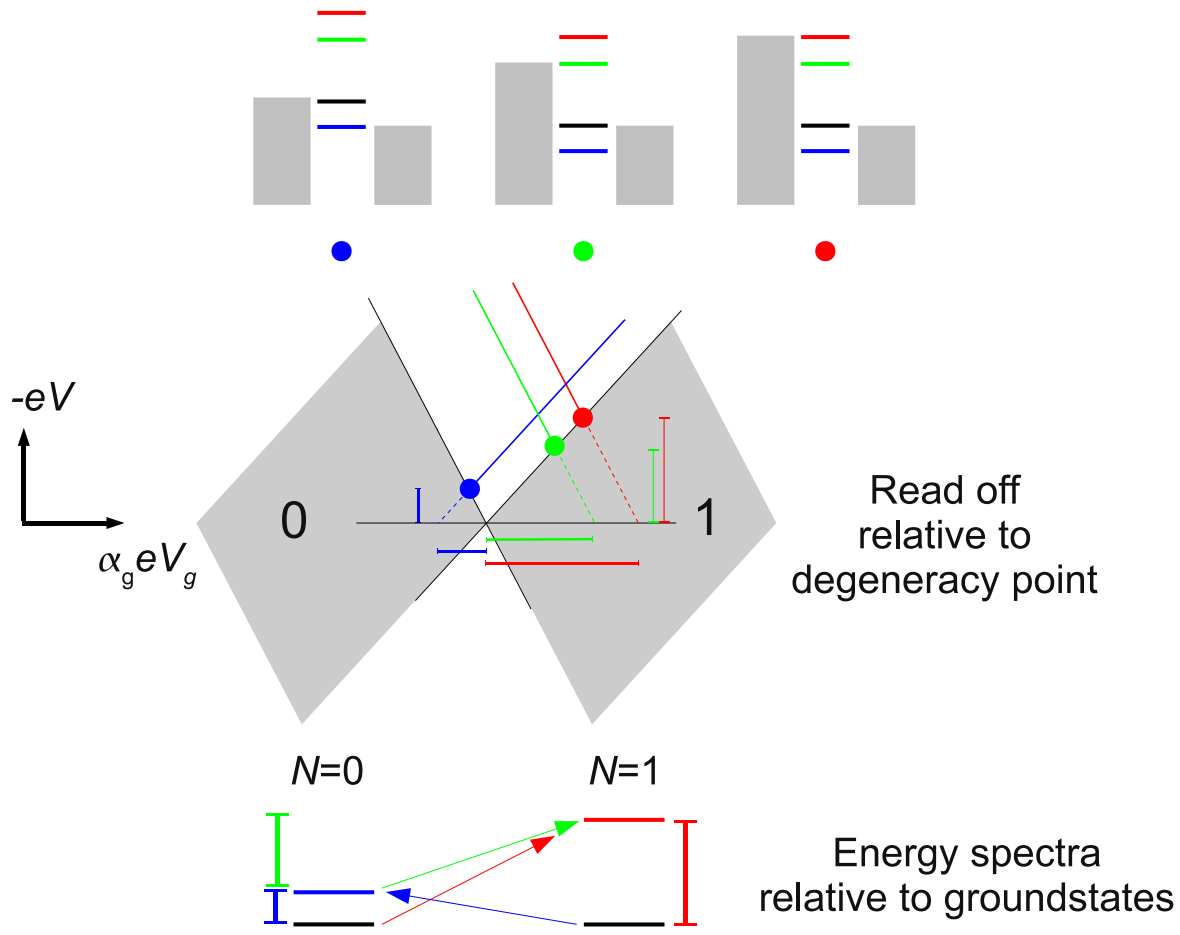


Fig. 5: *Center: Stability diagram in* $(\alpha_g eV_g, -eV)$ *plane for a quantum dot with excited many-electron states. Bottom: energy spectrum for* $N = 0, 1$ *electrons and transition induced by electron addition. Along the full colored lines in the SET region the current changes when the electrochemical potential corresponding to these transitions enters the bias window. The excitation energies, relative to the ground state, can be read off in two ways: (i) either vertically by the position of the colored dots marking the intersection with the SET regime boundary or (ii) horizontally by the position of the lines extrapolated to zero bias voltage (the dashed parts in the* $N = 0$ *or* $N = 1$ *Coulomb blockade regime do not show up in transport). Top: electrochemical potential diagrams for the corresponding colored dots at the SET boundary for method (i). The black "level" marks the electrochemical potential for the ground* $N = 0$ *to ground* $N = 1$*-electron transition (all such transitions for other electron numbers are not shown). The colored lines show the electrochemical potentials for transition involving an* excited state, *which are obtained by shifting the ground-ground level up (red) or down (blue) by each of the excitation energies relative to the ground state, or their difference (green). Note that the green level marks the transition between two excited states, which can only be observed in transport if relaxation mechanisms are weak [43], see [11] for an example of such transitions involving spin.*

3.4 Transport theory - occupations and rates

The aim of this section is twofold. On the one hand, we discuss some of the basic arguments we used above to construct stability diagrams. On the other hand, we indicate the microscopic origin of this picture and the general framework in which quantum transport through quantum dots can be calculated.

The crucial starting point is that the Coulomb interaction effects need to be included non-perturbatively: it cannot be used as an expansion parameter since it is one of the largest energy scales. We thus should treat the tunneling as a small parameter. From the previous sections it is clear that in general transport through a quantum dot is described by a model Hamiltonian containing three terms, $H = H_D + \sum_{r=L,R} H_r + H_T$, which account for the quantum dot, the electrodes and the tunneling between them, respectively. The strong classical electrostatic effects due to capacitive interaction with the electrodes and quantized orbitals we have been discussing so far enter into H_D, as well as further many-electron interactions on the dot (e.g. spin-spin, spin-orbit interactions, etc.), see [41] for a systematic discussion. The metallic electrodes with good screening properties are described by a single particle models H_r, which are fully specified by the density of states ρ_r (which can be considered as constant in nearly all cases of interest). The electrodes remain in equilibrium at their own electrochemical potentials $\mu_r = \epsilon_F - eV_r$ and temperature T. In order to calculate the quantum transport current which can be compared with experimental results, one needs to additionally specify the matrix elements of the microscopic coupling H_T accounting for electron tunneling between the electrodes and the dot.

The calculation of the non-linear current from this well-defined model is in general a non-trivial problem which is still under intense study. Importantly, since the electrodes impose conflicting boundary conditions (different electrochemical potentials) on the microscopic quantum system the occupations of the states are *not* given by a grand-canonical distribution. The density matrix of the dot P is thus in general *not known* and must be calculated from a kinetic equation, which in general takes the form [44]:

$$\frac{dP}{dt}(t) \;=\; -i[H_D, P(t)] + \int_{-\infty}^{t} dt' W(t-t') P(t') \tag{11}$$

The first term on the right hand side derives from the Liouville equation for the density operator P for an isolated quantum dot, $\dot{P}(t) = -i[H_D, P(t)]$ (equivalent to the Schrödinger equation). The second term describes the effect of the electrodes on the quantum dot and includes non-trivial retardation effects through the time integral extending into the past time-evolution. In the long-time limit, $t \to \infty$, where the dot has reached a time-independent, non-equilibrium stationary state $P(t) \to P$ this reduces to

$$\frac{dP}{dt}(\infty) = 0 = -i[H_D, P] + WP \tag{12}$$

where $W = \lim_{\eta\to 0} \int_0^\infty e^{i\eta\tau} W(\tau) d\tau$ includes generalized *transport rate*. This is the zero-frequency component of the memory kernel appearing in the "collision term" on the far right in the kinetic equation (11). From this equation one can find the density matrix P which we need to calculate the expectation value of the transport current.

For comparison with experiments the limit of weak tunnel coupling and non-degenerate quantum dot states is an important starting point. In this limit only the occupations P_a of the many-body quantum dot states labeled by a (diagonal elements of the density matrix P) matter, which

obey a stationary *master equation*:

$$\frac{dP_a}{dt}(\infty) = 0 = -\left(\sum_b W_{ba}\right) P_a + \sum_b W_{ab} P_b \tag{13}$$

Here we sum over quantum dot states b (electron number N_b) with one electron more or less than state a (electron number N_a). In this limit the rates simplify to those obtained by Fermi's golden rule:

$$W_{ba} \;=\; \sum_r \Gamma_r^{ba} \begin{cases} f_r^+(E_b - E_a) & N_b = N_a + 1 \\ f_r^-(E_a - E_b) & N_a = N_b + 1 \end{cases} \tag{14}$$

Here $f_r^{\pm}(x) = 1/(e^{\pm(x-\mu_r)/T} + 1)$ is the Fermi-distribution function for an electron tunneling out / into the electrode r and $f^+ = 1 - f^-$. The rate constant Γ_r^{ba} for electrode $r = L, R$ depends on the density of states: the quantity

$$\hbar\Gamma_r^{ba} = 2\pi \left(T_r^{ba}\right)^2 \rho_r \tag{15}$$

defines characteristic energy scales for quantum charge fluctuations due to tunneling. The matrix elements T_r^{ba} for an electron tunneling onto the dot from electrode r, thereby inducing a transition from state a to b on the dot, completely specify the tunneling part of the model, H_T. The expectation value for the current $I_r(t) = (-e)\langle -\dot{N}_r(t)\rangle$ flowing into the dot from electrode r (electron number N_r) is obtained from the probabilities $P(t)$ through

$$I_r(t) = \mathrm{Tr} \int_{-\infty}^{t} dt' W_{I_r}(t - t') P(t'). \tag{16}$$

Here Tr denotes the trace over the quantum dot eigenstates. In the stationary limit the current follows in a similar fashion from *current rates* W_{I_r}:

$$I_r = \mathrm{Tr} W_{I_r} P. \tag{17}$$

The key insight which Eq. (13) brings is that the stationary occupations follow from a balancing of probability flows. The loss of probability due to transitions from a to all states b (first term) must cancel the gain of probability of state a due to the decay of the latter states (second term). Whenever a positive energy change $E_b - E_a$ falls below the electrochemical potential μ_r due to a change in the bias voltage, the transition rate Eq. (14) increases by an amount Γ_r^{ba} since the tunneling onto the dot becomes energetically allowed. This implies that the probabilities and also the current change: we obtain a non-zero dI/dV and a resonance line in the stability plot. The width of this resonance is set by the thermal energy kT due to the Fermi-function in Eq. (14).[12] One can thus sketch the stability diagrams by considering the available quantum dot states for each electron number and the directions of the processes with non-zero transition rates based on the positions of the chemical potentials (10) relative to the energy bias window

[12] Since the change in the occupations cannot exceed one, the change in the current is on the order of the typical values of $e\Gamma_r^{ba}$ which we loosely denote by $e\Gamma$. The differential conductance is now estimated as $dI/dV = \Delta I/\Delta(\mu_L - \mu_R) \sim e^2\Gamma/(kT)$. From this one easily sees that the charge quantization condition (1), $dI/dV \ll e^2/h$ requires that the life-time broadening due to quantum fluctuations (15) stays much smaller than that due to thermal fluctuations: $\hbar\Gamma \ll kT$. This small ratio is the formal expansion parameter used in perturbation theory, although in special cases higher order effects may complicate matters e.g. due to the Kondo effect.

$\mu_L > \mu_R$. See for instance Fig. 5. Depending on the rate constants one can even in special cases find the qualitative dependence of the occupations and the current on the voltage.
As an explicit example, for the Anderson model the master equation for the probabilities for $N = 0, 1, 2$ electrons reads

$$\frac{d}{dt}\begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix} = \sum_r \Gamma_r \begin{pmatrix} -2f_r^+(\epsilon)P_0 + f_r^-(\epsilon)P_1 \\ 2f_r^+(\epsilon)P_0 - \left(f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)P_1 + 2f_r^-(\epsilon+U)P_2 \\ +f_r^+(\epsilon+U)P_1 - 2f_r^-(\epsilon+U)P_2 \end{pmatrix}. \quad (18)$$

Here $P_1 = \sum_\sigma P_\sigma$ is the occupation of the 1-electron state, irrespective of its spin value $\sigma = \uparrow, \downarrow$ and the probability is conserved, i.e. $P_0 + P_1 + P_2 = 1$. Plugging the solution of these equations into the expression

$$I_r = \sum_r (-e)\Gamma_r \left[2f_r^+(\epsilon)P_0 + \left(-f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)P_1 - 2f_r^-(\epsilon+U)P_2\right] \quad (19)$$

one obtains the current. This produces, for example, the curve labeled $p = 0$ in Fig. 8(a) in the next section.
Finally, we mention that corrections to the above single electron transport picture can be incorporated by calculating non-diagonal density-matrix elements (related to quantum superpositions of quantum dot states), which turn out to be crucial for the single electron spin-valve transistor in Sect. 4. Similarly, higher order tunnel processes can be included, such as for instance "cotunneling" processes involving pairs of electrons using perturbation theory [13] and the Kondo effect [14] using renormalization group methods [15]. Such effects play an important and sometimes even a dominant role in experiments and are key spectroscopic tools for the characterization and control of quantum dots, see Sect. 4.3. However, these issues are beyond the scope of this lecture.

4 Single electron spin-valve

Effects of a spin-polarization of conduction electrons in a tunnel junction have been of interest since the seminal papers of Julliere [45] and Slonczewski [46]. Consider two *ferromagnetic* electrodes connected by a tunnel junction as sketched in Fig. 6(a). The ferromagnets are characterized by a spin-dependent density of states $\rho_{r\sigma}, \sigma = \uparrow, \downarrow$ and a normalized polarization vector $\mathbf{n}_r$ along which the majority spin $\uparrow$ is pointing in electrode r. The *magnetic spin-valve effect* [46] refers to the reduction of the linear tunneling conductance when the polarization vectors $\mathbf{n}_r$ of the two ferromagnetic electrodes $r = L, R$ are at a angle θ defined by $\cos\theta = \mathbf{n}_L \cdot \mathbf{n}_R$:

$$\left.\frac{dI}{dV}\right|_{V=0}(\theta) \propto (1 + p_L p_R \cos\theta) \quad (20)$$

Here the relative spin-polarization of the spin-dependent density of states of each electrode $\rho_{r\sigma}$ is introduced:

$$p_r = \frac{\rho_{r\uparrow} - \rho_{r\downarrow}}{\rho_{r\uparrow} + \rho_{r\downarrow}} \quad (21)$$

i.e. $\rho_{r\sigma} = \frac{1}{2}(1 \pm p_r)\rho_r$ for $\sigma = \uparrow, \downarrow$ where $\rho_r = \sum_{\sigma=\uparrow,\downarrow} \rho_{r\sigma}$ is the total density of states. The angle dependence in Eq. (20) predicted by Slonczewski has been demonstrated experimentally

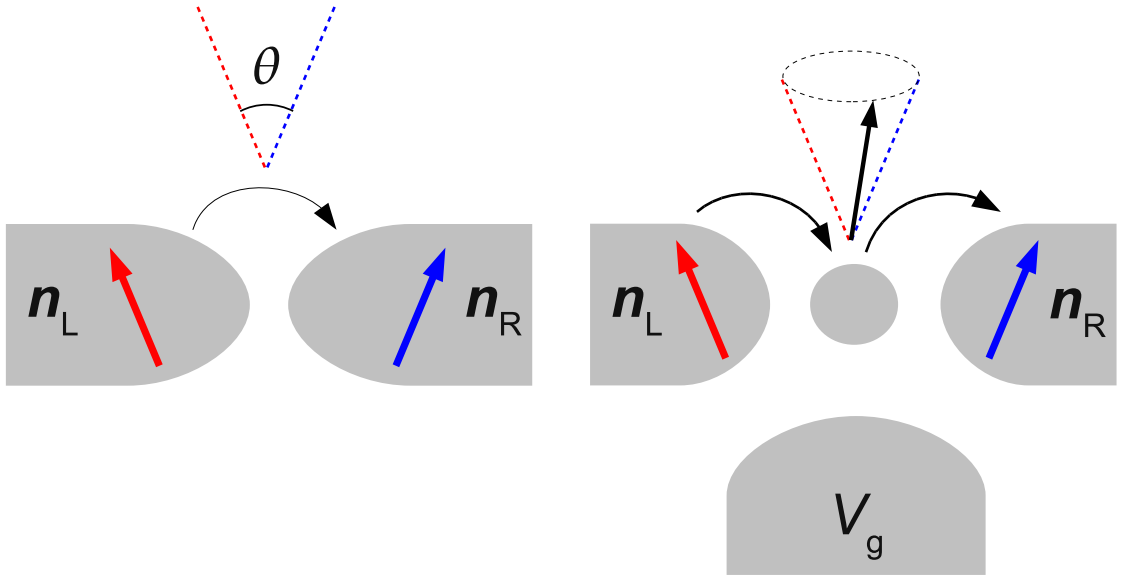


Fig. 6: *(a) Tunnel junction between magnetic electrodes. (b) Quantum dot connected to magnetic electrodes and a normal gate electrode. The electron spin can precess on the dot, even in the absence of an external magnetic field, thereby affecting the transport.*

in detail [47, 48]. It arises from the fact that an electron polarized along e.g. $\mathbf{n}_L$ has spinor component $\cos(\theta/2)$ ($\sin(\theta/2)$) when quantized along $\mathbf{n}_R$ and hence the probability depends on twice the angle i.e. on θ. For perfect ferromagnets, for which $p_r = 1$, the resistance is infinite (finite) for $\theta = \pi$ (0) i.e. we can switch between two resistance states by reversing the magnetization. This *tunnel-magneto resistance* (TMR) effect is based on the quantum mechanical spin degree of freedom and has been rapidly developed into a technologically relevant effect (e.g. MRAM).

A double tunnel junction offers new possibilities due to the role of Coulomb interaction and the tunable level in the quantum dot between the two ferromagnets. This *single-electron spin-valve transistor* is sketched in Fig. 6(b). The basic effects can again be understood from the Anderson model (9) i.e. a single level with Coulomb interaction U, simply generalized to account for a spin-dependent density of states in the electrodes which enter into the tunnel rates through Eq. (15).[13] The *charge* transport is now strongly influenced by the electron *spin* dynamics on the dot, which we now discuss.

4.1 Spin dynamics due to transport

We explain the basic physics of the spin dynamics in a heuristic fashion, arguing that the kinetic equation for the expectation value of the spin of the quantum dot $\mathbf{S}(t) = \langle \mathbf{S}(t) \rangle$ reads

$$\frac{d\mathbf{S}}{dt} = \left(\frac{d\mathbf{S}}{dt}\right)_{\text{acc}} + \left(\frac{d\mathbf{S}}{dt}\right)_{\text{rel}} + \left(\frac{d\mathbf{S}}{dt}\right)_{\text{pre}} \tag{22}$$

This equation can be derived rigorously from the general quantum kinetic equation (11) by accounting for the full density matrix, see the reviews [51, 52] and the original papers [53, 54].

[13] Additional spin effects related to the interface with the ferromagnets can be incorporated as well, see [49, 50].

Equation (22) is in fact a continuity equation for the spin, where the total change in time derives from the injected spin-currents (first 2 terms) and a 3rd precession term. We now discuss the three competing influences in turn.
First, due to transport on average a single electron with a spin $1/2$ will reside on the dot for a finite time. For non-magnetic electrodes the average local spin $\mathbf{S}$ vanishes due to rotational symmetry (contributions with opposite polarization cancel). If at least one electrode is polarized ($p_r \neq 0$) this symmetry is broken and a finite spin value results. This *spin-accumulation* derives from the spin-injection contribution

$$\left(\frac{d\mathbf{S}}{dt}\right)_{\text{acc}} = \sum_r \frac{\hbar}{2}(I_r/e)p_r\mathbf{n}_r \tag{23}$$

Only when both $\mathbf{n}_L = \mathbf{n}_R$ and $p_L = p_R$ does the accumulation term vanish. This follows from Eq. (23) using the current conservation $\sum_r I_r = 0$ in the stationary limit. An extreme example of spin-accumulation occurs for complete ($p_r = 1$) anti-parallel polarizations $\mathbf{n}_L = -\mathbf{n}_R$ where a single electron with spin parallel to the left injecting electrode gets "stuck" on the quantum dot, unable to exit to the right electrode at lower electrochemical potential with opposite polarization.
The transport however, also counteracts the spin accumulation due to the negative *spin-relaxation* term

$$\left(\frac{d\mathbf{S}}{dt}\right)_{\text{rel}} = -\sum_r \Gamma_r/\hbar \left(f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)\left(\mathbf{S} - p_r^2(\mathbf{S}\cdot\mathbf{n}_r)\mathbf{n}_r\right) \tag{24}$$

where $\Gamma_r\left(f_r^-(\epsilon) + f_r^+(\epsilon+U)\right)$ is the Golden Rule rate of tunneling from the one electron state with spin to the zero- and two- electron state with zero spin (singlets). Notably, the relaxation is anisotropic: for a perfect ferromagnet ($p_r = 1$) and an average spin parallel to $\mathbf{n}_r$ one of the relaxation terms vanishes, whereas for opposite average spin this term is maximal.
The third term in Eq. (22) describes a *spin-precession* of the electron spin about a magnetic field as described by the standard Bloch equation[14]

$$\left(\frac{d\mathbf{S}}{dt}\right)_{\text{pre}} = \gamma\mathbf{S}\times\sum_r \mathbf{B}_r. \tag{25}$$

This is surprising since there is no external field applied locally to the quantum dot. In fact, the electron, by virtually tunneling to the magnetic electrodes induces a field on the quantum dot, which acts on its own spin. This local magnetic *exchange field* $\mathbf{B}_r$ induced by the electrode r on the dot reads

$$\gamma\mathbf{B}_r = p_r\mathbf{n}_r\frac{\Gamma_r}{\pi\hbar}P\int_{-\infty}^{\infty} d\omega\left(\frac{f_r^-(\omega)}{\omega-\epsilon} + \frac{f_r^+(\omega)}{\omega-\epsilon-U}\right) \tag{26}$$

where $\gamma = -g\mu_B$ is the negative gyromagnetic ratio determined by the electron g-factor ($-g$) and the Bohr magneton μ_B [15]. The integral is a Cauchy principal value integral i.e.

[14] The Bloch equation follows simply from the Zeeman interaction $H_Z = -\gamma\mathbf{S}\cdot\mathbf{B}$ and the Heisenberg equation of motion: $\dot{\mathbf{S}} = i[H_Z,\mathbf{S}] = \gamma\mathbf{S}\times\mathbf{B}$ using the spin algebra $S_i = i\sum_{jk}\epsilon_{ijk}S_jS_k$.

[15] Note that in many of the original publications the negative gyromagnetic factor γ is absorbed into the exchange field, thereby reversing its direction.

$P\frac{1}{x} = \mathrm{Re}\frac{1}{x+i0}$ with $i0$ denoting an infinitesimal imaginary part. This field is induced by virtual coherent tunnel processes ($\propto \Gamma_r$) to the magnetic electrode r which shift the quantum dot states with opposite spin by a different amount i.e. they split in energy. This splitting, written as a local Zeeman term $g\mu_B \mathbf{S}\cdot\mathbf{B}_r$, defines the exchange field.[16] An important implication is that the exchange field can be very large for strong tunneling Γ_r. Indeed, the exchange field has been detected experimentally in carbon-nanotube [55] and single-molecule quantum dot spin-valves [56], where in the latter case fields on the order of 70 T were reported.

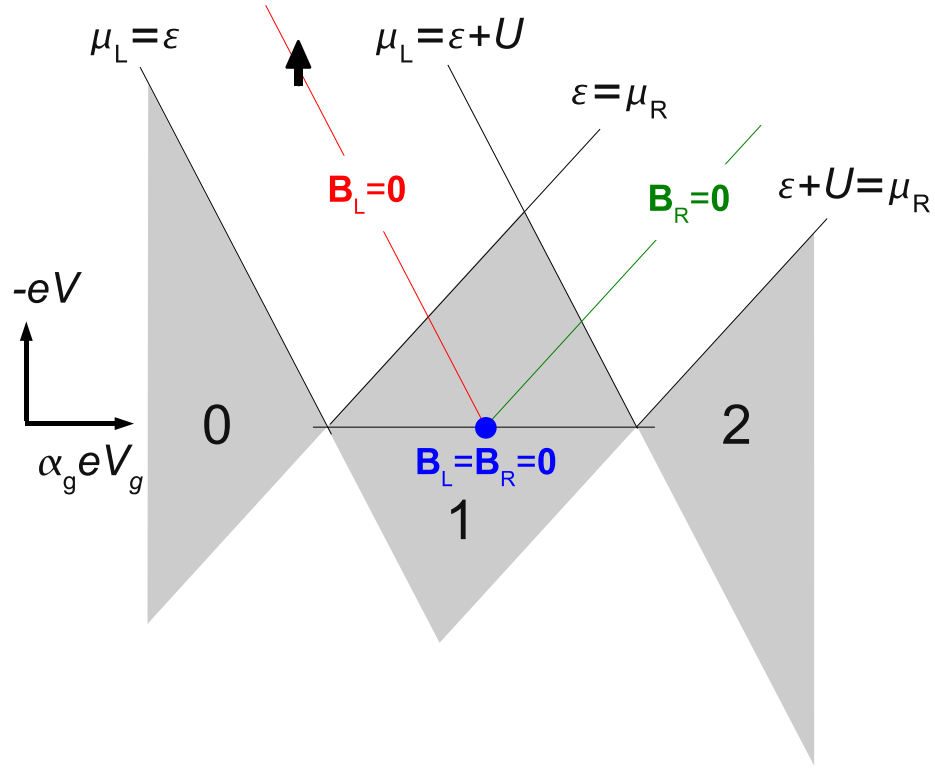


Fig. 7: *Stability diagram for the Anderson model: the boundaries of the finite bias SET regions are defined by the resonance conditions $\epsilon = \mu_r$ and $\epsilon + U = \mu_r$ where $r = L, R$. Midway between these resonances the exchange field induced by the left (right) ferromagnetic electrode vanishes.*

An interesting aspect is that the exchange field is a *Coulomb interaction effect*: it vanishes for $U = 0$ in which case the two terms under the integral cancel exactly (due $f^- + f^+ = 1$ and $P\int d\omega \frac{1}{\omega-\epsilon} = 0$). Here, the first (second) term corresponds to a virtual process where the electron tunnels out of (into) the dot. The finite interaction prevents this destructive interference between these electron and hole processes. The exchange field is also a *many-electron effect* since it involves an integral over the entire spectrum of the electrodes (cut off by the band widths). Furthermore, the exchange field can be *electrically controlled* by the bias (μ_r) and gate (ϵ) voltages, adding a transistor functionality to the device. Due to the interference already mentioned, one of the exchange fields $\mathbf{B}_r$ can even be switched off completely by tuning to

[16] For the simple case where one excludes the $N = 2$ state (i.e. $U = \infty$), the energy shift due to the tunneling in second order perturbation theory reads $E_\sigma - \epsilon = \sum_{k\in\sigma} \frac{t^2}{\epsilon-\omega_k} f_r^-(\omega)$ where the sum extends over the electron states in the band of spin σ. Here $f^- = 1 - f^+$ occurs since the electron fluctuates into unoccupied states in the electrode. Converting the sum into an integral with the density of states $\rho_{r\sigma} = \frac{1}{2}(1 \pm p_r)\rho_r$ and $\Gamma_r = 2\pi\rho_r t^2$, and subtracting the results for spin $\sigma = \uparrow$ and $\downarrow$ one obtains the Zeeman shift leading to the exchange field (26).

the condition $\epsilon - \mu_r = -U/2$ i.e. the electrochemical potential is midway between the two resonances of the Anderson model, $\epsilon = \mu_r$ and $\epsilon + U = \mu_r$. This defines a line in the stability diagram indicated in Fig. 7, along which the two terms in the integral cancel exactly since $\epsilon - \mu_r = -(\epsilon - \mu_r + U)$ and $P\int d\omega f^-(\omega)/(\omega - x) = P\int d\omega f^+(\omega)/(\omega + x)$ for $f^{\pm}(\omega) = (e^{\pm\omega/T} + 1)^{-1}$. The exchange fields of the electrodes can even vanish simultaneously for zero bias $\mu_r = 0$ at the special gate voltage point for which $\epsilon = U/2$ (particle-hole symmetry point). Based on the kinetic equation for the spin (22) we now have the following picture: spin accumulation and relaxation due to transport compete to establish a stationary, non-zero spin value on the dot. The crucial new feature of the spin-valve transistor is that through the exchange fields one can electrically control the *orientation* of this accumulated spin. This one can detect this in the charge transport.

4.2 Charge dynamics and non-linear transport

Having understood the spin dynamics we now turn to the charge dynamics in order to understand the transport current. The occupations of the charge states, $P_N, N = 0, 1, 2$, obey a *generalized* master equation:[17]

$$\frac{d}{dt}\begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix} = \frac{d}{dt}\begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix}_{\text{normal}} + \sum_r p_r\Gamma_r \begin{pmatrix} 2f_r^-(\epsilon) \\ -2f_r^-(\epsilon) + 2f_r^+(\epsilon + U) \\ -2f_r^+(\epsilon + U) \end{pmatrix} \mathbf{n}_r \cdot \mathbf{S}/\hbar. \quad (27)$$

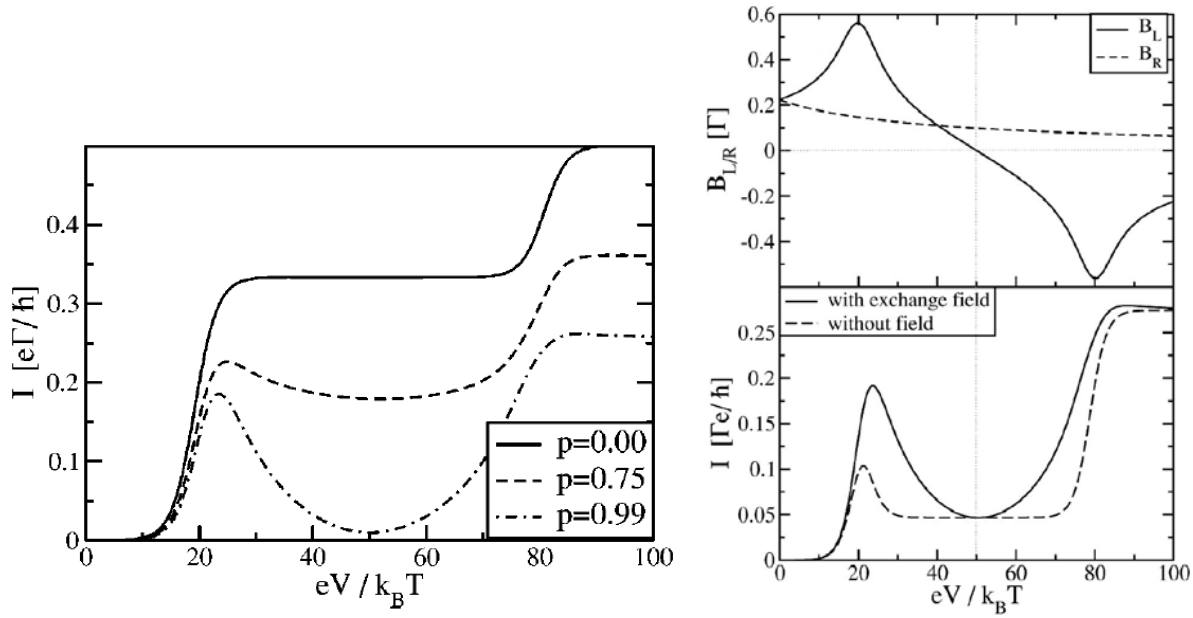
The first term on the right hand side is given by the right hand side of Eq. (18) and does not explicitly depend on the spin. The spin does however, influence it indirectly by changing the occupations. The second term explicitly depends on the spin and arises only for ferromagnetic electrodes: it may cause a gain or loss of probability depending the orientation of spin relative to the polarization vectors. The charge current has a similar structure:

$$I_r/(-e) = (I_r)_{\text{normal}}/(-e) - p_r\Gamma_r\left[f_r^-(\epsilon) + f_r^+(\epsilon + U)\right]\mathbf{n}_r \cdot \mathbf{S}_r/\hbar \quad (28)$$

where $(I_r)_{\text{normal}}$ is the expression for normal electrons on the right hand side of Eq. (19). The second term contains the rate of decay of the one-electron state with non-zero spin.
The exchange field modifies the transport in several ways. For instance, in the linear regime $-eV \ll kT$, it weakens the normal spin-valve effect for a single tunnel junction (see Eq. (20)).[18] Here we focus on the non-linear transport regime where a single electron and therefore a free spin can be present on the dot. In Fig. 8(a) an $I-V$ curve is shown for the case of perpendicular polarizations, i.e. $\theta = \pi/2$. With increasing polarization, the first current plateau is suppressed. In this regime either 0 or 1 electron is present on the dot ($P_2 = 0$) and the current directly is proportional to the occupation P_0. The striking shape of the $I - V$ curve arises due to a kind of *spin blockade* which is counteracted by the precession of the spin around the exchange fields. This is seen by setting $\mathbf{B}_L = \mathbf{B}_R = 0$, in which case the kinetic equation for the spin (22) gives $\mathbf{S} = p\left[(\Gamma_L/\Gamma_R)P_0\mathbf{n}_L - P_1/2\mathbf{n}_R\right]$. One furthermore finds that the dot will be occupied most of the time i.e. $P_1 \approx 1 \gg P_0$. Thus the spin is anti-parallel to the right (drain) electrode,

[17] The second term on the right hand side comes from the non-diagonal elements of the density matrix i.e. Eq. (18) is not a standard master equation which involves only occupancies.

[18] After tunneling through the first junction, the electron spin can precess and thereby adjust itself to the polarization orientation of the drain electrode. This increases the probability of escaping relative to the single junction case.



(a) (b)

Fig. 8: *Results of Braun et. al. [54]. (a) Current-voltage trace for perpendicular polarization directions i.e.* $\theta = \pi/2$*. As the degree of polarization* $p_L = p_R = p$ *is increased there is an overall suppression of the current due to spin blockade, which is counteracted by the precession about the exchange field away from the center of the plateau (*$eV = 50kT$*). The interaction strength is* $U = 30kT \gg \Gamma_L = \Gamma_R = \Gamma/2$ *and* $\epsilon = 10kT$*. (b) Upper panel: magnitude of the exchange fields as function of bias voltage:* $\mathbf{B}_L$ *vanishes at the center of the plateau (*$eV = 50kT$*), while* $\mathbf{B}_R$ *remains finite (which is irrelevant since the spin is parallel to it). Lower panel: comparison of the full result and the result obtain when setting the exchange field to zero "by hand". In the latter case the current uniformly drops to the minimum value due to the exponential dependence the Fermi-functions in the rates, see Eq. (14). In contrast, the much weaker voltage dependence of the exchange field, which lifts the blockade, dominates in the full result. The parameters are the same as in (b) and* $p = 0.95$*.*

$\mathbf{S} \approx -p/2\mathbf{n}_R$. The electron cannot exit due to the low density of minority spin $\downarrow$ in the right (drain) electrode when $p_R \approx 1$. No additional electrons can be transported due to the Coulomb blockade effect and the current is suppressed. This spin blockade would however suppress the current uniformly across the entire plateau as Fig. 8(b) shows. Instead, due to the exchange fields a broad minimum is found. Since the spin blockade causes the spin to be parallel to the exchange field of the right (drain) electrode it is not rotated by it. At the center of the plateau, $\mu_L = V/2 = \epsilon + U/2$, the exchange field of the left (source) electrode vanishes, $\mathbf{B}_L = 0$: only here the spin blockade is thus complete, see Fig. 8(b). However, away from this point the spin will precess around the non-zero $\mathbf{B}_L$ of the source electrode, generating a spin component parallel to the drain polarization $\mathbf{n}_R$ which increases the current.

4.3 Outlook: mapping the exchange field with higher order tunneling

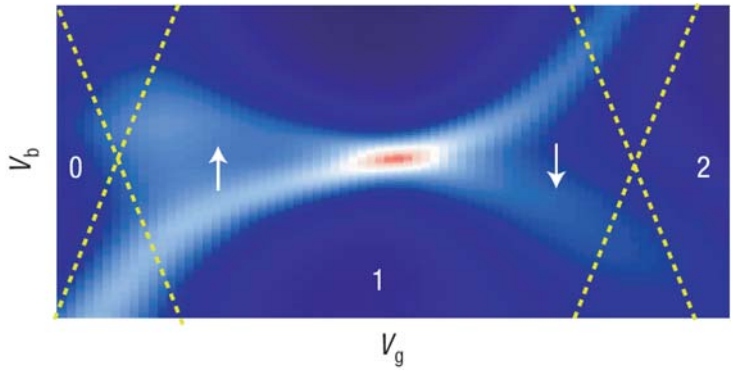


Fig. 9: *Result from Hauptmann et. al. [55]. Experimental stability diagram i.e. differential conductance in color scale as function of applied voltages. The yellow dashed lines indicate the SET regions which are hardly visible here due to the strong tunnel coupling. The white/red color indicates a conductance peak due to higher order tunnel processes. This so-called Kondo peak normally lies at zero bias (i.e. a horizontal line). Due to the exchange field induced by the ferromagnetic electrodes this peak splits and shifts to a finite bias equal to the Zeeman splitting. Clearly, this exchange field goes to zero approximately at the center of the Coulomb diamond, in line with Fig. 7. Away from this point the dependence on the bias follows the same qualitative behavior of $|\mathbf{B}_L|$ as plotted in the upper panel of Fig. 8(b).*

As mentioned above, the exchange field arises due to a splitting induced by virtual tunneling processes. The single electron tunneling discussed so far cannot directly resolve this splitting since thermal fluctuations $\sim kT$ dominate over any effects on the order of the tunnel rates $\hbar\Gamma$, such as the exchange field. However, transport effects involving higher order tunnel processes, such as cotunneling [57] and the Kondo effect [58, 59], are sensitive this splitting. These are beyond the scope of this introduction, see the reviews [51, 52]. As a concluding remark, we merely mention that the exchange field was demonstrated experimentally by measuring the predicted Zeeman shift of the Kondo effect [58, 59] in C_{60} molecular quantum dots [56]. Furthermore, recently [55] this Zeeman splitting was used to directly map out the magnitude of exchange field as function of a level position (gate voltage) in a carbon-nanotube quantum dot contacted with ferromagnetic electrodes, as shown in Fig. 9.

Acknowledgement F. Reckermann, S. Das and M. Baumgärtel are acknowledged for proof reading the manuscript.

References

[1] R. A. Millikan, Phys. Rev. (Series I) **32**(4), 349 (1911).

[2] C. J. Gorter, Physica **15**, 777 (1951).

[3] I. O. Kulik and R. I. Shekhter, Sov. Phys. JETP **4**, 308 (1975).

[4] T. A. Fulton and G. J. Dolan, Phys. Rev. Lett. **59**, 109 (1987).

[5] H. Grabert and M. H. Devoret, eds., *Single Charge Tunneling*, NATO ASI Series (New York, Plenum Press, 1992).

[6] D. K. Ferry and S. M. Goodnick, *Transport in Nanostructures* (Cambridge University Press, 1997), chap. Quantum dots and single electron phenomena.

[7] L. Kouwenhoven, C. Marcus, P. McEuen, S. Tarucha, R. Westervelt, and N. Wingreen, in *Mesoscopic electron transport*, edited by K. Sohn, L. Kouwenhoven, and G. Schön (Kluwer, 1997), chap. Electron Transport in Quantum Dots.

[8] K. K. Likharev, Proc. IEEE **87**, 606 (1999).

[9] L. P. Kouwenhoven, D. G. Austing, and S. Tarucha, Rep. Prog. Phys. **64**, 701 (2001).

[10] W. G. van der Wiel, S. D. Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, and L. P. Kouwenhoven, Rev. Mod. Phys. **75**, 1 (2003).

[11] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. Vandersypen, Rev. Mod. Phys. **79**, 1217 (2007).

[12] D. V. Averin and Y. V. Nazarov, in *Single Charge Tunneling*, edited by H. Grabert and M. H. Devoret (Plenum, New York, 1992).

[13] M. Leijnse and M. R. Wegewijs, Phys. Rev. B **78**, 235424 (2008).

[14] M. Pustilnik and L. Glazman, J. Phys. Condens. Matter **16**, R513 (2004).

[15] H. Schoeller, Eur. Phys. Journ. B (2009), (in press).

[16] T. Fujisawa, D. G. Austing, Y. Tokura, Y. Hirayama, and S. Tarucha, J. Phys. Cond. Mat. **15**, R1395 (2003).

[17] C. Meyer, J. Elzerman, and L. Kouwenhoven, Nano Letters **2**, 295 (2007).

[18] Y. Alhassid, Rev. Mod. Phys. **72**, 895 (2000).

[19] J. A. van Dam, Y. V. Nazarov, E. P. A. M. Bakkers, S. D. Franceschi, and L. P. Kouwenhoven, Nature **442**, 667 (2006).

[20] F. A. Zwanenburg, C. E. W. M. van Rijmenam, Y. Fang, C. M. Lieber, and L. P. Kouwenhoven, *Spin states of the first four holes in a silicon nanowire quantum dot* (2009), arXiv:0811.2914.

[21] S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs, and C. Dekker, Nature **386**, 474 (1997).

[22] M. Biercuk, S. Ilani, C. Marcus, and P. McEuen (2008), vol. 111 of *Topics in Applied Physics*, p. 455.

[23] E. Osorio, T. Bjornholm, J.-M. Lehn, M. Ruben, and H. van der Zant, Jour. Phys. Cond. Mat. **20**, 374121 (2008).

[24] C.W.J. Beenakker and H. van Houten, Solid State Phys. **44**, 1 (1991).

[25] K. Flensberg, S. M. Girvin, M. Jonson, D. R. Penn, and M. D. Stiles, Phys. Script. **T42**, 189 (1992).

[26] H. Grabert, Z. Phys. B **85**, 319 (1991).

[27] G. Schön, *Quantum transport and dissipation* (VCH Verlag, 1997), chap. Single-electron tunneling, p. 105, URL `http: //www.tfp.uni-karlsruhe.de/schoen/setrev4.pdf`.

[28] A. N. C. J. M. Schmidt and J. Clarke, Phys. Rev. Lett. **64**(13), 1565 (1990).

[29] G. Ingold and Y. V. Nazarov, in *Single charge tunneling*, edited by H. Grabert and M. H. Devoret (Plenum Press, New York, 1992), vol. B 294 of *NATO ASI Series*, chap. Charge tunneling rates in ultrasmall junctions.

[30] Yu. V. Nazarov, JETP Lett. **49**, 105 (1989).

[31] M. H. Devoret, D. Esteve, H. Grabert, G.-L. Ingold, H. Pothier, and C. Urbina, Phys. Rev. Lett. **64**, 1824 (1990).

[32] S. M. Girvin, L. I. Glazman, M. Jonson, D. R. Penn, and M. D. Stiles, Phys. Rev. Lett. **64**, 3183 (1990).

[33] S. Bednarek, B. Szafran, and J. Adamowski, Phys. Rev. B **61**, 4461 (2000).

[34] L. D. Hallam, J. Weis, and P. A. Maksym, Phys. Rev. B **53**, 1452 (1996).

[35] E. A. Osorio, K. O'Neill, M. R. Wegewijs, N. Stuhr-Hansen, J. Paaske, T. Bjørnholm, and H. S. van der Zant, Nanolett. **7**, 3336 (2007).

[36] K. Kaasbjerg and K. Flensberg, Nano Lett. **8**, 3809 (2008).

[37] M. Macucci, K. Hess, and G. J. Iafrate, Phys. Rev. B **48**, 17354 (1993).

[38] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1984).

[39] J. D. Jackson, *Classical Electro-dynamics* (Wiley, New York, 1975).

[40] G. Lientschnig, *Simulating Hybrid Circuits of Single-Electron Transistors and Field-Effect Transistors*, Ph.D. thesis, TU Delft (2003), (available from http://www.tudelft.nl).

[41] D. Weinmann, *Quantum transport in nanostructures*, Ph.D. thesis, University of Hamburg (1994), URL `http://www-ipcms.u-strasbg.fr/IMG/pdf/diss.pdf`.

[42] J. von Delft and D. C. Ralph, Phys. Rep. **345**, 61 (2001).

[43] E. A. Osorio, K. O'Neill, N. Stuhr-Hansen, O. F. Nielsen, T. Bjørnholm, and H. S. van der Zant, Adv. Mater. **19**, 281 (2007).

[44] H. Schoeller, *Mesoscopic Electron Transport* (Kluwer, 1997), chap. Transport through interacting quantum dots, p. 291.

[45] M. Julliere, Phys. Lett. A **54**, 225 (1975).

[46] J. C. Slonczewski, Phys. Rev. B **39**, 6995 (1989).

[47] J. S. Moodera and L. R. Kinder, J. Appl. Phys. **79**, 4724 (1996).

[48] H. Jaffrès, D. Lacour, F. Nguyen Van Dau, J. Briatico, F. Petroff, and A. Vaurès, Phys. Rev. B **64**, 064427 (2001).

[49] W. Wetzels, G. E. W. Bauer, and M. Grifoni, Phys. Rev. B **74**, 224406 (2006).

[50] H. Saarikoski, W. Wetzels, and G. E. W. Bauer, Phys. Rev. B **75**, 075313 (2007).

[51] J. König, J. Martinek, J. Barnas, and G. Schön, in *CFN Lectures on Functional Nanostructures*, edited by K. Busch et al. (Springer, 2005), Lecture Notes in Physics, chap. Quantum dots attached to ferromagnetic leads: exchange field, spin precession, and Kondo effect, p. 145.

[52] M. Braun, J. König, and J. Martinek, *Manipulating Single Spins in Quantum Dots Coupled to Ferromagnetic Leads* (Springer, 2006), Lecture Notes in Physics.

[53] J. König and J. Martinek, Phys. Rev. Lett. **90**, 166602 (2003).

[54] M. Braun, J. König, and J. Martinek, PRB **70**, 195345 (2004).

[55] J. R. Hauptmann, J. Paaske, and P. E. Lindelof, Nat. Phys **4**, 373 (2008).

[56] A. N. Pasupathy, R. C. B. and1 J. Martinek, J. E. Grose, L. A. K. Donev, P. L. McEuen, and D. C. Ralph, Science **306**, 86 (2004).

[57] I. Weymann and J. Barnas, Phys. Rev. B **75**, 155308 (2007).

[58] J. Martinek, M. Sindel, L. Borda, J. Barnas, J. Konig, G. Schön, and J. von Delft, Phys. Rev. Lett. **91**, 247202 (2003).

[59] J. Martinek, Y. Utsumi, H. Imamura, J. Barnas, S. Maekawa, J. Konig, and G. Schön, Phys. Rev. Lett. **91**, 127203 (2003).

D 1 Micromagnetism

Riccardo Hertel
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Representation of ferromagnets on different length scales **3**
2.1 Electron theory 4
2.2 Heisenberg Model 5
2.3 Micromagnetic approximation 5
2.4 Multiscale models 7
2.5 Domain theory 7
2.6 Macroscopic modelling 7
2.7 The macrospin model 8

3 Micromagnetic energy terms **8**
3.1 Exchange energy 9
3.2 Magnetostatic energy 10
3.3 Magnetocrystalline anisotropy 11
3.4 Shape anisotropy 12
3.5 Further energy terms 13

4 Equations for the magnetization dynamics **13**
4.1 Magnetization precession 13
4.2 Effective field terms 14
4.3 The Landau-Lifshitz equation 15
4.4 The Gilbert equation 15
4.5 The Landau-Lifshitz-Gilbert equation 16

5 Magnetic domain walls **17**
5.1 Bloch wall 18
5.2 Néel wall 19
5.3 Exchange lengths 21
5.4 Two-dimensional domain walls 21
5.5 Domain walls in thin strips 24

6 Magnetic domains in soft-magnetic thin film elements **25**
6.1 Flux closure domain patterns 25
6.2 The Van den Berg Scheme 26
6.3 High-remanence states 28

7 Magnetization dynamics in nanostructures **29**
7.1 Magnetization reversal of thin-film elements 30
7.2 Magnetic vortex dynamics 31
7.3 Magnetic Vortex Core switching 32

8 Current-induced magnetization dynamics **36**

1 Introduction

Small ferromagnets are essential elements of spintronic devices: The impact of the magnetization state of small ferromagnets on the electric conductivity due the spin-polarization of electrical currents is exploited in spin valves, magnetic tunnel junctions, GMR elements [1, 2] and MRAM cells, to name only a few of examples [3, 4, 5].
Apart from their technical importance due to their functionality in modern spin-based devices, ferromagnetic nano- and microstructures are also interesting from a fundamental point of view because of the non-trivial static and dynamic behavior that the magnetization exhibits in such small confined geometries. Several characteristic magnetic properties of ferromagnets of size below about one micron can in fact display a strong dependence on the particle size and shape. A prominent example is the impact of the nanopatterning on the coercive field of magnets, which loosely corresponds to the switching field, *i.e.*, the value of the external field required to reverse the magnetization. While the coercive field of bulk magnets is mostly determined by the magnetic *material* [1], it becomes crucially dependent on the *shape* of the ferromagnet in the case of small thin-film elements [7]. This difference is particularly striking for soft-magnetic materials, *i.e.*, materials with a low value of the magneto-crystalline anisotropy: in those materials, for which the coercive field of the bulk material is comparatively low (in the order of a few mT), the switching field can become significantly larger in elongated magnetic nanoelements (up to about 100 mT) due to a magnetostatic effect known as *shape anisotropy*, see Section 3.4. Magnetic nanostructures are therefore not just "small magnets" as their physical properties are very different from bulk samples.
Obviously, understanding the reasons for this strong impact of the nanopatterning on the magnetic properties of ferromagnets is of utmost importance for the development of spintronic devices. Ideally, magnetic nanostructures could be fabricated such that they are tailored to specific shapes and sizes in order to display the desired magnetic properties. A precise control of the magnetic properties of nanostructures through the patterning process is however difficult to obtain, since various different metastable magnetic structures can occur in identical samples [8].
The theory of micromagnetism provides the necessary framework to describe the effects on the magnetization induced by the nanopatterning. In fact, in a somewhat simplified definition, micromagnetism could be considered as the theory on the influence of finite-size effects in mesoscopic ferromagnets.
The finite-size effects discussed here result from classical magnetostatics: due to the general tendency of ferromagnets to reduce the energy of the magnetic field that is emanating from them, the magnetization often arranges in alternating magnetic domain patterns or in other, more or less complicated ways inside confined geometries. The magnetostatic effects leading to these arrangements of the magnetization should not be mistaken for finite-size effects as they occur on atomistic length scales, where changes in the electronic structure at surfaces strongly affects the magnetic structure of ultrathin magnetic surfaces or monoatomic chains [9, 10, 11, 12].
A central aspect of the theory of micromagnetism consists in describing the result of competing interactions influencing the magnetic structure. More specifically, this statement means that the impact of the sample size and shape on the magnetic structure is of central importance in micromagnetism. The size dependence arises from the ferromagnetic exchange interaction, which tends to avoid inhomogeneities of the magnetization on a small length scale, while the shape dependence results from the magnetostatic self-energy, which is minimized if the magnetiza-

[1] More specifically, the coercive field depends on the ratio of the magneto-crystalline anisotropy K and the saturation magnetization M_s [6]; two material constants that will be discussed in section 3.

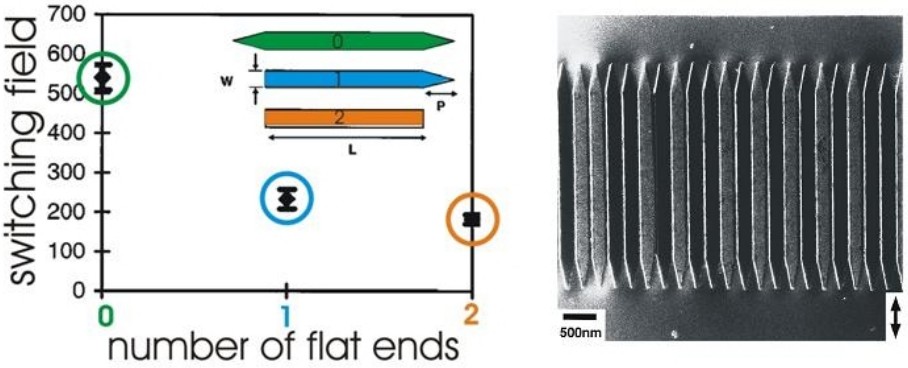


Fig. 1: *The magnetic properties of patterned thin-film elements can depend very sensitively on their precise shape. The diagram on the left displays the measured switching field (in Oersted) for elements with different shape. In the thin-film elements with two tapered ends the switching field is more than twice as large as in the case of rectangular elements with flat ends [7]. The panel on the left displays the array of elements with tapered ends used in the study by Kirk* et. al.

tion aligns with the surfaces and boundaries of the sample. Out of this competition, interesting magnetization structures with particular static and dynamic properties can result which can be analyzed theoretically in the framework of micromagnetism.

Many basic concepts of micromagnetism like the general behavior of the magnetization in samples of special, simple geometries can be understood from analytic considerations. In addition several important fundamental micromagnetic structures like elementary magnetic domain walls can be described analytically. However, for modern applications of micro- and nanomagnetic problems, numerical simulations are of decisive importance. Such simulations are generally required to obtain quantitative predictions for the magnetic properties of nanomagnets and to provide explanations for non-trivial magnetization structures observed in experiments. Owing to the tremendous progress in this field over the past years – in terms of both, numerical techniques and increased computational power – micromagnetic simulations are now generally considered to have a strong predictive power. Perfect agreement between simulated and observed magnetization structures has been demonstrated in several cases (e.g., [13, 14]).

2 Representation of ferromagnets on different length scales

Before introducing the basic equations that constitute the basis of micromagnetism, it is useful to clarify the definition of the most important quantity of the theory: the magnetization. Terms like “spin”, “magnetic moment”, and “magnetization” are usually connected to different length scales on which the magnetic properties are considered. Magnetism on these different length scales is generally described by different theoretical frameworks. Therefore, in order to avoid possible misunderstanding connected with these terms, it is helpful to separate the different theories necessary to describe the magnetization on different length scales.

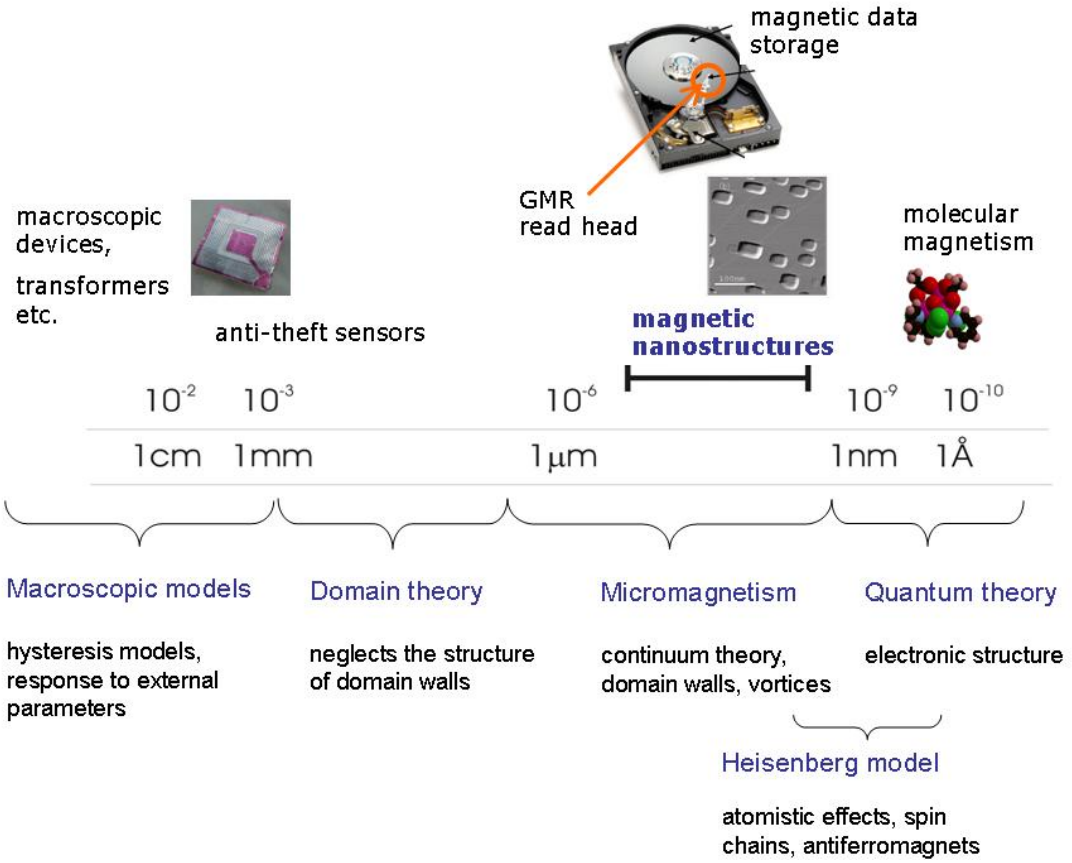


Fig. 2: *Depending on their size, ferromagnetic materials display different phenomena and are used for different applications. The bar on the bottom gives a coarse schematic overview of the different theories involved in the description of magnetism on different length scales.*

2.1 Electron theory

The smallest, most detailed level to study the magnetic properties of solids is their electronic structure: Even if magnetism can display large-scale phenomena connected with huge magnetic fields and macroscopic domain structures, it is ultimately the electron spin that gives rise to these effects. The electron spin can thus be regarded as the fundamental entity of magnetism in solids. This "electronic" level of description is governed by the quantum theory of solids and is a topic of intense research. It is however obviously not possible to describe an entire ferromagnetic particle including its magnetic domain structure on a purely electronic level, even if the element is only a few 100 nm large. In fact, theoretical studies on an electronic level often require several simplifications, such as the approximation of periodic boundary conditions, which states that it is sufficient to consider one elementary cell and construct a magnetic solid by a repetition of such cells. Moreover the dipolar energy is usually neglected, because it is much weaker than the other energy terms involved in these calculations.

Important achievements have been obtained with numerical simulations within the framework of the quantum theory of solids. Density functional theory, *e.g.*, made it possible to obtain material properties based on first-principles calculations [15, 16]. However, it is obviously not possible to capture all relevant aspects of magnetism in solids with only one single theory. For

instance, the approximations mentioned before preclude the consideration of such important aspects of ferromagnets as their magnetic domain structure or the impact of magnetostatic effects. The detailed consideration of such magnetostatic effects is of crucial importance for the correct consideration of the previously mentioned shape dependence of the magnetic properties of elements on the nanoscale. In fact, several magnetic phenomena develop on length scales larger than the ones accessible by electron theory, and their description correspondingly requires different theories.

2.2 Heisenberg Model

The next level of approximation is given by the atomistic theories, *e.g*, the Heisenberg model. In this approximation each atom of the ferromagnet is assumed to carry a magnetic moment, and those magnetic moments are interacting with each other in the lattice of the solid. Note that although the moments originate in the electron spin, here they are ascribed to the lattice sites. The exchange interaction between these moments, J_{ij}, has a quantum-mechanical origin (Pauli principle), but is here assumed to be just a constant prefactor to the multiplication of the magnetic moments $\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j$:

$$\mathcal{H} = \sum_{i \neq j} J_{ij} \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j \tag{1}$$

Extensions of the Heisenberg model can contain further energy terms like the anisotropy and the dipolar interaction. Various approximations are used in the atomistic Heisenberg models in order to address larger length scales: A common approximation used in calculations based on the Heisenberg model consists in considering only nearest-neighbor interactions, since the exchange interaction is short-ranged.
As a result of these simplifications, it becomes possible to access larger length scales, thereby making it possible to calculate complicated magnetic structures on the length scale of several tens of nanometers. The dipolar field is usually not included, since it is generally the weakest of all interactions in a ferromagnet. Due to its long range of interaction, however, it can become the dominating energy term in mesoscopic and macroscopic ferromagnets. The situation can be compared to the effect of inter-particle gravitational effects in the case of classical, macroscopic mechanics: This force can generally be completely neglected in common mechanical problems, but it becomes of utmost importance on an astronomic length scale due to its cumulative effect arising form the long range of the interaction.
The atomistic Heisenberg model can serve to describe spin structures on atomistic level, e.g., in monoatomic chains or on ultrathin magnetic films and surfaces. In many cases, however, the relevant length scales for magnetic structures are much larger than the atomic lattice constants. One of the most fundamental magnetization structures, the magnetic domain wall, typically extends over several tens of nanometers in bulk material. In thin films the domain wall width can increase up to several hundreds of nanometers. For such large problems, an atomic representation is not meaningful. Structures of this size are treated within the framework of micromagnetism.

2.3 Micromagnetic approximation

The transition from the electronic to the atomistic scale can be viewed at as a *coarse graining* procedure, in which the relevant electronic properties are grouped into effective spins and the

addition over all electronic states is replaced by a summation over magnetic moments. In contrast to this, the transition from the atomistic to the micromagnetic approximation is a transition from a discrete to a continuous representation. In this sense it involves a qualitative change of the equations describing the problem.

In the micromagnetic approximation, the microscopic magnetic moments are replaced by an averaged quantity: the magnetization. The magnetizationindexmagnetization is defined as the *density* of magnetic moments. Correspondingly, the summations over dipoles or over magnetic moments occurring in atomistic representations are replaced by volume integrals containing the magnetic moment density, *i.e.*, the magnetization. By means of a Taylor expansion, the Heisenberg exchange interaction is converted into an energy term that contains gradients of the magnetization in the continuum approximation [17].

The fundamental difference between the micromagnetic representation of ferromagnets and atomistic Heisenberg model lies in the fact that the magnetic structure is represented by a *continuous vector field* in the micromagnetic approximation while it is considered as the ensemble of discrete magnetic moments in atomistic models. The vector field of the magnetization is a *directional* field, meaning that at any point in space $\boldsymbol{r}$ it is free to assume any direction (which it can also change in time t), but it preserves everywhere its magnitude, $|\boldsymbol{M}| = M_\mathrm{s}$. The saturation magnetization (sometimes also called spontaneous magnetization) is a material property which is characteristic for ferromagnetic elements. The task of micromagnetic problems consists in calculating the vector field of the magnetization $\boldsymbol{M}(\boldsymbol{r})$ and sometimes also its temporal evolution $\mathrm{d}\boldsymbol{M}/\mathrm{d}t$.

Typical magnetic structures studied in the framework of micromagnetism are magnetic domain walls, magnetic vortices and domain patterns in mesoscopic ferromagnets, but also dynamic effects like spin waves, magnetic normal modes and magnetization reversal processes. On the mesoscopic level addressed by micromagnetic theory, atomistic effects are completely neglected, at least in any explicit form. The impact of the crystal lattice on the magnetic properties is treated only phenomenologically by means of energy terms describing the so-called magneto-crystalline anisotropy. The magneto-crystalline anisotropy accounts for the existence of preferential crystalline axes towards which the magnetization tends to align.

The theory of micromagnetism can provide detailed understanding for a large amount of technologically important and physically interesting phenomena of magnetic nanostructures. When the loss of detailed information on atomistic and electronic level is not decisive, the continuum approximation can be very powerful. The fundamental concepts of micromagnetism were developed by Landau and Lifshitz [18], although the major contribution to the theory was probably given by W. F. Brown Jr.[19, 20]. Many other scientist have provided important contributions to nowadays' understanding of micromagnetic processes. Significant progress was achieved, *e.g.* by F. Bloch [21], L. Néel [22], Ch. Kittel [23] and by the work of Stoner and Wohlfarth [24], to name a few. In recent years, the developments in micromagnetism are strongly influenced by the dramatic progress in experimental resolution in both space [25] and time [26, 27, 28] as well as by the new possibilities given by increasingly powerful computer simulations.

It might be noteworthy that the term "micromagnetism" that was coined when the theory was developed was referring to the importance of very small, *i.e.*, *micro*scopic effects on the magnetic properties of materials. However, the relevant length scale of these microscopic magnetic structures like domain walls or magnetic vortices rather lies on the nanometer range than in the micron range. As far as the length scales are concerned, a more appropriate term for the theory of micromagnetism might therefore be "nanomagnetism", but this term is nowadays mostly used to describe magnetism on an atomistic length scale.

2.4 Multiscale models

In spite of the huge differences between the continuum and the discrete representation, the transition between these models can be rigorously performed in an analytic way without any "trick" or doubtful assumptions. Nevertheless, the difference between the equations describing these different models is important, and it poses a certain problem for the task of *multiscale* modelling. Such kind of "hybrid" modelling is important in some particular cases where very different length scales interact. In exceptional situations it can occur that significant inhomogeneities of the magnetic structure develop on atomistic length scales while the overall magnetic properties are influenced by much larger structures on a mesoscopic length scale. The modelling of such cases requires a combination of both, the continuum representation and atomistic models (and possibly also electronic calculations) in order to describe system correctly. Such situations are important, *e.g.*, in the vicinity of topological defects such as Bloch points, which are also known as micromagnetic singularities or Feldtkeller singularities. A detailed discussion on the difficult task of multiscale modelling of magnetic materials cannot be given in this chapter, but it should be kept in mind that in some cases the theories suitable for different length scales might need to be combined, and that this represents a significant challenge for theory and for computational aspects.

2.5 Domain theory

After the micromagnetic approximation, the next largest scale for the theoretical description of ferromagnets is the domain theory. This approximation is to a certain extent similar to the theory of micromagnetism since it considers that the magnetic structure of a ferromagnet is subdivided into magnetic domains, which are separated by domain walls. However, on the length scale relevant for domain theory, the details of the transition regions in which the magnetization changes its direction (domain walls or vortices) are neglected. The magnetization is assumed to change its direction abruptly between neighboring domains. The transition regions are treated as infinitely thin entities (lines or surfaces). Their contribution to the total energy is considered by assigning an exchange energy density to them, such that the total exchange energy of the sample results as an integration over the domain wall area. Many fundamental works that have lead to the theory of micromagnetism are connected with domain theory. The original work of Landau and Lifshitz, in which the famous Landau pattern was derived [18], also made use of approximations that later became part of domain theory. The results deduced using approximations from domain theory have therefore played a very important role for the development of micromagnetism. But today, especially if compared with (numerical) micromagnetism, domain theory is only of marginal importance for modern problems in magnetism.

2.6 Macroscopic modelling

The last, largest length scale, the macroscopic modelling, is only mentioned for completeness here. It refers to phenomenological constants or functions to represent the response of a ferromagnet to externally applied fields. On this length scale, every microscopic aspect is neglected, and the magnetic material is exclusively described by averaged quantities like its susceptibility or the shape of the magnetic hysteresis. As an example, the task of the Preisach modelling essentially consists in reproducing the shape of the hysteresis curve of ferromagnetic materials by means of distribution functions of so-called *hysterons* (an elementary unit that displays hys-

teretic behavior and that is characterized by the coercivity and an offset value). It is sometimes criticized in these models that the modelling is not based on the averaging over known microscopic effects and that it is not clear how the concept of hysterons is connected to the material properties. Such macroscopic models are mostly of importance for technical applications in engineering sciences, like transformer coils or magnets in electric motors.

2.7 The macrospin model

So far, various theories and the models applicable to different *length* scales have been discussed in this section. For the completeness of this overview a further frequently used approximation shall be mentioned: the macrospin model. In contrast to the previously mentioned models, this approximation or assumption refers to the *dynamics* of the magnetization. It essentially assumes that the magnetization in a small nanomagnet remains homogeneous even after an external perturbation or driving force (like a magnetic field pulse or an electric current pulse). Due to this homogeneity, it is not necessary to consider the distribution of the magnetization inside the sample. Instead, it is sufficient to represent the magnetization state by one single spin, the so-called macrospin. This macrospin model, which will be discussed again in the section on magnetization dynamics, is particularly appealing because of its simplicity.
However, the macrospin model is presumably one of the most critical and doubtful approximations commonly used in modern magnetism. The reason is that it is very difficult – if not impossible – to estimate *a priori* the range of validity of this model. If a small ferromagnet is homogeneously magnetized in zero field, it is of course reasonable to represent its magnetization by a macrospin. But it is not at all obvious that this property of homogeneous magnetization is preserved during dynamic processes occurring on the nano- and picosecond time scales. In fact, several numerical simulations show that significant inhomogeneities of the magnetization can develop temporarily during, *e.g.* magnetic switching processes of elements which in their zero-field equilibrium state are homogeneously magnetized. The macrospin *model* is often called the macrospin *approximation*. This terminology suggests that the macrospin model at least coarsely captures the essential features of the magnetization dynamics of the problems it is applied to. However, if the magnetization temporarily becomes strongly inhomogeneous, the macrospin model may lead to qualitatively wrong results which cannot even be considered as a meaningful approximation. The model can become particularly misleading in cases where it appears possible to fit experimental data to results obtained with macrospin models, even though the magnetization dynamics is completely different. So far, performing accurate micromagnetic simulations is the best way to validate the applicability of the macrospin model, which *a priori* can only be safely assumed to hold for very small particles of only a few nm size.

3 Micromagnetic energy terms

The energy of a ferromagnet depends on its magnetic structure [20]. In a static equilibrium state, the magnetization field $\boldsymbol{M}(\boldsymbol{r})$ arranges such as to minimize the total energy. This minimum can either be a local or a global minimum. If several different minima can be achieved, the magnetic history of the sample is decisive for the selection of the equilibrium configuration. One possibility to calculate such equilibrium states consists in first establishing an energy functional which provides the energy for any given magnetization structure $\boldsymbol{M}(\boldsymbol{r})$. The energy can then be minimized by means of a variational calculation, yielding the equilibrium structure of the

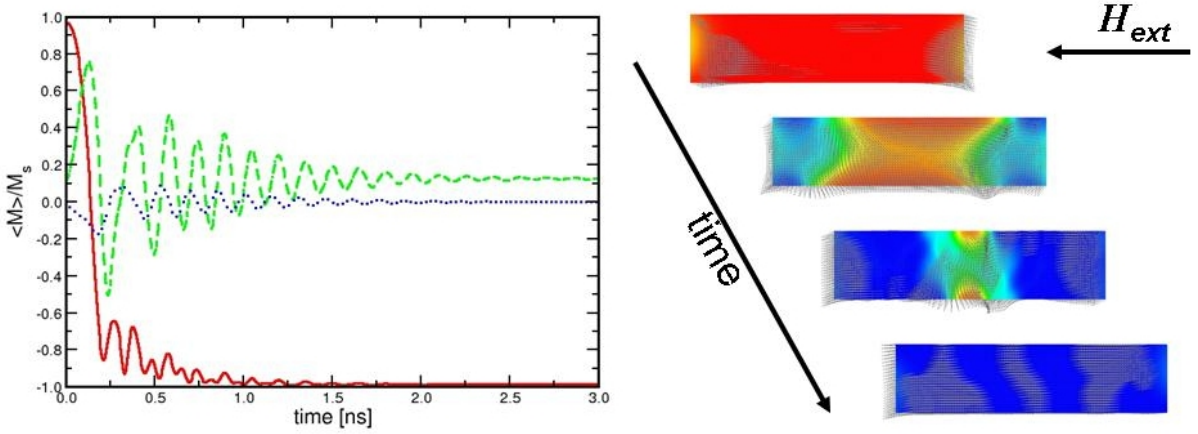


Fig. 3: *Simulated magnetization reversal of a permalloy platelet (500 nm × 125 nm × 3 nm) in an externally applied field H_{ext}. The simulated case is a micromagnetic standard problem used to compare results of different groups [29]. Frames at different time steps of the reversal are shown on the right; the diagram on the left displays the average magnetization components m_x, m_y, and m_z as a function of time t after application of the field. It can be seen that the magnetization oscillates around the new equilibrium position for about 2 ns after the switching (magnetic ringing). Even though the platelet is a magnetic single-domain particle, strong magnetic inhomogeneities develop during the switching process.*

magnetization. Alternatively, one can use the dynamic equation of motion of the magnetization (the Landau-Lifshitz-Gilbert equation, discussed in section 4.3) and calculate the converged solution as the limiting case at which the magnetization does not change further in time. Let us first consider the energy terms:
In micromagnetic problems, the most important contributions to the total energy are usually the ferromagnetic exchange energy, the dipolar or magnetostatic energy and the magneto-crystalline anisotropy energy and the Zeeman energy in an external magnetic field. These energy terms will be briefly described in the next paragraphs. A more detailed discussion of the micromagnetic energy terms can be found in various textbooks on micromagnetism [20, 30, 17, 6, 31].

3.1 Exchange energy

The characteristic property of ferromagnetic materials is their tendency to keep neighboring magnetic moments parallel to each other. The short-range exchange interaction prevents strong inhomogeneities of the magnetization on small length scales. In other words, any increase of inhomogeneity of the magnetization field increases the exchange energy. The simplest and in most cases perfectly sufficient representation of the exchange energy density is given by

$$e_{\mathrm{exc}} = \sum_{i=x,y,z} A \cdot (\boldsymbol{\nabla} m_i)^2 \quad , \tag{2}$$

where A is the exchange constant and $\boldsymbol{m} = \boldsymbol{M}/M_{\mathrm{s}}$ is the reduced or normalized magnetization. This expression can also be derived from a Taylor expansion of the Heisenberg term assuming small-angle deviations between neighboring moments [17, 6].

3.2 Magnetostatic energy

Each magnetic moment in a ferromagnetic sample represents a magnetic dipole and therefore contributes to a total magnetic field $\boldsymbol{H}_\mathrm{s}$ inside the sample. The energy connected with this field is known as the stray field energy or the magnetostatic energy. The local stray field energy density depends on the orientation of the magnetic moments with respect to this field,

$$e_\mathrm{st} = -\frac{\mu_0}{2}\boldsymbol{H}_\mathrm{s}\cdot\boldsymbol{M} \quad . \tag{3}$$

In the literature, different terms are used for the field $\boldsymbol{H}_\mathrm{s}$. It is either called the magnetic stray field, the dipolar field, the demagnetizing field or the magnetostatic field. The factor $1/2$ in Eq. (3) is required for self-energy terms. This energy contribution arises from the long-range magnetostatic interaction between the magnetic moments in the sample. .
The stray field $\boldsymbol{H}_\mathrm{s}(\boldsymbol{r})$ at one point $\boldsymbol{r}$ results from a summation of the form

$$\boldsymbol{H}_\mathrm{s}(\boldsymbol{r}) = \frac{1}{4\pi}\left(\sum_i \frac{\boldsymbol{\mu}_i}{|\boldsymbol{r}-\boldsymbol{r}_i|^3} - 3\cdot\frac{(\boldsymbol{\mu}_i\cdot\boldsymbol{r})\cdot\boldsymbol{r}}{|\boldsymbol{r}-\boldsymbol{r}_i|^5}\right) \quad , \tag{4}$$

which extends over all the magnetic moments inside the sample.
The magnetic dipole moments $\boldsymbol{\mu}_i$ can be assumed to be localized on an atomic length scale. It is not practicable to perform a calculation of the stray field in mesoscopic particles by means of the summation over these point dipoles. In the framework of micromagnetism, such dipolar sums are therefore not considered. The micromagnetic calculation of demagnetizing fields involves a transition from the discrete sum (4) to a continuum integration. By converting the sum (4) over the individual magnetic dipoles into an integral over the sample volume [20], the stray field can be obtained from

$$\boldsymbol{H}_\mathrm{s}(\boldsymbol{r}) = -\frac{1}{4\pi}\int \frac{(\boldsymbol{r}-\boldsymbol{r'})\rho(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|^3}\,\mathrm{d}V' + \frac{1}{4\pi}\oint \frac{(\boldsymbol{r}-\boldsymbol{r'})\sigma(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|^3}\,\mathrm{d}S' \,, \tag{5}$$

where $\boldsymbol{n}$ is the surface normal vector. The magnetic volume charges $\rho = -\boldsymbol{\nabla}\cdot\boldsymbol{M}$ and the surface charges $\sigma = \boldsymbol{M}\cdot\boldsymbol{n}$ are the sources of the demagnetizing field. It can be shown that the total stray field energy, which results from a volume integration of Eq. (3), is always positive definite [19, 17]. It is therefore not possible to arrange the magnetization in a way so that the stray field lowers the total energy. In order to minimize the stray field energy of a ferromagnet, the sources of the field $\boldsymbol{H}_\mathrm{s}$ have to be minimized. This is known as Brown's pole avoidance principle [19]
It is easy to see that the field $\boldsymbol{H}$ calculated according to equation (5) can be represented as a gradient field of a scalar potential, $\boldsymbol{H} = -\boldsymbol{\nabla}U$ by comparing the right-hand side with the general solution of the Poisson equation

$$\Delta U(\boldsymbol{r}) = -\rho(\boldsymbol{r}) = \boldsymbol{\nabla}\boldsymbol{M} \quad . \tag{6}$$

It follows from this that the field $\boldsymbol{H}$ is *irrotational*. In many cases it is more convenient to solve Poisson's equation instead of performing the integration according to equation (5).
Another way of obtaining the result that the magnetic field $\boldsymbol{H}$ can be identified as the gradient field of the potential U that solves the Poisson equation is the following: The connection between the magnetization $\boldsymbol{M}$, the magnetic field $\boldsymbol{H}$ and the magnetic induction $\boldsymbol{B}$ is

$$\boldsymbol{B} = \mu_0\boldsymbol{H} + \mu_0\boldsymbol{M} \tag{7}$$

This fundamental equation can be interpreted according to the Helmholtz theorem, which states that any sufficiently smooth vector field (here: $\mu_0 \boldsymbol{M}$) can be decomposed into an irrotational ($\mu_0 \boldsymbol{H}$) and a solenoidal vector field ($\boldsymbol{B}$). Since $\boldsymbol{B}$ is solenoidal, and $\boldsymbol{M}$ is a general vector field, it follows that $\boldsymbol{H}$ is irrotational, *i.e.*, that it can be represented as a gradient field. It thus follows directly from Eq. 7 that

$$\boldsymbol{\nabla} \boldsymbol{H} = -\boldsymbol{\nabla} \boldsymbol{M} \quad . \tag{8}$$

with $\boldsymbol{H} = -\boldsymbol{\nabla} U$ it follows that U satisfies Poisson's equation:

$$\Delta U = \boldsymbol{\nabla} \boldsymbol{M} \tag{9}$$

For the unique solution of Poisson's equation it is obviously necessary to have knowledge of the boundary conditions of U. The application of the boundary conditions can be problematic in practical cases, since this is an "open boundary" problem, meaning that the boundary conditions for the potential U and its gradients are defined at infinity,

$$\lim_{x \to \pm\infty} U(x) = 0 \quad . \tag{10}$$

Several numerical algorithms have been developed to solve such open boundary problems efficiently.

3.3 Magnetocrystalline anisotropy

The magnetocrystalline anisotropy energy results from the crystalline structure of a ferromagnet. According to the crystal symmetry, the direction of the magnetization favors energetically an alignment towards certain axes. In the simplest case of a uniaxial magnetocrystalline anisotropy, the energy density connected with this term is to first non-vanishing order given by

$$e_{\mathrm{an}} = K_{\mathrm{u}} \left[1 - (\boldsymbol{m} \cdot \boldsymbol{k})^2 \right] \quad , \tag{11}$$

where K_{u} is the uniaxial anisotropy constant and $\boldsymbol{k}$ is a unit vector parallel to the easy axis. The easy axis represents the preferential orientation of the magnetization in the crystal. In some cases, the anisotropy energy is including a higher-order term, the second anisotropy constant K_{u2},

$$e_{\mathrm{an}} = K_{\mathrm{u1}} \left[1 - (\boldsymbol{m} \cdot \boldsymbol{k})^2 \right] + K_{\mathrm{u2}} \left[1 - (\boldsymbol{m} \cdot \boldsymbol{k})^2 \right]^2 \tag{12}$$

Usually it is sufficient to consider only the first-order term.
In cubic crystals, the anisotropy energy density is described by two cubic anisotropy constants, K_{c1} and K_{c2}. Unlike the uniaxial anisotropy, the constant of the second non-vanishing anisotropy term is not a higher-order contribution. The coefficients K_{c1} and K_{c2} refer to terms which are generally of similar importance.

$$K_{\mathrm{c}} = K_{c1} \left(m_x^2 m_y^2 + m_x^2 m_z^2 + m_y^2 m_z^2 \right) + K_{c2} m_x^2 m_y^2 m_z^2 \tag{13}$$

Ferromagnetic materials can loosely be categorized as magnetically soft or hard, depending on the strength of their magneto-crystalline anisotropy. A material with a huge anisotropy is "hard" in the sense that it changes its magnetization state only under the influence of a strong external field, whereas the magnetization state of a "soft" material is easily affected by external

Fig. 4: *The shape anisotropy of a homogeneously magnetized, elongated particle is a magnetostatic effect: An alignment of the magnetization with the long axis (sketch on the right) minimizes the magnetostatic surface charges. The configuration therefore has a lower magnetostatic energy than a transversal orientation (left), thereby giving rise to a shape-dependent preferential direction ("easy axis") of the magnetization.*

fields. For uniaxial materials, the *quality factor* Q can be introduced as a useful measure for the magnetic hardness. It is defined as

$$Q = \frac{K_u}{\frac{1}{2}\mu_0 M_\mathrm{s}^2} \tag{14}$$

and thus represents the ratio of the uniaxial anisotropy to the maximum value of the "shape anisotropy" $K_d = \mu_0 M_s^2/2$ (assumed in the case of a thin film) that this material can have. If $Q \ll 1$, the magnetostatic fields prevail over the anisotropy, whereas for $Q \simeq 1$ or larger the anisotropy of the material is of significant importance. Correspondingly, the terms magnetically "hard" and "soft" can be assigned to materials with $Q \geq 1$ and $Q \ll 1$, respectively.

3.4 Shape anisotropy

The term "shape anisotropy" used before describes how the particle shape can give rise to preferred directions of the magnetization orientation. The effect of shape anisotropy can be similar to that of the magnetocrystalline anisotropy. For example, an elongated ferromagnetic particle with negligible magnetocrystalline anisotropy can have properties similar to those of a spherical particle with uniaxial magnetocrystalline anisotropy. However, the physical origin is completely different and the term shape anisotropy can therefore be misleading. In fact, the effect of shape anisotropy is already fully included in magnetostatics. The impact of the shape is due to a reduction of the magnetic surface charges $\sigma = \boldsymbol{m} \cdot \boldsymbol{n}$, in order to minimize the stray field energy. The magnetostatic energy of a *homogeneously magnetized* ellipsoidal particle with axes $a > b > c$ is smallest when the magnetization aligns with the a axis. A quantitative description of the shape anisotropy requires the calculation of demagnetization factors [32]. Strictly speaking, the concept of shape anisotropy is only applicable to uniformly magnetized ellipsoidal particles. It is nevertheless often used in a more loose context to describe the general tendency of the magnetization to adapt to the sample shape, *e.g.*, to align in the film plane in the case of thin films, even if their magnetic structure is inhomogeneous. It should however be kept in mind that the concept of shape anisotropy is merely a simplified description of magnetostatic effects. The shape anisotropy is therefore not one of the other, fundamental energy terms described in this section.

3.5 Further energy terms

If a ferromagnet is exposed to an externally applied magnetic field $\boldsymbol{H}_{\mathrm{ext}}$, its energy also depends on the orientation of the magnetization with respect to the field. The Zeeman energy of the system is then

$$e_{\mathrm{ext}} = -\mu_0 \boldsymbol{H}_{\mathrm{ext}} \cdot \boldsymbol{M} \quad . \tag{15}$$

Further energy terms that may be relevant in certain cases are the surface anisotropy and the energy connected with magnetoelastic effects. These energy terms are discussed in detail elsewhere [20, 30, 17, 6]. Magnetoelastic effects can be ignored in most practical cases. The consideration of the latter would require the solution of additional differential equations, thus remarkably complicating the overall calculation.

4 Equations for the magnetization dynamics

The energy terms described in the previous section can be applied to any given magnetization structure. In the micromagnetic approximation, the total energy of the ferromagnetic system is thus uniquely defined by the magnetization state. Due to the tendency of the magnetic system to reach a minimum-energy state, the equilibrium structure of the magnetization can be calculated from a variational calculus in which the total energy is minimized[2]. It is thus possible to determine *static* magnetization structures which represent an equilibrium configuration by means of energy minimization. Energy considerations however do not provide information on the temporal evolution of a system. If, *e.g.*, a short perturbation like a magnetic field pulse is applied to a nanomagnet, the magnetization will respond to this perturbation and will undergo a more or less complex transformation before finding again an equilibrium state. These dynamic processes typically occurs on the time scale of pico- and nanosecond. From the viewpoint of technology it is important to gain detailed knowledge about the dynamics of the magnetization since it plays an important role in many applications, like the writing speed in magnetic data storage. The fundamental equation to describe the magnetization dynamics is the Landau-Lifshitz-Gilbert (LLG) equation, which will be discussed in this section.

4.1 Magnetization precession

A magnetic field can have different effects on a magnetic dipole: While an inhomogeneous magnetic field exerts a *force* on a dipole, a homogeneous magnetic field exerts a *torque* on it. For the internal dynamics of the magnetization, only the torque term is decisive. In a field $\boldsymbol{H}$ a magnetic moment $\boldsymbol{\mu}$ experiences a torque $\boldsymbol{L}$ according to

$$\boldsymbol{T} = \boldsymbol{\mu} \times \boldsymbol{H} \tag{16}$$

By definition, the torque $\boldsymbol{T}$ is the change of angular momentum $\boldsymbol{L}$ per unit time, $\mathrm{d}\boldsymbol{L}/\mathrm{d}t = \boldsymbol{T}$. The angular momentum $\boldsymbol{L}$, in turn, is connected to the magnetic moment by the *gyromagnetic ratio* γ_0:

$$\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}t} = \gamma_0 \boldsymbol{\mu} \times \boldsymbol{H} \tag{17}$$

[2] More strictly, the minimum of the free energy represents the equilibrium state, but entropy contributions can usually safely be neglected.

with $\gamma_0 = -g|e|\mu_0/2m_e$, where μ_0 is the vacuum permeability, m_e is the electron mass and g is the Landé factor.
Equation (16) describes the precessional motion of the magnetization $\boldsymbol{M}$ in an external field

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = -\gamma_0 \boldsymbol{M} \times \boldsymbol{H} \tag{18}$$

It is easy to see that this equation preserves the magnitude of the magnetization $|\boldsymbol{M}|$.
Note that this fundamental equation describing the precession of the magnetization in an external field can also be derived directly from the quantum mechanical expression of the dynamics of the expectation value of a magnetic moment $\langle \boldsymbol{\mu} \rangle$ in a magnetic field $\boldsymbol{H}$.

4.2 Effective field terms

The magnetization precession according to Eq. (18) describes the dynamic effect of an external field $\boldsymbol{H}$ on the magnetization $\boldsymbol{M}$ However, an external field $\boldsymbol{H}$ is not the only possible source of a torque on the magnetization in a ferromagnet. Interaction with the local magnetization due to, *e.g.*, the ferromagnetic exchange, the dipolar field or the magnetocrystalline anisotropy can also give rise to torques on the local magnetization. These internal effects are considered by introducing an *effective field* $\boldsymbol{H}_{\mathrm{eff}}$. Each micromagnetic energy term is connected with an effective field contribution $\boldsymbol{H}_{\mathrm{eff}}$ according to

$$\mu_0 \boldsymbol{H}_{\mathrm{eff}} = -\frac{1}{M_\mathrm{s}} \frac{\partial e}{\partial \boldsymbol{m}} \quad , \tag{19}$$

with $\boldsymbol{m} = \boldsymbol{M}/M_\mathrm{s}$. The effective field $\boldsymbol{H}_{\mathrm{eff}}$ is defined as the negative variational derivative of the micromagnetic energy density e with respect to the magnetization [3].
The usual form of the effective field in micromagnetics is

$$\boldsymbol{H}_{\mathrm{eff}} = \frac{2A}{\mu_0 M_\mathrm{s}} \Delta \boldsymbol{m} + \boldsymbol{H}_{\mathrm{ext}} + \boldsymbol{H}_{\mathrm{stray}} - \frac{1}{\mu_0 M_\mathrm{s}} \frac{\delta e_\mathrm{K}}{\delta \boldsymbol{m}} \tag{20}$$

The effective field contains all effects from external and internal fields or energy contributions. It is generally a complicated function of the magnetization distribution in the sample, such that it can only be calculated numerically. A formal derivation of the effective field can be found, *e.g.* in the textbooks of W. F. Brown Jr. [19, 20].
An equation for the precessional motion of the local magnetization around the local effective field is obtained by simply replacing the external field $\boldsymbol{H}_{\mathrm{ext}}$ with the effective field $\boldsymbol{H}_{\mathrm{eff}}$ in Eq. (18). Note that only the component of the effective field perpendicular to $\boldsymbol{M}$ has an influence on the magnetization dynamics. The effective field is thus not uniquely defined, since any additional component parallel to $\boldsymbol{M}$ would not affect the magnetization dynamics.
Interestingly, in spite of the previous discussion according to which the micromagnetic energy terms are primarily relevant for determining static equilibrium structures, they are evidently also important for the magnetization dynamics. More precisely, the *derivatives* of the energy terms with respect to the magnetization are the effective fields and therefore play a crucial role for the dynamics.

[3] This definition of the effective field can be compared with a similar definition used in mechanical systems, where the local force density can be obtained as the negative gradient of the energy density.

4.3 The Landau-Lifshitz equation

Obviously equation (18) cannot represent a complete description of the magnetization dynamics. In this form, a dynamic process of the magnetization once initiated would last forever: The equation does not contain any dissipation ("friction") term. This is an unrealistic situation which contradicts the observation according to which the magnetization of a ferromagnet can rapidly align with an external field and equilibrate within a few nanoseconds. In order to account for these dissipation effects, Landau and Lifshitz introduced a phenomenological dissipation term to the precessional motion. For a realistic description a dissipation term should modify the equation of motion of the magnetization in a way to fulfil the following requirements:

1. In the limit of long time, the magnetization should align with the effective field $\boldsymbol{H}_{\mathrm{eff}}$. When this is reached, $\mathrm{d}\boldsymbol{M}/\mathrm{d}t$ is equal to zero and the magnetization comes to a halt.

2. The magnitude of the magnetization $M_{\mathrm{s}} = |\boldsymbol{M}|$ should be preserved at any time. This is a basic requirement of the micromagnetic approximation.

These conditions are realized by a damping term of the form

$$\left.\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t}\right|_{\mathrm{damping}} = -\frac{\eta}{M_{\mathrm{s}}}\left[\boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{H}_{\mathrm{eff}})\right] \tag{21}$$

where η is the phenomenological damping constant introduced by Landau and Lifshitz. The famous Landau-Lifshitz equation describing the magnetization dynamics is thus

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = \gamma_{\mathrm{L}}\left[\boldsymbol{M} \times \boldsymbol{H}_{\mathrm{eff}}\right] - \frac{\eta}{M_{\mathrm{s}}}\left[\boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{H}_{\mathrm{eff}})\right] \quad . \tag{22}$$

4.4 The Gilbert equation

A different way of obtaining an equation that describes the necessary combination of precession and damping while preserving the magnitude $|M_{\mathrm{s}}|$ has been derived by Gilbert. The Gilbert equation has been presented much later than the Landau-Lifshitz equation, namely in 1950. Nevertheless, it has attracted much attention and it is often considered as the more fundamental of these two equations describing the magnetization dynamics.
The Gilbert equation is given in an *implicit* form as the change of magnetization in time $\mathrm{d}\boldsymbol{M}/\mathrm{d}t$ appears on either side of the equation:

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = -\gamma\left(\boldsymbol{M} \times \boldsymbol{H}_{\mathrm{eff}}\right) + \frac{\alpha}{M_{\mathrm{s}}}\left(\boldsymbol{M} \times \frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t}\right) \tag{23}$$

here α is a phenomenological damping constant, known as the Gilbert damping constant. Note that the Gilbert damping constant α is *not* identical with the Landau-Lifshitz damping constant η. The dimensionless Gilbert damping α is the only value used to describe the damping in the recent literature. The Landau-Lifshitz damping constant can be found in older publications.
Physically, the difference between the Gilbert damping and the Landau-Lifshitz damping lies in the nature of the damping. The Gilbert equation describes the commonly applicable situation of *viscous* damping: the damping increases with increasing "magnetic speed" $|\mathrm{d}\boldsymbol{M}/\mathrm{d}t|$, regardless of the character of the motion. In the Landau-Lifshitz equation, however, only the relaxation term is damped. The precession term is not affected by the damping - a situation that appears

to describe an unphysical property since any motion of the magnetization should be subject to damping – regardless of whether it describes a precession or relaxation motion. Due to these and other formal aspects of the interpretation of the equations of motion, the Gilbert equation is often considered as the more fundamental one.

As will be discussed in the following section, there is no need for a distinction between the Gilbert and Landau-Lifshitz equation in their original form. However, it has recently been found that additional torque terms should be added to describe the dynamics of the magnetization induced by spin-polarized electric currents. Those spin-torque terms need to be treated separately, *i.e.*, they cannot be included into an effective field term. In these cases it can make an important difference whether the additional torque terms are added on the right-hand side of the Gilbert equation or of the Landau-Lifshitz equation. While the correct implementation of these terms is partly still a subject of debate, most authors prefer including these additional terms into the Gilbert equation instead of the Landau-Lifshitz equation, for the aforementioned reasons.

4.5 The Landau-Lifshitz-Gilbert equation

In spite of its apparent differences, the Gilbert equation is just a different representation of the Landau-Lifshitz equation. The Landau-Lifshitz and the Gilbert equation can be identically converted into each other, so that a distinction is unnecessary from a mathematical point of view. It can in fact be shown rather easily that if the Gilbert equation is rewritten into an explicit form, it assumes precisely the form of the Landau-Lifshitz equation, with the damping given by a term of the form $[\boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{H})]$. In this explicit form, the equation is known as the *Landau-Lifshitz-Gilbert* (LLG) equation

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = -\frac{\gamma}{1+\alpha^2}\left(\boldsymbol{M} \times \boldsymbol{H}_{\mathrm{eff}}\right) - \frac{\alpha\gamma}{M_s(1+\alpha^2)}\left[\boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{H}_{\mathrm{eff}})\right] \quad . \tag{24}$$

The connection between the Gilbert equation and the Landau-Lifshitz equation is given by two equations relating the damping coefficients and the prefactors of the precession terms in these two equations:

$$\gamma_{\mathrm{L}} = \frac{\gamma_{\mathrm{G}}}{1+\alpha^2}, \qquad \eta = \frac{\alpha\gamma_{\mathrm{G}}}{(1+\alpha^2)M_{\mathrm{s}}} \tag{25}$$

As already mentioned, practically only the Gilbert damping α is used in the literature.

Analytic solutions of the Gilbert equation are usually only possible in the macro-spin approximation [33], where the magnetic structure of the sample is assumed to be homogeneous throughout the reversal process. Usually, the macro-spin approximation is only valid for special cases or for very small magnetic particles of up to about 10 nm size [34]. To calculate the dynamics in the case of inhomogeneous, three-dimensional magnetization distributions, numerical methods are generally required, with which the LLG equation is integrated in time. A few examples on dynamic effects of the magnetization calculated with the LLG equation will be discussed in section 7.

The micromagnetic model, the relevant energy terms, and the equation of motion of the magnetization described in the previous sections provide the basis for the description of micromagnetic structures and effects that will be described in the rest of this chapter.

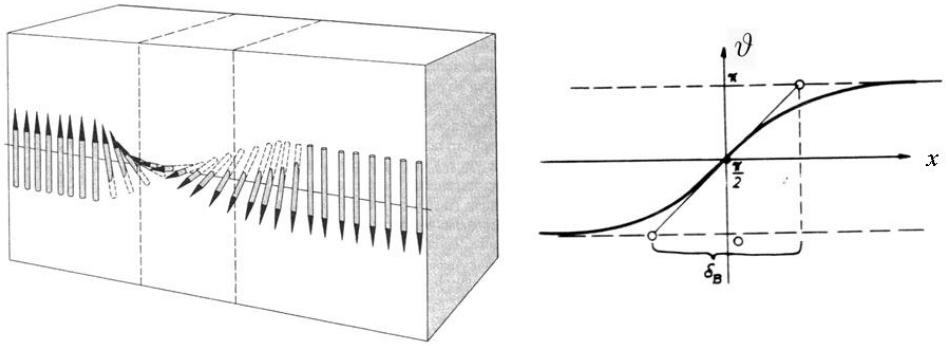


Fig. 5: *Left: Sketch of a Bloch domain wall [23]. The magnetization rotates smoothly by 180° between the magnetic domains. The width and the profile of the domain-wall transition result from a balance between the exchange energy and the anisotropy energy. Right: Profile $\vartheta(x)$ of the domain wall transition.*

5 Magnetic domain walls

Inhomogeneities of the magnetization are at the core of micromagnetic studies. Even the most complex magnetization patterns with a rich variety of features [30] can usually be decomposed into only a few fundamental magnetization structures, such as magnetic domains, domain walls or vortices. Perhaps the simplest and at the same time the most fundamental magnetic structures in micromagnetism are the $180°$ domain walls. These domain walls describe the magnetic structure in the transition region between two extended regions which are magnetized homogeneously and in opposite direction, see Fig. 5.

In the limit of domain theory, where the internal structure of domain walls is neglected, the domain wall is the line (in two dimensions) or surface (in the three-dimensional case) separating the mutually antiparallel domains. In the framework of micromagnetism the continuous transition of the magnetization is studied. Accordingly, the domain wall can there be defined as the line or the surface on which the magnetization is oriented perpendicular to both domains, i.e., the region in which the magnetization has performed half of the reorientation between one domain wall direction and the other. The classical case, which will be discussed in the following, is that the domain wall is oriented parallel to the magnetization direction. The famous Bloch wall and the Néel wall are both of this type. A different category of domain wall which has been discussed intensively in the last years are various forms of head-to-head domain walls, where the magnetization in the domains is perpendicular to the domain wall.

For simplicity it may be assumed that the magnetization does not display any variation along the domain wall: The orientation of the magnetization then only depends on the distance from the domain wall, such that the description of the domain wall structure is a one-dimensional problem. There are two fundamentally different possibilities for the magnetization to rotate continuously from the direction of one domain to the other, known as Bloch wall and Néel wall. These one-dimensional domain walls are the simplest domain wall types. Needless to say that they only represent idealized situations, and that in reality a much larger spectrum of domain wall types can be found. In particular, two-dimensional domain walls like cross-tie walls, asymmetric Bloch walls or asymmetric Néel walls are important and fascinatingly complex “hybrid” domain walls, which display combinations of both fundamental types: the Bloch wall and the Néel wall.

5.1 Bloch wall

The Bloch wall describes the rotation of the magnetization in a small region between domains with antiparallel magnetization. Bulk ferromagnetic material with uniaxial anisotropy is assumed. Effects connected with the surface of the sample are not considered and the magnetization in the domains is aligned with the easy axis of the anisotropy. In the case of a Bloch wall the magnetization always remains perpendicular to the wall normal, around which it rotates by 180° in the transition region.
To describe this situation it is useful to select a specific coordinate frame in which the x-component of the magnetization is the component perpendicular to the domain wall. In the case of a Bloch wall, this m_x component is everywhere equal to zero. The domain wall profile $(m_y(x), m_z(x))$ only depends on the x coordinate and can be conveniently described by polar coordinates in the form

$$m_z(x) = \cos\vartheta(x), \qquad m_y(x) = \sin\vartheta(x) \tag{26}$$

With the given boundary conditions

$$\vartheta(-\infty) = 0, \qquad \vartheta(\infty) = \pi \tag{27}$$

and the selected sense of rotation of the magnetization, the static magnetic structure describing the domain wall profile $\vartheta(x)$ can be obtained from *energy minimization*. The involved energy terms are the exchange energy and the anisotropy energy. These are competing interactions in the given situation: The exchange energy is minimized if the domain wall is wide, *i.e.*, if the transition occurs very smoothly over a large distance, since then the inhomogeneities $(\boldsymbol{\nabla} m_i)^2$ are small. On the other hand, the anisotropy energy tends to reduce the width of the domain wall as much as possible in order to align the magnetization with the easy axes. The result of this competition is obtained by minimizing the domain wall energy γ_{B}:

$$\gamma_{\mathrm{B}} = \int_{-\infty}^{\infty} \left[K_{\mathrm{u}} \sin^2\vartheta + A \left(\frac{\mathrm{d}\vartheta}{\mathrm{d}x} \right)^2 \right] \mathrm{d}x \tag{28}$$

The minimization $\delta\gamma_w = 0$ can be performed analytically and yields

$$\cos\vartheta = \tanh(x/\sqrt{A/K_{\mathrm{u}}}) \tag{29}$$

The profile of the magnetization of this wall type is sketched in Fig. 5. A frequent definition of the domain wall width δ_{B} is given by the distance between the points at which the tangent at $x = 0$ crosses the $\vartheta = 0$ and the $\vartheta = \pi$ lines. With this definition, that is due to Lilley [35], the domain wall width of a Bloch wall is

$$\delta_{\mathrm{B}} = \pi\sqrt{A/K_{\mathrm{u}}} \quad . \tag{30}$$

The energy density γ_B of a Bloch wall (defined as the energy per unit area of the domain wall) results to be

$$\gamma_{\mathrm{B}} = 4\sqrt{AK} \quad . \tag{31}$$

An important property of the Bloch wall is that the magnetization distribution is free of divergence[4]. Therefore, no bulk charges $\rho = -\boldsymbol{\nabla}\boldsymbol{M}$ are formed, which would be sources of

[4]This property is not a result of the energy minimization. It follows immediately from the choice of the rotation of the magnetization according to Eqs. (26).

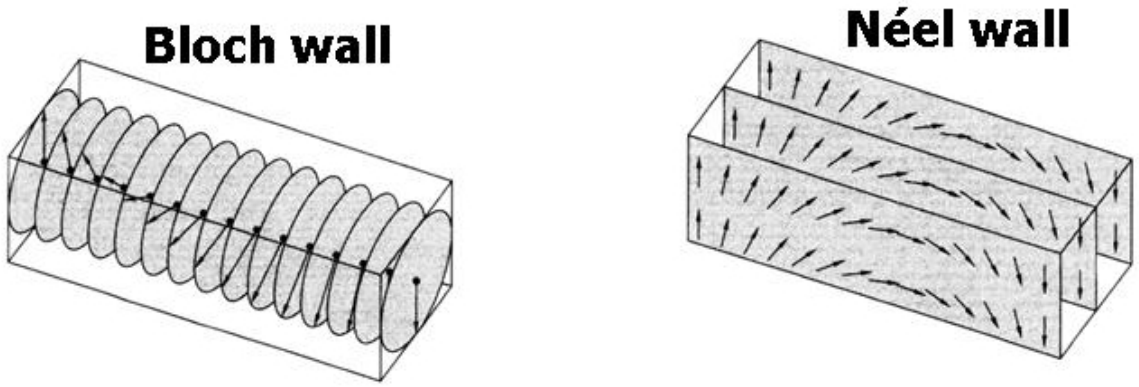


Fig. 6: *This is the schematics of a Néel wall transition. The Bloch wall is characterized by a rotation of the magnetization around the normal of the domain wall,* i.e.*, the magnetization always remains parallel to the plane separating the two domains. In the case of a Néel wall the magnetization rotates in the film plane, such that at the central part of the transition it is oriented perpendicular to the domain wall (adapted from Ref. [37]).*

magnetostatic stray fields. Since, by definition of the problem, surface charges are neglected, there are no sources of the magnetostatic field present and it is thus legitimate to neglect the dipolar energy term in the energy minimization.
The Bloch wall has been derived under the assumption of bulk ferromagnetic material with uniaxial anisotropy. The term "bulk" means that the sample is very thick, such that surface effects can be neglected. In contrast to this, the Néel wall is the classical domain wall type occurring in thin, soft magnetic films.

5.2 Néel wall

The difference of the domain wall type between the thin-film case and the bulk case is due to magnetostatic effects. In bulk samples (*i.e.*, thick samples of typically about 100 nm thickness and above) it may be quite safely neglected that a small amount of surface charges $\sigma = \boldsymbol{m} \cdot \boldsymbol{n}$ occurs in a Bloch-type domain wall transition as the magnetization rotates by 180°. But in thin films, the relative impact of the magnetostatic energy connected with the field created by these charges is much larger. Louis Néel has demonstrated [36] that in these cases it is energetically more favorable for the magnetic system to perform the change of the magnetization direction by a rotation in the film plane. This transition is sketched in Fig. 6.
The Néel wall profile can be calculated by the minimization of a one-dimensional energy functional, analogous to the case of the Bloch wall described before. In the case of Néel walls the competing interactions are given by the magnetostatic energy and the exchange energy. The tendency of the magnetostatics to keep the domain wall as small as possible is balanced by the tendency of the exchange term to avoid strong inhomogeneities of the magnetization. The magnetostatic field in the case of a Néel wall is calculated using Eq. (8). The one-dimensional energy functional to be minimized is then

$$\gamma_{\mathrm{N}} = \int_{-\infty}^{\infty} \left[\frac{\mu_0}{2} M_s^2 \cos^2 \vartheta x + A \left(\frac{\mathrm{d}\vartheta}{\mathrm{d}x} \right)^2 \right] \mathrm{d}x \tag{32}$$

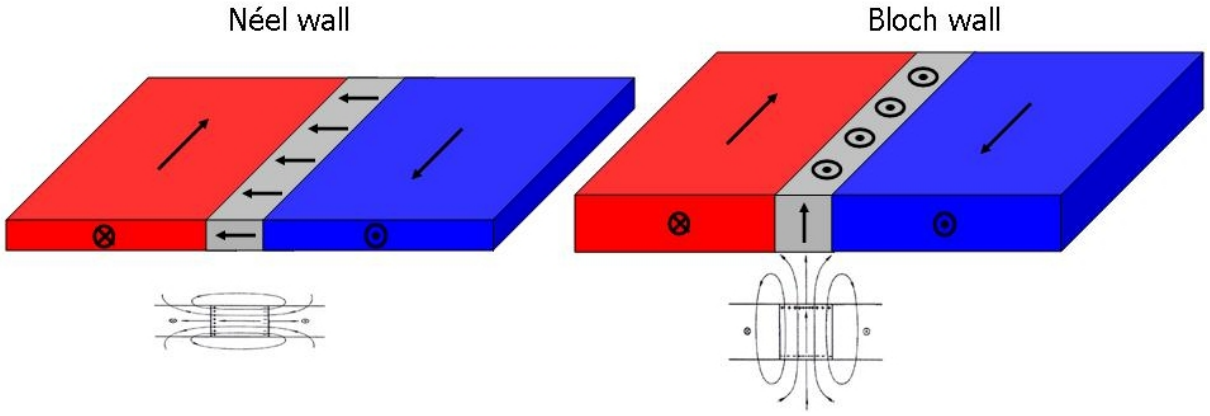


Fig. 7: *Magnetic stray fields are created by both, the Néel and the Bloch wall: In the former case volume charges $\rho = -\nabla \boldsymbol{M}$ occur, in the latter surface charge $\sigma = \boldsymbol{m} \cdot \boldsymbol{n}$. Similar to the arguments used in the case of shape anisotropy (3.4) it can be argued that the minimization of magnetostatic field energy gives rise to a thickness-dependent transition between these two types of domain walls.*

The result of this minimization yields the profile of the Néel domain wall

$$\cos \vartheta(x) = \tanh\left(\frac{x}{\sqrt{2A/\mu_o M_s^2}}\right) \tag{33}$$

which has form that is very similar to that previously derived for the Bloch wall. The domain wall width of a Néel wall is accordingly

$$\delta_{\mathrm{N}} = \pi\sqrt{\frac{2A}{\mu_0 M_s^2}} \tag{34}$$

and the energy density of the Néel wall is

$$\gamma_{\mathrm{N}} = 4\sqrt{A\mu_0 M_s^2/2} \quad . \tag{35}$$

Evidently, the role played by the stray field constant $K_d = \mu_0 M_s^2/2$ for Néel walls is analogous to that of the anisotropy constant K_u in the case of Bloch walls. The ratio of anisotropy constant K_u to stray field constant K_d is the definition of the quality factor Q that indicates the magnetic hardness. In this context it is not surprising that the quantities K_u and K_d appear in the comparison between Bloch an Néel wall, since the Bloch wall has been calculated for hard magnetic materials with uniaxial anisotropy while the Néel wall is derived for ideally soft materials ($K_u = 0$). However, the selection between Néel and Bloch type domain wall is not primarily given by the magnetic hardness, but rather by the film thickness. In thick samples Bloch walls are favored, in thin films Néel walls. Of course the terms “thin” and “thick” are only relative terms which require a comparison with a well-defined length scale to become meaningful. As far as the distinction between Néel and Bloch wall type is concerned, it can coarsely be estimated that a film can be considered to be thin if the domain wall width is larger than the film thickness, whereas in thick films it is the opposite.
It should be noted that Néel type domain walls are magnetically charged: Contrary to the Bloch wall, the magnetic structure of Néel walls displays a non-vanishing divergence, so that magnetic volume charges $\rho = -\nabla \boldsymbol{M}$ are connected with Néel walls. The magnetic fields created by these volume charges are relatively small as long as the film thickness is low.

5.3 Exchange lengths

The calculation of the Bloch and Néel wall profiles in the previous sections has provided the characteristic length scales, which describe the width of the wall. Those so-called exchange lengths are material specific length scales on which magnetic inhomogeneities can develop in ferromagnetic materials. The magnetostatic exchange length is defined as

$$l_s = \sqrt{A/K_d} = \sqrt{2A/(\mu_0 M_s^2)} \tag{36}$$

and the magnetocrystalline exchange length is

$$l_K = \sqrt{A/K_u} \quad . \tag{37}$$

They differ from Lilley's definition of the domain wall width by a factor of π. For a magnetic material, the smallest of these quantities is usually the decisive one. Typical values of the exchange lengths are of the order of about 10 nm. Detailed tables of exchange lengths including also other definitions (e.g., in the presence of external fields) are given in the textbook of Kronmüller and Fähnle [6].
The exchange lengths are important quantities for two reasons: firstly, they represent good estimates for the typical extension of magnetic inhomogeneities in equilibrium magnetization states. Since these length scales are found to be always of the order of at least several nanometers (*i.e.*, much larger than the lattice constant), they represent an *a posteriori* justification of the micromagnetic model: the smoothening effect of the ferromagnetic exchange prevents inhomogeneities of the magnetization on an atomistic level, which is a necessary condition for a meaningful continuum representation of the magnetization. The second important aspect of exchange lengths refers to computational micromagnetism. In micromagnetic simulations, the continuous vector field of the magnetization is represented in a concrete form: it is calculated at discrete points, or within cells of a certain volume. The size of these cells (or the distance between the points) is the size of the discretization cells. Since the exchange lengths represent the "feature size" of typical micromagnetic structures, they also indicate the discretization size required to resolve important inhomogeneities of the magnetization, like domain walls, vortices and vortex cores. But the impact of the exchange length on the required discretization size goes beyond the possible loss of information on a few micromagnetic details that can occur due to too large cells. Using discretization cells with sizes larger than the exchange length may cause important discretization errors that can lead to unreliable results of the overall magnetization structure. It is therefore necessary to use cell sizes which are not larger than the length scale given by the exchange length in micromagnetic simulations, even if one is not interested in the details of the micromagnetic structure. Since, particularly in three-dimensional simulations, the number of cell sizes (and consequently also the required computation time and memory requirements) increases drastically with increasing problem size, it is not possible to apply standard micromagnetic simulations to large-scale macroscopic elements.

5.4 Two-dimensional domain walls

The Bloch wall and the Néel wall are the fundamental types of domain walls which represent idealized cases. They represent the minimum energy configuration of domain walls in the limit of thin films on one hand and in bulk materials on the other hand. Interesting situations occur at intermediate film thicknesses, where neither the pure Bloch wall nor the pure Néel wall represents an optimal transition. In these cases mixed types of domain walls occur, which contain

both, a Néel and a Bloch type contribution in their transition. These domain walls are necessarily two-dimensional, sometimes even three-dimensional, meaning that the magnetization direction is not only a function of the distance from the domain wall but also a function of the position along the domain wall or along the thickness, or both.
In soft magnetic materials, the thickness-dependent transition from the Néel type to the Bloch type domain wall can be observed by the onset of a famous domain wall known as *cross-tie* wall. The cross-tie wall was first reported by Huber *et al.* in 1958 [40]. In spite of the low resolution

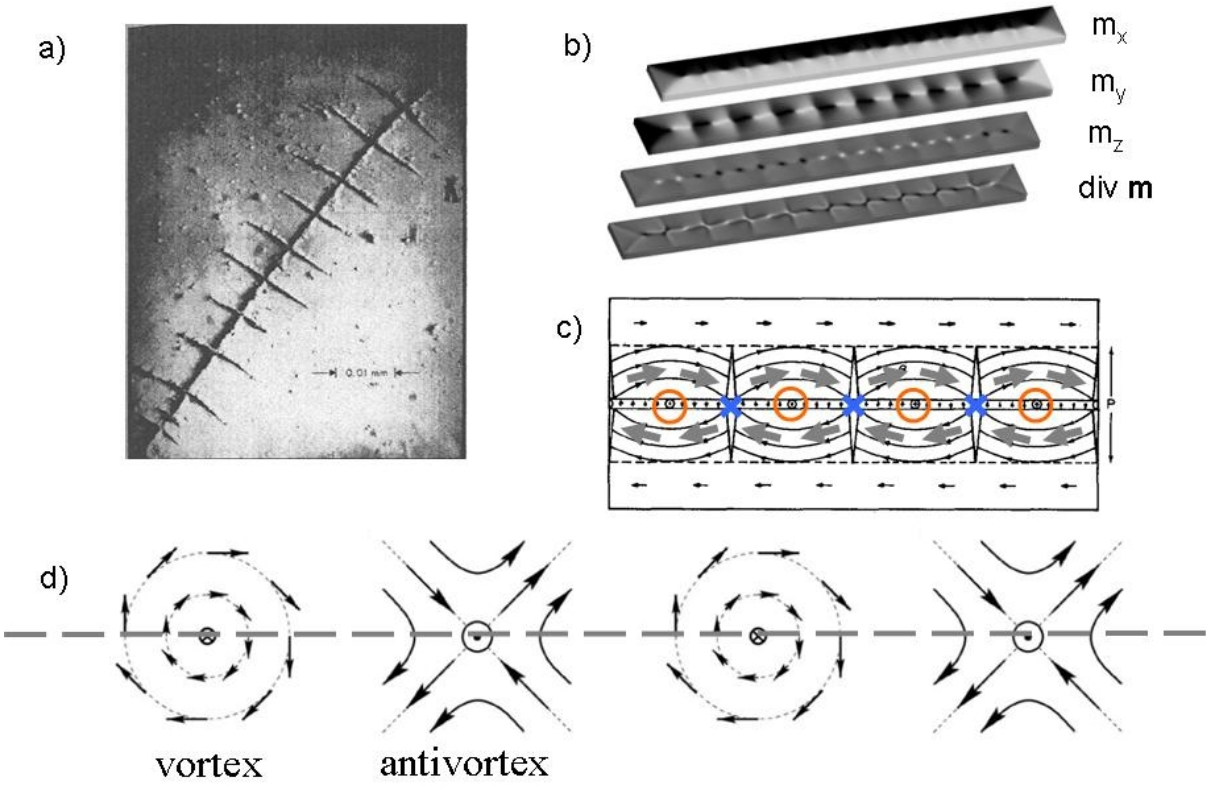


Fig. 8: *a) Image of a cross-tie domain wall in a Permalloy film of 30 nm thickness, observed with the Bitter technique [38]. Small colloidal Fe particles are deposited in regions of strong local magnetic fields. The characteristic spikes can extend over several microns into the domains. b) Micromagnetic simulation of a cross-tie domain structure in a 1000 nm $\times$ 200 nm $\times$ 30 nm Permalloy element. The first three images (from the top) are greyscale images displaying the magnetization component along the long edge (x), the short axis (y) and the thickness (z). Note the black and white spots in the z component of the magnetization, representing the (anti-)vortex core regions. The fourth image shows the distribution of volume charges $\rho = -\nabla \boldsymbol{m}$, showing the characteristic cross-tie spikes [A. Kakay* et al.*, FZ Jülich, unpublished]. c) Magnetic structure of the cross-tie domain wall as described by Middelhoek [39]. d) Schematic representation of the fine structure of a cross-tie wall as an alternating series of vortices and antivortices.*

of the magnetic imaging techniques available at that time, the fine structure of this type of domain wall was quickly interpreted correctly by various authors [40, 38, 39]. It can be regarded as an alternating series of vortex and antivortex structures. The characteristic spikes that were observed in the original images obtained with the Bitter technique [40] are due to the volume

charges connected with the antivortex. Similar to Bloch and Néel domain walls, vortices and antivortices are also fundamental magnetization structures which represent the fine structure of several micromagnetic domain states. Since the magnetization in the center of the vortex and the antivortex is perpendicular to the surface, the magnetization performs a Bloch-like transition from one domain to the other along a line perpendicular to the domain wall through the (anti-)vortex core (see. Fig. 8). In the region between the vortex and the antivortex cores, the magnetization remains in the film plane and the domain wall has a Néel character. Therefore, the Bloch character of the domain wall is increased gradually with increasing number of vortex-antivortex pairs. In the absence of these pairs (*i.e.*, the distance between vortex and antivortex goes to infinity), the domain wall has a pure Néel character. The amount of Bloch type transition increases with increasing the cross-tie density. According to the previous considerations according to which Néel walls are favored in thin films while Bloch walls are preferred in thick films, a continuous increase of cross-tie density is in fact observed experimentally with increasing film thickness (see Fig. 9). In spite of this relatively simple qualitative explanation for the increase of cross-tie density with increasing thickness, understanding the details of the energetic balance determining the equilibrium density of vortex-antivortex pairs in a cross-tie domain wall as a function of the thickness remains a highly complicated question, which has recently attracted much interest also among mathematicians [41].

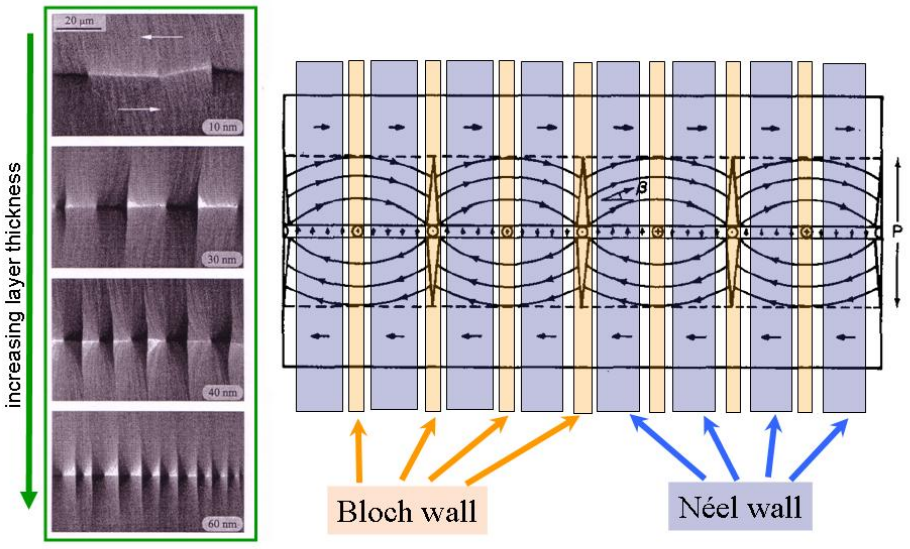


Fig. 9: *Left: In extended magnetic films the cross-tie density increases with increasing film thickness [30]. The schematics of the cross-tie wall structure on the right indicates that the cross-tie contains both, a Bloch- and a Néel-like transition. Changes in the density of cross-ties increase or reduce the Bloch (Néel) character of the transition region.*

Another type of two-dimensional domain walls is given by the so-called asymmetric Bloch wall. The asymmetric Bloch wall occurs in soft magnetic films of elevated thickness of about ten exchange lengths or more. It also describes a combination of Néel and Bloch domain wall, but here the variations do not occur along the domain wall (as they do in a cross-tie wall) but along

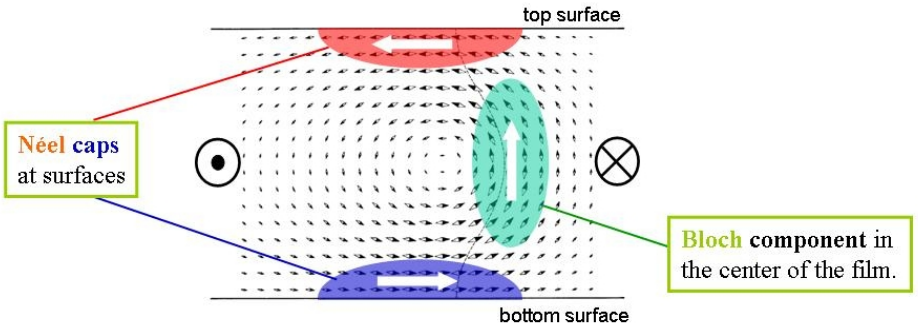


Fig. 10: *The asymmetric Bloch wall is a hybrid domain wall structure which contains Bloch- and Néel components to reduce magnetostatic fields: Near the surface the magnetization rotates in the plane (the so-called Néel caps) to prevent the formation of surface charges* $\sigma = \boldsymbol{m} \cdot \boldsymbol{n}$*; whereas in the center of the film a Bloch-like transition occurs, thereby avoiding volume charges* $\rho = -\boldsymbol{\nabla} \boldsymbol{m}$ *(schematics adapted from [42]).*

the film thickness. Interestingly, the asymmetric Bloch wall was the first micromagnetic configuration that was predicted by micromagnetic simulations. In 1969, Hubert [43] and LaBonte [42] obtained a domain wall structure as displayed in Fig. 10. The results were very surprising at that time, since such a type of hybrid domain wall consisting of both, Néel and Bloch transition regions was not known before. The structure of the asymmetric Bloch wall can be understood as an attempt of the magnetization to reduce magnetostatic fields as far as possible. The arrangement is practically free of stray fields, since the transition minimizes volume charges as well as surface charges. From the viewpoint of magnetostatic energy minimization, the asymmetric Bloch wall combines the advantage of a divergenceless transition in the case of a Bloch wall with the advantage of the surface-charge free transition of a Néel wall. As can be seen in Fig. 10, this domain wall has a Néel character near the surface (the so-called Néel caps), while in the bulk part of the film the transition is given by a Bloch-type rotation of the magnetization. The existence of asymmetric Bloch walls has been verified experimentally shortly after their theoretical prediction [44]. Many aspects of these domain walls have been investigated in detail in various studies [45, 14, 46], including the particular effects occurring when the direction of the Bloch component, of the Néel caps, or both, changes along the film thickness ([30], p.445).

5.5 Domain walls in thin strips

Yet another class of domain walls has recently moved into the focus of interest of many studies in nano- and micromagnetism, namely domain walls occurring in soft-magnetic thin strips. The strips are typically up to a few 100 nm wide and a few tens of nanometers thick. Domain walls forming in such strips are so-called head-to-head or tail-to-tail walls. Also in this case, the two fundamental structures of this type (transverse wall and vortex wall) have first been predicted by simulations [47]. It has been suggested that such domain walls could serve as units of information that can be shifted along the strips. The magnetic strips act as guides of these domain walls which can display a particle-like behavior [48]. Based on the domain wall propagation along strips, novel data storage devices (“race track memory” [49, 50]) and new concepts for devices performing logical operations [51] have been developed. Also in the recently suggested concept of spin wave logics [52] such head-to-head domain walls have

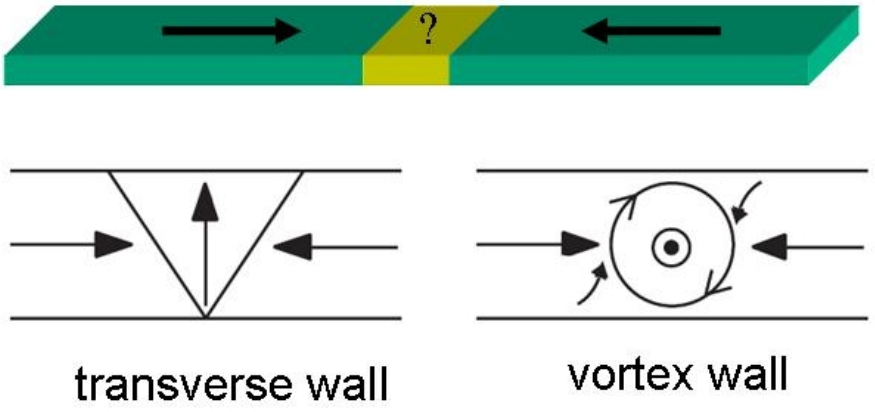


Fig. 11: *The two most important types of head-to-head domain walls in thin strips are the transverse wall (left) and the vortex wall (right). Such head-to-head type domain walls are qualitatively different from Bloch walls or Néel walls (sketches adapted from Ref. [53]).*

been proposed as phase shifters for travelling spin waves. A much more detailed discussion of these types of domain walls can be found in Ref. [48]. It should be stressed that these types of domain walls are neither Bloch nor Néel type transitions. They are qualitatively different since the domain wall is oriented perpendicular, and not parallel to the magnetization direction in the adjacent domains.

6 Magnetic domains in soft-magnetic thin film elements

The enormous variety of magnetic domain structures makes it an absolutely impossible task to give a satisfactory overview of magnetic domains within this chapter. An impression of the diversity of magnetic domains can probably best be obtained from the hundreds of figures in the textbook by Hubert and Schäfer [30]. The domain structures discussed in the following should therefore be considered as only a small selection of possible structures, without any claim to completeness. To further limit the enormous size of the topic of magnetic domains, only domain structures in soft-magnetic thin film elements shall be discussed here. Such soft-magnetic thin-film elements are particularly important for spintronic applications.

6.1 Flux closure domain patterns

An appealing aspect of domain structures in soft-magnetic elements is that they can be understood in a simple way even though the physics leading to their formation is complicated.
If the material is sufficiently soft-magnetic and the sample sufficiently large, the magnetic structure is completely determined by the need to minimize the dipolar energy. As mentioned before, it is not possible to arrange the magnetization in such a way that the resulting charges (poles) create a negative magnetostatic energy. The best minimization of the dipolar energy is therefore obtained by avoiding the sources of the field. This behavior is known as Brown's pole avoidance principle. In such large and soft-magnetic thin-film elements the magnetic structure thus arranges in a way to avoid magnetic charges as far as possible. The first obvious consequence

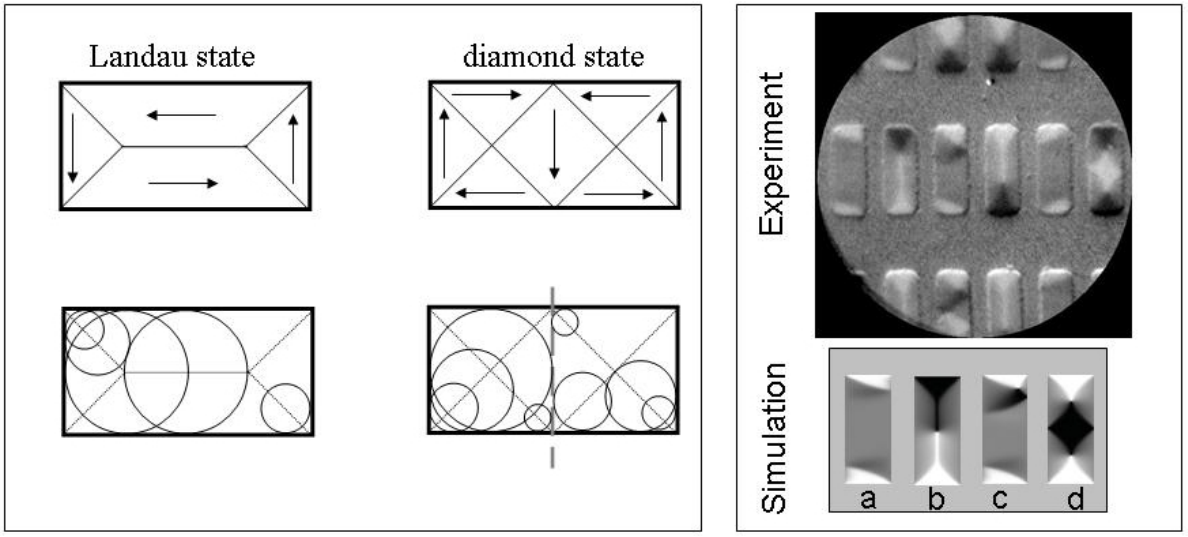


Fig. 12: *Left panel: The Landau state and the Diamond state are the most frequent flux-closure domain patterns in rectangular thin-film elements. The Landau domain structure can be constructed easily with the Van den Berg scheme: The domain walls are given by the sets of central points of circles that touch the sample boundary at least twice. The diamond state can be constructed with the same scheme if the sample is virtually cut into two parts along the dashed line. Right panel: Virtually identical magnetic thin-film elements samples (here: sub-micron sized Co platelets) can have very different domain structures. In all cases, practically perfect agreement is obtained between simulation and experiment [8]. High-remanence states (a), (c) can be found as well as the Landau (b) or the diamond structure (d).*

is that the magnetization lies perfectly in the film plane to avoid surface charges[5]. A further minimization of surface charges is obtained by aligning the magnetization parallel to the *edges* of the thin-film element. If it is further assumed that, at least in the domain-theory limit of vanishing domain wall width, the structure is free of volume charges, $\nabla \boldsymbol{M} = 0$, it is possible to construct a few elementary domain structures frequently observed in thin-film elements, such as the famous Landau structure or the diamond structure.

6.2 The Van den Berg Scheme

For such *demagnetized states* in soft-magnetic thin film elements as the ones shown in Fig. 12, in which the spatially averaged magnetization is exactly zero and the structure is completely free of stray fields, there exists a simple graphical construction scheme that is due to H. A. M. Van den Berg [54, 55, 56]. The simplicity of the graphical scheme is remarkable in view of the relatively complicated mathematical background from which these rules have been derived. The construction of the domain structure can be performed for arbitrarily shaped two-dimensional elements, and provides patterns that often correspond perfectly to the experimental observation. In a simplified way, it can be said that the Van den Berg (VdB) scheme essentially consists in drawing circles: more precisely, circles that touch the boundary of the element at least twice. The set of the central points of these circles constitutes the magnetic domain walls (which are

[5]Topological peculiarities, like the perpendicular component of the magnetization in the case of vortex cores or some surface charges created by head-to-head walls can be neglected here.

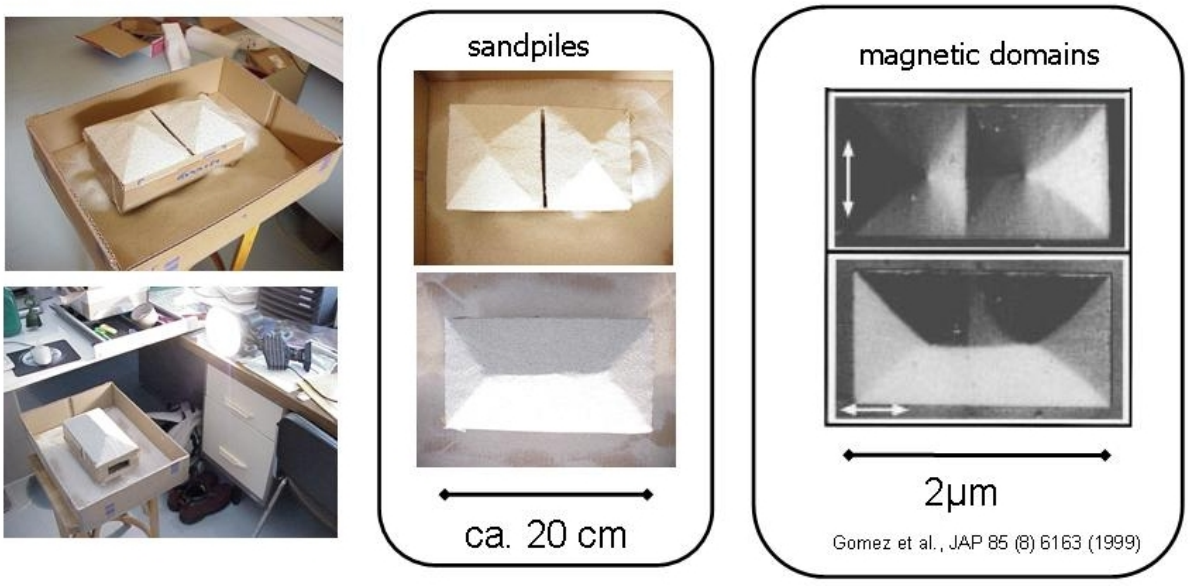


Fig. 13: *The equilibrium structure of carefully piled-up sandpiles depends on the shape of the plate on which they are formed. Since there exists a mathematical analogy between the height profile of such sandpiles and the domain structure of flux-closure states in soft-magnetic thin film elements, suitably illuminated sandpiles resemble magnetic domain structures. (Sandpiles and photos by O. Fruchart, Institut Néel, Grenoble, France; experimental image on the right taken from Ref. [57]).*

assumed to be lines in this model). The magnetization is perpendicular to the radial lines of the circles that touch the boundary. Examples for domain patterns obtained with the Van den Berg construction scheme is shown in Fig.12 The agreement of experimentally observed demagnetized states in patterned thin films with the Van Den Berg scheme can often be recognized immediately.

It should be mentioned that the domain structures that can be constructed with the VdB scheme are not unique. It is possible to virtually divide an element into different parts and apply the VdB scheme to each of those segments. The re-assembly of these parts will then of course lead again to a demagnetized magnetic domain structure, which is different from the one obtained when the VdB scheme is applied directly to the sample. The lack of uniqueness is not a weakness of the scheme, since this corresponds to the experimentally observed situation: identical samples can display very different magnetic domain structures, even on the sub-micron scale [8].

There is a remarkable connection between the domain structures constructed according to the VdB and the shape of sandpiles, as discussed in the book of Hubert [30] on p. 177. Without giving the relatively simple explanation for this analogy, a similarity between illuminated sandpiles on a macroscopic plate and magnetic domain structures in microscopic magnetic elements is shown in Fig. 13.

A more detailed understanding of magnetization structures in thin film elements cannot be achieved with this simple model that is exclusively based on the assumption of perfect stray field minimization. Important aspects such as the details of the domain walls, magnetic domain structures in non-demagnetized states and calculations of the energy of the magnetic domain structures requires combinations of experimental and numerical studies.

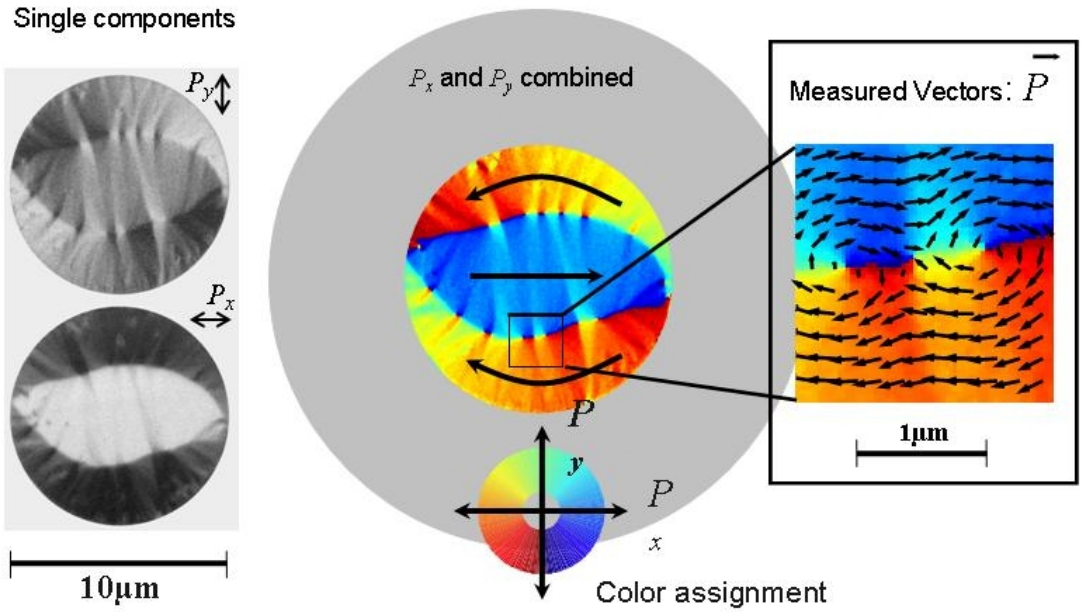


Fig. 14: *Particularly complex magnetic domain structures can develop in thin film elements when several characteristic length scales interact. In this example from Ref. [58], a flux-closure magnetic domain structure in a 10µm Co disk is analyzed using scanning electron microscopy with polarization analysis (SEMPA), which can measure simultaneously two magnetization components (panel on the left). Using a 360° color-coding, the information can be assembled into one image displaying the in-plane magnetization direction (middle). On a large length scale, the domain structure is a demagnetized state according to the VdB scheme. The smaller details of the structure display complicated cross-tie domain walls with several vortices and antivortices (right).*

6.3 High-remanence states

The technically more relevant situation than the demagnetized configuration are the high-remanence states. In a first approximation, the high-remanence magnetization states of thin-film elements can be regarded as homogeneously magnetized states in which the magnetization points into one specific direction. If the magnetization can be switched in the opposite direction and if these are the only two stable states, this represents a suitable unit for binary magnetic data storage. Although such high-remanence states can be prepared and the shape of magnetic elements can be chosen such that demagnetized states are unstable, the notion of a homogeneously magnetized thin-film element is not correct and disregards potentially important details. Strictly speaking, a homogeneous magnetization structure is only possible in an ellipsoidal particle [17]. It can therefore not be achieved in a two-dimensional thin-film element, which at best can have an elliptical shape to reduce stray-field induced inhomogeneities near the ends of the element.

The characteristic inhomogeneities near the ends are particularly well known for rectangular elements, where the so-called S-state and the C-state are possible high-remanent states. Also in magnetic discs and rings typical high-remanence states are known, the most common of which is usually labelled the “onion state”. Cube-shaped particles may display a high-remanence state known as “flower state” [61, 59, 62]. From an application point of view, it would appear

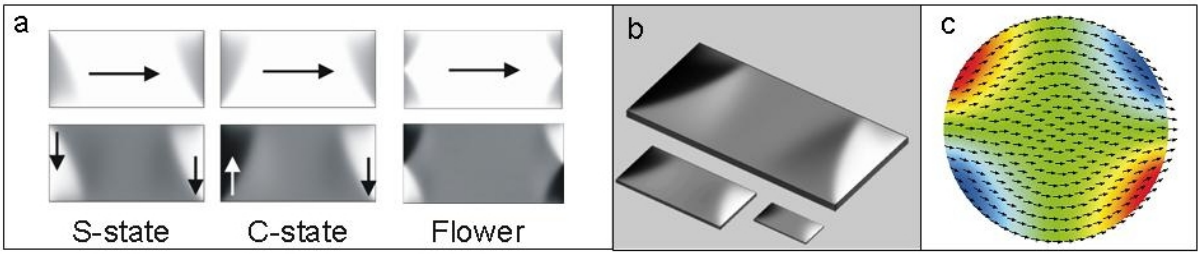


Fig. 15: *High-remanence states in thin-film elements are not homogeneously magnetized. In rectangular platelets the most important cases are the S-state, the C-state and the Flower state (a). In these states the magnetization near the edges and corners displays characteristic inhomogeneities. To a large extent, the magnetization structure,* e.g.*, of the C-state is the same for elements of different size. The Permalloy rectangles in frame (b) are between 250 nm and 1000 nm large. In disks, the onion state (c) is the typical magnetization structure of the high-remanence state (figures from Refs. [59] and [60])*

to be desirable to have homogeneous magnetization states without any significant asymmetry (like in the case of the S- or the C-state) or reduction of the remanence. However, as already mentioned, it is not possible to completely suppress such magnetic inhomogeneities near the boundary thin-film elements by tailoring their shape or size. Moreover, the inhomogeneities at the particle's end, sometimes called "end domains", can have beneficial effects on the magnetic properties. They may represent nucleation sites for the magnetization reversal process such that the dynamic switching occurs in a reproducible fashion and at well-defined fields. The elements with the tapered end presented at the beginning of this chapter (Fig. 1) are an example where end domains and, in particular, any possible asymmetric magnetization state has been largely suppressed by the particular element shape. This results in very high switching fields and, as simulations have shown, in switching processes that are not well reproducible, in some cases starting in the middle of the element and in other cases at the ends.
The reason for the formation of the characteristic end domains is the attempt of the magnetic structure to minimize the magnetic surface charges as far as possible. A partial alignment of the magnetization with the short edge, as is the case for the S- and the C-state already leads to a significant reduction of surface charges while maintaining a largely homogeneous magnetic structure, so that the exchange energy remains small. The precise details of the shape of a thin-film element can strongly affect the formation of end-domains. The role of end domains in high-remanence states and their important impact on the switching field of magnetic thin film elements is an example of the sensitive dependence of the particle shape on the magnetic properties which was mentioned in the beginning of this chapter.

7 Magnetization dynamics in nanostructures

Over about the last ten years, investigations on the fast and ultrafast magnetization dynamics have attracted constantly growing interest. Progress in experimental techniques for time-resolved magnetic imaging and the increasing capabilities of numerical simulations have made it possible to access this complex topic and study the properties on the magnetization dynamics on the pico- and nanosecond time scale. In fact, similar to the situation described before for magnetic domains, the numerous phenomena on magnetization dynamics that have been

reported in a very large amount of studies cannot possibly be summarized in this chapter. Thus, only a small selection of examples on the magnetization dynamics in nanostructures is presented here. More comprehensive treatments of this topic can be found elsewhere [63, 31].

7.1 Magnetization reversal of thin-film elements

The most important aspect of magnetization dynamics for spintronic applications usually refers to the magnetic switching process in thin-film elements. A prominent example for this are MRAM cells [4, 64], where information is stored in small magnetic thin film elements that can be switched between two high-remanence states. This magnetic thin-film element, the so-called free layer, is embedded in a multilayer magnetic structure, which changes its resistivity depending on the relative orientation of the magnetization direction of the free layer with respect to a "fixed" magnetic layer within the memory cell. The fixed layer is a magnetic element with antiferromagnetically pinned magnetization direction that can only be switched by applying very large fields. For the application it is obviously important to explore how fast the free layer can be switched and how this reversal process occurs if an external magnetic field is applied towards which the magnetization should eventually align. It was found by both, simulations and experiments, that the switching of such an element generally occurs in a very inhomogeneous way [26, 65, 27].

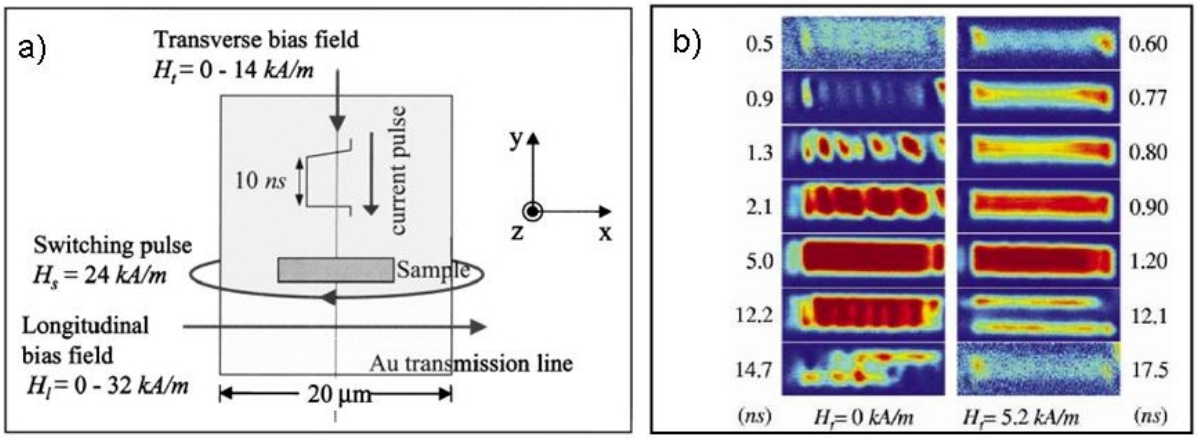


Fig. 16: *The magnetization dynamics of a reversal process can be remarkably changed if a bias field H_t is added in transversal direction (left panel). The dynamics of magnetization reversal in a Permalloy thin-film element of 10 µm × 2 µm and 15 nm thickness is shown in the right panel for the case of $H_t = 0$ (b, left part) and H_t=5.2 kA/m. The transversal field leads to a faster and more coherent magnetization reversal (taken from Ref. [66])*

Even in the case of single domain particles, *i.e.* magnetic elements that are so small that their only stable magnetization structures are high-remanence states, the magnetization reversal generally occurs in a very inhomogeneous fashion, with a nucleation of the reversal starting either in the ends or in the middle of the sample and propagating along the element. The inhomogeneous switching is connected with the creation of spin waves which persist over several nanoseconds, an effect known as magnetic ringing. The reason for this ringing can be understood from energy considerations: It is connected with the difference in Zeeman energy between the initial state, in which the magnetization is antiparallel to the external field and the final state. Therefore, immediately after the switching, the element remains in an energetically excited state in which the

Zeeman energy is converted into spin waves. The dissipation of energy with time is described by the damping constant α in the Landau-Lifshitz-Gilbert equation. The value of the damping constant α (typically in the order of about 0.01) is too low to allow for the dissipation of this energy difference within less than a few nanoseconds. This behavior of inhomogeneous switching and spin wave generation is obviously not desirable for applications in which ideally the magnetic element should be switched as fast as possible, preferably by means of a homogeneous rotation and without the undesired ringing after the reversal.
A switching process fulfilling almost all of these properties has in fact been discovered a few years later. First, it was reported that a transverse bias field applied together with the longitudinal switching field pulse had favorable effects on the magnetization dynamics, since it largely suppresses the magnetic ringing and leads to a more coherent dynamic switching (cf. Fig. 16) Even better properties of the magnetization reversal dynamics have been obtained shortly later with the precessional switching mode, which comes very close to the characteristics of an "ideal" switching process. In the *precessional* switching mode [67, 68, 69] the magnetization of a thin-film element rotates very rapidly and coherently by 180° after a suitably shaped magnetic field pulse is applied in the plane, perpendicular to the initial magnetization direction. The pulse excites a precessional motion of the magnetization, initially around the applied field and subsequently around the demagnetizing field created by the out-of-plane component of the magnetization. If the pulse is shaped correctly, the magnetization comes to a halt after having switched exactly by 180°.
The difficulties of this switching process lie mostly in the necessity of perfectly shaping the field pulse in terms of duration and amplitude. Only small margins of variations of the pulse shape are allowed in order to achieve the switching. Due to these complications, the precessional switching is more interesting from the point of view of fundamental research than for practical applications. Apart from the specific difficulties of the precessional switching, the magnetic-field induced reversal of magnetic elements is generally problematic because of unfavorable scaling properties: as the sample size becomes smaller and the integration density increases, the tolerable limits within which the amplitude of the applied field can be varied become very small. On one hand, the switching probability of the selected element should be close to 100%, on the other hand neighboring elements should not be switched. The Savtchenko toggle switching mode [70] is probably the most powerful scheme to solve this so-called selectivity problem in the case of field-induced switching. It is however probably not as promising for applications as the recently discovered possibility of utilizing electrical currents instead of magnetic fields to switch the magnetization. This new approach for the switching of magnetic elements, which is particularly suitable for densely packed arrays of nanomagnets, will be briefly discussed in section 8.

7.2 Magnetic vortex dynamics

Every flux-closure magnetic domain pattern in a singly-connected thin film element, like, *e.g.* the Landau structure discussed in section 6.1, contains at least one magnetic vortex. The fundamental nature of vortices as elementary micromagnetic structures has already been mentioned in the discussion of the cross-tie domain walls in section 5.4. For studies of the magnetization dynamics, the magnetic vortex state is very interesting because it displays a very rich variety of dynamic properties. The magnetic vortex dynamics has therefore been studied very intensively over the past years. Contrary to the previously described cases of irreversible switching processes in thin-film elements, these studied on the vortex refer to oscillations of the magneti-

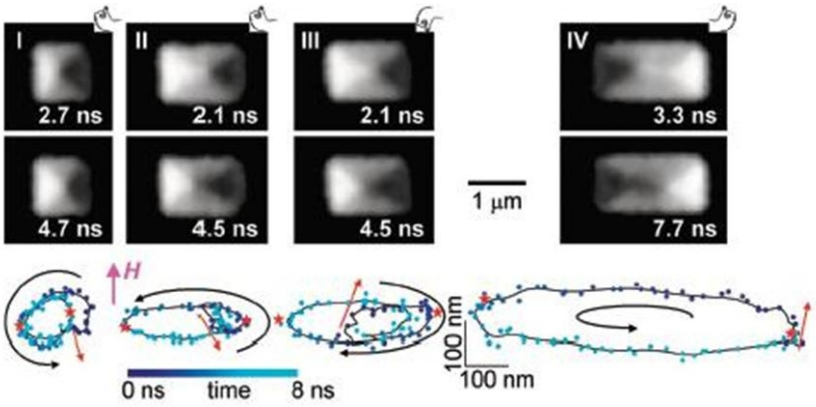


Fig. 17: *The gyrotropic motion of vortices around the equilibrium position was imaged by Choe* et al. *[71]. Even though the spatial resolution was not sufficient to image the tiny vortex core, the position of the vortex could easily be identified as the intersection of the domain walls or the junction between adjacent domains. In the same work, it was also shown that the vortex core magnetization significantly influences the vortex dynamics.*

zation around the equilibrium vortex state. Only very recently a magnetic switching process of vortices has been discovered, which has attracted much attention and that will be discussed in the next section. The lowest-frequency oscillatory mode of vortices is the gyrotropic motion: when a vortex is displaced from its equilibrium position, it relaxes back to its initial position on a spiralling orbit. This gyration of the vortex occurs with a characteristic frequency that is typically of the order of a few 100 MHz. The value of this resonance frequency depends on the aspect ratio (width / thickness) of the thin-film element containing the vortex [72]. The gyrotropic motion of vortices was predicted theoretically in 1974 by Thiele [73]. An experimental verification of the gyrotropic motion of a vortex due to a resonant excitation tuned at the resonance frequency was already reported in 1984 [74]. The gyrotropic motion of the vortex was recently imaged directly by means of scanning magnetic X-ray transmission microscopy [71, 75]. The vortex structure can give rise to several further oscillatory modes. In the case of square platelets with Landau structure, where four 90° domain walls are connected in the middle by a magnetic vortex, also the domain walls and the triangular domains can provide oscillations with characteristic frequencies after a short-pulse excitation [78]. In magnetic disks with vortex configuration, higher frequency excitations consist in radial and azimuthal standing wave modes [79, 76, 77].

7.3 Magnetic Vortex Core switching

The rich variety of dynamic properties connected with a structure as simple and fundamental as a magnetic vortex has inspired several studies on the vortex dynamics over the past years. This interest has further increased since the recent discovery of the dynamic vortex core switching. Before explaining the importance of the core for the vortex dynamics, it is useful to describe shortly the static properties of magnetic vortex cores.

As mentioned in the section on the cross-tie domain wall structure, vortices (and antivortices) have a core region where the magnetization is perpendicular to the plane. This out-of-plane rotation of the magnetization in the core region allows for a smooth magnetic structure. If the magnetization was aligned everywhere in the film plane, it would abruptly change its orientation

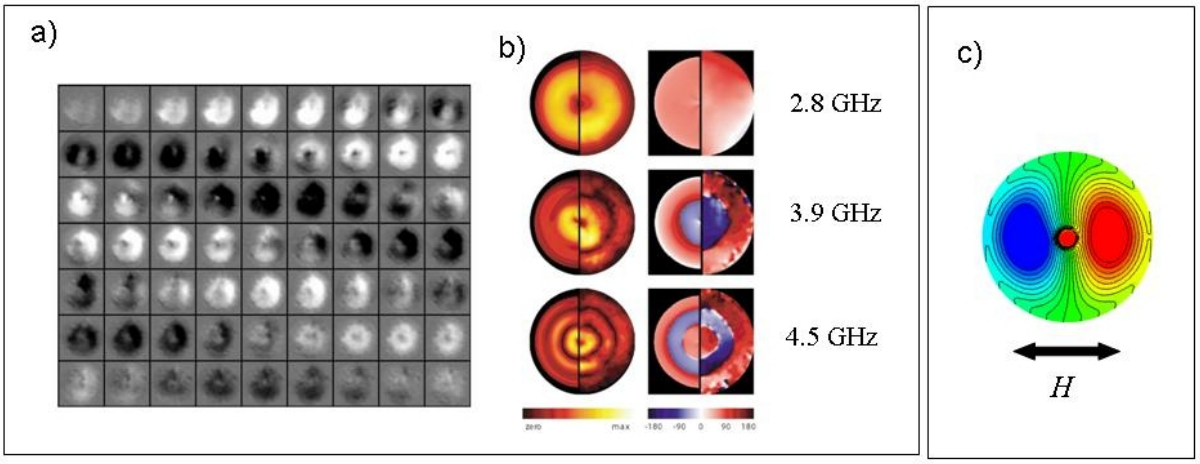


Fig. 18: *Higher-frequency modes of the magnetization in vortex states are standing-wave patterns in the order of several GHz. Buess et al. [76] performed spatial Fourier analysis of the time-resolved measurements of the magnetization dynamics in disks with vortex configuration (a) and found characteristic modes which could be compared with simulation results (b). Such standing-wave modes in vortex structures can also be obtained if an external oscillating in-plane field is applied with suitable frequency (c), see Ref. [77].*

by 360° in the region near the vortex center, leading to a singularity in the exchange energy density. The existence of the vortex core was predicted by Feldtkeller [80], who also provided analytic calculations on the vortex core size. The vortex core region is very small: its size is about 10 nm. Due to its small size, which for a long time was below the resolution limit of magnetic imaging techniques, the vortex core could only be observed experimentally about 40 years after its theoretical prediction [81]. With spin-polarized scanning tunnelling microscopy a magnetic imaging technique with ultra-high resolution has become available which has made it possible to measure precisely the profile of a magnetic vortex core [25]. The experimental result corresponds perfectly to the structure obtained by micromagnetic simulations.

The orientation of the magnetization in the vortex core (the vortex *polarization*) is not connected to the sense of rotation of the magnetization in the film plane (the *circulation*). Therefore, both quantities, the circulation and the polarization, are required to describe the static structure of a vortex. The perfectly bistable magnetic behavior of a vortex core (the magnetization in the core can point either “up” or “down”), its high stability, its small size combined with the fact that this structure forms spontaneously in simple disk-shaped elements makes vortex cores interesting candidates for binary magnetic data storage [82, 83]. However, as a result of the high stability of the vortex core, large magnetic fields of about 300 mT need to be applied in order to switch the core [84]. As far as the magnetization dynamics is concerned, the core region has a significant impact on the gyrotropic mode: Whether the sense of rotation of the gyrating motion of a vortex around the equilibrium is clockwise or counterclockwise depends exclusively on the vortex core polarization [73, 71].

It was recently discovered that a vortex that is resonantly excited to perform a gyroscopic rotation can change its sense of rotation after a short perturbation with a small field pulse [85], see Fig. 20. This is an indirect, but unambiguous evidence that the vortex core has switched as a result of the small perturbation. This possibility of *dynamically* reversing a vortex core with small field in-plane pulses (in the case reported in Ref. [85] it was a field pulse of only 1.5

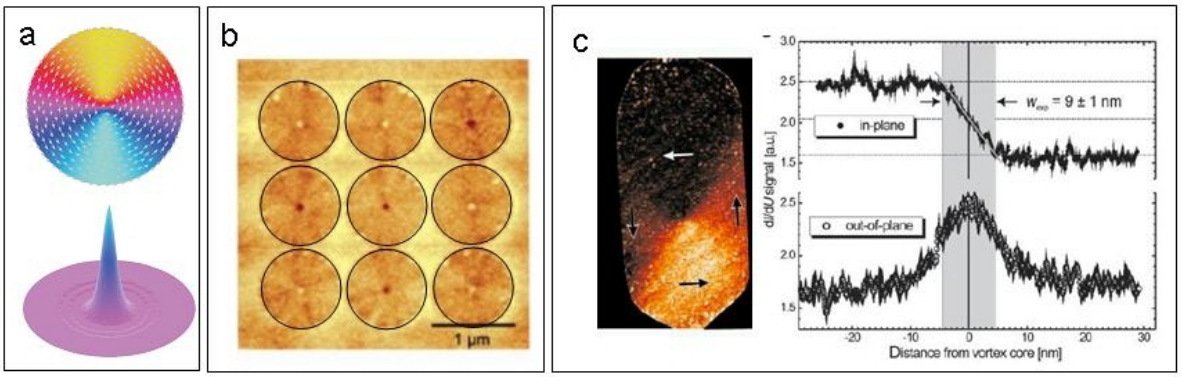


Fig. 19: *(a): A magnetic vortex is characterized by an in-plane circulation of the magnetization and by a small perpendicularly magnetized core region. (b): The existence of magnetic vortex cores was first evidenced directly by Shinjo* et al. *[81]. (c): The detailed profile of a magnetic vortex core was measured with spin-polarized scanning tunnelling microscopy by Wachowiak* et al. *[25].*

mT field) is very surprising, especially if compared with the large fields required in the static case. The micromagnetic explanation for the core switching was also suggested in Ref. [85]: The reversal of the vortex core is achieved by the temporary formation of a vortex-antivortex pair and a subsequent annihilation of the original vortex with the temporarily created antivortex, see Fig. 20c). The dynamics of vortex-antivortex annihilation processes had previously been investigated with micromagnetic simulations [87], from which it could be concluded that the vortex core reversal process should be accompanied by a burst-like generation of spin waves. Shortly after the experimental observation of the dynamic vortex core switch, micromagnetic simulations predicted that the resonant excitation of the core should not be necessary to obtain the core switch: a suitably shaped single, short and strong in-plane field pulse should lead to the same chain of events, *i.e.*, the temporary creation of a vortex-antivortex pair and the annihilation of the original vortex with the new antivortex. The fact that the new vortex always has a polarization opposite to the original one was attributed to the influence of the dipolar field created by the core [88]. The simulations provide detailed information on this new micromagnetic switching process, and very recently experimental studies seem to confirm these details [89].

In addition to this vortex core switching process triggered by field pulse excitation, it was predicted by simulations [86] and observed experimentally [90] that a single electric current pulse through the plane of the sample should lead to the same results (Fig. 21). The current-induced *resonant* switching of vortex cores was demonstrated by Yamada *et al.* [91].

Within short time, this new micromagnetic process has given rise to a large number of studies with important contributions from many different groups [92, 93, 94]. The interest is not only due to the novelty and the appealing complexity of this dynamic process: it is also of interest for the topic of ultra-fast magnetization dynamics. Until recently, the precessional switching was considered to be the fastest magnetization reversal process, but now the non-resonant magnetic vortex core switching [95] has taken over this position. With a switching time of only about 50 ps the non-resonant vortex core switch is the fastest field-induced magnetic switching process known so far.

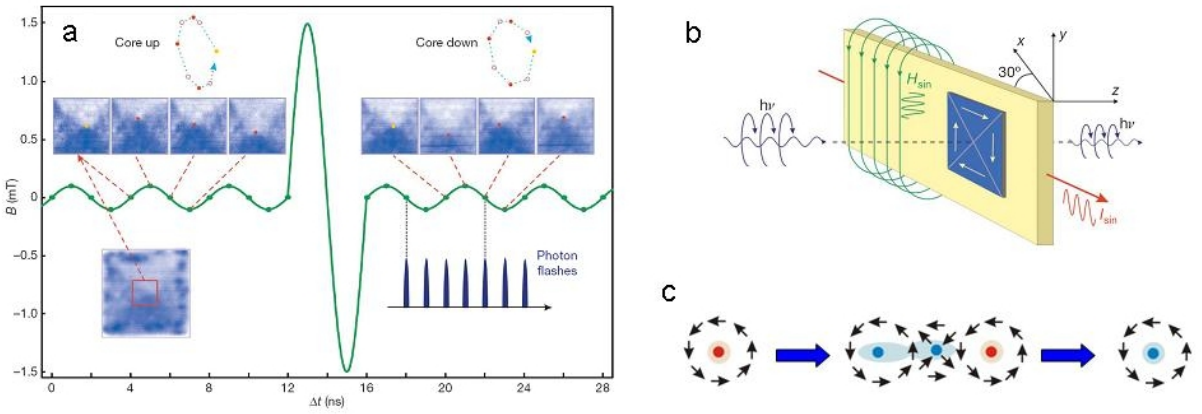


Fig. 20: *Vortex core switching as first observed by VanWaeyenberge et al. [85]. The vortex is in a dynamic equilibrium state, gyrating around its equilibrium position due to an externally applied low-amplitude resonant field(b). After a short and small increase of the exciting field over one period, the vortex core was found to rotate in gyrate direction, which demonstrates that the vortex core magnetization has changed sign (b). The micromagnetic processes leading to the reversal are sketched in frame (c): The switching is mediated by a vortex-antivortex pair creation and annihilation process.*

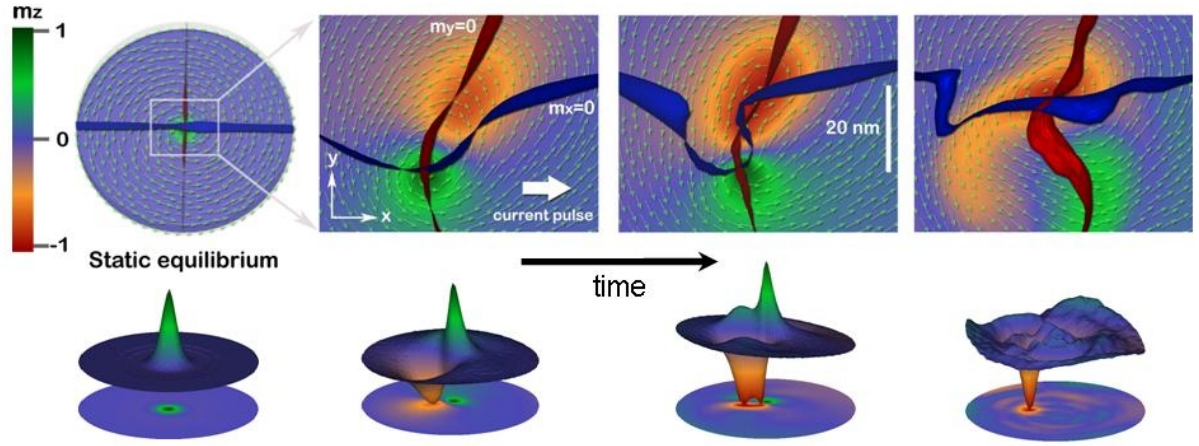


Fig. 21: *Micromagnetic simulation of the vortex core switching in a 200 nm Permalloy disk induced by a short electrical in-plane current pulse (from Ref. [86]). The three frames on the top show snapshots of the dynamic core switching process. The stripes in these frames are iso-surfaces displaying, respectively, the regions in which the in-plane magnetization components m_x and m_y are equal to zero. Their crossing points indicate the precise location of the (anti-)vortices [87]. The additional crossings of these "ribbons" in the second frame thus highlight the temporary vortex-antivortex pair formation. The colors refer to the out-of-plane magnetization component m_z, the topography of which is displayed in the snapshots on the bottom.*

8 Current-induced magnetization dynamics

Current-induced magnetization processes like the vortex core switching [86, 91, 90] are due to the transfer of electron spin from the conducting electrons to the local magnetization. This spin-transfer torque effect shall be discussed briefly in this section. Much more detailed treatments on this topic are given by the contributions of D. Bürgler and S.S.P, Parkin in this book.

It was predicted theoretically by Berger [96] and Slonczewsky [97] that spin-polarized electric currents flowing through a ferromagnet can exert a torque on the magnetization, the so-called spin-transfer torque, which can be strong enough to generate spin waves or even to switch the magnetization. The principle is that electric currents flowing through ferromagnets are *spin-polarized* according to the local magnetization. As conducting electrons flow through differently magnetized regions, they change their spin-polarization. Conservation of angular momentum requires that angular momentum is correspondingly transferred to the magnetization, resulting in the spin-transfer torque (STT) effect. The STT can be considered as an additional torque term in the Gilbert equation. The influence of a spin-polarized electric current on the magnetization is qualitatively different from that of a magnetic field. Its influence correspondingly cannot be represented as an effective field term; and there is no micromagnetic energy term connected with the STT effect.

Two different cases of current-induced magnetization dynamics are usually considered. The first refers to the so-called "pillar" geometry, in which a spin-polarized electric current flows perpendicularly through a magnetic thin-film element. The polarization of the current is achieved by a "fixed" ferromagnetic layer, the polarizer, which is part of the pillar and located next to the free layer. As the electrons enter the free layer, a torque acts on the local magnetization if the spin-polarization is not aligned with the local magnetization direction. The switching of the free layer parallel or antiparallel to the fixed layer can be achieved by changing the direction of the electron flow along the pillar [98].

In micromagnetics, the STT effect in pillar geometries is described by the torque term

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = \frac{\chi}{M_\mathrm{s}} \boldsymbol{M} \times (\boldsymbol{M} \times \boldsymbol{p}) \tag{38}$$

where $\boldsymbol{p}$ is the unit vector along the spin polarization direction given by the fixed layer and χ is a function of $\boldsymbol{M} \cdot \boldsymbol{p}$ as described by Slonczewski. This torque term should be added to the *Gilbert* equation.

A particular effect of current-induced magnetization dynamics in the pillar geometry is the possibility of exciting persistent oscillations of the magnetization with DC currents. The oscillations are in the GHz range and their frequency can be tuned, to a certain extent, by varying the current density. Such tunable nano-oscillators are supposed to have high potential for future technological applications in nanoscale communication devices. The excitation of persistent oscillations with DC currents is a new effect without analogy in the field-driven magnetization dynamics. It can occur in situations where the STT effect would favor the switching of a nanomagnet but an external magnetic field prevents the reversal.

Several studies, experimental and theoretical ones, have been performed on such current-induced nano-oscillators [99, 100]. Frequently the experimental results are interpreted in the framework of macrospin models. Even though the nanomagnets representing the free layer are very small, it is not certain that the macrospin model may be applicable. Various simulations have in fact reported the occurrence of highly inhomogeneous structures [101], indicating that a precise theoretical understanding of these dynamic effects can only be achieved with accurate simulations [100].

The second case in which current-induced dynamics plays an important role is the displacement of domain walls along ferromagnetic strips [50]. In this case, the torque is exerted by electrons flowing along an inhomogeneously magnetized ferromagnet. The change of spin-polarization that the conducting electrons experience as they pass through differently magnetized regions results in a torque on the magnetization. This torque is described by *two* additional terms: the adiabatic and the non-adiabatic term [102, 103].

$$\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t} = -\left(\boldsymbol{u}\boldsymbol{\nabla}\right)\boldsymbol{M} + \frac{\beta}{M_\mathrm{s}}\boldsymbol{M}\times\left[\left(\boldsymbol{u}\boldsymbol{\nabla}\right)\boldsymbol{M}\right] \tag{39}$$

Here, $\boldsymbol{u}$ is a vector pointing in the electron flow direction with the amplitude $u = jPg\mu_\mathrm{B}/(2eM_\mathrm{s})$ [103], where j is the current density, P is the degree of electron polarization, g is the Landé splitting factor, μ_B is the Bohr magneton, and e is the electron charge. Also the previously mentioned electric vortex-core switching [86, 91] is due to the STT effect described by this term.

The STT effect provides interesting new possibilities of manipulating and controlling the magnetization on the nanoscale. The technological advantages of using electric currents instead of magnetic fields for magnetic switching are, firstly, the unproblematic selectivity of individual elements within dense arrays of identical magnetic elements, and, secondly, the good scaling properties of the parameters required for current-induced switching.

When current-induced magnetization processes are studied, it should not be overlooked that it is not only the spin-polarization of the conducting electrons that can affect the magnetization, but also the Oersted field created by the current. It depends on the individual situation whether the Oersted field represents only a negligible perturbation of the overall magnetization dynamics or if it significantly affects it. The precise impact of the Oersted field can be analyzed with numerical simulations [104, 100], which unlike experimental studies allow to study the artificial situation of electrical currents without Oersted field.

References

[1] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, Phys. Rev. B **39**, 4828 (1989).

[2] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friedrich, and J. Chazelas, Phys. Rev. Lett. **61**, 2472 (1988).

[3] G. Prinz, Science **282**, 1660 (1998).

[4] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnár, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, Science **294**, 1488 (2001).

[5] *Handbook of Magnetism and Advanced Magnetic Materials Vol. 5: Spintronics and Magnetoelectronics*, edited by H. Kronmüller and S. P. P. Parkin (John Wiley & Sons Ltd., Chichester, West Sussex, England, 2007).

[6] H. Kronmüller and M. Fähnle, *Micromagnetism and the microstructure of ferromagnetic solids* (Cambridge University Press, Cambridge, New York, 2003).

[7] K. J. Kirk, J. N. Chapman, and C. D. W. Wilkinson, Appl. Phys. Lett. **71**, 539 (1997).

[8] S. Cherifi, R. Hertel, J. Kirschner, H. Wang, R. Belkhou, A. Locatelli, S. Heun, A. Pavlovska, and E. Bauer, J. Appl. Phys. **98**, 0430901 (2005).

[9] M. Bode, M. Heide, K. Bergmann, P. Ferriani, S. Heinze, G. Bihlmayer, A. Kubetzka, O. Pietzsch, S. Blügel, and R. Wiesendanger, Nature **447**, 190 (2007).

[10] S. Lounis, P. H. Dederichs, and S. Blügel, Phys. Rev. Lett. **101**, 107204 (2008).

[11] *Ultrathin Magnetic Structures I*, edited by J. A. C. Bland and B. Heinrich (Springer, Berlin, Heidelberg, New York, 2005).

[12] P. Gambardella, A. Dallmeyer, K. Maiti, M. C. Malagoli, W. Eberhardt, and C. Carbone, Nature **416**, 301 (2002).

[13] C. Dietrich, R. Hertel, M. Huber, D. Weiss, R. Schäfer, and J. Zweck, Phys. Rev. B **77**, 174427 (2008).

[14] R. Hertel, O. Fruchart, S. Cherifi, P. Jubert, S. Heun, A. Locatelli, and J. Kirschner, Phys. Rev. B **72**, 214409 (2005).

[15] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

[16] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).

[17] A. Aharoni, *Introduction to the Theory of Ferromagnetism* (Oxford Science Publications, Clarendon Press, Oxford, 1996).

[18] L. Landau and E. M. Lifshitz, Phys. Z. Sowjet. **8**, 153 (1935).

[19] W. F. Brown, Jr., *Magnetostatic Principles in Ferromagnetism* (North-Holland publishing company, Amsterdam, 1962).

[20] W. F. Brown, Jr., *Micromagnetics* (Interscience Publishers, John Wiley & Sons, New York, London, 1963).

[21] F. Bloch, Z. Physik **74**, 295 (1932).

[22] L. Néel, C. R. Acad. Sci. Paris **241**, 533 (1955).

[23] C. Kittel, Rev. Mod. Phys. **21**, 541 (1949).

[24] E. C. Stoner and E. P. Wohlfarth, Phil. Trans. Roy. Soc. **240**, 599 (1948).

[25] A. Wachowiak, J. Wiebe, M. Bode, O. Pietzsch, M. Morgenstern, and R. Wiesendanger, Science **298**, 577 (2002).

[26] W. K. Hiebert, G. E. Ballentine, L. Lagae, R. W. Hunt, and M. R. Freeman, J. Appl. Phys. **92**, 392 (2002).

[27] C. M. Schneider, A. Kuksov, A. Krasyuk, A. Oelsner, D. Neeb, S. A. Nepijko, G. Schonhense, I. Monch, R. Kaltofen, J. Morais, C. de Nadai, and N. B. Brookes, Appl. Phys. Lett. **85**, 2562 (2004).

[28] J. Vogel, W. Kuch, R. Hertel, J. Camarero, K. Fukumoto, F. Romanens, S. Pizzini, M. Bonfirm, F. Petroff, A. Fontaine, and J. Kirschner, Phys. Rev. B **72**, 220402(R) (2005).

[29] R. D. McMichael, *Standard problem #4* (μMAG Micromagnetic Modeling Activity Group, `http://www.ctcms.nist.gov/~rdm/mumag.html`, 2000).

[30] A. Hubert and R. Schäfer, *Magnetic Domains - The Analysis of Magnetic Microstructures* (Springer, Berlin, New York, Heidelberg, 1998).

[31] *Handbook of Magnetism and Advanced Magnetic Materials Vol. 2: Micromagnetism*, edited by H. Kronmüller and S. P. P. Parkin (John Wiley & Sons Ltd, Chichester, West Sussex, England, 2007).

[32] A. Aharoni, J. Appl. Phys. **83**, 3432 (1998).

[33] R. Kikuchi, J. Appl. Phys. **27**, 1352 (1956).

[34] W. Wernsdorfer, E. B. Orozco, K. Hasselbach, A. Benoit, B. Barbara, N. Demoncy, A. Loiseau, H. Pascard, and D. Mailly, Phys. Rev. Lett. **78**, 1791 (1997).

[35] B. A. Lilley, Phil. Mag. **41**, 792 (11950).

[36] L. Néel, C. R. Acad. Sci., Paris **237**, 1468 (1953).

[37] R. Skomski, *Simple Models Of Magnetism* (Oxford University Press, (United Kingdom), 2008).

[38] R. M. Moon, J. Appl. Phys. **30**, 82S (1959).

[39] S. Middelhoek, J. Appl. Phys. **34**, 1054 (1963).

[40] E. E. Huber Jr., D. O. Smith, and J. B. Goodenough, J. Appl. Phys. **29**, 294 (1958).

[41] A. De Simone, R. V. Kohn, S. Müller, and F. Otto, Mult. Sc. Mod & Sim. **1**, 57 (2003).

[42] A. E. LaBonte, J. Appl. Phys. **40**, 2450 (1969).

[43] A. Hubert, phys. stat. sol. **32**, 519 (1969).

[44] S. Tsukahara and H. Kawakatsu, J. Phys. Soc. Japan **32**, 1493 (1972).

[45] M. R. Scheinfein, J. Unguris, J. L. Blue, K. J. Caokley, D. T. Pierce, R. J. Celotta, and P. Ryan, Phys. Rev. B **43**, 3395 (1991).

[46] R. Hertel and H. Kronmüller, Phys. Rev. B **60**, 7366 (1999).

[47] R. D. McMichael and M. J. Donahue, IEEE Trans. Magn. **33**, 4167 (1997).

[48] M. Kläui, J. Phys.: Condens. Matter **20**, 313001 (2008).

[49] S. S. P. Parkin, U. S. Patent 6834005 (2004).

[50] M. Hayashi, L. Thomas, C. Rettner, and S. S. P. Parkin, Science **32**, 209 (2008).

[51] D. A. Allwood, G. Xiong, M. D. Cooke, C. C. Faulkner, D. Atkinson, N. Vernier, and R. Cowburn, Science **296**, 2003 (2002).

[52] R. Hertel, W. Wulfhekel, and J. Kirschner, Phys. Rev. Lett. **93**, 257202 (2004).

[53] Y. Nakatani, A. Thiaville, and J. Miltat, J. Magn. Magn. Mater. **290**, 750 (2005).

[54] H. A. M. Van Den Berg, J. Appl. Phys. **57**, 2168 (1985).

[55] H. A. M. Van Den Berg, J. Appl. Phys. **60**, 1104 (1986).

[56] H. A. M. Van Den Berg and A. H. J. Van Den Brandt, J. Appl. Phys. **62**, 1952 (1987).

[57] R. D. Gomez, T. V. Luu, A. O. Pak, K. J. Kirk, and J. N. Chapman, J. Appl. Phys. **85**, 6163 (1999).

[58] H. P. Oepen, G. Steierl, and J. Kirschner, J. Vac. Sci. Technol. B **20**, 2535 (2002).

[59] R. Hertel, Z. Metallkd. **93**, 957 (2002).

[60] J. K. Ha, R. Hertel, and J. Kirschner, Phys. Rev. B **67**, 224432 (2003).

[61] W. Rave and A. Hubert, IEEE Trans. Magn. **36**, 3886 (2000).

[62] M. E. Schabes and H. N. Bertram, J. Appl. Phys. **64**, 1347 (1988).

[63] *Spin Dynamics in Confined Structures I*, edited by B. Hillebrands and K. Ounadjela (Springer, Cambridge, New York, 2002).

[64] G. Prinz and K. Hataway, Physics Today **48**, 24 (1995).

[65] R. H. Koch, J. G. Deak, W. Abraham, P. L. Trouilloud, R. A. Altman, Y. Lu, W. J. Gallagher, R. Scheuerlein, and S. S. P. Parkin, Phys. Rev. Lett. **81**, 4512 (1998).

[66] B. C. Choi, M. Belov, G. E. Ballentine, and M. R. Freeman, Phys. Rev. Lett. **86**, 728 (2001).

[67] T. Gerrits, H. A. M. Van den Berg, J. Hohlfeld, L. Bär, and T. Rasing, Nature **418**, 509 (2002).

[68] H. W. Schumacher, C. Chappert, R. C. Sousa, P. P. Freitas, and J. Miltat, Phys. Rev. Lett. **90**, 017204 (2003).

[69] T. Devolder and C. Chappert, J. Appl. Phys. **95**, 1933 (2004).

[70] L. Savtchenko, U.S. patent 6,545,906 B1 (2003).

[71] S. B. Choe, Y. Acremann, A. Scholl, A. Bauer, A. Doran, J. Stohr, and H. A. Padmore, Science **304**, 420 (2004).

[72] K. Y. Guslienko, B. A. Ivanov, V. Novosad, Y. Otani, H. Shima, and K. Fukamichi, J. Appl. Phys. **91**, 8037 (2002).

[73] A. A. Thiele, J. Appl. Phys. **45**, 377 (1974).

[74] B. E. Argyle, E. Terrenzio, and J. C. Slonczewski, Phys. Rev. Lett. **53**, 190 (1984).

[75] M. Bolte, G. Meier, B. Krüger, A. Drews, R. Eiselt, L. Bocklage, S. Bohlens, T. Tyliszczak, A. Vansteenkiste, B. Van Waeyenberge, K. W. Chou, A. Puzic, and H. Stoll, Phys. Rev. Lett. **100**, 176601 (2008).

[76] M. Buess, R. Höllinger, T. Haug, K. Perzlmaier, U. Krey, D. Pescia, M. R. Scheinfein, D. Weiss, and C. H. Back, Phys. Rev. Lett. **93**, 077207 (2004).

[77] R. Hertel and J. Kirschner, J. Magn. Magn. Mater. **272-276**, 655 (2004).

[78] M. H. Park, Y. K. Hong, S. H. Gee, D. W. Erickson, and B. C. Choi, Appl. Phys. Lett. **83**, 329 (2003).

[79] Y. Acremann, C. Back, M. Buess, O. Portmann, A. Vaterlaus, D. Pescia, and H. Melchior, Science **290**, 492 (2000).

[80] E. Feldtkeller, Z. Angew. Phy. **19**, 530 (1965).

[81] T. Shinjo, T. Okuno, R. Hassdorf, K. Shigeto, and T. Ono, Science **289**, 930 (2000).

[82] N. Kikuchi, S. Okamoto, O. Kitakami, Y. Shimada, S. G. Kim, Y. Otani, and K. Fukamichi, J. Appl. Phys. **90**, 6548 (2001).

[83] R. Höllinger, A. Killinger, and U. Krey, J. Magn. Magn. Mater. **261**, 178 (2003).

[84] T. Okuno, K. Shigeto, T. Ono, K. Mibu, and T. Shinjo, J. Magn. Magn. Mater. **240**, 1 (2002).

[85] B. Van Waeyenberge, A. Puzic, H. Stoll, K. W. Chou, T. Tyliszczak, R. Hertel, M. Fähnle, H. Brückl, K. Rott, G. Reiss, I. Neudecker, D. Weiss, C. H. Back, and G. Schütz, Nature **444**, 461 (2006).

[86] Y. Liu, S. Gliga, R. Hertel, and C. M. Schneider, Appl. Phys. Lett. **91**, 112501 (2007).

[87] R. Hertel and C. M. Schneider, Phys. Rev. Lett. **97**, 177202 (2006).

[88] S. Gliga, R. Hertel, and C. M. Schneider, Physica B **403**, 334 (2008).

[89] A. Vansteenkiste, K. W. Chou, M. Weigand, M. Curcic, V. Sackmann, H. Stoll, T. Tyliszczak, G. Woltersdorf, C. H. Back, G. Schütz, and B. Van Waeyenberge, http://arxiv.org/abs/0811.1348v2 (2009).

[90] K. Yamada, S. Kasai, Y. Nakatani, K. Kobayashi, and T. Ono, Appl. Phys. Lett. **93**, 152502 (2008).

[91] K. Yamada, S. Kasai, Y. Nakatani, K. Kobayashi, H. Kohno, A. Thiaville, and T. Ono, Nature Mater. **6**, 269 (2007).

[92] Q. F. Xiao, J. Rudge, B. C. Choi, Y. K. Hong, and G. Donohoe, Appl. Phys. Lett. **89**, 262507 (2006).

[93] K.-S. Lee, S.-K. Kim, Y.-S. Yu, Y.-S. Choi, K. Y. Guslienko, H. Jung, and P. Fischer, Phys. Rev. Lett. **101**, 267206 (2008).

[94] D. D. Sheka, Y. Gaididei, and F. G. Mertens, Appl. Phys. Lett. **91**, 082509 (2007).

[95] R. Hertel, S. Gliga, M. Fähnle, and C. M. Schneider, Phys. Rev. Lett. **98**, 117201 (2007).

[96] L. Berger, Phys. Rev. B **54**, 9353 (1996).

[97] J. C. Slonczewski, J. Magn. Magn. Mater. **159**, L1 (1996).

[98] H. Dassow, R. Lehndorff, D. E. Bürgler, M. Buchmeier, P. A. Grünberg, C. M. Schneider, and A. D. van der Hart, Appl. Phys. Lett. **89**, 222511 (2006).

[99] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, Nature **425**, 380 (2003).

[100] D. V. Berkov and J. Miltat, J. Magn. Magn. Mater. **320**, 1238 (2007).

[101] K.-J. Lee, A. Deac, O. Redon, J.-P. Nozières, and B. Dieny, Nature Mater. **3**, 877 (2004).

[102] S. Zhang and Z. Li, Phys. Rev. Lett. **93**, 127204 (2004).

[103] A. Thiaville, Y. Nakatani, J. Miltat, and Y. Suzuki, Europhys. Lett. **69**, 990 (2005).

[104] K. Ito, T. Devolder, C. Chappert, M. J. Carey, and J. A. Katine, J. Phys. D: Appl. Phys. **40**, 1261 (2007).

D 2 Spin waves in confined systems

S. O. Demokritov
Institut für Angewandte Physik
Westfälische Wilhelm Universität Münster

Contents

1 Introduction **2**

2 Basics of spin waves in bulk materials **3**

3 Spin waves in films **4**

4 Brillouin light scattering technique: a tool for investigation of spin waves **6**

5 Spin waves in magnetic double layers **9**

6 Spin waves in magnetic stripes **11**

7 Conclusions **16**

1 Introduction

The concept of dynamic eigenmodes of unconfined magnetic media called spin waves was introduced by Bloch in 1930 [1]. As shown in Fig. 1, a spin wave represents a wave of spin precession propagating in a magnetically ordered medium. Early experimental evidence for the existence of spin waves came from measurements of thermodynamic properties of ferromagnets, in particular the temperature dependence of their saturation magnetization. The first direct observation of spin precession was made using ferromagnetic resonance by Griffiths in 1946 [2] for the case of uniform precession which can be treated as a spin wave with the zero wave vector. Later inelastic light scattering experiments performed by Fleury, Porto, Cheesman, and Guggenheim [3] in 1966 have confirmed the existence of spin waves.

The history of the spin-wave conception is inseparably linked with the reduction of the dimensions. Being first introduced for bulk media, spin waves in magnetic films have attracted enormous interest 60s and 70s of the last century. It was found that the confinement caused by the finite thickness of a film results in a variety of new effects and even new types of spin waves. The so-called surface spin waves were introduced theoretically by Damon and Eshbach in 1961 [4] and than observed experimentally by Grünberg and Metawe in 1977 [5], which is followed by the investigation of spin waves in magnetic double layres and the experimental observation of the interlayer coupling by Grünberg [6]. This study on spin waves in films was a basic for discovery of the giant magnetoresistance effect by Grünberg [7] and Fert [8] and of a wide application of magnetic films in the information technology.

The study of quantized and localized spin-wave modes of small magnetic elements is vital for understanding of dynamic magnetic properties of magnetic storage and sensors devices. For example, thermal magnetic noise limits the figures of merit of a magnetic reading head, which can be considered as a magnetic resonator. The thermal noise appears at the resonance frequencies of such a resonator which are in the microwave frequency range. Recently, it has been found [9, 10, 11] that the frequency spectrum of the magnetic noise is quantized, i.e. shows well defined maxima at certain frequencies. To control the noise in order to improve the properties of such laterally confined magnetic elements one needs to study the properties of thermally excited spin-wave modes in those systems.

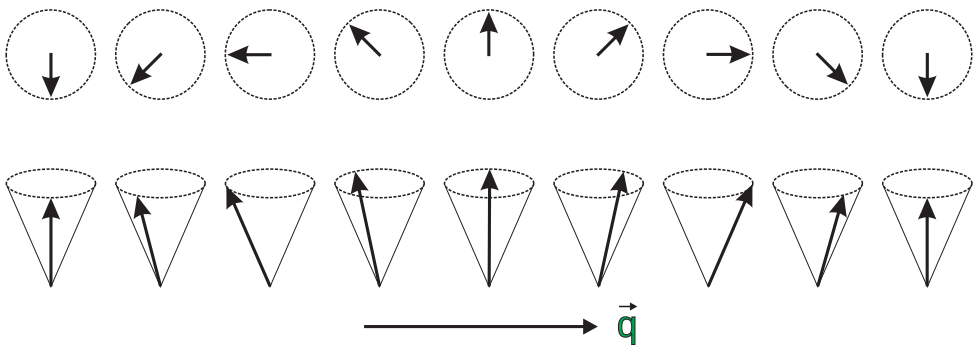


Fig. 1: *Propagating spin wave*

2 Basics of spin waves in bulk materials

In many aspects spin waves can be considered as a magnetic analogue of sound. However, as it will be discussed in this lecture, the main difference between spin waves and sound lies in the very different dispersion laws. Contrary to sound or light, spin waves usually have a gap in their dispersion spectrum which depends on the applied magnetic field.
The magnetization dynamics of a ferromagnet is described by the Landau-Lifshitz torque equation [12]

$$-\frac{1}{\gamma}\frac{d\vec{M}}{dt} = \vec{M} \times \vec{H}_{\text{eff}} \quad , \tag{1}$$

where $\vec{M}$ is the total magnetization, γ is the modulus of the gyromagnetic ratio for the electron spin ($\gamma/2\pi = 2.8\,\text{MHz/Oe}$), and $\vec{H}_{\text{eff}} = -\delta W/\delta\vec{M}$ is the effective magnetic field calculated as a variational derivative of the energy function W, where all the relevant interactions in the magnetic substance have been taken into account.
We assume that due to applied external magnetic field $\vec{H}_{\text{e}} = H_{\text{e}}\vec{e}_z$ the static magnetization in the elements is uniform $\vec{M}_{\text{stat}} = M_{\text{S}}\vec{e}_z = \text{const}$ everywhere, except the narrow regions close to the edges of the elements (edge domains). The total magnetization can be written as follows

$$\vec{M}(\vec{r},t) \approx M_{\text{S}}\vec{e}_z + m_x(\vec{r},t)\vec{e}_x + m_y(\vec{r},t)\vec{e}_y \ , \tag{2}$$

where $\vec{m}$ is the dynamic magnetization. We will neglect the nonlinear effects and we restrict ourselves by the case that $m \ll M_{\text{S}}$.
In general, two important interactions - the exchange interaction and the magnetic dipole interaction - determine the dispersion spectrum of a spin wave. In this case the effective magnetic field can be expressed as

$$\vec{H}_{\text{eff}}(\vec{r},t) = \vec{H}_{\text{e}} + \frac{2A}{M_{\text{S}}^2}\nabla^2\vec{M}(\vec{r},t) + \vec{H}_{\text{dip}}(\vec{r},t);, \tag{3}$$

where A is the exchange constant, ∇^2 is the Laplace operator, $\vec{H}_{\text{dip}}(\vec{r},t)$ is the dipole field, which should be calculated based on the Maxwell equations. Let us emphasize that $\vec{H}_{\text{dip}}(\vec{r},t)$ in Eq. 3 describes both the static and dynamic demagnetizing fields. The effective field $\vec{H}_{\text{eff}}(\vec{r},t)$ can be also presented as a sum of a static and a dynamic part similar to the presentation of the magnetization Eq. 2,

$$\vec{H}_{\text{eff}}(\vec{r},t) = H_{\text{i}}(\vec{r})\vec{e}_z + h_{\text{eff},x}(\vec{r},t)\vec{e}_x + h_{\text{eff},y}(\vec{r},t)\vec{e}_y \ . \tag{4}$$

with $h_{\text{eff}} \ll H_{\text{i}}$.
The dipole-exchange spin-wave dispersion spectrum in an unlimited ferromagnetic medium can be easily calculated using the above approach. In this case H_{i} is uniform over the medium. The spectrum is given by the Herrings-Kittel formula

$$\left(\frac{\omega}{\gamma}\right)^2 = \left(H_{\text{i}} + \frac{2Aq^2}{M_{\text{S}}}\right)\left(H_{\text{i}} + \frac{2Aq^2}{M_{\text{S}}} + 4\pi M_{\text{S}}\sin^2\varphi_q\right) \ . \tag{5}$$

where q is the wavevector of the spin wave, and φ_q is the angle between the direction of the wavevector and the magnetization. Both crystalline and surface anisotropy of the material are neglected in this equation, which is a good approximation for NiFe (permalloy). It is important

to note here that the spin-wave frequency depends on both the absolute value and the direction of its wavevector.

3 Spin waves in films

In a magnetic film with a finite thickness d the spin wave spectrum is modified due to the fact that the translational invariance of an infinite medium is broken in the vicinity of the film surfaces. In the following we assume that the applied magnetic field is in the plane if the film and use a Cartesian coordinate system oriented in such a way that the film normal is along the x-axis and the magnetic field is along the z-axis. The detailed derivation of the equations describing the spectrum of spin waves in films including both the dipole and the exchange interaction for an arbitrary direction of the wavevector based on the approach proposed in [13] can be found in [14]. In a long-wavelength limit ($q_{\|}d < 1$) an approximate expression for the spin wave frequencies of a film can be written in the form, analogous to Eq. 5

$$\omega_p = 2\pi\nu_p = \gamma \left[\left(H + \frac{2A}{M_{\mathrm{S}}} q^2 \right) \left(H + \frac{2A}{M_{\mathrm{S}}} q^2 + 4\pi M_{\mathrm{S}} \cdot F_{pp}(q_{\|}d) \right) \right]^{1/2}, \tag{6}$$

where

$$q^2 = q_x^2 + q_y^2 + q_z^2 = \left(\frac{p\pi}{d}\right)^2 + q_y^2 + q_z^2 = \left(\frac{p\pi}{d}\right)^2 + q_{\|}^2 \tag{7}$$

Here $q_{\|}$ is the continuously varying in-plane wavevector, $F_{pp}(q_{\|}, d)$ is the matrix element of the magnetic dipole interaction, and p = 0,1,2,... is a quantization number. Equation 7 is obtained under boundary conditions of "unpinned" spins on the film surfaces for the dynamic part $\vec{m}$ of the magnetization:

$$\left| \frac{\partial \vec{m}}{\partial x} \right|_{x=\pm d/2} = 0. \tag{8}$$

For a general description of the spin wave modes one can use a more complicated boundary condition instead of the above one [15]:

$$\pm \frac{\partial \vec{m}}{\partial x} + D\vec{m}|_{x=\pm d/2} = 0. \tag{9}$$

with the so-called pinning parameter D determined by the effective surface anisotropy, k_S, and the exchange stiffness constant A: $D = k_S/A$. However, Eq. 8 is applicable if the material of the film does not posses a large surface anisotropy.

To analyze the dispersion Eq. 6 let us divide the spin-wave modes into two groups:

(i) modes corresponding to p = 1,2,... which are called the perpendicular standing spin waves (PSSW). The have a strongly non-uniform distribution of $\vec{m}$ along the film thickness, since even for p = 1 $q_x = \pi/d \gg q_{\|}$ for most reasonable film thicknesses. The frequencies of PSSW are determined by the isotropic exchange interaction and the external magnetic field and can be obtained from Eq. 6 by substituting $F_{pp}(q_{\|}d) = 1$. Note here that the frequencies of the PSSW modes does not depend on the orientation of $q_{\|}$.

(ii) p = 0 modes with quasi-uniform profile along the film thickness. For these modes the exchange interaction can be neglected and their properties are mainly determined by the dipole interaction. Since the dipole interaction is anisotropic, the spin-wave dispersion spectrum in a

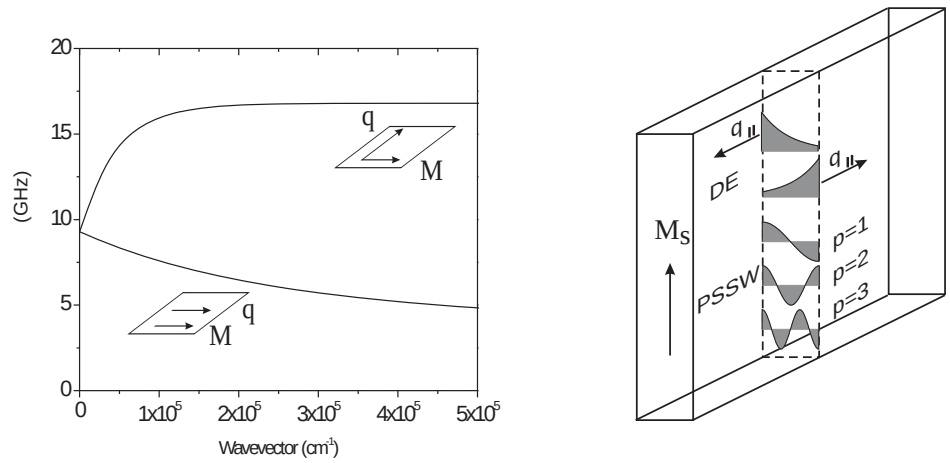


Fig. 2: *Left: Dispersion of spin wave modes in a magnetic film for different mutual orientations of the magnetization, $\vec{M}_{\rm S}$, and the in-plane wavevector, $\vec{q}_{\|}$; Right: Distributions of dynamic magnetization in the Damon-Eshbach-mode with two opposite $\vec{q}_{\|}$ and in the PSSW-modes with different p.*

film is anisotropic as well. For two mostly used limit cases $\vec{q}_{\|} \parallel \vec{M}_{\rm S}$ (the so-called backward volume (BV) geometry) and $\vec{q}_{\|} \perp \vec{M}_{\rm S}$ (the so-called Damon-Eshbach (DE) geopmetry) [4] one obtains the expressions for the spin-wave frequencies:

$$\omega_B V = 2\pi\nu_B V = \gamma \left[H \left(H - 4\pi M_{\rm S}^2 \cdot \frac{1 - exp(-q_{\|} d)}{q_{\|} d} \right) \right]^{1/2}, \tag{10}$$

and

$$\omega_{DE} = 2\pi\nu_{DE} = \gamma \cdot \left[H \cdot (H + 4\pi M_{\rm S}) + (2\pi M_{\rm S})^2 \cdot \left(1 - e^{-2q_{\|} d} \right) \right]^{1/2}. \tag{11}$$

Left panel of Fig. 2 illustrates the spin wave dsipersion for different orientations between the in-plane wavevector, $\vec{q}$, and the magnetization, $\vec{M}_{\rm S}$. Two different geometries are illustrated. If $\vec{q}$ and $\vec{M}_{\rm S}$ are both in the film plane, and if $\vec{q}$ is perpendicular to $\vec{M}_{\rm S}$, the Damon-Eshbach mode exists. If $\vec{q}$ and $\vec{M}_{\rm S}$ are collinear in the film plane, a BV-mode with negative dispersion. The group velocity of the BV-mode is antiparallel to its wavevector.

The distributions of the dynamic magnetization for different modes are illustrated by the right panel of Fig. 2. In the unpinned PSSW-modes the cos-type distribution is determined by $q_x = p\pi/d$. More interesting is the distribution for the Damon-Eshbach-mode. Although $p = 0$ for this mode, the distribution is non-uniform and assymteric: this mode has a larger amplitude at one surface that at another. Therefore it is called the surface mode. As shown in Figure, the assymetry inverses if $\vec{q}_{\|}$ inverses. The assymetry is unimportant if $q_{\|} d \ll 1$. For thin magnetic films with the thickness of about 10 nm this is usually the case.

4 Brillouin light scattering technique: a tool for investigation of spin waves

After getting a basic information about spin waves in magnetic films let us consider Brillouin light scattering (BLS) spectroscopy which is widely used for spin-wave studies since decades. Together with time-resolved Kerr-microscopy and x-ray microscopy BLS recently became one of the most powerful techniques for experimental investigations of inhomogeneous magnetization dynamics in confined systems. BLS has a number of advantages over other techniques. It provides a possibility to investigate spin waves with different absolute values and orientations of their wave vectors in a very broad frequency range from 2 GHz to 500 GHz. Moreover, an important advantage of BLS is that its sensitivity allows for detection of thermally excited spin waves, i.e. there is no need for external excitation sources. However, up to recently, the application of BLS for studies of spin waves in small magnetic elements was restricted by its poor lateral resolution defined by the size of the probing laser spot, which was usually 30-40 μm in diameter. The invention of micro-BLS (μ-BLS), where the laser beam is focused almost down to the theoretical diffraction limit of 200-250 nm and the scattered light can be effectively collected from this small area for further spectral analysis, has opened new perspectives for BLS. Magnetic dynamics of single elements with sub-micrometer sizes is accessible now.
BLS process is illustrated by the left panel of Fig. 3. Photons of energy $\hbar\omega_I$ and momentum $\hbar\vec{q}_I$ creates the elementary quanta of spin waves $(\hbar\omega, \hbar\vec{q})$, which are magnons. The scattered photon looses a part of its energy and changes its momentum:

$$\begin{aligned} \hbar\vec{q}_S &= \hbar(\vec{q}_I - \vec{q}) \\ \hbar\omega_S &= \hbar(\omega_I - \omega), \end{aligned} \tag{12}$$

The process corresponding to creation of the magnon is called the Stokes process. From Eq. 12 it is evident, that the wave vector $\vec{q}_I - \vec{q}_S$, transferred in the scattering process, is equal to the

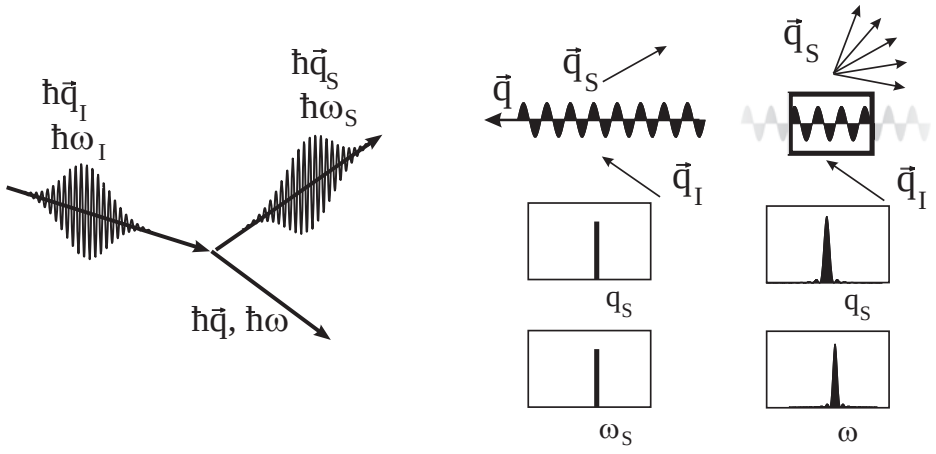


Fig. 3: *Left: Brillouin light scattering process from spin waves (magnons). The wave vectors of the incident and the scattered photons as well as that of the magnon are shown. If the magnon is absorbed, the energy of the photon increases; if the magnon is radiated, the energy of the photon is decreases; Right: Schematics of the light scattering process for infinite scattering volume (plane spin wave) and confined spin-wave*

wave vector $\vec{q}$ of the created magnon. A magnon can also be absorbed by the photon, which in the scattered state has the energy $\hbar(\omega_I + \omega)$ and momentum $\hbar(\vec{q}_I + \vec{q})$. The corresponding process is called the anti-Stokes scattering. For room temperature ($T \gg \hbar\omega/k_B \approx 1K$) both processes have about the same probability. In a classical treatment the scattering process can be understood as follows: Due to magneto-optical effects a phase grating is created in the material by a spin wave. Since the grating propagates with the phase velocity of the spin wave, light is Bragg-reflected from the phase grating with its frequency Doppler-shifted by the spin-wave frequency.

The conservation law Eq. 12 for wave vectors is a consequence of the translational invariance of an infinite medium. If the translational invariance of the scattering medium is broken by confinement effects in one or more directions, the corresponding components of the wave vector, $\vec{q}$, are not fully conserved in the light scattering process, as it is illustrated by the right panel of Fig. 3. In this case the technique looses the resolution with respect to one or another component of the wave vector of the scattering wave. The confinement of a spin wave itself causes uncertainty of its wave vector, which breaks the conservation law Eq. 12 as well. The uncertainty in q_i is, apparently, inversely proportional to the confinement length in the i-direction t_i. For example, for the case of a film the two-dimensional in-plane wavevector is conserved only, the confinement length in the third direction being equal to the thickness of the films, d. In the case of a long stripe, the component of $\vec{q}$ along the stripe axis is conserved only.

The BLS intensity from a spin wave in a magnetic film is determined by the integral:

$$I(q_\parallel) \propto \left| \int_{-\infty}^{\infty} m(\vec{r}) \times \exp(-i\vec{q}_\parallel \vec{r}) d\vec{r} \right|^2 \tag{13}$$

where the two-dimensional integration is performed over the entire film. Thus, the BLS intensity is determined by the Fourier-component of $m(\vec{r})$ at $\vec{q}_\parallel$.

Let us illustrate the above consideration by a BLS spectrum from spin waves in a film. Figure 4 shows a BLS spectrum of a NiFe film obtained at the direction of the transferred in-plane wavevector perpendicular to the applied in-plane field. This arrangement corresponds to the Damon-Eshbach geometry. One sees two peaks in both Stokes and anti-Stokes parts of the spectrum. The low-frequency peaks (about 7 GHz) are connected with the dipole-dominated Damon-Eshbach mode, whereas the high frequency peaks (about 17 GHz) are due to PSSW-modes with $p = 1$. PSSW-modes with higher values of p are beyond the measured spectral interval. One can see the assymetry between the peak intensities of the Damon-Eshbach mode for the Stokes and anti-Stokes parts of the spectrum due to the surface nature of the mode. In fact, at given experimental conditions, i.e., at fixed orientations of $\vec{q}_I$ and $\vec{q}_S$ spin waves with opposite wavevectors participate at the Stokes ($\vec{q}_I - \vec{q}_S$) and anti-Stokes ($\vec{q}_S - \vec{q}_I$) processes. Since, as shown in Fig. 2 the corresponding spin waves have different amplitudes at the films surface, where the light scattering is studied, the intensities of the Stokes and anti-Stokes peaks are different as well.

Let us now consider a magnetic stripe with its axis aligned along the z-axis. It was shown in [16], that by calculating the intensity of the light scattered by spin-wave modes in a stripe for a transferred wavevector along the stripe axis (the z-axis) an one-dimensional equivalent of the integral Eq. 13 can be used. However, for a transferred wavevector along the y-axis the BLS intensity is connected with the profile of the dynamic magnetization, $m(y)$ of the spin-wave

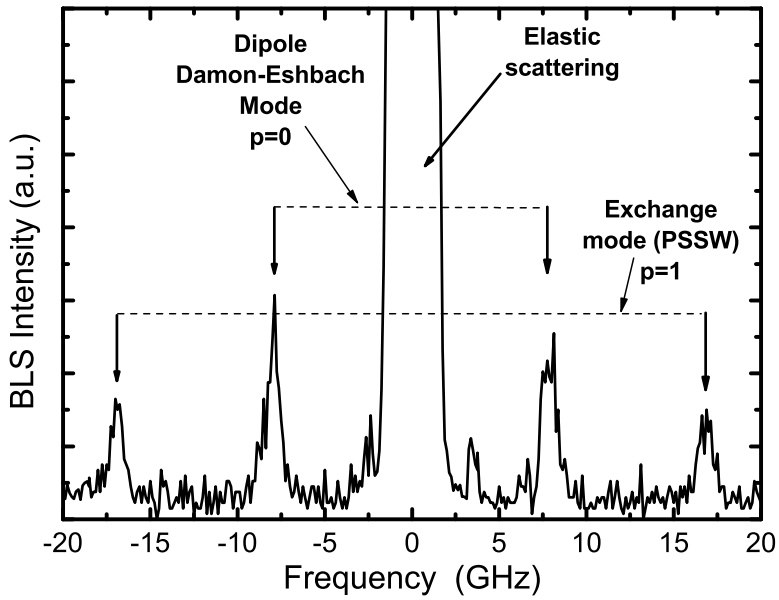


Fig. 4: *BLS spectrum of a NiFe film with a thickness of 30 nm in the Damon-Eshbach geometry for $q_{\|} = 0.2 \cdot 10^5 cm^{-1}$ and $H = 600$ Oe. The part of the spectrum with negative frequencies occurs due to creation of a spin wave in the BLS process (Stokes), whereas positive frequencies correspond to absorption of a spin wave (anti-Stokes). The exchange- and the dipole-dominated modes are detected.*

mode confined in the stripe.

$$I(q) \propto \left| \int_{-w/2}^{w/2} m(y) \times \exp(-iqy) dy \right|^2 \tag{14}$$

where w is the stripe width. Thus, the light scattering intensity is proportional to the squared Fourier transform of the dynamic magnetization, $m(y)$.

Technically, to study BLS light of a single-frequency laser is focused onto the sample by an objective lens. The light scattered from the sample (elastic and inelastic contributions) is collected and sent to a Fabry-Perot interferometer [17, 18]. In order to obtain the high contrast necessary to detect the weak inelastic signals, the light passes the interferometer several times guided by a system of retroreflectors and mirrors. The frequency selected light transmitted by the interferometer is detected by a photomultiplier. Data collection is performed by a multichannel analyzer realized using a personal computer.

A recent development of the BLS-technique is the so-called micro-BLS, allowing direct spatial resolution of spin-wave modes close to 250 nm. [19]. The sample is mounted on a xyz piezoelectric stage, which provides the sample positioning along all three dimensions. Taking into account the uncertainty in the wavevector of the scattered light caused by the spatial confinement, the micro-BLS setup does not provide any wave vector selectivity. Instead the contributions from all spin waves up to a certain wavevector are automatically integrated.

5 Spin waves in magnetic double layers

Let us consider a magnetic layered system consisting of two ferromagnetic layers separated by a nonmagnetic spacer. Interaction between two magnetic layers can be phenomenologically described by:

$$W = W_1 + W_2 - J_1 \cos\phi - J_2 \cos^2\phi \tag{15}$$

where W_1 and W_2 describe the energies per surface unit of each layer and the last two terms correspond to the magnetic coupling interface energy between the two layers. Angle ϕ is the angle between the magnetizations of two magnetic layers and the parameters J_1 and J_2 represent the strength of the bilinear and biquadratic coupling, respectively [20]. If J_1 dominates and negative, the antiparallel orientation of two magnetization vectors is energetically favorable. If J_2 dominates and is negative, it promotes perpendicular ($90°$) orientation of the two magnetization vectors. The microscopic origin of the bilinear coupling is a long-range interaction between the magnetic moments via conduction electrons of the spacer. For smooth interfaces the strength of the coupling oscillates as a function of the spacer thickness [21], whereas the biquadratic terms is negligible. Below we consider the bilinear coupling only.

The interlayer coupling not only influences the magnetic ground state of the system, but together with the dipole interaction it mixes the spin-wave modes in two magnetic layers as well. One can imagine collective spin-wave modes of two different symmetries in such a double layer as it is shown in Fig. 5. The dynamic magnetizations of two ferromagnetic layers in the symmetric (acoustic) mode at the interface oscillate in phase, whereas in the anti-symmetric (optic) mode they oscillate out-of-phase.

In contrast to intralayer bulk exchange which contributes to $\vec{H}_{\text{eff}}$ (cf. Eq. 3), the interlayer exchange coupling, due to its local, interface charachter, appears only in boundary conditions at the interface between the ferromagnetic layers.

If one assumes for simplicity that the saturation magnetizations and the intralayer exchange constants A_1 and A_2 are the same and the interface anisotropy can be neglected, the boundary conditions for the dynamic magnetizations of two layers at the interface can be written as

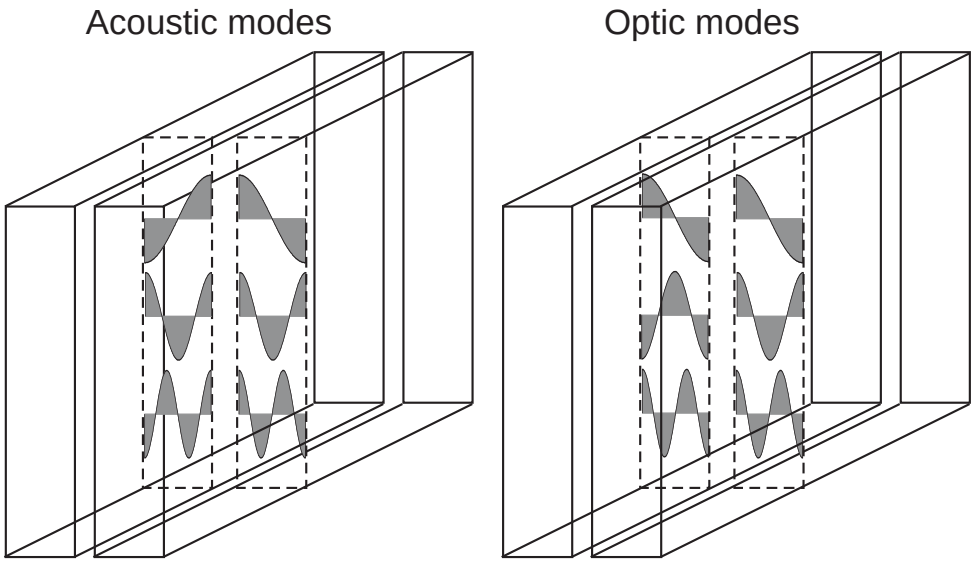


Fig. 5: *Magnetic layered system consisting of two ferromagnetic layers. Profiles of symmetric (acoustic) and anti-symmetric (optic) spin-wave modes are illustrated.*

follows:

$$\frac{\partial \vec{m}_1}{\partial x} + D_{\rm int}(\vec{m}_1 - \vec{m}_2)|_{x=0} = 0.$$

$$\frac{\partial \vec{m}_2}{\partial x} + D_{\rm int}(\vec{m}_2 - \vec{m}_1)|_{x=0} = 0. \qquad (16)$$

where $D_{\rm int} = J_1/A$. As indicated in Fig. 5 there exist two kind of solutions fulfilling the boundary conditions Eq. 16:
(i) the symmetric in-phase mode at which the mode profiles are continuous at the interface $(\vec{m}_1 - \vec{m}_2)|_{x=0} = 0$. This solution does not depend on the strength of the interlayer coupling J_1 and coincides with a spin-wave mode of a single magnetic film with the thickness equal to $2d$.
(ii) the antisymmetric, out-of-phase solution, where two magnitizations are oscillating out-of-phase. This mode strongly depends on the strength of the interlayer coupling. For $J_1 > 0$ (ferromagnetic coupling) the frequency of the out-of-phase mode increases, whereas for $J_1 < 0$ (antiferromagnetic coupling) the frequency of the out-of-phase mode decreases.
One should also mentioned that even for $J_1 = 0$ the frequencies of the symmetric and anti-symmetric modes are different. The reason of that is the difference in the dynamic dipole field. This field is weaker for the out-of-phase mode, since the out-of-phase dynamic magnetizations of two layers effectively compensate the dipole fields of each other.
One can conclude from Eq. 16 that the interlayer coupling causes pinning at the interface. For strong coupling this pinning becomes essential. For $J_1 \simeq A/a$, where a is the lattice constant,

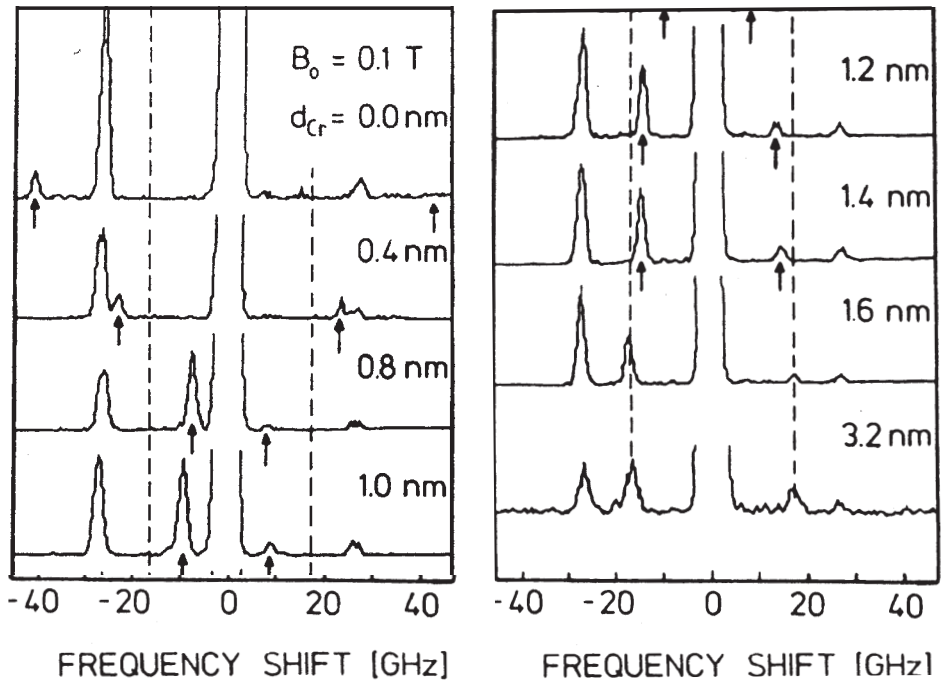


Fig. 6: *BLS spectra of Fe/Cr/Fe layered system with different thickness of the Cr spacer as indicated. The arrows point out the and anti-symmetric (optic) spin-wave modes, which are sensititve to the interlayer coupling. The dash lines mark the position of the mode for vanishing interlayer coupling (adapted from [6]).*

the profile of the out-phase-mode is close to the profile of the exchange PSSW mode of the single layer with the doubled thickness.

The dependence of the out-of-phase mode on the interlayer coupling provides an elegant way to investigate the interlayer coupling by studying spin waves. In real layered magnetic systems a large direct ferromagnetic-type interlayer coupling for a relatively thin non-magnetic spacer is often observed. This occurs due to magnetic 'bridges' in the spacer, i.e. direct contacts between two magnetic films. Recent achievements in growth methods make it possible to avoid 'bridges' in the best samples for spacer thicknesses more than 0.2-0.4 nm. Therefore, the interlayer exchange coupling observed for thicker spacers is thought to be connected with an indirect coupling between two ferromagnetic layers transferred by the carriers of the non-magnetic spacers. The antiferromagnetic type of coupling across a non-magnetic spacer (which is obviously not connected with 'bridges') was discovered by means of BLS in the Fe/Cr/Fe system [6]. BLS spectra for Fe/Cr/Fe samples with different thicknesses of the spacer are shown in Fig. 6.

As expected, two spin-wave modes were detected by means of light scattering. The first one is the in-phase mode. Its frequency is independent of the spacer thickness. The second mode corresponds to the out-of-phase precession of magnetic moments in two Fe films and its frequency is sensitive to the interlayer exchange coupling. The dash lines mark the position of the mode for vanishing interlayer coupling. A clear frequency downshift of the second mode is pronounced for d_{Cr} = 0.8-1.2 nm. The fact that this mode is observed at a frequency which is lower than the value corresponding to zero interlayer coupling can only be interpreted as indicating that the interlayer exchange coupling has become antiferromagnetic.

6 Spin waves in magnetic stripes

The problem of the calculation of a spin wave spectrum for an axially magnetized infinite ferromagnetic stripe with a rectangular cross-section has not been solved analytically yet. However, in the particular case of a thin stripe with $d \ll w$, where d is the thickness of the stripe and w is its width, the spectrum of long-wavelength magnetic excitations can be calculated approximately using the theory of dipole-exchange spin waves in a magnetic film [13], mentioned above.

Let us consider a magnetic stripe magnetized in plane along the z-direction and having a finite width w along the y-direction as shown in Fig. 7. A boundary condition similar to Eq. 9 at the lateral edges of the stripe should be imposed:

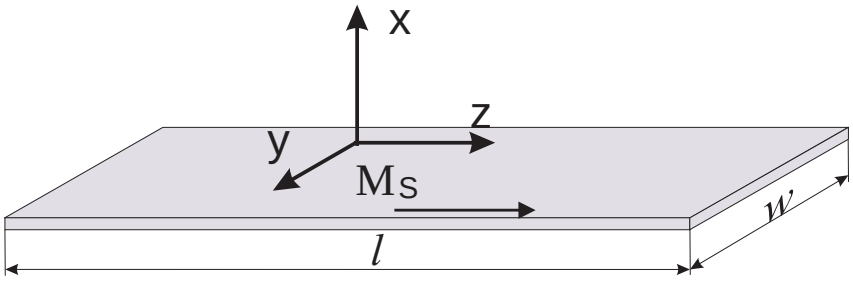


Fig. 7: *Used coordinate system for the longitudinally magnetized stripes. Note that the z-axis is parallel to the direction of the static magnetization $\vec{M}_S$. The lateral dimensions of the stripe $w \ll l$ are indicated as well.*

$$\left| \frac{\partial \vec{m}}{\partial y} + D_{dip} \vec{m} \right|_{y=\pm w/2} = 0. \tag{17}$$

Note here, that contrary to the boundary conditions on the film surfaces Eq. 9 the inhomogeneous dynamic dipole field mainly contributes to the pinning coefficient D_{dip} at the stripe edges [22], which depends on the aspect ratio w/d

$$D_{dip} = \frac{2\pi w/d}{1 + 2\ln(w/d)} \tag{18}$$

For $w \gg d$ this pinning is rather strong.
An additional quantization of the y-component of $\vec{q}_{\|}$ is then obtained:

$$q_{y,n} = \frac{n\pi}{w}, \tag{19}$$

where $n = 1, 2, \ldots$. Using Eqs. 6, 11, and the quantization expression Eq. 19 one can calculate the frequencies of these so-called width- (or laterally quantized) modes. The profile of the dynamic part of the magnetization $\vec{m}$ in the n-th mode can be written as follows:

$$\vec{m}_n(y) = \vec{a}_n \cdot \sin\left(q_{y,n}\left(y + \frac{w}{2}\right)\right), \qquad -\frac{w}{2} < y < \frac{w}{2}. \tag{20}$$

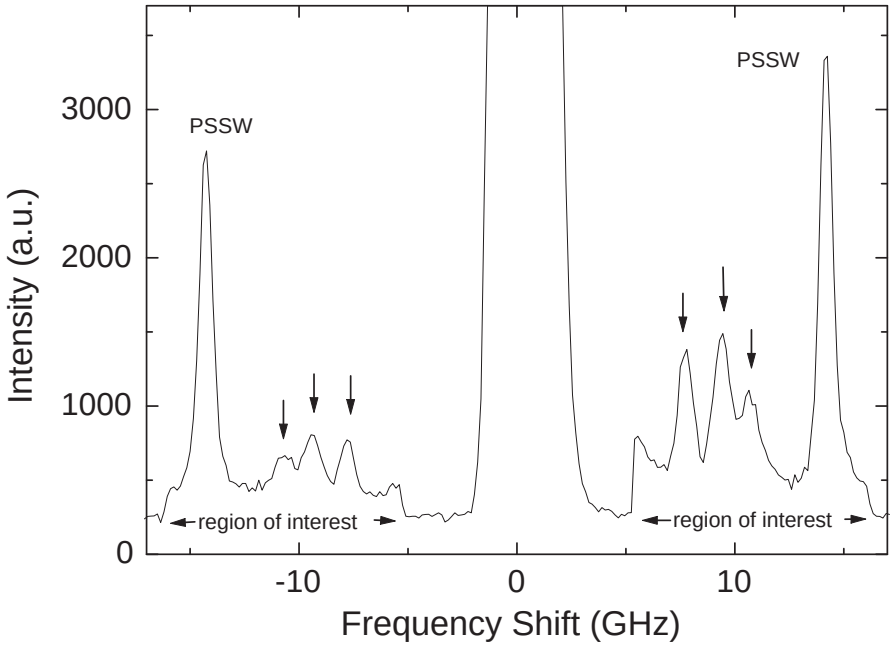


Fig. 8: *Experimental BLS spectrum obtained from the stripe array with a stripe thickness of 40 nm, a stripe width of 1.8 μm. The applied field is 500 Oe orientated along the stripe axis. The transferred wavevector of $q_{\|} = 0.3 \cdot 10^5$ cm^{-1} is oriented perpendicular to the stripes. The discrete spin-wave modes are indicated by arrows. In the "region of interest" (5-17 GHz) the scanning speed was reduced by a factor of three increasing the number of recorded photons by the same factor.*

Equation 20 describes a standing mode consisting of two counterpropagating waves with quantized wavevenumbers, $q_{y,n}$. Note here, that due to the truncation of the sin-function at the stripe boundaries the modes are no more infinite plane waves and the quantized values $q_{y,n}$ are not true wavevectors.

Mathieu et al. [23] and Jorzick et al. [16] investigated spin wave excitations by BLS in arrays of FeNi stripes and observed the above quantization of the spin wave mode. This quantization is caused by a confinement effect of the spin waves in each stripe. It was not seen in the geometry when the wavevector of the detected mode was parallel to the stripes.

The samples were made of 20-40 nm thick $Ni_{81}Fe_{19}$ films deposited in UHV onto a Si(111) substrate by means of e-beam evaporation. Several types of periodic arrays of stripes with stripe widths 1-1.8 μm were prepared. The length L of the stripes was 500 μm. The patterned area was 500×500 μm.

In a BLS experiment with backscattering geometry the in-plane wavevector $\vec{q}_{\parallel} = (\vec{q}_S - \vec{q}_I)_{\parallel}$, transferred in the light scattering process, was oriented perpendicular to the stripes, and its value was varied by changing the angle of light incidence, θ, measured from the surface normal: $q_{\parallel} = (4\pi/\lambda_{Laser}) \cdot \sin\theta$. It is important to mention here again, that, strictly speaking, due to spin wave confinement in a stripe the transferred wavevector $\vec{q}_{\parallel}$ cannot be considered as the wavevector of the spin wave mode taking part in the scattering process and, thus, tested in the experiment. Because of the confinement the spin wave mode does not possess a well-defined wavevector.

Figure 8 shows a typical BLS spectrum obtained in the stripe array. The spectrum contains four distinct modes near 7.8, 9.3, 10.4, and 14.0 GHz, the latter being the PPSW mode, corresponding to $p = 1$. Note here, that in the region of interest (5-17 GHz) the scanning speed of the interferometer was reduced by a factor of three to increase the accumulation time in this region and, thus, to improve the signal-to-noise ratio.

Figure 9 shows BLS spectra obtained for an array of 1 μm wide and 35 nm thick permalloy stripes in a longitudinal magnetic field of 500 Oe for different transversal components of the

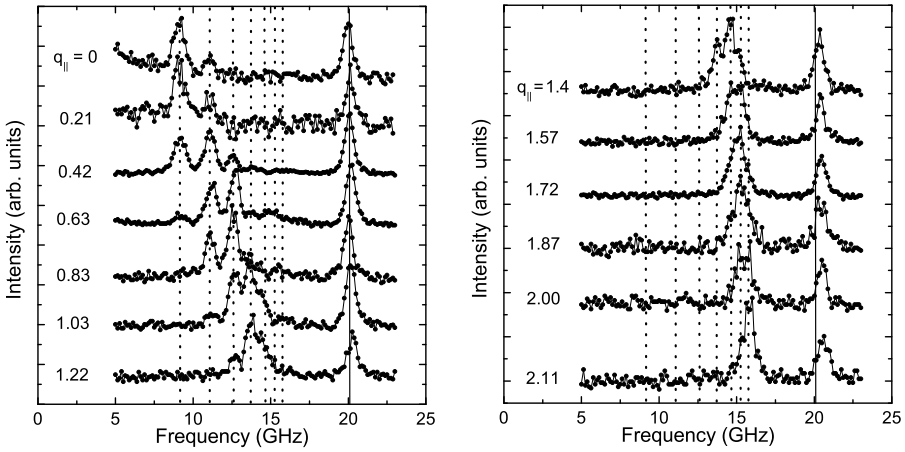


Fig. 9: *BLS spectra of longitudinally magnetized 1 μm wide and 35 nm thick stripes for different transferred wavevectors in an external field of 500 Oe. The transversal component of the wavevector $q_{\parallel}$ is given in units of $10^5 cm^{-1}$. The vertical solid line indicates the nearly constant frequency of the PSSW mode, the vertical dotted lines indicate the fixed frequencies of the quantized spin-wave modes.*

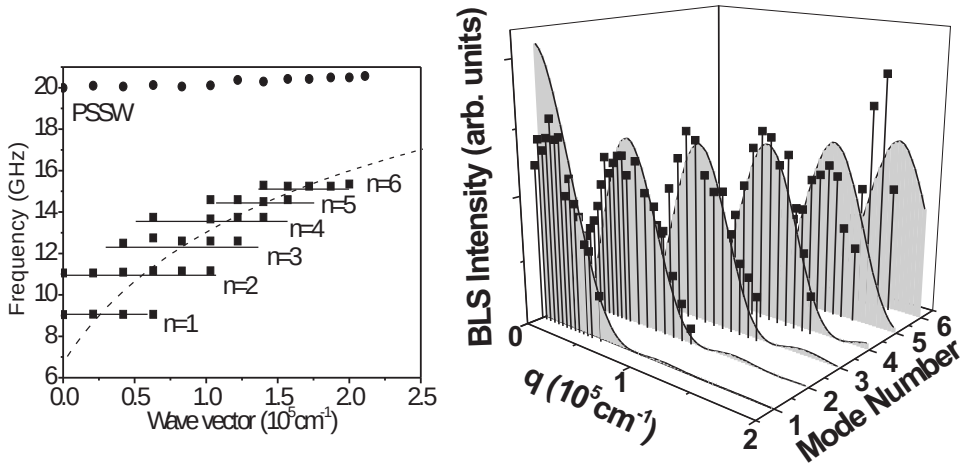


Fig. 10: *Left: Frequencies of the spin waves in longitudinally magnetized 1 μm wide and 35 nm thick stripes in an external field of 500 Oe. The squares correspond to the quantized Damon-Eshbach modes, the circles to PSSW. The horizontal solid lines are the frequencies of quantized spin-wave modes calculated using Eq. 19. Right: Measured BLS-intensities for different quantized modes (black squares) as a function of the transferred wavevector $q_{\|}$ and the mode quantization number m for 1 μm wide stripes magnetized by the static magnetic field of 100 Oe in comparison to the results of calculations based on Eq. 14(gray colored curves).*

wavevector as indicated. One can see from the figure that the PSSW mode (note a growth of the PSSW-frequency with respect to that shown in Fig. 8 due to smaller thickness of the stripe) is observed for the entire investigated range of $q_{\|}$.

Having a more close look into the low-frequency part of the spectra one can conclude that contrary to a continuous film (cf. Fig. 2) the frequencies of the observed peaks are not varying continuously when the wavevector transferred in the BLS-process is growing. Instead the frequencies are quantized, i.e., the modes behave like discrete standing wave resonances. Every discrete mode is observed over a continuous range of the transferred wavevector $q_{\|}$. The lowest mode has a maximum of its intensity close to $q_{\|} = 0$, the higher modes reach their maximam at higher values of $q_{\|}$.

The above conclusions are corroborated by Fig. 10 (left panel), where the measured frequencies of the quantized modes together with that of the PSSW mode are shown as a function of the wavevector . Horizontal solid lines represent the calculated frequencies of the Damon-Eshbach modes quantized due to the confinement caused by the finite width of the stripe. The quantization conditions Eq. 19 were used for the calculation. The dashed line shows the dispersion for spin waves in an unpatterned permalloy film calculated using Eq. 11.

The theory predicts not only the frequencies of the quantized modes, but their profiles as well. The comparison between the experimentally measured BLS intensities and the results of the calculation based on Eqs. 20 and 14 is presented on the right panel of Fig. 10. Shown are the BLS intensities from different quantized modes normalized to the intensity of the PSSW mode. Squares in the figure correspond to the experimentally measured intensities. The gray curves represent the calculated intensities using Eq. 14. The only used fitting parameter is the scaling factor along the intensity axis which is the same for all modes. The mode profiles have been calculated taking into account the dipolar pinning. The very good agreement between the experimental data and the results of the calculation confirms that the observed spin-wave modes

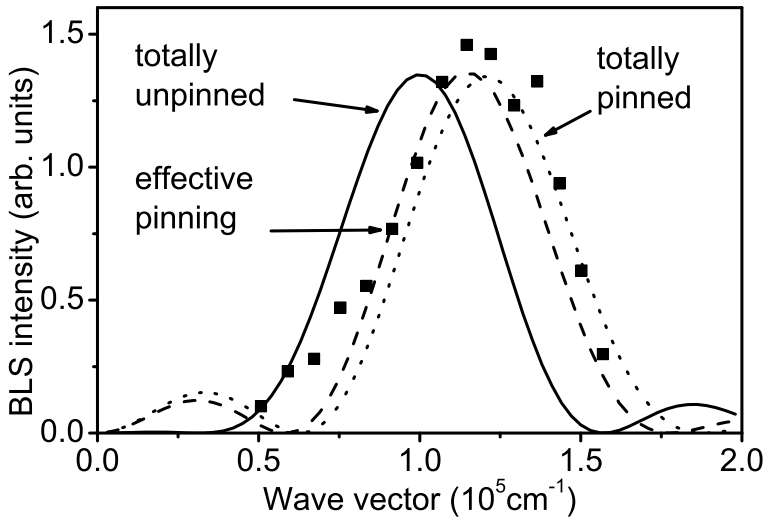


Fig. 11: *BLS intensity as a function of the transferred wavevector for the mode $m = 4$. The squares indicate the experimental data while the lines correspond to the Fourier transforms of three different mode profiles: totally unpinned (full line), totally pinned (dotted line), effectively pinned (dashed line).*

are in fact quantized Damon-Eshbach modes and justifies the used boundary conditions Eq. 17. Figure 11 addresses the BLS intensity for a given mode ($m = 4$) in more detail. Shown are the

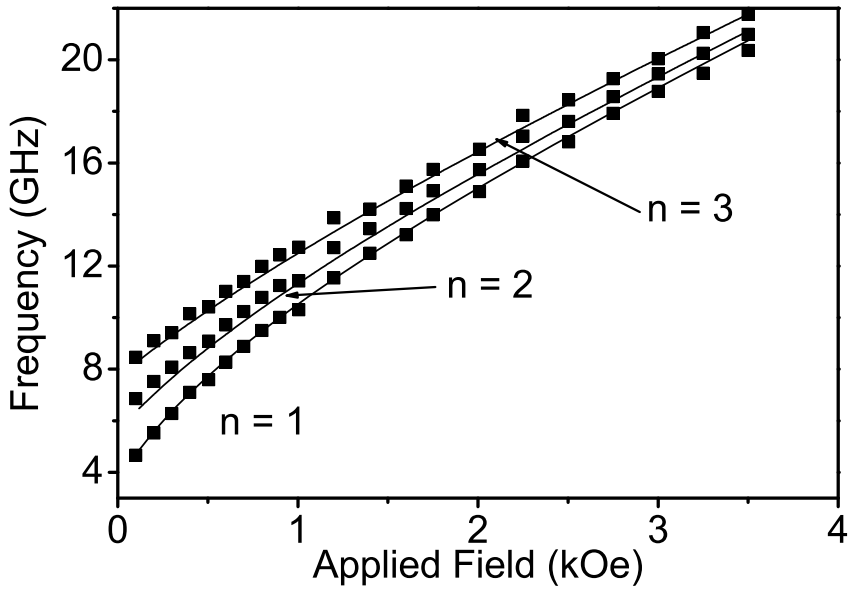


Fig. 12: *Spin-wave frequency of longitudinally magnetized 1.8 μm wide and 40 nm thick stripes as a function of the applied field. The transferred wavevector is $q = 0.3 \times 10^5$ cm^{-1}.*

calculated intensities corresponding to three pinning conditions: free spins ($D = 0$), effective dipole pinning from Eq. 18, and complete pinning ($D = \infty$). It is evident from Fig. 11, that the unpinned boundary condition does not describe the experimental data. However, it is impossible to distinguish experimentally between the effective dipolar pinning and the complete pinning. This is due to a large aspect ratio of the stipe $w/d \gg 1$ causing a strong pinning.
The field dependence of the frequencies of the quantized spin-wave modes is shown in Fig. 12. The solid lines shown in the figure are calculated based on the above described approach. The figure is another demonstration of a nice agreement between the theory and the experiment.

7 Conclusions

We have considered spin waves, which are dynamic eigen-excitations of a magnetic media, for ferromagnets of different geometries. Their properties are determined by the two main interactions: the exchange and the magnetic dipole interaction. Independently of the geometry of the medium, the main difference of a spin wave with respect to light or sound is its dispersion spectrum with a gap at the zero wavevector. The gap is mainly determined by the applied magnetic field. Due to the anisotropic nature of the dipole interaction the dispersion of spin waves is anistropic as well. For unrestricted bulk ferromagnets spin waves represent plane waves with well defined wavevectors.
Reducing the dimensionality of the problem, one inevitably modifies the spin waves. In the case of a film spin-wave quantization due to finite thickness of the films occurs. The corresponding perpendicular standing spin-wave modes posses a strongly non-uniform mode profile along the film thickness. In addition, for the wavevectors which are parallel to the applied in-plane magnetic field spin-wave modes with a negative dispersion exists. Moreover, a new surface-localized Damon-Eshbach spin-wave mode appears. This mode has larger amplitude of the dynamic magnetization at one surface that at another. Considering two films separated by a thing non-magnetic spacer, one obtains hybridized two types spin-wave modes, mixing the dynamic magnetizations of the two films.
Considering quasi-one-dimensional stripes, one find, that lateral quantization due to finite width of the stripe further modifies the spin-wave properties. Laterally quantized spin-wave modes appear.
The above results were experimentally confirmed by means of the Brillouin light scattering spectroscopy, which is a powerful tool for investigation of spin-wave dynamics in different geometries.

References

[1] F. Bloch , Z. Phys. **61** 206 (1930).

[2] J. H. E. Griffits, Nature **158** 670 (1946).

[3] P. A. Fleury, S. P. S. Porto, L. E. Cheesman, and P. Guggenheim, Phys. Rev. Lett. **17** 658 (1966).

[4] R. W. Damon and J. R. Eshbach, J. Phys. Chem. Solids **19**, 308 (1961).

[5] P. Grunberg and F. Metawe, Phys. Rev. Lett. **39** 1561 (1977).

[6] P. Grunberg, R. Schreiber, Y. Pang, M. B. Brodsky, and H. Sowers, Phys. Rev. Lett. **57** 2442 (1986).

[7] G. Binasch, P. Grunberg, P. Saurenbach, and W. Zinn, Phys. Rev. **B39** 482 (1989).

[8] M. N. Baibich, J. M. Broto, A. Fert, F. Nguen Van Day, F. Petroff, P. Etienne, G. Creuzet, A. Frederich, and J. Chazelas, Phys. Rev. Lett. **61** 2472 (1988).

[9] N. Smith, P. Arnett, Appl. Phys. Lett. **78** 1448 (2001).

[10] J.-G. Zhu, J. Appl. Phys. **91** 7273 (2002).

[11] Y. Zhou, A. Roesler, J.-G. Zhu, J. Appl. Phys. **91** 7276 (2002).

[12] L. Landau and E. Lifshitz, Phys. Z. Sowjetun. **8** 153 (1935).

[13] B. A. Kalinikos and A. N. Slavin, J. Phys. C: Solid State Phys. **19** 7013 (1986).

[14] S. O. Demokritov, B. Hillebrands, and A. N. Slavin, Phys. Rep. **348** 441 (2001).

[15] G. T. Rado and J. R. Weertman, J. Phys. Chem. Solids **11** 315 (1959).

[16] J. Jorzick, S. O. Demokritov, C. Mathieu, B. Hillebrands, B. Bartenlian, C. Chappert, F. Rousseaux, A. N. Slavin, Phys. Rev. **B 60** 15194 (1999).

[17] J. R. Sandercock, *Trends in Brillouin scattering: Studies of opque materials, supported films, and central modes, in: Light Scattering in Solids III* (M. Cardona and G. Güntherodt (eds.), Springer Verlag, Berlin, Heidelberg, New York, 1982).

[18] B. Hillebrands, Rev. Sci. Instrum. **70** 1589 (1999).

[19] V. E. Demidov, S. O. Demokritov, B. Hillebrands, M. Laufenberg, P.P. Freitas, Appl. Phys. Lett. **85** 2866 (2004).

[20] M. Ruhrig, R. Schäfer, A. Hubert, R. Mosler, J. A. Wolf, S. Demokritov, P. Grünberg, Phys. Stat. Sol(a) **125** 635 (1991).

[21] S. S. P. Parkin, Phys. Rev. Lett. **67** 3558 (1991).

[22] K. Yu. Guslienko, S. O. Demokritov, B. Hillebrands, A. N. Slavin, Phys. Rev. **B 66** 132402 (2002).

[23] C. Mathieu, J. Jorzick, A. Frank, S. O. Demokritov, A. N. Slavin, B. Hillebrands, B. Bartenlian, C. Chappert, D. Decanini, F. Rousseaux, E. Cambrill, Phys. Rev. Lett. **81** 3968 (1998).

D 3 Spin-Transfer Torque Dynamics

Daniel E. Bürgler
Institut für Festkörperforschung
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Spin-transfer torque (STT) **3**
2.1 Phenomenology 3
2.2 Physical picture: Absorption of the transverse spin current component 5
2.3 Role of diffusive transport 7
2.4 Experimental confirmation of physical picture 8

3 Extended Gilbert equation **9**
3.1 Basics of magnetization dynamics 9
3.2 STT extension of the Gilbert equation 11

4 Current-driven magnetization dynamics **12**
4.1 Steady-state oscillatory modes 12
4.2 Spin-torque nano-oscillators 16

5 STT in epitaxial Fe/Ag/Fe(001) nanopillars **17**
5.1 Sample fabrication 18
5.2 Slonczewski's model 19
5.3 Two-step switching process 21
5.4 Zero-field excitations in the 90°-state 23

6 Spin-transfer torques in non-uniform magnetization structures **25**
6.1 Current-driven domain wall motion 25
6.2 Current-driven vortex motion and switching 26
6.3 GMR and current-induced torques in metallic antiferromagnets 26

7 Conclusions **29**

1 Introduction

Magnetization dynamics and magnetization reversal as one particular dynamic process have become fundamental issues of the physics of magnetic (nano)particles. The research field is driven by the needs of current and perspective technologies in magnetic data storage and spintronics. The magnetic random-access memory (MRAM) and patterned, ultra-high-density perpendicular recording media are examples, where small magnetic elements have to be magnetically switched. Furthermore, most envisaged spintronic devices, *e.g.* the spin transistor, rely on spin-polarized currents generated or analyzed by ferromagnetic electrodes. Control of these current polarizer and analyzer elements involves manipulating their magnetization directions by means of external stimuli.

In the conventional approach, magnetization dynamics, which may eventually trigger a magnetization reversal, is induced by applying an external magnetic field. As the structures get smaller and smaller it becomes increasingly more difficult to focus an external magnetic field to a single nanoobject because magnetic fields are of dipolar nature and, thus, long ranged. Furthermore, the required currents to generate these fields increase and the process becomes inefficient. Therefore, one would prefer a excitation scheme based on an electric current or voltage, which can be applied very locally by means of lithographically defined electrodes.

Any new concept to manipulate magnetization has to fulfill certain criteria in order to be relevant for applications. The ever increasing clock frequencies in devices for information technology require cycle times below 1 ns. Thus, we deal here with a timescale of roughly 100 ps. Ultra-fast magnetization dynamics on even faster timescales down to the fs-regime, which come into play when the electronic and hence the magnetic system is highly excited and driven away from the equilibrium by ultra-short high-power laser pulses, is currently the subject of intense research and is discussed in Theo Rasing's lecture "Optically induced spin dynamics". As a further requirement, the switching processes should dissipate as little energy as possible in order to reduce the heat load on the devices and the power consumption of (battery-powered) equipment. Additionally, a clear potential for down-scaling and compatibility with semiconductor technology are also criteria for competitiveness.

All these requirements –speed, low energy dissipation, local addressing, scalability and integrability into semiconductor technology– demand for new and advanced concepts to induce magnetization dynamics, in particular magnetization switching.

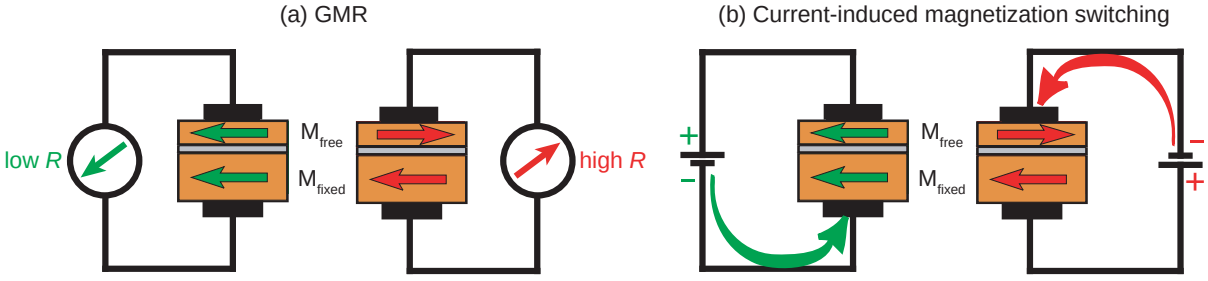


Fig. 1: *Phenomenology of (a) GMR and (b) current-induced magnetization switching. (a) The electric resistance of a trilayer structure consisting of two ferromagnets separated by a non-magnetic, metallic interlayer depends on the alignment of the layer magnetizations. (b) The stable alignment of the magnetizations depends on the polarity, i.e. the direction, of the current flowing perpendicularly through the trilayer.*

The giant magnetoresistance effect (GMR), which is introduced in the lecture "Magnetoresistance in layered structures" of Peter Grünberg, describes the phenomenon that the alignment of the layer magnetizations in a trilayer controls the electric resistance, *i.e.* the current flow [Fig. 1(a)]. According to Newton's third law "*actio* equals *reactio*" there should also exist an inverted effect, for which the current flow controls the magnetization alignment [Fig. 1(b)]. Such an effect is of outmost interest as it represents a novel, alternative concept to induce magnetization switching or more generally magnetization dynamics, in particular in nano-scale magnets.
As we will see below, the current-induced magnetization switching as sketched in Fig. 1(b) can be traced back to the transfer of spin momentum from one magnetic layer to the other, which is carried by the current. The transferred momentum acts as a torque on the local magnetization and may excite a dynamic response of the magnetization. Therefore, we talk about "Spin-transfer torque dynamics". It turns out that the dynamics that can be excited reaches far beyond simple switching events. Steady-state periodic motions of the magnetization with a large dynamic component $|\vec{M_{dynamic}}(t)|$ comparable to or larger than the static component $|\vec{M_{static}}|$ can be excited. For instance, precessional-type motions with opening angles exceeding 90° or even 180° can appear, which are inaccessible by field excitation. Therefore, spin-transfer torques enable novel, highly non-linear magnetization dynamics, which is of interest from a fundamental scientific point of view as well as in the context of applications as spin-transfer-torque microwave oscillators and spin-wave emitters. A comprehensible review article on spin-transfer torques and dynamics by Stiles and Miltat can be found in Ref. [1].

2 Spin-transfer torque (STT)

In 1996 Slonczewski [2] and Berger [3] predicted that a spin-polarized current propagating into a ferromagnetic layer exerts a torque on the magnetization of the layer, due to the exchange interaction between the electrons and the local magnetic moments. In layered metallic systems with alternating magnetic and non-magnetic layers, a current flowing perpendicular to the plane of the layers (CPP-geometry) is polarized by one ferromagnetic layer and transfers spin angular momentum to another ferromagnetic layer, where the transferred momentum acts as a torque on the magnetization, the so called *spin-transfer torque*. For this torque to be sufficient to perturb the magnetization from equilibrium, large current densities ($> 10^7$ A/cm^2) are required.

2.1 Phenomenology

If two stable equilibria for the magnetization exist (*e.g.* due to an uniaxial anisotropy), the spin-transfer torques due to currents of opposite polarity can reversibly switch the magnetization from one equilibrium position to the other. This process is called *current-induced magnetization switching* and does not require an external magnetic field to induce switching.
The phenomenology of current-induced magnetization switching is shown in Fig. 1(b). We consider two ferromagnetic layers separated by a non-ferromagnetic spacer with a thickness below its spin diffusion length. The ferromagnetic layers are different in such a way (*e.g.* thickness or coercive field), that one of them can be remagnetized more easily than the other. We distinguish the two layers in the following by calling them "free" and "fixed" and draw them as a thinner and ticker layer, respectively. When electrons flow[1] from the fixed to the free

[1]Whenever I refer in this lecture to the direction of a current, I mean the direction of the electron flux rather than the (opposite) technical current direction.

layer, the magnetization of the free layer aligns parallel to the magnetization of the fixed layer and this alignment is stabilized [Fig. 1(b) left side]. When the current direction is reversed, however, the antiparallel alignment is more stable and is therefore taken [Fig. 1(b) right side] as will be explained in Sect. 2.2. Thus, a magnetization reversal can be induced by reversing the *polarity* of the DC current flowing through the layers.

An experimental arrangement for the observation of current-induced switching is displayed in Fig. 2(a) taken from one of the pioneering experimental papers in this field by the Cornell group [4]. The sample is a nanopillar consisting of a thin, free Co layer (Co 1) with a thickness of 2.5 nm and a thick, fixed Co layer (Co 2) of 10 nm thickness. The Cu spacer in between is 6 nm thick and, thus, thinner than the spin diffusion length of Cu (about 100 nm). The lateral diameter of the pillar is 130 nm. A current can be applied between leads I^- and I^+, and the voltage drop is measured at V^- and V^+. The lateral restriction is required to obtain the necessary high current density (of the order of 10^7–10^8 A/cm^2 corresponding to 1–10 mA in a 100 nm-diameter pillar) to establish a steady (constant current) non-equilibrium situation. The relative orientation of the Co layers can be measured *via* the GMR effect of the Co 1/Cu/Co 2 trilayer system. Figure 2(b) displays the differential resistance (dV/dI) as a function of the applied current. At negative bias electrons flow from the fixed (thick) to the free (thin) Co layer and stabilize the parallel magnetization alignment which yields a low dV/dI [green curve in Fig. 2(b)]. At positive bias the parallel alignment is destabilized, Co 1 switches to the antiparallel alignment at a sufficiently large current, and dV/dI increases [red curve in Fig. 2(b)]. Upon reducing the current [thick black line in Fig. 2(b)] hysteretic behavior is observed such that Co 1 switches back to parallel at a negative current. An external magnetic field is applied to define and fix the magnetization direction of the Co 2 layer. Note, that the curve shows hysteretic behavior with two different stable states at zero applied current. Therefore, positive and negative current pulses allow to switch between states at zero current with parallel and antiparallel magnetization alignments. This mechanism constitutes the basis for an advanced switching mechanism, that fulfills all the requirements discussed above.

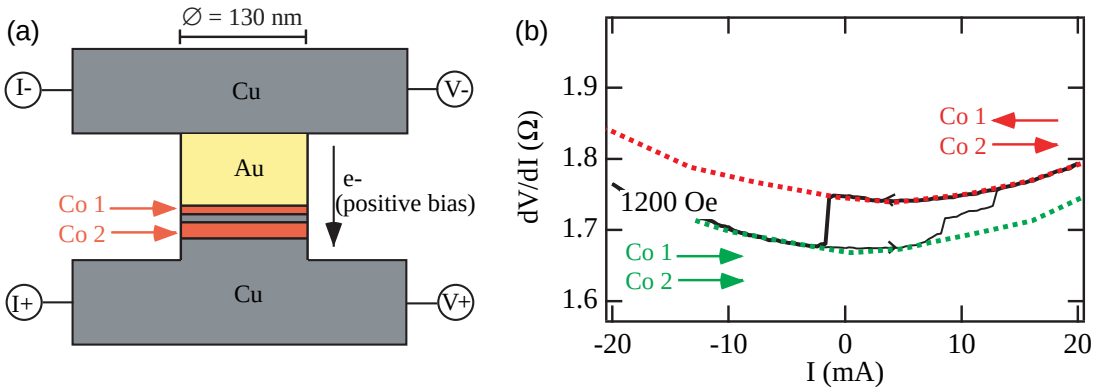


Fig. 2: *(a) Schematic pillar device with two Co layers (Co 1 and Co 2) separated by a 6 nm thick Cu layer. The pillar diameter is 130 nm. (b) The dV/dI measurement as a function of the current through the column device yields the relative alignment of the magnetic layers via the GMR effect. After [4].*

2.2 Physical picture: Absorption of the transverse spin current component

In view of the high current densities, one might suppose that the Oersted field generated by the current is responsible for the switching behavior. However, the Oersted field has the wrong symmetry. Its circular field lines lie in the plane of the sample and favor in the steady state (*i.e.* state formed a long time after switching on the current) a vortex-like magnetization state with the direction of rotation depending on the current polarity. These vortex states would appear symmetrically for both current directions in clear contrast to the experimentally observed behavior [see for instance Fig. 2(b)]. Furthermore, the strongest Oersted field occurs at the pillar circumference and scales like I/d, where I is the current and d the pillar diameter. The spin-transfer torque, on the other hand, scales like the current density I/d^2. Therefore, the spin-transfer torque effect becomes stronger below a certain structure size d_c. Theoretical estimates and available experiments suggest a d_c of the order of 100 nm. This fundamental size restriction fortunately coincides with the possibilities of e-beam lithography (see Sect. 5.1) and at the same time yields the needed current densities at technically convenient current amplitudes [10 mA flowing through an area of $(100\,\text{nm})^2$ correspond to 10^8 A/cm^2]. In practice one always has to be aware of the presence of the Oersted field and has to take into account its possible influence.

In order to develop a physical picture for the spin-transfer torque effects, we start by considering the fate of a polarized current that enters a ferromagnet from a metallic non-magnet. The situation is sketched in Fig. 3(a). We assume, that the incident current is polarized along an axis tilted by the angle θ with respect to the magnetization $\vec{M}$ of the ferromagnet. For simplicity we assume a polarization axis in the drawing plane. In experiments, due to the shape anisotropy, the polarization axis is usually in the plane of the layers. The general arguments given below are valid for both cases. The (normalized) spinor Ψ_{in} of an incident, accordingly polarized electron can be written as a superposition of spin-up and spin-down components with respect

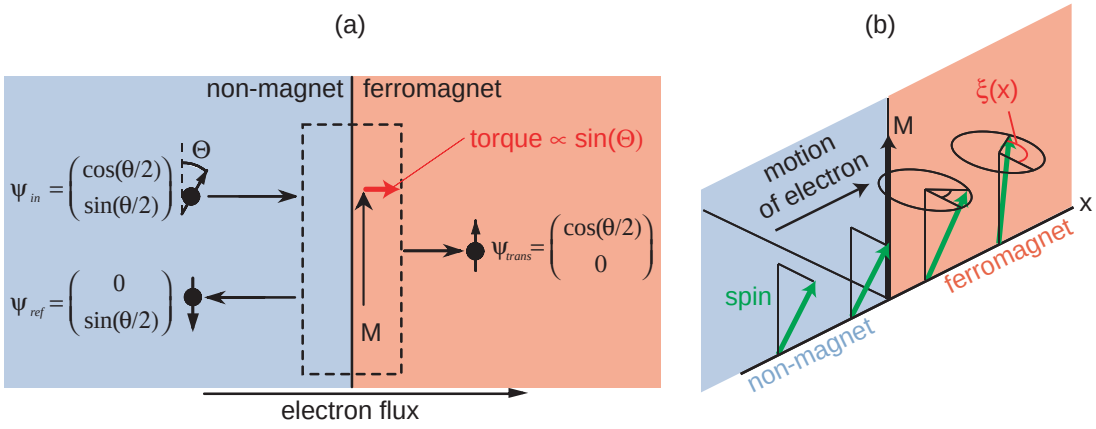


Fig. 3: *Two effects contributing to the absorption of the transversal spin current component in the interface region (dashed box) between a non-magnet and a ferromagnet. (a) Spin filtering: The incoming Ψ_{in}, transmitted Ψ_{trans}, and reflected Ψ_{ref} wave functions (spinors) for the idealized case of perfect spin filtering are indicated. The absorbed transversal spin current is proportional to $\sin(\theta)$ and acts as a torque on the interface magnetization. (b) Spatial precession of the spin in the ferromagnet: The phase ξ is constant in the non-magnet, but increases in the ferromagnet with distance x from the interface.*

to the quantization axis defined by $\vec{M}$. The amplitudes are $\cos(\theta/2)$ and $\sin(\theta/2)$, respectively, and correspond to a transverse component of the spin vector given by $\sin(\theta)$. At the interface to the ferromagnet the potential experienced by the electron changes and becomes spin-dependent. Inside the ferromagnet this gives rise to the spin-split density of states. At the interface it leads to spin-dependent transmission and reflection. Therefore, the transmitted and reflected spinors, Ψ_{trans} and Ψ_{ref}, are modified superpositions of spin-up and spin-down components compared to the incident spinor. This leads unavoidably to different transverse spin components and, thus, to a discontinuity in the transverse spin current. The "missing" transverse spin current is absorbed by the interface and acts as a current-induced torque on the magnetization. This effect occurs for each electron *individually* and is called *spin filtering* [2]. Figure 3(a) shows the spinors in the extreme case of perfect spin filtering. This means that only one spin orientation (here spin-up) can propagate in the ferromagnet, whereas the other cannot and therefore is completely reflected at the interface. For realistic ferromagnets like the 3d transition metals (*i.e.* Fe and Co) and their alloys, roughly 50% of the transversal component is absorbed, and the transmitted as well as reflected spinors still carry transversal components [5].

The actual *current polarization* of the transmitted and reflected currents is obtained by summing over all electronic states contributing to the transport. This introduces two additional effects. The first one arises because the reflection and transmission amplitudes at the interface may be complex. This means that the spin of an incoming electron rotates upon reflection and transmission. The angle of rotation is in general different for each state under consideration, as it depends on the k vector of the incoming electron. The cancellation, which occurs when we sum over all the resulting different spin vectors, reduces the net outgoing transverse spin current. This is an entirely quantum mechanical phenomenon, for which there is no classical analog. A second effect arises because spin-up and spin-down components of an electron have the same wave vector $k^{\uparrow\downarrow}$ in the non-magnet, but different wave vectors when they are transmitted into the ferromagnet, $\Delta k = k^{\downarrow} - k^{\uparrow} \neq 0$. This is a consequence of the spin-split density of states. The two components are coherent, and a spatial phase $\xi(x) = \xi_0 + \Delta k x$ builds up [Fig. 3(b)]. This corresponds to a precession of the spin vector in space rather than time. The precession frequency is different for electrons from different portions of the Fermi surface due to different Δk. Therefore, when we sum over all conduction electrons, almost complete cancellation of the transverse spin occurs after propagation into the ferromagnet by a few lattice constants.

Taking all three effects –(i) spin filtering, (ii) rotation of the reflected and transmitted spin, and (iii) spatial precession of the spin in the ferromagnet– together, to a good approximation, the transverse component of the transmitted and reflected spin currents are zero for most systems of interest (the completeness of the cancellations depends on the actual band structures). Thus, the incoming transverse spin current is absorbed by the interface and acts as a current-induced torque on the magnetization. A comprehensive theoretical treatment of these effects is given in Ref. [5].

Up to now we have assumed that the incident current is polarized. In the experiment this can be achieved by a second ferromagnetic layer with a slightly tilted magnetization (angle θ). This is possible when the two magnetic layers are separated by a non-magnetic spacer layer as shown in Fig. 4. The spin polarization process in this layer proceeds by the same mechanisms as described above. However, the polarization is *not* modified at the interface to the *non-magnetic* spacer layer because its density of states is *not* spin-split. The only requirement is that the spacer layer thickness is below its spin diffusion length to prevent significant depolarization by spin-flip scattering. Additionally, we reduce in Fig. 4 the extended ferromagnet of Fig. 3(a) to a thin film element and arrive at a situation very similar to the experimental setup of Fig. 2(a).

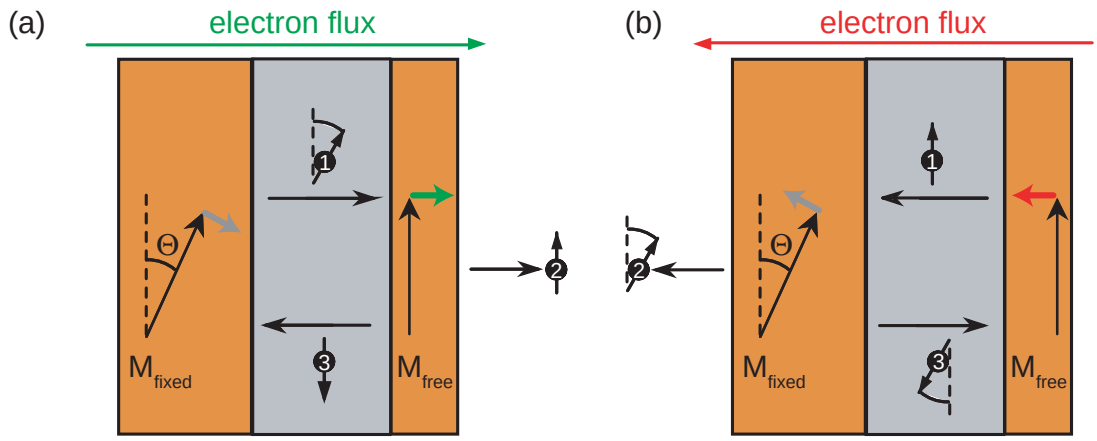


Fig. 4: *Physical picture of the current-induced magnetization switching. Orange regions represent the two ferromagnetic layers. Due to the assumed asymmetry $\vec{M}_{fixed}$ does not respond to the torque (short gray arrows) acting on it, whereas $\vec{M}_{free}$ can follow the torque (short green and red arrows). The numbers in the spins refer to the sequence of the description. (a) and (b) show the situation for opposite electron flux directions, which result in stable parallel or stable antiparallel alignment, respectively.*

In Fig. 4(a) the electrons flow from the fixed to the free layer. A current polarized by the fixed layer (1) hits the free layer and transfers its transversal component as a torque to the free layer. Part of the current is transmitted (2) and another part is reflected (3). This reflected current can now be considered as a polarized current impinging on the fixed layer. Again, the transversal component will be absorbed and acts as a torque on the fixed layer. However, due to the assumed large coercivity of the fixed layer, the torque will not be sufficient to induce a coherent rotation of the magnetization. It will rather generate incoherent spin waves to dissipate the energy and angular momentum associated with the exerted torque. Therefore, the fixed magnetization resists to the torque, and only $\vec{M}_{\text{free}}$ starts to rotate due to its lower coercivity in order to reach the stable parallel alignment with $\vec{M}_{\text{fixed}}$. For the opposite direction of the electron flux in Fig. 4(b), we obtain a similar situation but the resulting torques point in the opposite directions. Therefore, the stable state corresponds to the antiparallel alignment of $\vec{M}_{\text{free}}$ and $\vec{M}_{\text{fixed}}$. Note, that in this case the torque on $\vec{M}_{\text{free}}$ arises from the current which first has been reflected (3) from the fixed layer. Obviously, the asymmetry (fixed $\leftrightarrow$ free) plays an important role, which is very reasonable because ”left” and ”right” flowing current cannot be distinguished in the symmetric case.

2.3 Role of diffusive transport

The consideration given in the previous section is valid independent of what gives rise to the spin-polarized current. Obviously the picture holds for ballistic transport and for the drift motion of electrons due to an applied bias voltage. In the presence of spin-flip scattering there is also a diffusive contribution to the spin current. The origin is a deviation of spin density from the equilibrium spin density, the so-called *spin accumulation* δm. Spin accumulation occurs whenever a current crosses a region with unequal spin-dependent resistivities, *e.g.* a ferromagnetic layer with $\rho_\uparrow < \rho_\downarrow$ (Fig. 5). In this example spin-down electrons accumulate in front of this region (left of the Co layer in Fig. 5), and spin-up electrons prevail behind it (right of the

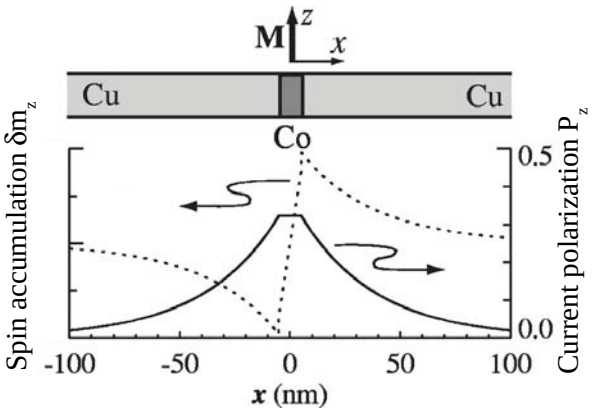


Fig. 5: *Spin accumulation δm_z (dotted, left ordinate) and current polarization P_z (solid, right ordinate) due to a diffusive current flowing in the $+x$-direction through a single ferromagnetic Co layer embedded between two semi-infinite Cu leads. After [6].*

Co layer in Fig. 5). Due to spin-flip scattering for instance from magnetic impurities, the spin accumulation exponentially decays with distance x from its source (in our case an interface between a ferromagnet and a non-ferromagnet). The characteristic length scale is the spin-flip scattering length λ, which is material-dependent and varies in the range from a few nanometers (*e.g.* $Ni_{80}Fe_{20}$ alloy, also called permalloy) up to several tens of nanometers (*e.g.* Co) for magnetic alloys and metals and exceeds 100 nm for non-magnetic metals (*e.g.* Cu). λ also depends on extrinsic properties like crystallinity and purity of the material. Any gradient in the spin accumulation gives rise to a spin current, very much in the same manner as a gradient in a particle density induces a diffusive particle current. In general, the polarization of the diffusive spin current is not collinear with the magnetization in a ferromagnet. Therefore, the processes discussed in the context of Figs. 3 and 4 apply: The transversal component of the diffusive spin current is also absorbed and acts on the magnetization like a torque [5, 7].

2.4 Experimental confirmation of physical picture

The critical current I_c (or critical current density J_c) needed for switching can be derived from the condition that the spin-transfer torque term must exceed the Gilbert damping torque [see Eq. (5) in Sect. 3.2]. The explicit expression for I_c depends on details of the geometry, anisotropies, *etc.* and is not displayed here. In general the critical current density for switching from parallel to antiparallel alignment J_c^+ is different from the critical current density J_c^- needed for the reversed switching direction.

Figure 6 displays the dependence of the critical current densities $J_c^\pm$ and the CPP-GMR times area product ΔRA of Co/Cu(d_{Cu})/Co nanopillars as a function of the spacer thickness $d_{Cu} = 6 \ldots 50$ nm [8]. All three quantities are normalized to their value at $d_{Cu} = 6$ nm. Due to the rather large spacer thicknesses one expects spin-flip scattering to play a role, at least for the largest d_{Cu}. CPP-GMR exponentially decreases like $\Delta RA \propto \exp(-d_{Cu}/\lambda)$, where λ is the spin diffusion length of Cu at RT. The fit of the CPP-GMR data (solid circles) in Fig. 6 yields $\lambda = 190 \pm 20$ nm. Spin-flip scattering also reduces the efficiency of spin-transfer torque. Therefore, the critical current densities should increase with the total distance the electrons travel in Cu before they exert the switching torque on the free layer. According to Fig. 4(a) the electrons have to traverse the Cu spacer once from the fixed to the free layer to switch to the parallel state.

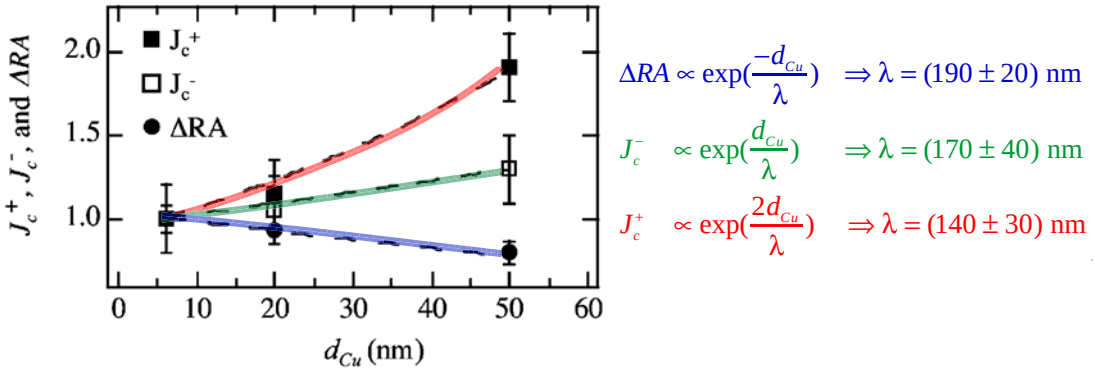


Fig. 6: *Normalized critical current densities J_c^- (open squares), J_c^+ (filled squares), and magnetoresistance ΔRA (filled circles) of Co/Cu(d_{Cu})/Co nanopillars as a function of d_{Cu}. Dashed lines are fits to the exponential functions given on the right hand side. After [8].*

Hence, $J_c^- \propto \exp(d_{Cu}/\lambda)$. For the switching to the antiparallel state, however, the electron travel first from the free to the fixed layer, where they are reflected, and then back to the free layer to induce the switching [Fig. 4(b)]. Hence, we expect for J_c^+ a factor of 2 in the exponent, $J_c^+ \propto \exp(2d_{Cu}/\lambda)$. The dashed lines in Fig. 6 are fits according to these expectations and yield $\lambda = 170 \pm 40$ nm for J_c^- and $\lambda = 140 \pm 30$ nm for J_c^+. If the factor of 2 is not taken into account, the J_c^+ data yields $\lambda = 70 \pm 20$ nm, which is not in agreement with the values from the ΔRA and J_c^- data. Therefore, this experiment nicely confirms the physical picture introduced in Sect. 2.2.

3 Extended Gilbert equation

Now that we have developed a physical picture for the origin of spin-transfer torque on a quantum-mechanical level, we address the question how this torque influences the dynamics of the macroscopically observable magnetization. In order to do so we have to consider the micromagnetism of the magnetic elements, on which the torque acts.

3.1 Basics of magnetization dynamics

A comprehensive introduction to magnetization dynamics is given in the lecture "Micromagnetism" by Riccardo Hertel. This section is intended to be a brief reminder of the basic ideas and to define the quantities and symbols to be used later on in this lecture.

The micromagnetic approach is a non-atomistic, continuum description of magnetization dynamics, which assumes that the amplitude of the magnetization attributed to a certain volume of magnetic material is constant. Therefore, the magnetization dynamics of a volume element is restricted to changes of the magnetization direction, which can be described by the Gilbert equation of motion for a magnetic moment:

$$\frac{d\vec{m}}{dt} = \underbrace{-\gamma \vec{m} \times \vec{H}_{eff}}_{\propto \frac{d\vec{M}_P}{dt}} + \underbrace{\alpha \vec{m} \times \frac{d\vec{m}}{dt}}_{\propto \frac{d\vec{M}_D}{dt}}, \tag{1}$$

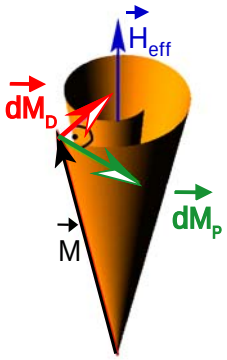


Fig. 7: *Motion of a magnetization vector $\vec{M}$ in an effective field $\vec{H}_{eff}$. The first term in Eq. (1) gives rise to the tangential torque $d\vec{M}_P/dt$ driving the precession and the second term $d\vec{M}_D/dt$ causes the damping.*

where $\vec{m}$ is the normalized magnetization vector of the considered volume element, γ is the gyromagnetic ratio, and α is the phenomenological Gilbert damping constant. The effective field $\vec{H}_{\text{eff}}$ is the negative variational derivative of the total areal energy density E_{tot} with respect to the magnetization $\vec{M}$

$$\vec{H}_{\text{eff}} = -\frac{1}{\mu_0}\frac{\delta E_{\text{tot}}}{\delta \vec{M}}. \tag{2}$$

The total energy E_{tot} comprises contributions from the exchange energy, anisotropy energy, stray field energy, and the Zeeman energy due to an external field. The first term in Eq. (1) describes the precessional motion of $\vec{m}$ about $\vec{H}_{\text{eff}}$, and the second term the damping, which forces $\vec{m}$ to relax to the lowest energy configuration, $\vec{m}||\vec{H}_{\text{eff}}$ (Fig. 7).

Obviously, the temporal evolution of a remagnetization process is intimately related to the magnetization dynamics described by the Gilbert equation. Note that both terms in Eq. (1) have their own timescales. The precession period (given by the Larmor frequency) is determined by the effective field $\vec{H}_{\text{eff}}$. The stronger $\vec{H}_{\text{eff}}$, the higher the Larmor frequency. Typical frequencies for magnetic materials like Fe, Co, Ni, or permalloy are of the order of several GHz yielding precession periods τ of fractions of a nanosecond. The timescale of the damping term, on the other hand, is governed by the phenomenological damping parameter α. Typical values $\alpha \approx 0.001$ result in relaxation times of several nanoseconds. Therefore, relaxation usually occurs over several precessional revolutions. The magnetization component along $\vec{H}_{\text{eff}}$ approaches exponentially saturation magnetization M_S, and components perpendicular to $\vec{H}_{\text{eff}}$ show an exponentially damped oscillatory behavior, the so-called magnetic ringing.

In order to describe mesoscopic magnetic objects properly, micromagnetic simulations are employed: The object to be described is geometrically subdivided into small volume elements with a size smaller than the exchange length of the material (typically a few nm), within which the assumption of constant magnetization is justified. The magnetization dynamics of each element is described by the Gilbert equation. The direct exchange with neighboring volume elements and the demagnetizing field due to all other elements are taken into account through the effective field. The resulting set of coupled equations is solved by means of finite-element computer codes, see lecture “Micromagnetism” by Riccardo Hertel.

3.2 STT extension of the Gilbert equation

The current-induced spin-transfer torque $d\vec{M}_{\text{free}}/dt$ acting on the free layer as introduced in Sect. 2.2 has been calculated by Slonczewski [2] as

$$\frac{1}{M_S}\frac{d\vec{M}_{\text{free}}}{dt} = \frac{d\vec{m}_{\text{free}}}{dt} = \frac{I}{A}\cdot g(\theta)\cdot \vec{m}_{\text{free}} \times (\vec{m}_{\text{free}} \times \vec{m}_{\text{fixed}}), \tag{3}$$

where I/A is the current density. $g(\theta)$ is the material dependent STT efficiency function, which is a measure for the conversion of current into spin-transfer torque. In general it depends on the angle θ between $\vec{M}_{\text{free}}$ and $\vec{M}_{\text{fixed}}$. The materials enter *via* the spin polarization P, volume and interface resistivities and other transport properties. In particular, the spin accumulation effects discussed in Sect. 2.3 modify the functional form of $g(\theta)$. An example will be given in Sect. 5. The double cross product is indeed proportional to $\sin(\theta)$ and, thus, the absorbed transversal component of the spin current. The linear dependence on I yields the reversed torque upon reversing the current direction. In order to study the influence of the spin-transfer torque effect on the magnetization dynamics the additional torque in Eq. (3) must be included into the Gilbert equation (1)

$$\frac{d\vec{m}}{dt} = \underbrace{-\gamma\vec{m}\times\vec{H}_{eff}}_{\propto\frac{d\vec{M}_P}{dt}} + \underbrace{\alpha\vec{m}\times\frac{d\vec{m}}{dt}}_{\propto\frac{d\vec{M}_D}{dt}} + \underbrace{\frac{I}{A}\cdot g(\theta)\cdot\vec{m}\times(\vec{m}\times\vec{m}_{fixed})}_{\propto\frac{d\vec{M}_{STT}}{dt}}. \tag{4}$$

The subscript *free* is dropped for clarity. This equation can again be separated into one part of precessional character and a second part of damping character:

$$\begin{aligned}\frac{1}{\gamma}\frac{d\vec{m}}{dt} &= -\beta\vec{m}\times(\vec{H}_{eff} + \frac{I\cdot g(\theta)}{A\gamma}\alpha\vec{m}_{fixed}) \\ &\quad -\beta\vec{m}\times\left[\vec{m}\times(\alpha\vec{H}_{eff} - \frac{I\cdot g(\theta)}{A\gamma}\vec{m}_{fixed})\right], \\ \text{with}\quad \beta &= \frac{1}{1+\alpha^2}.\end{aligned} \tag{5}$$

The damping constant is usually much smaller than unity, $\alpha \ll 1$, and, thus, $\beta \approx 1$. Due to the smallness of α the impact of the STT on the precessional term, *i.e.* the precession frequency,

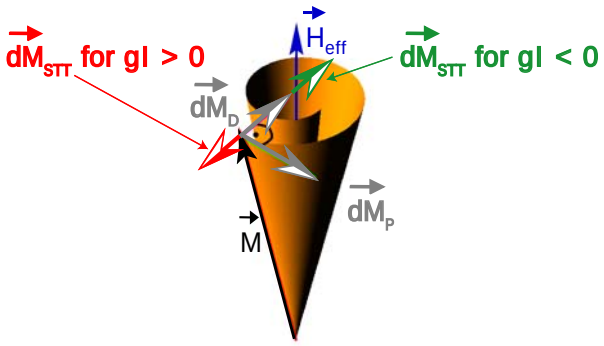


Fig. 8: *The spin-transfer torque $d\vec{M}_{STT}/dt$ can point along the Gilbert damping $d\vec{M}_D/dt$ or opposite to it. In the latter case it can destabilize $\vec{M}$ and induce switching or microwave oscillations, see Sect. 4.*

in the first term of Eq. (5) is only weak. However, the STT also appears in the damping term, where it "competes" with $\alpha\vec{H}_{eff}$. Depending on the sign of I the damping due to the STT can be negative or positive, see Fig. 8. The latter case is more interesting, because the conventional damping torque $d\vec{M}_D/dt$ may be compensated by the STT term $d\vec{M}_{STT}/dt$, and the precession amplitude is amplified, which leads to a destabilization of $\vec{M}_{\text{free}}$, *i.e.* switching or steady-state oscillatory modes.

4 Current-driven magnetization dynamics

The extended Gilbert equation (5) allows for a more detailed analysis of the impact of the STT on the magnetization dynamics. It turns out that the current-induced magnetization dynamics is very rich and in some respects exceeds the limits of the dynamics that can be stimulated by magnetic fields alone.

4.1 Steady-state oscillatory modes

In the phenomenological description of current-induced magnetization switching, *e.g.* in the context of Fig. 1, we have considered the STT and damping terms of the Gilbert equation, but neglected the precessional term. A more complete analysis taking all terms into account shows that the switching process after applying a DC current of the correct polarity starts with the excitation of a precessional motion about the initial state [for an example see Fig. 22(a) below]. The cone angle of the trajectory increases steadily under the action of the STT, which opposes the restoring Gilbert torque. When the magnetization reaches the position, where a potential maximum separates the initial and the final states, switching occurs and the magnetization relaxes towards the final state, now on a precessional trajectory with decreasing cone angle. The details of the transition from excitation to damping (increasing to decreasing cone angle) depend on the energy landscape $E_{tot}(\vec{M})$ and the functional form of $g(\theta)$ in Eq. (5).
Up to now we have assumed that the external field is lower than the coercive field of the free layer. Therefore, the presence of (uniaxial) shape or magnetocrystalline anisotropy gives rise to at least two stable states, and the current-induced torque can cause switching between these states. If the external field exceeds the coercivity, the energy landscape $E_{tot}(\vec{M})$ is modified in

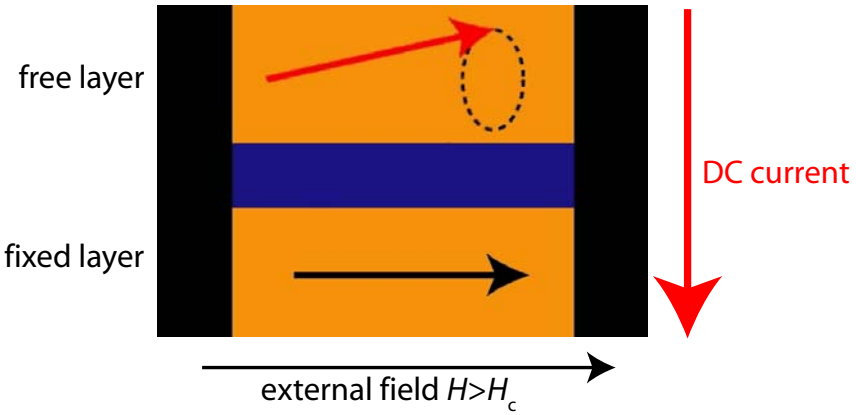


Fig. 9: *A DC current excites a steady-state oscillatory motion, if the external field exceeds the coercive field H_c of the free layer.*

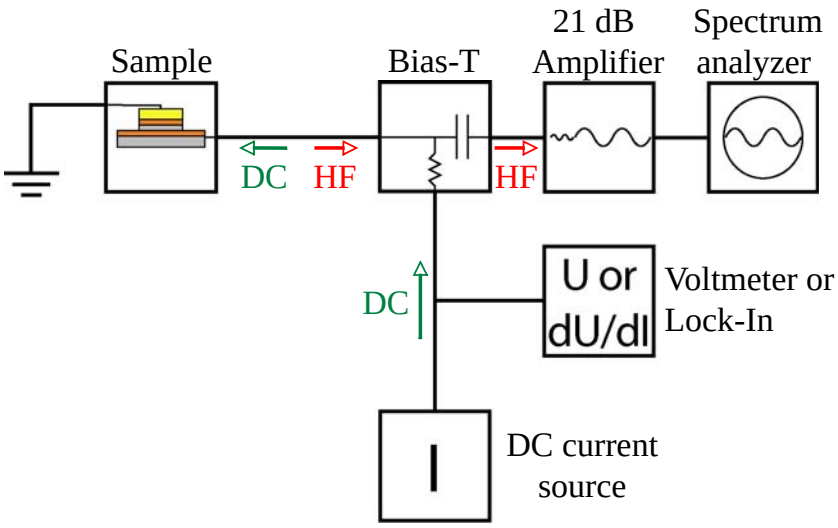


Fig. 10: *Experimental setup for the measurement of microwave signals generated by the spin-transfer torque in the nanopillar shown on the left side. The bias-T separates the applied DC current and the HF output signal. The voltmeter or lock-in amplifier allow for a simultaneous measurement of (differential) DC resistance changes in four-point geometry.*

such a way that only one stable magnetization state exists, namely parallel to the external field. In this situation, switching is not possible for either current polarity. For one polarity the system is not excited at all, whereas for the other polarity it enters a steady-state oscillatory motion, which is characterized by an equilibrium between the Gilbert damping torque and the STT due to the applied current (Fig. 9). This is a very interesting case because the magnetization of the free layer can be driven into new types of oscillatory dynamic modes, which are not attainable with magnetic fields alone. An example are the large-angle precessional modes with excitation angles as large as 180°. Any oscillatory motion of the free layer with respect to the fixed layer results, due to the GMR effect, in a variation of the resistance. Therefore, the DC current that

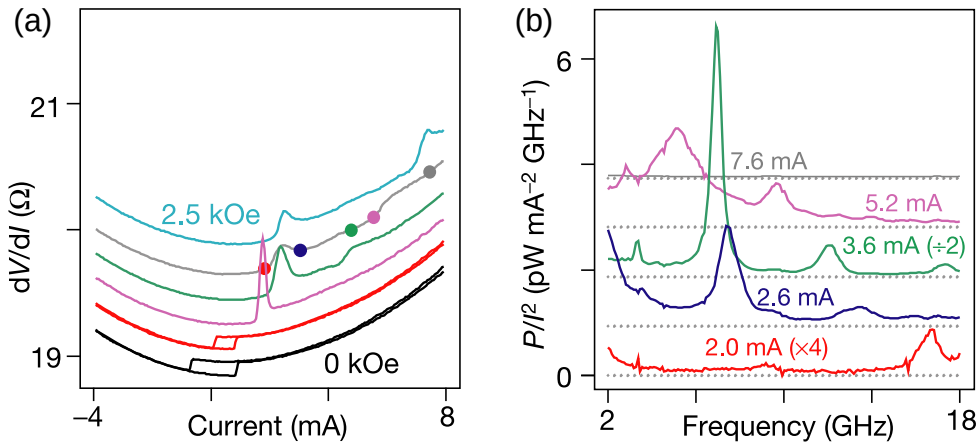


Fig. 11: *(a) Differential resistance versus current for different fields. The hysteresis still visible in the black and red curves vanishes for external fields exceeding the coercivity and peaks appear instead, e.g. purple curve. (b) Microwave spectra measured at 2 kOe [gray curve in (a)] at different current amplitudes as marked by colored dots in (a). After [9].*

gives rise to the STT effect generates a time-varying voltage with typical frequencies in the GHz range. Figure 10 shows a measurement setup that allows direct electrical measurements of the microwave-frequency dynamics. The microwave voltage signal generated due to an oscillatory excitation and the GMR inside the nanopillar is separated from the DC current using a bias-T, amplified, and fed into a spectrum analyzer for GHz frequencies. Figure 11(a) shows resistance *versus* current plots, similar to Fig. 2(b), for different external fields. With increasing field (from bottom to top) the hysteretic behavior (black and red) gives way for peaks, because external fields larger than 1 kOe (purple curve) exceed the coercivity of the free layer. Microwave spectra taken under the current and field conditions marked in Fig. 11(a) by colored dots are displayed in Fig. 11(b). Rather sharp peaks at frequencies of several GHz are resolved. Figure 12(a) shows the dynamic stability diagram determined from such spectra. The basic features can be reproduced by numerically solving the extended Gilbert equation (5) for a macrospin, which is supposed to describe the complete free layer, as demonstrated in Fig. 12(b). Here, P and AP stand for stable parallel and antiparallel alignment, and P/AP is the region of bistability, where hysteretic switching is possible. S marks the small-amplitude precessional regime and L the large-amplitude dynamic regime. Region W in Fig. 12(a) cannot be described by a macrospin simulation. Recent micromagnetic simulations beyond the Stoner-Wohlfarth approximation have identified region W to correspond to the formation and annihilation of dynamic vortices through the interplay of the current-induced Oersted field and the STT effect [10].

Stability diagrams like these in Fig. 12 show the richness of the current-driven magnetization dynamics, although the comparison to simulations is based on a simple macrospin model. The fact that region W cannot be reproduced in macrospin models indicates that the dynamic behavior of real samples is even more complex and can only be understood in the framework of full-scale micromagnetic simulations.

Figures 13 and 14 show an example of a micromagnetic simulation of current-driven magnetization dynamics in an Fe disc of 150 nm diameter and 2 nm thickness. The sample is assumed to be monocrystalline with cubic magnetocrystalline anisotropy. If an external field of 50 mT is applied in the film plane, the magnetization in the disc is essentially homogeneous and aligned with the external field H. Due to the disc shape and the inhomogeneous demagnetizing field in any non-ellipsoidal particle, some small inhomogeneities occur at the perimeter of the disc. The

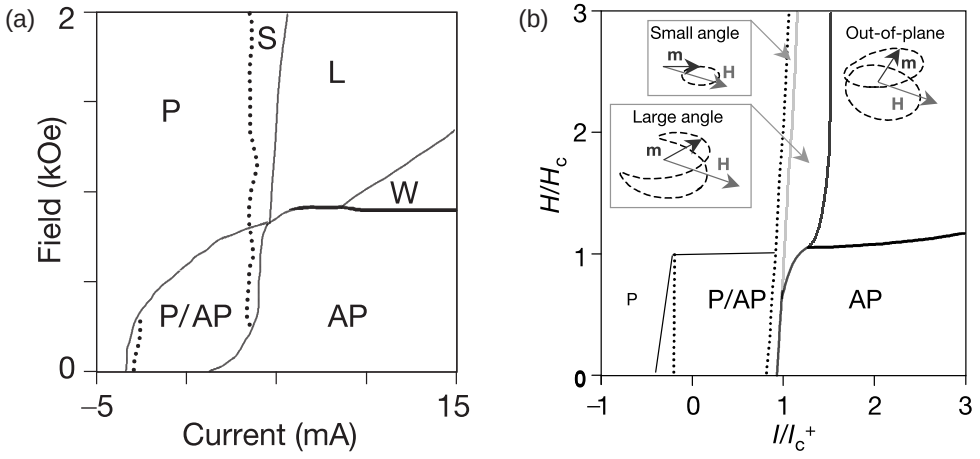


Fig. 12: *(a) Experimental and (b) calculated dynamic stability diagram. Different regions are explained in the text. Current and field axes in (b) are normalized to the critical current I_c^+ and the coercive field H_c, respectively. After [9].*

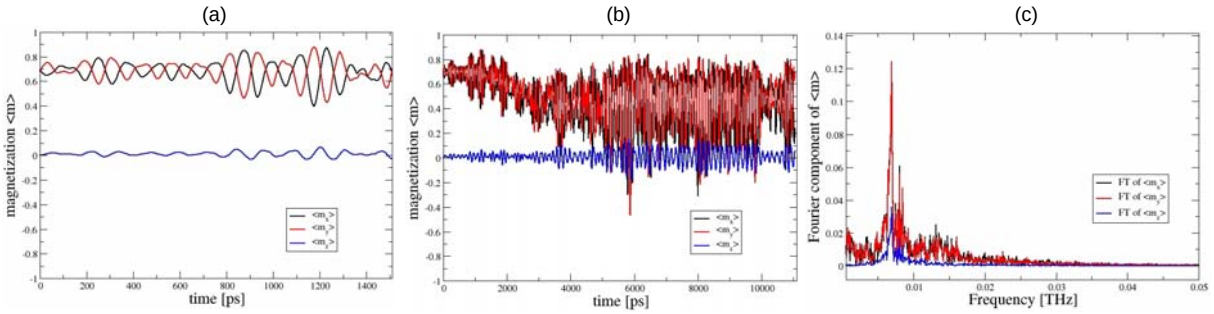


Fig. 13: *Simulated current-driven oscillations of the magnetization in a disc-shaped nanomagnet ($d = 150\,nm$). After applying the current, the initial oscillations (a) quickly evolve into a noisy signal (b). At least one clear peak can be seen in the Fourier spectrum (c). From [11].*

magnetization of the disc is in a so-called onion state in order to reduce magnetostatic surface charges. H is applied at 45° with respect to the easy axis and a polarized current with homogeneous current density $j = 5 \times 10^7$ A/cm^2 runs perpendicularly through the sample plane. The polarization direction of the current is parallel to one of the easy axes and encloses an angle of 135° with H as sketched in Fig. 14(a). The dynamics of the magnetization can be monitored by plotting the volume-averaged normalized Cartesian components of the magnetization as a function of time. In the first 1.5 ns after switching on the electric current, only relatively small oscillations occur [Fig. 13(a)]. After about 3 ns, these small oscillations are soon replaced by a noisier signal. While the frequency of the oscillations remains conserved to a good extent, the amplitude varies strongly in a seemingly chaotic way [Fig. 13(b)]. The Fourier transform of this signal shows a sharp peak at about 7 GHz and some additional features at about 13 GHz, which are less pronounced [Fig. 13(c)]. Snapshots of the magnetic structure in the disc during this process are shown in Fig. 14. Surprisingly, the magnetization structures are very inhomogeneous and do not display any clear pattern. In contrast to this result, more well-defined features like resonant modes, spin waves or domain wall displacements usually occur in the case of ordinary, field-driven magnetization dynamics. These inhomogeneities can be attributed to the constant supply of energy provided by the electric current [12]. The system apparently converts

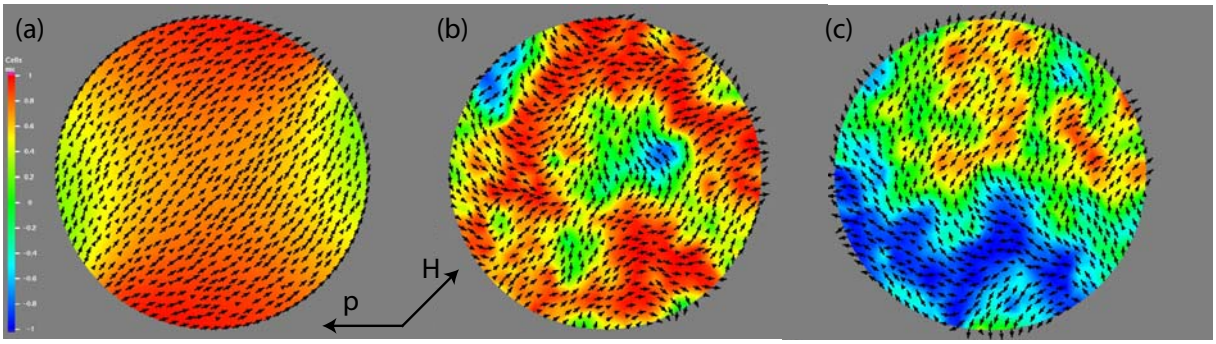


Fig. 14: *Typical snapshots of the simulated magnetic structure in the disc at different times after applying the current. (a): 0 ns; (b): 5.4 ns; (c): 8.3 ns. The magnetization structures are strongly inhomogeneous and are very dissimilar to structures known from field-driven magnetization dynamics. The directions of the current polarization p and the external field H are indicated. From [11].*

this energy into a large number of spin waves, which superimpose incoherently and lead to this type of noise [13]. The value of the intrinsic damping constant α, which was set to 0.01 in this example, has a strong influence on the balance between energy dissipation and energy pumping [12, 13]. In spite of the pronounced magnetic inhomogeneities, the Fourier signal is remarkably clear in the sense that the average magnetization oscillates mainly at a well-defined frequency. It is reasonable to assume, that it is this dominant frequency that is measured experimentally [*e.g.* in Fig. 11(c)].
Simulations of this sort yield the following general conclusions: (i) The macrospin approximation does not describe all the details of the real dynamics of the magnetization. It is only suited as a first simple toy model to obtain qualitative understanding that then needs to be confirmed by full-scale micromagnetic simulations. (ii) A clear signal in the frequency (as it is observed in experiments) does not necessarily result from a homogeneous spin precession. (iii) The magnetization dynamics induced by STT is qualitatively very different from the field-driven dynamics.

4.2 Spin-torque nano-oscillators

Nanomagnets driven by spin-polarized currents have the potential to serve as nanoscale, on-chip microwave sources or oscillators, tunable by field and current over a wide frequency range. These devices are called *spin-torque nano-oscillators (STNO).* Possible fields of application are communication technology, *e.g.* for inter- and intra-chip communication. At present the major drawback of STNOs is the low output power level of the order of 1 nW.
An obvious route to increase the output power of STNOs is to take advantage of the much larger magnetoresistive effects in TMR structures (see lecture on “Tunneling Magnetoresistance” by Daniel Wortmann). Tunneling junctions based on epitaxial MgO barriers can reach TMR ratios above 500% at room temperature and, thus, exceed typical GMR ratios by about two orders of magnitude. Since STNOs are operated under constant-current conditions, the output power scales with the square of the magnetoresistance ratio. Therefore, TMR-based STNOs readily have a 3 to 4 orders of magnitude larger output power.
A recent example of a CoFeB/MgO/CoFeB tunneling junction with a 1.2 nm thick MgO barrier is shown in Figs. 15 and 16. The junction with an elliptical cross-section of $160 \times 70\,\mathrm{nm}^2$ yields a TMR ratio of 110% [Fig. 15(a)]. Compared to the value of 500% mentioned above, this lower TMR ratio results from a compromise between a large TMR ratio and a high bar-

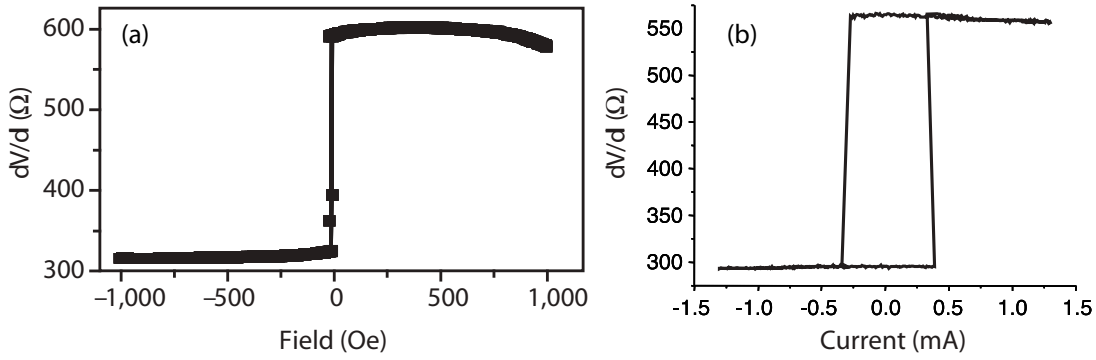


Fig. 15: *Field (a) and current (b) sweep loops of a CoFeB/MgO/CoFeB tunnel junction with a resistance times area product $RA = 4\,\Omega\mu\mathrm{m}^2$ and an elliptical cross-section of 160×70 nm^2. The critical current density for switching is about 3×10^6 A/cm^2. After [14].*

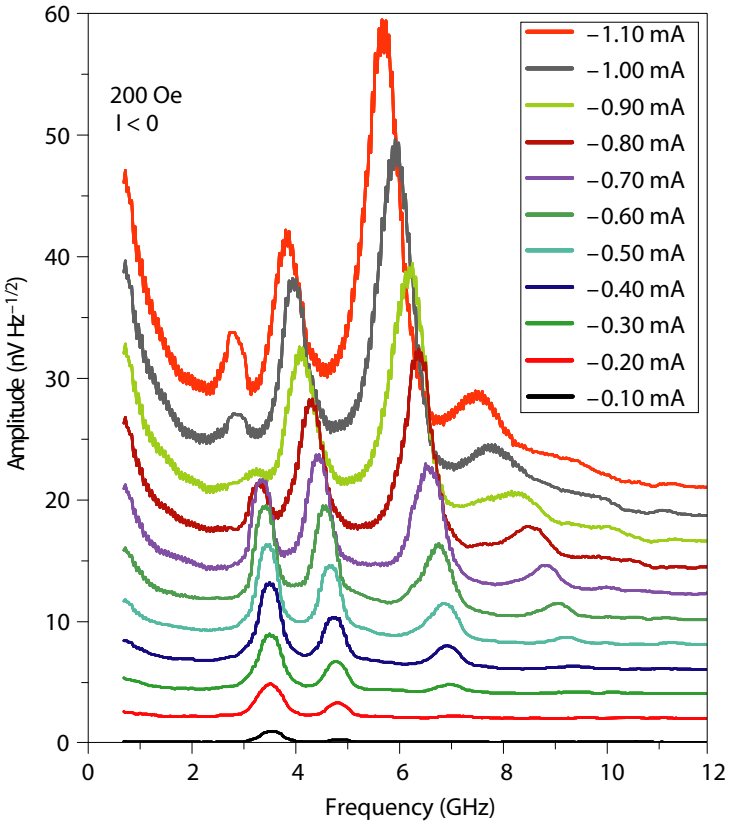


Fig. 16: *Microwave spectra for the same tunneling junction as in Fig. 16 for increasing negative DC current. After [14].*

rier transparency, which is required to obtain sufficiently large current densities. The resistance times area product of this junction is only $RA = 4\ \Omega\mu\text{m}^2$ and the critical current density for switching is about 3×10^6 A/cm^2 [Fig. 15(b)]. Figure 16 shows the evolution of the microwave spectra with increasing negative DC current. The different peaks correspond to first and second harmonic of modes located in the center and at the edge of the elliptical magnetic element. Samples of this type reach maximum integrated output power of up to 0.48 μW although a significant fraction of the generated power is not detected due to the impedance mismatch between the nanopillar and the 50 Ω leads [14]. Thus, the μW range needed for applications [15] is within reach.

Another route to increase the output of STNOs is to arrange a large number of STNOs in an array and to synchronize the oscillatory motion in all oscillators. Due to the coherent generation of the microwave signal, one expects that N coupled STNOs yield about an N^2-fold power output. The coupling of STNOs can be achieved *via* spinwaves in a common ferromagnetic layer [15] or *via* microwave cross-talk in contact leads [16]. The synchronization of ensembles of STNOs and other aspects of STNOs will be discussed in Albert Fert's lecture on "Spin transfer oscillations and synchronization of spin transfer oscillators".

5 STT in epitaxial Fe/Ag/Fe(001) nanopillars

In this section I want to present experiments performed with a specific set of samples with the aim to discuss how the sample design and the choice of the materials can influence the spin-

transfer torque effects. We have chosen to study single-crystalline, epitaxial nanopillars because they can be considered as model systems with well-defined interfaces and homogeneous magnetic as well as electric properties. The absence of grain boundaries (which are present in the usually employed polycrystalline samples) ensures a unperturbed current distribution across the pillar area. The crystalline order of the magnetic layers gives rise to magnetocrystalline anisotropy, which in our case is a four-fold, in-plane effective anisotropy with the magnetic easy axes along Fe[100] and Fe[010] directions. The presence of a four-fold anisotropy is distinctly different from the usually employed two-fold shape anisotropy induced by elliptically shaped nanopillars. We will see that the four-fold anisotropy gives rise to a novel two-step switching behavior as well as to a novel oscillatory mode in very low or even zero external magnetic fields.

5.1 Sample fabrication

As already mentioned in Sect. 2.2, the observation of STT effects requires small cross sections for the current flow in order to minimize the influence of the Oersted field and to obtain high enough current densities. As an example for the fabrication of samples with suitable properties, the preparation of single-crystalline Fe/Ag/Fe(001) nanopillars according to the process developed in Jülich [17] is introduced below.
In order to achieve single-crystalline growth the magnetic multilayers are deposited in a standard molecular-beam epitaxy (MBE) system. The native oxygen layer of the GaAs(001) substrates ($10 \times 10\,\mathrm{mm}^2$) is desorbed by annealing for 60 min at $580^\circ\mathrm{C}$ under ultra-high vacuum (UHV) conditions. We deposit 1 nm Fe and 150 nm Ag at $100^\circ\mathrm{C}$ to get a flat buffer system after annealing at $300^\circ\mathrm{C}$ for $1\,\mathrm{h}$. The Ag buffer also acts as a bottom electrode for the transport measurements [Figs. 17(a) and (b)]. The following layers are then deposited at room temperature: Fe(20 nm)/Ag(6 nm)/Fe(2 nm). We check the crystalline surface structure after each deposited layer by low-energy electron diffraction (LEED). The spots characteristic of (001) surfaces slightly broaden with increasing total thickness, but still indicate high crystalline quality, even for the final 50 nm Au(001) capping layer. Thicknesses are controlled by quartz crystal monitors. The bottom Fe(20 nm) layer is magnetically harder with respect to the top Fe(2 nm) layer.

In order to measure the STT effects in the CPP-geometry we have developed a combined process of optical and e-beam lithography. First, we define the leads and contact pads of the

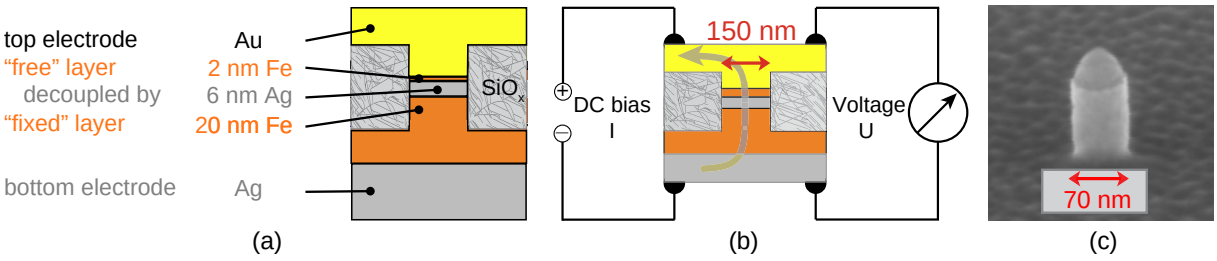


Fig. 17: *(a) Sequence, thicknesses, and functions of the layers within the multilayer stack. (b) Scheme of the junction geometry and the contacts for transport measurements. The DC current is confined to a diameter of $d \approx 70 - 150\,\mathrm{nm}$ by the nanopillar. The voltage drop is measured across the pillar in 4-point geometry. (c) SEM micrograph of a free-standing nanopillar after ion-beam etching.*

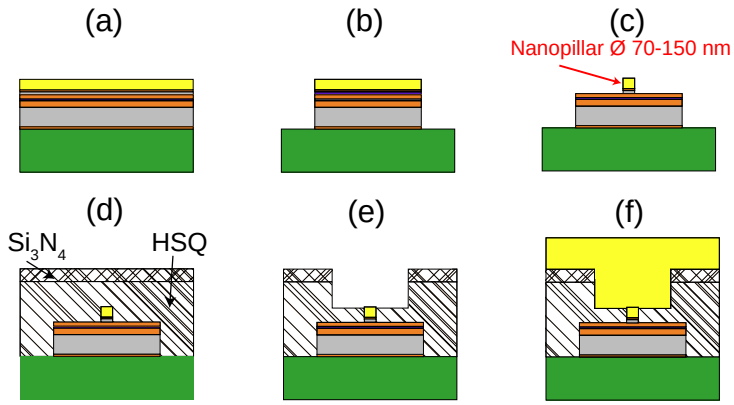


Fig. 18: *Lithographic process: (a) Extended epitaxial multilayer grown by MBE, (b) definition of 10 μm-wide bottom electrodes by optical lithography and IBE, (c) definition of nanopillars of 70-150 nm diameter by e-beam lithography and IBE, (d) planarization by HSQ and additional insulation by Si_3N_4; e-beam exposure converts HSQ into insulating SiO_x, (e) opening of a 8×6 μm^2 window to the top of the nanopillar by IBE, and (f) definition of the top electrodes by optical lift-off. The colors of different materials correspond to those of Fig. 17.*

bottom electrode by using AZ5206 photoresist and ion beam etching (IBE) [Fig. 18(b)]. We then employ HSQ (hydrogen silsesquioxane) as negative e-beam sensitive resist and a Leica EBPG 5HR e-beam writer to define small nanopillars. The resist structures are circular and transferred into the magnetic layers by IBE [Fig. 18(c)]. The timed etching process is stopped inside the magnetic multilayer. Typical dimensions of the developed resist structures are 50 - 150 nm (measured with an atomic force microscope). Due to redeposition of etched material during IBE, the nanopillars broaden to 70 - 200 nm. An SEM micrograph of the free-standing nanopillars in stage (c) of Fig. 18 is shown in Fig. 17(c). The pillars are planarized by spin-coating HSQ [Fig. 18(d)]. Subsequent e-beam exposure turns HSQ into SiO_x, which electrically insulates the pillars. In order to improve the insulation, especially at the side walls of the bottom electrodes, a 50 nm-thick Si_3N_4 layer is deposited by plasma enhanced chemical vapor deposition (PECVD). We open the insulation above the nanopillars by IBE and use an optical lift-off process of 300 nm Au for the preparation of the top electrode for the 4-point resistance measurements.

5.2 Slonczewski's model

We have chosen Fe and Ag not only for their good epitaxial match, but also for the specific spin-dependent transport properties of Ag/Fe(001) interfaces. According to Stiles and Penn [18] the spin-dependent interface resistances $R^{+,-}$ differ by more than one order of magnitude, $AR^{+}_{Fe/Ag(100)} = 1.07\times10^{-15}\,\Omega\mathrm{m}^2$ and $AR^{-}_{Fe/Ag(100)} = 12.86\times10^{-15}\,\Omega\mathrm{m}^2$, where A is the pillar cross-section. Therefore, we can apply Slonczewski's theory [19] for a unified description of GMR and STT, which is based on two main assumptions: (i) Negligible interfacial reflection for majority-spin electrons and (ii) electrically symmetric trilayer system. Assumption (i) is fulfilled because $AR^{+}_{Fe/Ag(100)} \ll AR^{-}_{Fe/Ag(100)}$, and (ii) is fulfilled even for trilayers with unequal Fe thicknesses because the interface resistances are much larger than the volume resistance.
In Slonczewski's theory, the angular dependences of the GMR effect with the current flowing perpendicular to the plane (CPP-GMR) and the STT $L_{STT} = g(\theta)\sin(\theta)$ [compare Eq. (3)] on

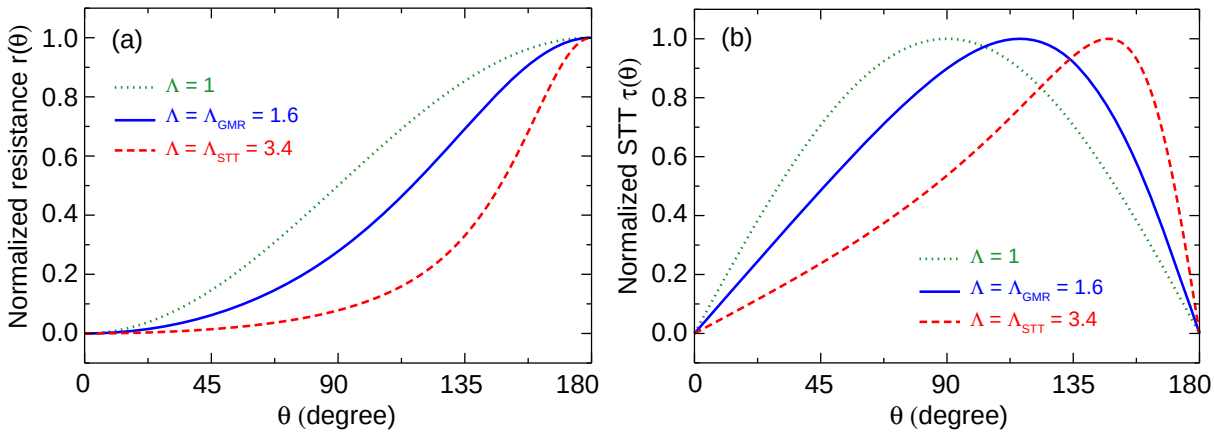


Fig. 19: *Angular variations of the GMR signal (a) and the spin-transfer torque (b) for three different* Λ *values according to Eqs. (6)-(8). The dotted line shows the symmetric behavior. The solid (*$\Lambda_{GMR} = 1.6$*) and broken (*$\Lambda_{STT} = 3.4$*) line is what we determine from GMR and STT experiments, see below. After [20].*

the angle θ between the two layer magnetizations are given by [19]

$$r(\theta) = \frac{R(\theta) - R(0°)}{R(180°) - R(0°)} = \frac{1 - \cos^2(\theta/2)}{1 + \chi \cos^2(\theta/2)} \quad (6)$$

$$L_{STT}(\theta) = \frac{\hbar I P \Lambda}{4Ae} \frac{\sin(\theta)}{\Lambda \cos^2(\theta/2) + \Lambda^{-1} \sin^2(\theta/2)} \quad (7)$$

$$\text{with} \quad \Lambda^2 = \chi + 1 = AG\frac{R^+ + R^-}{2}. \quad (8)$$

$R(\theta)$ is the dependence of the resistance on θ, $G = e^2 k_f^2/\sqrt{3}\pi h$ the conductance of the interlayer, $R^{+(-)}$ the total (interface and bulk) resistance for spin-up (spin-down) electrons for one half of the symmetric system, and $P = (R^- - R^+)/(R^- + R^+)$ is the spin polarization. The parameters Λ (or χ) is a measure for the deviation from the so-called symmetric behavior, which is given by $\Lambda = 1$ ($\chi = 0$). Λ^2 expresses the ratio of the mean resistance of one magnetic layer (including the lead) to the intrinsic interlayer resistance. If it deviates from unity, the averaged spin-dependent resistance is different from the interlayer resistance, which leads to an enhanced spin accumulation. The expectations according to Eqs. (6) and (7) are plotted in Fig. 19 for various values of Λ. The dotted lines for $\Lambda = 1$ represent the symmetric case. The GMR effect is then given by $r(\theta) = \sin^2(\theta/2)$ with point symmetry about the value at $\theta = 90°$ as known for current-in-plane (CIP) GMR. The angular dependence of the STT is reduced to $L_{STT}(\theta) \propto \sin(\theta)$ and is thus mirror symmetric about $\theta = 90°$. Having $\Lambda \neq 1$ breaks this mirror symmetry.

For the Fe/Ag(001) system we expect based on theoretical calculations [18] an asymmetry parameter $\Lambda = 4$ yielding a strongly asymmetric behavior as well as a large polarization $P = 0.85$, which should result in strong STT effects. The asymmetric behavior can directly be seen in the GMR data of Fig. 20, where we show $r(\theta)$ according to Eq. (6) as a function of the magnetic field applied along a hard axis of the Fe layers. The red lines are simulations for which we assume for single domain behavior for both Fe layers (Stoner-Wohlfarth model). The experimental curve is well reproduced and yields the relative alignment of the magnetizations as a function of the external field (icons with pairs of arrows). An interesting magnetic configuration

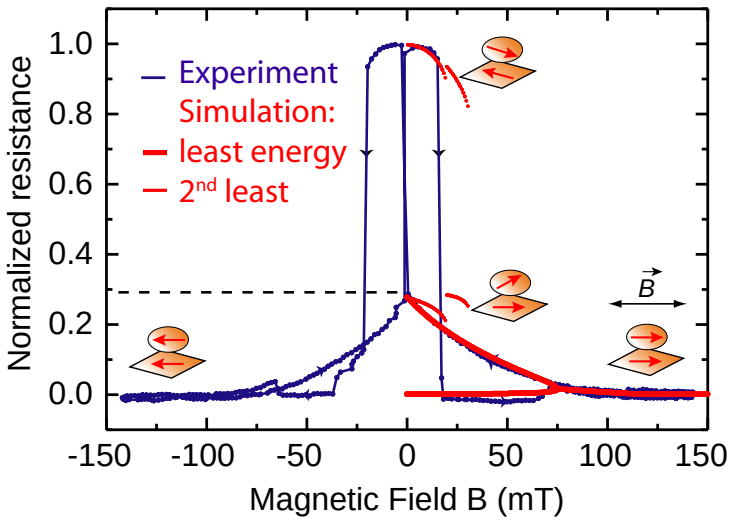


Fig. 20: *CPP-GMR data (blue) measured at 5 K with the magnetic field being applied along a hard axis of the single-crystalline Fe layers. Thick and thin red lines are least and second least energy solutions of a Stoner-Wohlfarth fit, respectively. Pairs of red arrows indicate the relative alignment of magnetizations as derived from the fit. After [20].*

occurs after decreasing the field to zero, where the two magnetizations rest in two different easy axes and, thus, include an angle of $\theta = 90°$. For a standard, symmetric angular dependence of the GMR, $r(\theta) = \sin^2(\theta/2)$, $r(90°)$ would be 0.5. Instead we find a much lower value of 0.3 (dashed line). This deviation originates from enhanced spin accumulation at the Fe/Ag(001) interfaces and yields with Eqs. (6) ans (8) and asymmetry parameter $\Lambda_{GMR} = 1.6$.

5.3 Two-step switching process

Figure 21 shows the differential resistance through the pillar *versus* DC current taken at a magnetic field of 7.9 mT applied roughly along a hard axis. This field is much weaker than the anisotropy field of about 40 mT. Therefore, both magnetizations are aligned along an easy axis and the measurement starts at zero DC current in a low resistive state. In contrast to the data in Fig. 2(b) the switching occurs here in two steps *via* an intermediate resistance level. At a positive current I_{c1} the free layer starts to rotate with respect to the fixed layer. The anisotropy energy minimum at $90°$ stabilizes the orthogonal state. Only at an even higher current I_{c2} the local energy minimum is overcome and the free layer switches to the antiparallel alignment. Upon reversing the current, a similar behavior is observed. At first glance, it is surprising that there are two switching events at different critical currents, because the anisotropy energy barriers to overcome are in both cases the same. The different current densities arise from the angular dependence of the STT efficiency function $g(\theta)$ as can be seen in the macrospin simulations of Fig. 22. The trajectory of the first switching step from parallel to perpendicular with respect to the fixed magnetization is shown in Fig. 22(a). The direction of the damping torque (red) and the STT (blue) during the switching are shown in Fig. 22(c). The viewing direction is along the $-x$-direction. As expected, the STT always points outward and, thus, acts as an excitation. Therefore, the cone angle of the precession around the initial state increases, until the anisotropy energy barrier between the initial and the 90°-state is overcome. Fig. 22(d) shows the torques after the switching; now with the viewing direction in $-y$-direction [note the different abcissae

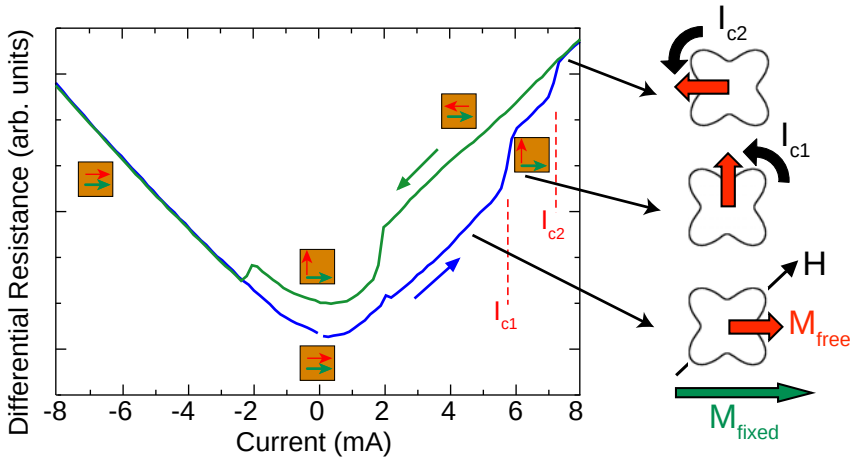


Fig. 21: *Two-step current-induced magnetization switching of a single-crystalline Fe/Ag/Fe(001) nanopillar measured at 5 K. The free layer magnetization switches first from parallel to perpendicular and then to antiparallel alignment relative to the fixed layer magnetization and vice versa for the decreasing current branch (green). A static magnetic field of 7.9 mT, which is weaker than the in-plane anisotropy field, is applied roughly along a hard axis. The diagrams looking like cloverleaves represent the in-plane anisotropy energy landscape. The four minima correspond to easy axis directions. After [20].*

of Figs. 22(c) and (d)]. The symmetry of the STT after switching is completely different. For $m_x < 0$ [right half of Fig. 22(d)] the STT is still pointing outward and is an exciting torque. For $m_x > 0$ [left half of Fig. 22(d)], however, the STT points towards the precession axis and damps the oscillatory motion. The total action of the STT along one revolution tends to cancel out. Therefore, the magnetization relaxes towards the $+x$-direction after the first switching step as can also be seen in Fig. 22(a). The simulation also reproduces the second switching step at a higher DC current from the 90°-state to the antiparallel state [Fig. 22(b)]. Again, the angle of the precession around the initial state increases until the switching occurs. (In this specific simulation the switching is a bit more complicated, because the magnetization overcomes the energy barrier on the "wrong" side of the trajectory, which requires it to reach the final state *via* a "detour".) But how can the STT in Fig. 22(d) at a higher current excite the magnetization and induce the switching? The answer can be found in the magnitude of the STT, which is not constant along the trajectory. In fact, the STT for $m_x < 0$ [right half of Fig. 22(d)] is stronger than for $m_x > 0$ [left half of Fig. 22(d)]. Therefore, the total action along one revolution does not completely cancel out and a small exciting net torque remains. For a large enough critical current I_{c2} this net torque is sufficient to overcome the damping torque and induces the second switching step. This asymmetry of the magnitude of the STT is a consequence of Slonczewsi's theory: For an asymmetry parameter $\Lambda > 1$ the maximum of the STT occurs for angles larger than $\theta = 90°$ (see blue or red curve in Fig. 19(b)]. From a comparison between the experimentally determined ratio of the critical currents I_{c2}/I_{c1} with corresponding values extracted from simulations with varying asymmetry parameter Λ, we obtain $\Lambda_{STT} = 3.4$. These experiments represent the first direct determination of the asymmetry parameter Λ for GMR and STT in the same sample. The deviations between the calculated value of Λ (4.0) and those determined from GMR and STT data (1.6 and 3.4, respectively) most likely arise from imprecise knowledge of material parameters and from simplifying assumptions of the model, such as the complete neglect of the minority channel, which is only approximately fulfilled for Fe/Ag(001).

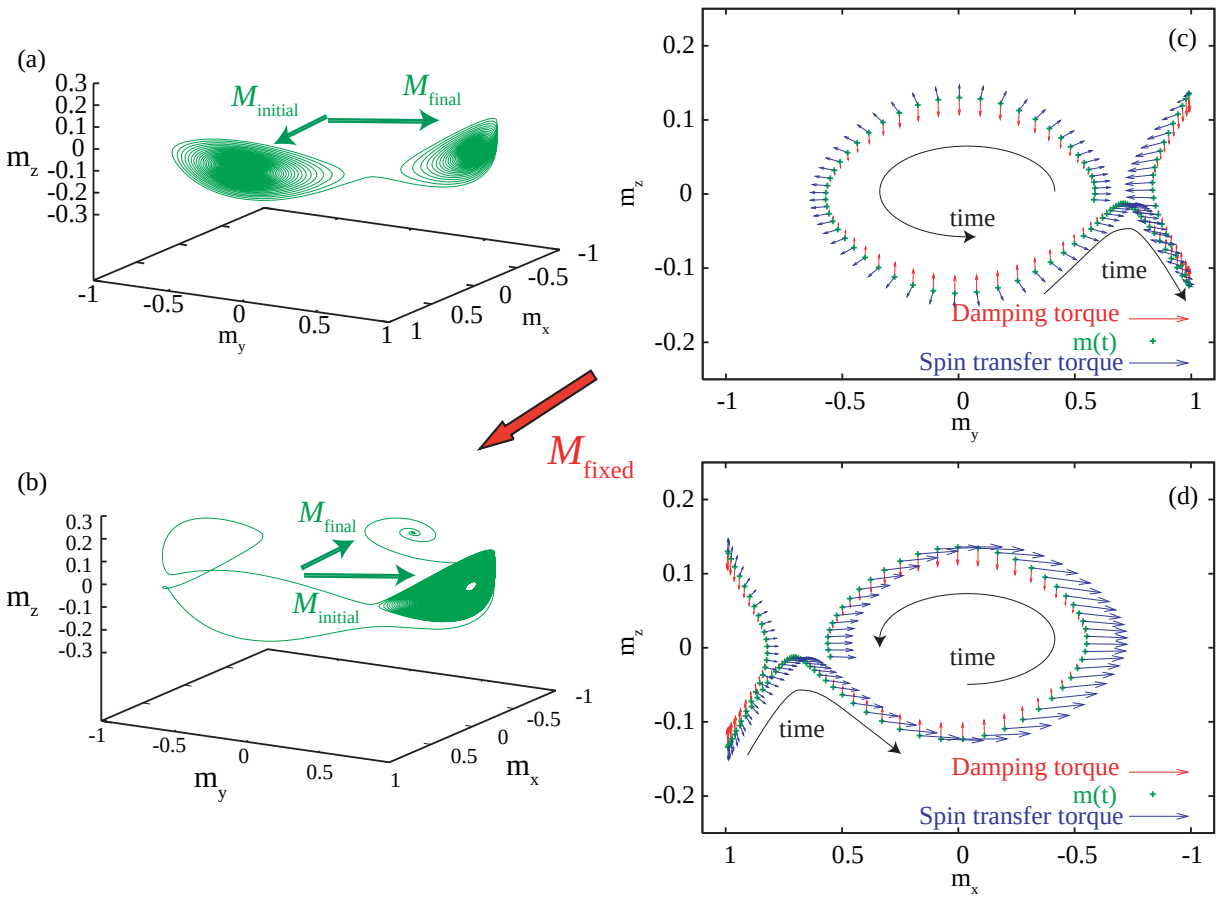


Fig. 22: *STT-induced switching of a macrospin in the presence of cubic magnetocrystalline anisotropy and demagnetizing field. The magnetization switches under the influence of a persistent DC current first from parallel ($+x$-direction) to a 90°-orientation ($+y$-direction) with respect to the fixed layer magnetization and then from the 90°-orientation to the antiparallel alignment ($-x$-direction). (a,b) Trajectories of the two switching events. (c,d) representation of the STT (blue arrows) and damping torque (red arrows) viewed along (b) the initial, parallel and (c) the 90°-orientation of the macrospin. Only a fraction of the trajectory in the immediate vicinity of the switching event (a) is shown in (c) and (d). After [21].*

5.4 Zero-field excitations in the 90°-state

The simulations in Fig. 22 suggest that a steady-state oscillatory mode can be excited at a low external field, if the system is prepared in the 90°-state and a DC current between I_{c1} and I_{c2} is applied. Figure 23 shows microwave spectra measured under these conditions using the setup described in Fig. 10. The frequencies of the observed modes slightly shift with increasing current strength to higher frequencies.

The corresponding macrospin simulation is shown in Fig. 24. We indeed find a precessional trajectory around the $+y$-direction, which is the static direction of the free magnetization in the 90°-state. The analysis of the torques in Fig. 24(b) yields a similar picture as for the second switching step in Fig. 22: The integrated action of the STT along one revolution almost cancels out. Only a relatively small net excitation remains due to the asymmetry of the STT magnitude as a function of the angle θ. The excitation of the STT on the right hand part of the trajectory is larger than the damping on the left hand part. The relatively small cone angle of the trajectory

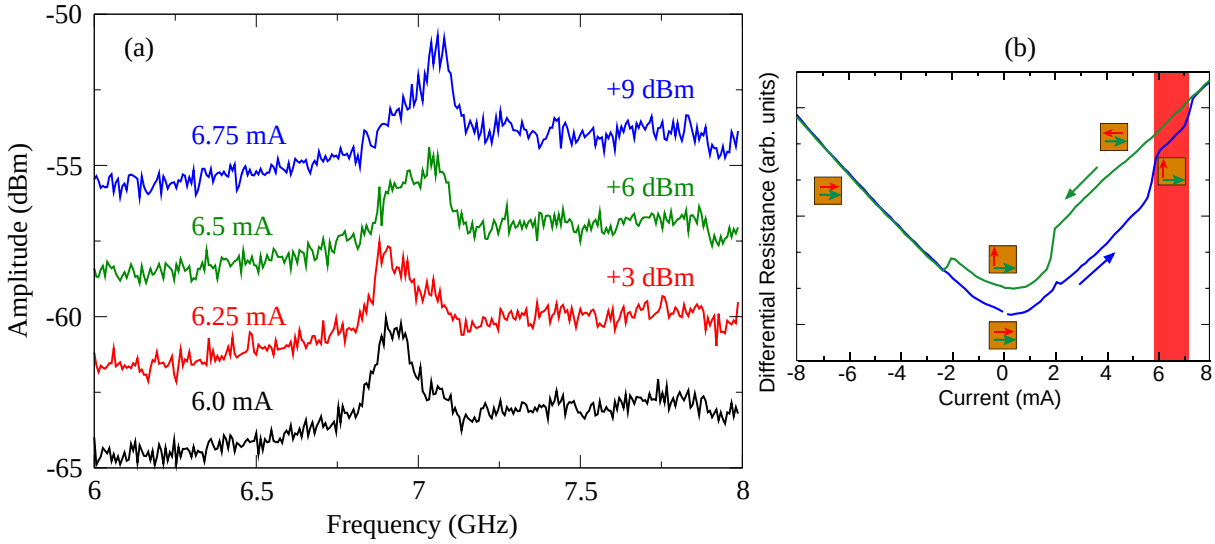


Fig. 23: *(a) Low-field precession: Microwave spectra recorded at 5 K with a weak field of 5 mT applied at an angle of 15° relative to an easy axis and at different DC currents as indicated. All curves are taken at DC currents that correspond to the 90°-state [red range in (b)]. After [21].*

in Fig. 24(a) and the weak peaks in Fig. 23(a) reflect the weak excitation.

The observation of a steady-state precession at low external or even zero field is of importance from the application point of view, because the usual need for an external field exceeding the coercivity of the oscillating layer is disadvantageous as it increases the complexity and cost of STNOs. In our case, the role of the external field is taken by an internal field, namely the magnetocrystalline anisotropy field. Therefore, the observation of low-field excitations in the 90°-state is – as is the case for the two-step switching process presented in Fig. 21 – a direct consequence of the interplay between magnetocrystalline anisotropy and the STT.

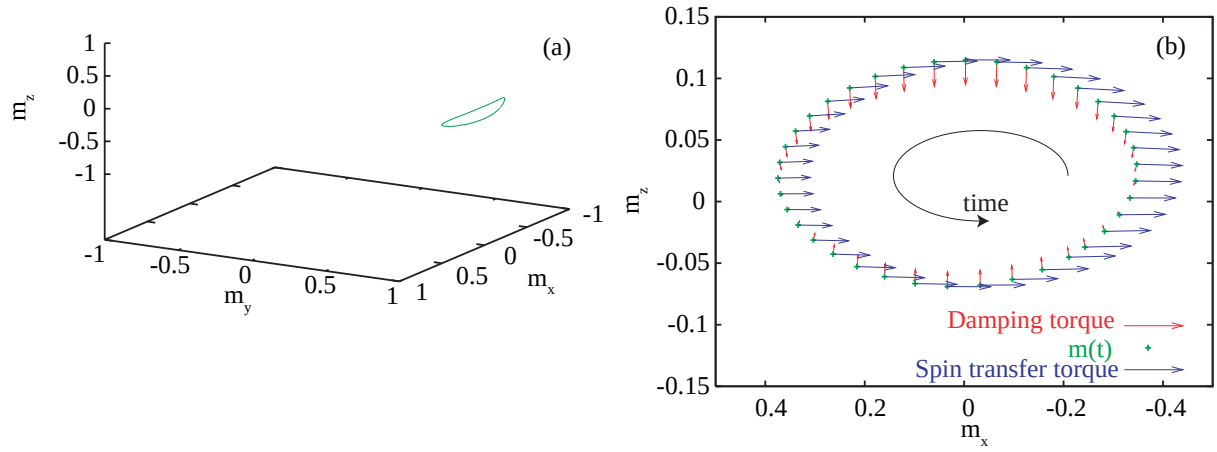


Fig. 24: *STT-induced excitation of low-field precession in the 90°-state in the presence of cubic magnetocrystalline anisotropy and demagnetizing field. (a) Simulated trajectory and (b) representation of the STT (blue arrows) and damping torque (red arrows) viewed along the 90°-orientation ($-y$-direction) of the macrospin. After [21].*

6 Spin-transfer torques in non-uniform magnetization structures

Up to now we have assumed that the magnetic element, which is subject to the STT, displays in the static state –apart from minor deviations, such as the edge effects shown in Fig. 14(a)– a uniform magnetization pattern. However, STT also occurs when a current passes through a non-uniformly magnetized object or when a spin-polarized current impinges on a non-uniform magnetization pattern. In fact, STT in non-uniform structures has recently attained a lot of scientific interest because it reveals novel concepts for the control and manipulation of the magnetic state in nanomagnets. Some of these phenomena are expected to have a strong impact on applications.

6.1 Current-driven domain wall motion

A domain wall is per definition a non-uniform magnetization pattern. In a thin and narrow magnetic wire domain walls divide the system into sections of opposite magnetization directions. A current flowing along the wire has to repolarize each time after passing a domain wall in order to adjust to the local magnetization. This repolarization is associated with a transfer of spin momentum and generates a STT on the local magnetization in such a way that the domain walls move in the direction of the electron flux. The first observation of current-driven domain wall motion is shown in Fig. 25.
In fact, the situation in a wire with a domain wall is very similar to Fig. 4, except that the domain wall plays to role of the spacer layer. Since the magnetization direction is varying

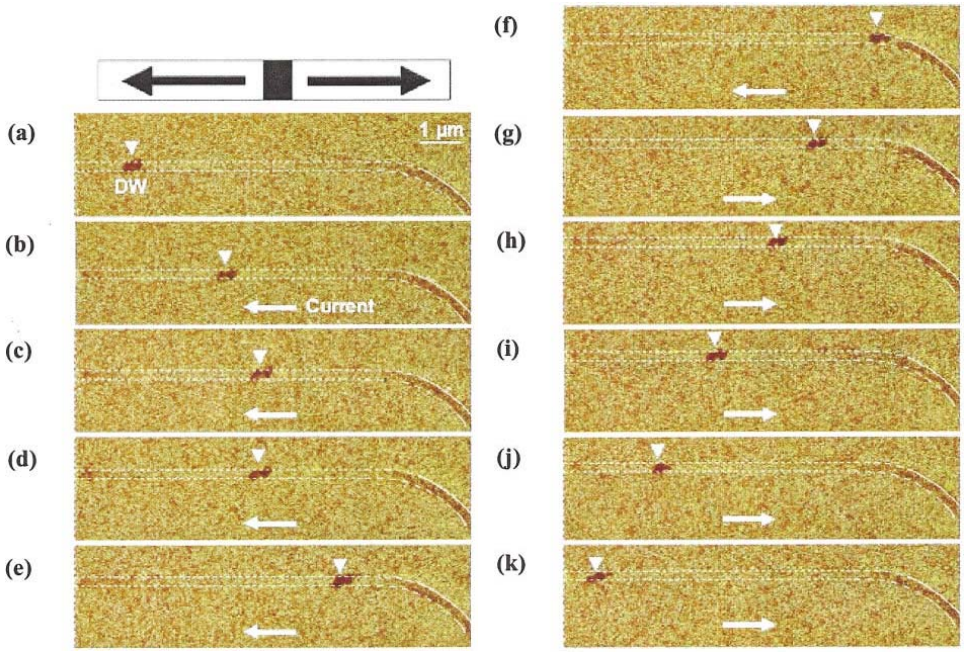


Fig. 25: *Successive magnetic force microscopy (MFM) images with one current pulse applied between each consecutive image. The current density and the pulse duration were* 1.2×10^{12} *A/m*^{2} *and* $0.5\,\mu$*s, respectively. Note that the white arrows indicate the technical current direction; the electron flux is in the opposite direction. After [22].*

continuously within a domain wall, the corresponding torque term to be included in the Gilbert equation needs to be formulated in the continuos limit. The Slonczewski STT term in Eq. (3) becomes the so-called adiabatic STT term, for which polarization adiabaticity (the spin of the charge carrier is always aligned with the local magnetization) and full transfer of angular momentum to the local magnetization is assumed. It turns out [23] that a second, so-called non-adiabatic STT term is needed to achieve agreement with experiments. Several mechanisms, such as linear momentum transfer [24] or spin accumulation in combination with s-d exchange interaction [25], are discussed as the origin of the non-adiabatic term. The relative strength of the adiabatic and non-adiabatic term is an important, yet unresolved issue. In spite of this lacking of a complete understanding of the underlying physics, current-driven domain wall motion is a crucial ingredient for the concept for a novel non-volatile magnetic storage device, the so-called magnetic domain-wall racetrack memory. It will be discussed in detail along with recent developments in the controlled movement of domain walls in magnetic nanowires in the lecture "MRAM and Domain Wall Memories" by Stuart Parkin.

6.2 Current-driven vortex motion and switching

A magnetic vortex is a flux-closure domain pattern with circular magnetization lines in the plane of the sample. Towards the center the magnetization lifts out of the plane in order to avoid a divergence of the exchange energy. This so-called vortex core has typically a diameter of the order of 10 nm, whereas the whole vortex structure is micron-sized. With the exception of a small contribution of the core, no stray field is emanating from a vortex. Therefore, it represents a highly stable magnetization configuration. A vortex is characterized by the vorticity (clockwise or anti-clockwise) and the core polarization (up or down), which are independent and thus allow for four different configurations. The stability together with these two binary quantities renders magnetic vortices interesting objects for magnetic storage. The controlled switching of the polarity or the vorticity is a indispensable requirement for the writing process of vortex-based memory cells. Recent micromagnetic simulations by Liu *et al.* [26] predicted a current-induced core switching process based on STT. The current is applied in the plane of the vortex and is polarized while passing from the edge to the central part of the vortex. Spin momentum is then transferred to the second half, which initially is magnetized in the opposite direction. Details of this process as well as the excitation of the gyrotropic motion of the vortex core by STT are discussed in the lecture "Micromagnetism" by Riccardo Hertel.

6.3 GMR and current-induced torques in metallic antiferromagnets

Recent calculations [27, 28, 29] suggest that current-induced torques and GMR also occur in antiferromagnetic metals, *i.e.* in antiferromagnet/paramagnet/antiferromagnet structures. The basic result is the following: A current impinging on a paramagnet/antiferromagnet interface is partly transmitted into the antiferromagnet and partly reflected. The transmitted electrons conserve their spins. Spin filtering is not effective because there is no exchange splitting of the bandstructure in an antiferromagnet. The spin orientations of the reflected electrons, however, are rotated around the order parameter $\vec{n}$ of the antiferromagnet in opposite senses depending on their direction of incidence. This leads to a spin polarization of the reflected current along the direction of $\vec{n}$. Next, we consider an antiferromagnet/paramagnet/antiferromagnet heterostructure as shown in Fig. 26, where the two antiferromagnets are characterized by the order parameters $\vec{n}_{1,2}$. If $\vec{n}_1$ and $\vec{n}_2$ are non-collinear, then multiple reflection processes from the antiferromag-

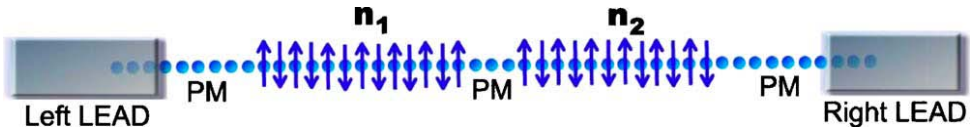


Fig. 26: *Antiferromagnet/paramagnet/antiferromagnet heterostructure. The antiferromagnets are assumed to be commensurable spin density wave systems. PM stands for paramagnet. After [27].*

netic layers lead to a spin-density with the quantization axis along $\vec{n}_\perp = (\vec{n}_1 \times \vec{n}_2)/|\vec{n}_1 \times \vec{n}_2|$. The spin-density is periodic with the same period as the static spin-density wave in the antiferromagnets. Figure 27 shows the Landau-Bütiker conductance calculated for this situation as a function of the angle θ between $\vec{n}_1$ and $\vec{n}_2$. $\theta = 0$ corresponds to the configuration sketched in Fig. 26, where the two spins next to the spacer layer are parallel. The angular dependence of this so-called antiferromagnetic giant magnetoresistance (AGMR) is similar to that of the GMR effect. The effect size in this particular calculation amounts to noticeable 12%. Since interference is involved, AGMR is expected to be restricted to spacer layers thinner than the phase coherence length. This condition is not very severe because it is similar to that for interlayer exchange coupling; another effect that arises from interfering wave functions in the spacer layer. Another important issue for the experimental realization are (lateral) domains in the antiferromagnets, which tend to reduce the AGMR due to destructive interference.

Next we consider the interaction of the current-induced spin-density wave with the antiferromagnetic layer. We distinguish between (i) the current-induced spin-density amplitude at lattice site i, $\vec{W}_i$, which is a non-equilibrium, steady-state quantity and is due to the transport electrons

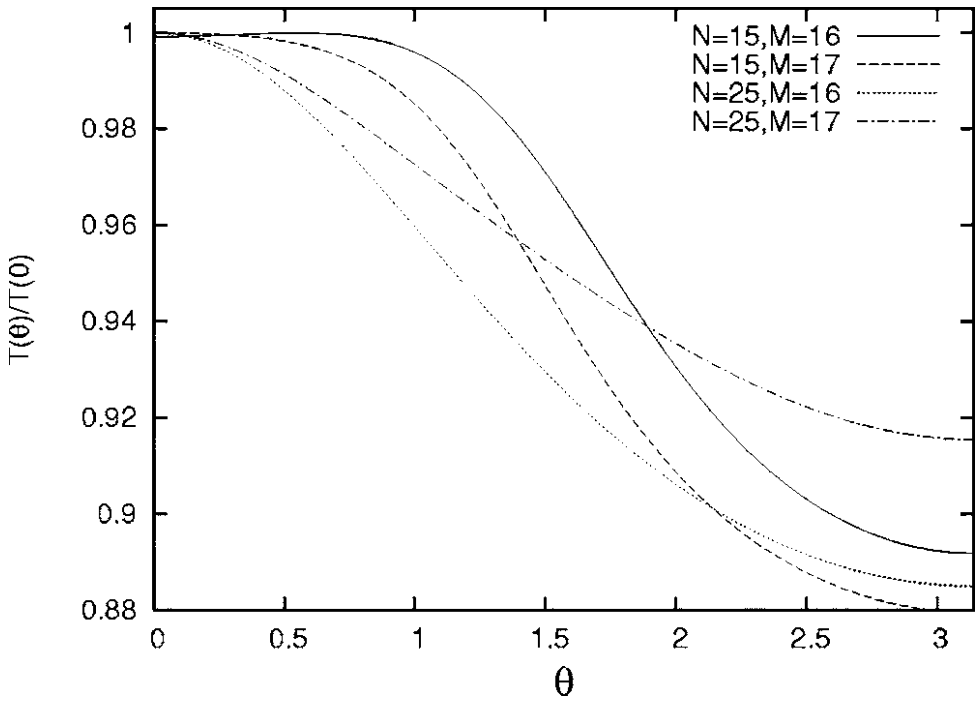


Fig. 27: *Landau-Büttiker conductance as a function the angle θ between $\vec{n}_1$ and $\vec{n}_2$ for the structure shown in Fig. 26. N and M denote the number of atomic layers of the two antiferromagnetic layers. After [27].*

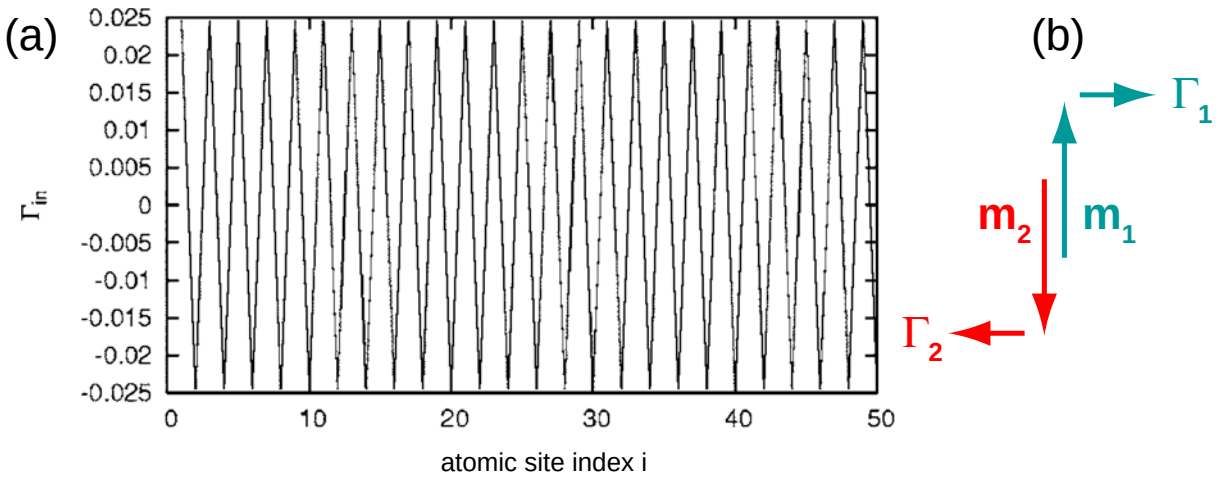


Fig. 28: *(a) Local current-induced torques $\vec{\Gamma}_i$ on different lattice sites i of the antiferromagnetic layer. The component Γ_{in} in the plane defined by $\vec{n}_1$ and $\vec{n}_2$ is shown. (b) Cooperative rotation of all local moments $\vec{m}_i$ in the antiferromagnet. After [27].*

with energies within the transport window and (ii) the local magnetization $\vec{m}_i$ due to the electrons outside the transport window. The two subsystems interact and can mutually precess. Writing down an equivalent equation to Eq. (3) for each lattice site i we obtain the torque $\vec{\Gamma}_i$ exerted on $\vec{m}_i$:

$$\vec{\Gamma}_i \propto \vec{m}_i \times \vec{W}_i \times \vec{m}_i. \tag{9}$$

The component Γ_{in} in the plane defined by $\vec{n}_1$ and $\vec{n}_2$ changes sign from each atomic site to the next [Fig. 28(a)]. Therefore, the net action of the current-induced torques in the antiferromagnet is a cooperative rotation of all local moments $\vec{m}_i$ with the same sense of rotation. For the example sketched in Fig. 28(a) the order parameter of the antiferromagnet rotates clockwise. This rotation does not cost exchange energy. Note that in Fig. 28(a) the amplitude of the torque does not decay with distance from the interface, where the current enters at $i = 0$. This is a remarkable difference to the situation in a ferromagnetic layer, where the STT rapidly decays over the Fermi wavelength (few lattice constants) due to the spatial precession of the transmitted spins as explained in Fig. 3(a). The spatial precession is not present in an antiferromagnet because there is no exchange splitting of the bandstructure, *i.e.* $\Delta k = k^{\downarrow} - k^{\uparrow} = 0$ (see Sect. 2.2). As a consequence the current-driven torque in an antiferromagnet is a bulk effect. In the presence of scattering it acts within an inelastic scattering length of the interface with the paramagnet. This volume is about one order of magnitude larger than for a ferromagnet. Therefore, the current-induced torque effect in metallic antiferromagnets is expected to be sizable. The absence of shape anisotropy for an antiferromagnet leaves the crystalline anisotropy as the only contribution to the effective field and yields low estimates for the critical current density for switching of the order of 10^5 A/m^2 [27]. Finally note, that I use the term "current-induced torque" instead of "spin-transfer torque" for the case of metallic antiferromagnets. The latter implies conservation of angular momentum, which indeed was a main ingredient for the explanation of STT presented in Sect. 2.2. Here, however, there is no transfer of spin angular momentum between the subsystems. This can directly be seen in Fig. 28(a), where the integrated torque and, thus, the transferred angular momentum, vanishes.

7 Conclusions

The field of spin-transfer torque effects is rapidly evolving and bears the potential for further exciting physics and applications, *e.g.* the realization of the magnetic analog of the injection laser. This device would provide *spin-wave amplification by stimulated emission of radiation*, and, accordingly, the name SWASER is already suggested [3].

Acknowledgment

I thank Ronald Lehndorff and Julius Mennig for critically discussing and carefully proofreading this manuscript.

References

[1] M. D. Stiles and J. Miltat, in *Spin Dynamics in Confined Magnetic Structures III*, Vol. 101 of *Topics in Applied Physics*, edited by B. Hillebrands and A. Thiaville (Springer, Berlin, 2006), Chap. Spin Transfer Torque and Dynamics.

[2] J. C. Slonczewski, J. Magn. Magn. Mater. **159**, L1 (1996).

[3] L. Berger, Phys. Rev. B **54**, 9353 (1996).

[4] J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Myers, and D. C. Ralph, Phys. Rev. Lett. **84**, 3149 (2000).

[5] M. D. Stiles and A. Zangwill, Phys. Rev. B **66**, 014407 (2002).

[6] M. D. Stiles and A. Zangwill, J. Appl. Phys. **91**, 6812 (2002).

[7] S. Zhang, P. M. Levy, and A. Fert, Phys. Rev. Lett. **88**, 236601 (2002).

[8] F. J. Albert, N. C. Emley, E. B. Myers, D. C. Ralph, and R. A. Buhrman, Phys. Rev. Lett. **89**, 226802 (2002).

[9] S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. E. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, Nature **425**, 380 (2003).

[10] K.-J. Lee, A. Deac, O. Redon, J.-P. Nozières, and B. Dieny, Nature Materials **3**, 877 (2004).

[11] R. Hertel, unpublished results.

[12] Z. Li and S. Zhang, Phys. Rev. B **68**, 024404 (2003).

[13] J.-G. Zhu and X. Zhu, IEEE Trans. Magn. **40**, 182 (2004).

[14] A. M. Deac, A. Fukushima, H. Kubota, H. Maehara, Y. Suzuki, S. Yuasa, Y. Nagamine, K. Tsunekawa, D. D. Djayaprawira, and N. Watanabe, Nature Physics **4**, 803 (2008).

[15] S. Kaka, M. Pufall, W. H. Rippard, T. J. Silva, S. E. Russek, and J. Katine, Nature **437**, 389 (2005).

[16] J. Grollier, V. Cros, and A. Fert, Phys. Rev. B **73**, 060409 (2006).

[17] H. Dassow, R. Lehndorff, D. E. Bürgler, M. Buchmeier, P. A. Grünberg, C. M. Schneider, and A. van der Hart, Appl. Phys. Lett. **89**, 222511 (2006).

[18] M. D. Stiles and D. R. Penn, Phys. Rev. B **61**, 3200 (2000).

[19] J. C. Slonczewski, J. Magn. Magn. Mater. **247**, 324 (2002).

[20] R. Lehndorff, D. E. Bürgler, A. Kakay, R. Hertel, and C. M. Schneider, Phys. Rev. B **76**, 214420 (2007).

[21] R. Lehndorff, D. E. Bürgler, A. Kakay, R. Hertel, and C. M. Schneider, IEEE Trans. Magn. **44**, 1951 (2008).

[22] A. Yamaguchi, T. Ono, S. Nasu, K. Miyake, K. Mibu, and T. Shinjo, Phys. Rev. Lett. **92**, 077205 (2004).

[23] A. Thiaville, Y. Nakatani, J. Miltat, and Y. Suzuki, Europhys. Lett. **69**, 990 (2005).

[24] G. Tatara and H. Kohno, Phys. Rev. Lett. **92**, 086601 (2004).

[25] S. Zhang and Z. Li, Phys. Rev. Lett. **93**, 127204 (2004).

[26] Y. Liu, S. Gliga, R. Hertel, and C. M. Schneider, Appl. Phys. Lett. **91**, 112501 (2007).

[27] A. S. Núñez, R. A. Duine, P. Haney, and A. H. MacDonald, Phys. Rev. B **73**, 214426 (2006).

[28] P. M. Haney, D. Waldron, R. A. Duine, A. S. Núñez, H. Guo, and A. H. MacDonald, Phys. Rev. B **75**, 174428 (2007).

[29] Y. Xu, S. Wang, and K. Xia, Phys. Rev. Lett. **100**, 226602 (2008).

D 4 Femtosecond opto-magnetism: ultrafast laser manipulation of spins

Theo Rasing, Alexey Kimel, Andrei Kirilyuk
Institute for Molecules and Materials
Radboud University Nijmegen, the Netherlands

Contents

1 Introduction **2**

2 Interaction between photons and spins **2**
2.1 Spin motion 2
2.2 Magneto-optics and opto-magnetism 3

3 Measurements of the inverse Faraday effect at femtosecond time-scales **5**

4 Coherent control of spins **9**
4.1 Double pump excitation 9
4.2 Pulse shaping 11

5 All optical magnetic recording **13**

6 Conclusions **16**

1 Introduction

Optical pulses could be an alternative stimulus for spins, see Figure 1. Indeed, due to tremendous developments of ultrafast laser sources over the past 20 years, a laser pulse has become one of the shortest ever man-made events. It has been recently demonstrated that excitation of a magnetically ordered material with an ultrashort (10^{-13} seconds and shorter) laser pulse may result in demagnetization [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15], spin-reorientation [16, 17], or even modification of magnetic structure [18, 19] and this all on a time-scale of 1 picosecond or less. Nevertheless, for all the above-mentioned experiments, the observed laser-induced magnetic changes were of *thermal* origin. It means that the magnetic changes were the result of optical absorption followed by a rapid temperature increase. This thermal origin of spin excitation considerably limits potential applications because the repetition frequency is limited by the cooling time [20]. Moreover, because of this thermal excitation mechanism, heat diffusion will also limit the recording density. The solution to both these problems could be a *nonthermal* optical control of magnetism.

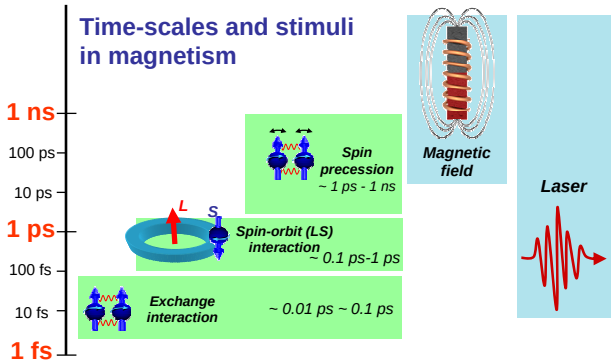


Fig. 1: *Time scales and stimuli in magnetism.*

Can light manipulate and control the magnetization? The essence of magnetization is angular momentum. A circularly polarized photon carries angular momentum too. If it would be possible, using this angular momentum, to affect spins of electrons directly, this would result in ultrafast and nonthermal laser control of magnetism, since right- and left-handed circularly polarized light-waves should affect spins as magnetic fields of opposite sign. This lecture summarizes the recent progress in the study of ultrafast nonthermal effects of light on magnetic materials.

2 Interaction between photons and spins

2.1 Spin motion

Let us discuss first the motion of spins in an external magnetic field. It is known that a spin $\mathbf{S}$ in a magnetic field $\mathbf{H}$ experiences a torque $\mathbf{T} = -|\gamma|\mathbf{S} \times \mathbf{H}$, where γ is the gyromagnetic ratio. From the fundamental law of the conservation of angular momentum one can find an equation for the spin motion in the external magnetic field

$$\frac{\delta \mathbf{S}}{\delta t} = -|\gamma|\mathbf{S} \times \mathbf{H}, \tag{1}$$

This equation shows that the time derivative of the spin $\mathbf{S}$ is perpendicular to both the magnetic field $\mathbf{H}$ and the spin $\mathbf{S}$ itself. Thus the spin $\mathbf{S}$ in an external magnetic field $\mathbf{H}$ will precess around this field. It can be shown that the frequency of this precession is a linear function of the external magnetic field. In ferro- and ferrimagnetic materials it is convenient to introduce the vector of the magnetization $\mathbf{M} = -|\gamma|\mathbf{J}$, where $\mathbf{J}$ is the total angular momentum of a unit volume. The equation of motion of the magnetization can be written in a similar way

$$\frac{\delta \mathbf{M}}{\delta t} = -|\gamma|\mathbf{M} \times \mathbf{H}. \tag{2}$$

If the spin experiences not only the action of the external magnetic field, but is also affected by the magneto-crystalline anisotropy, shape anisotropy, magnetic dipole interaction etc, the situation becomes more complicated. All these interactions will contribute to the thermodynamical potential Φ and the combined action of all these contributions can be considered as an effective magnetic field

$$\mathbf{H}^{eff} = -\partial\Phi/\partial\mathbf{M} \tag{3}$$

Thus the equation of motion of the magnetization vector can be written as

$$\frac{\delta \mathbf{M}}{\delta t} = -|\gamma|\mathbf{M} \times \mathbf{H}^{eff}, \tag{4}$$

This expression is also known as the Landau-Lifshitz equation in the nondissipative approximation [21, 3].

2.2 Magneto-optics and opto-magnetism

Let us now discuss the interaction between photons and spins using the energy consideration [22]. It can be shown that the thermodynamical potential Φ of an isotropic, nonabsorbtive, magnetically ordered medium with static magnetization $\mathbf{M}(0)$ and in a monochromatic light field $\mathbf{E}(\omega)$ includes a term:

$$F = \alpha_{ijk} E_i(\omega) E_j(\omega)^* M_k(0), \tag{5}$$

where α_{ijk} is the magneto-optical susceptibility [23, 4, 24, 25, 26]. In the electric dipole approximation the linear optical response of a medium to a field $\mathbf{E}(\omega)$ is defined by the optical polarization $\mathbf{P}(\omega)$=$\partial\Phi/\partial\mathbf{E}(\omega)^*$. From Eq. 5 one can easily see that the optical polarization $\mathbf{P}(\omega)$ should have a contribution $\mathbf{P}^{(m)}$ proportional to the magnetization $\mathbf{M}$:

$$P_i^{(m)}(\omega) = \alpha_{ijk} E_j(\omega) M_k(0). \tag{6}$$

From this equation one can find that when linearly polarized light is transmitted through a magnetized medium, the polarization plane of the light gradually rotates over the angle θ_F given by:

$$\theta_F = \frac{\alpha_{ijk} M_k(0) \cdot \omega L}{cn}, \tag{7}$$

where c is the speed of light in the vacuum, n is the refraction coefficient of the medium, and L is the propagation distance of the light in the medium [27]. Equation (7) describes the so-called magneto-optical Faraday effect discovered by M. Faraday in 1846 [28].

The magneto-optical Faraday effect clearly demonstrates that a magnetically ordered medium can indeed affect photons and change the polarization of light [27]. Is the inverse phenomenon

feasible, that is, is it possible that polarized photons affect the magnetization? From Eq. 5 one can see that an electric field of light at frequency ω will act on the magnetization as an effective magnetic field $\mathbf{H}^{eff}$ directed along the wave-vector of the light **k**:

$$H_k = -\frac{\partial F}{\partial M_k} = \alpha_{ijk} E_i(\omega) E_j(\omega)^*. \tag{8}$$

In isotropic media, α_{ijk} is a fully antisymmetric tensor with a single independent element α. Therefore, Eq. (8) could be re-written as

$$\mathbf{H} = \alpha[\mathbf{E}(\omega) \times \mathbf{E}(\omega)^*]. \tag{9}$$

From this it becomes obvious that right- and left-handed circularly polarized waves should act as magnetic fields of opposite sign [23, 4, 24, 25]. Therefore, it is seen from Eq. (9) that in addition to the well-known magneto-optical Faraday effect, where the polarization of light is affected by the magnetization **M**, the same susceptibility α also determines the inverse, opto-magnetic phenomenon: circularly polarized light affects the magnetization via the inverse Faraday effect. Thus in a thermodynamical approach, the effect of light on spins in a magnetically ordered material can be described by the Landau-Lifshitz equation, where the magnetic field is generated by light via the inverse Faraday effect.

However, it must be noted that Eq. (9) has been derived for a monochromatic optical excitation and using the approximation of thermal equilibrium. It is therefore interesting to consider the situation when the product $[\mathbf{E}(\omega) \times \mathbf{E}(\omega)^*]$ changes much faster than the fundamental time-scales in a magnetically ordered material, given by the spin precession period and the spin-lattice relaxation time [25]. Consider the excitation of spins by a laser pulse with duration $\Delta t =$ 100 fs. Using Fourier transformation one can see that such a laser pulse, being ultrashort in the time domain, is spectrally broad ($\Delta\omega \sim$5 THz). Such laser excitation of spins is conveniently described in the frequency domain (see Figure 2). Initially the electron is in the ground state $|1\rangle$ and its spin is up. If the state is non-degenerate, being an orbital singlet, the spin-orbit coupling for the electron in this state can be neglected. If we act on this electron with a photon, during the optical transition the wave-function of the electron becomes a superposition of several eigenstates. This will effectively increase the orbital momentum of the electron, leading to an increased spin-orbit coupling and thus resulting in an intensification of the spin-flip process. If the energy of the photon is smaller than the gap between the ground state $|1\rangle$ and the nearest excited state $|2\rangle$, the photon will not excite any real electronic transition, but just result in a spin-flip of the electron in the ground state. (As the excitation involves a virtual state, the de-excitation process is very fast). In other words, the spin-flip in the ground state is due to the fact that circularly polarized light mixes a fraction of the excited-state wave function into the ground state [25]. This process will be accompanied by the coherent re-emission of a photon of energy $\hbar\omega_2 = \hbar(\omega_1 - \Omega_m)$. In magnetically ordered materials $\hbar\Omega_{\mathrm{m}}$ corresponds to the energy of a magnon. Moreover, such a laser-induced spin-flip process can be coherently stimulated if both frequencies ω_1 and ω_2 are present in the laser pulse (see Fig. 2). The time of the spin-flip process τ_{sf} is given by the energy of the spin-orbit interaction in the perturbed ground state E_{SO}. For materials with a large magneto-optical susceptibility the energy of the spin-orbit coupling may exceed 50 meV [29] and thus the spin-flip process can be as fast as $\tau_{sf} \sim \hbar/E_{SO} \sim 20$ fs. Note that such a spin-flip process is allowed in the electric-dipole approximation [30]. In contrast to magnetic dipole transitions, this mechanism is much more effective and does not require annihilation of a photon. It means that the energy transfer from photons to spins (magnons) is

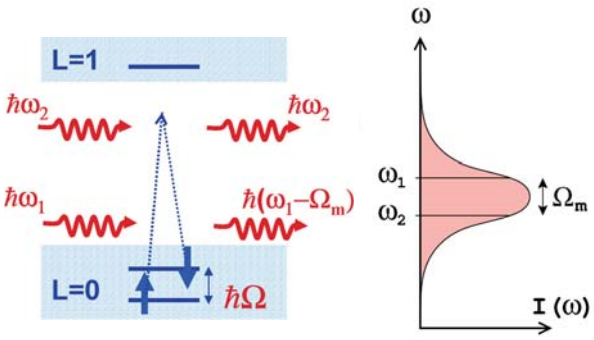


Fig. 2: *Ultrafast spin-flip via the process of the stimulated Raman scattering.*

realized via an inelastic scattering process. While some photons loose a small part of their energy, the total number of photons remains unchanged. It is important to note that in a spin-flip via stimulated Raman scattering, as described above, the stimulating and re-emitted photons have identical polarization, implying that such induced spin-flip is not accompanied by a loss of the angular momentum of the photons.

3 Measurements of the inverse Faraday effect at femtosecond time-scales

The inverse Faraday effect was already predicted more than 40 years ago in a theoretical analysis by Pitaevskii. He showed that circularly polarized light acts on a transparent dispersive medium as an effective magnetic field and this may result in magnetization of the medium [23]. Soon after the predicted inverse Faraday effect was observed in paramagnetic solids [25, 24] and in plasmas [31]. These earlier studies demonstrated that excitation of a medium with a circularly polarized laser pulse corresponded to the action of an effective magnetic field. For a 30 ns laser pulse with a fluence of 10^7 W/cm^2 the strength of the effective magnetic field was only 0.01 G. Modern ultrafast laser systems are able to generate pulses shorter than 100 fs (10^{-13} sec) and the fluence of the laser excitation may exceed 10^{12} W/cm^2. Consequently one may expect that a far stronger inverse Faraday effect is feasible with such pulses. Indeed, recent experimental studies of the inverse Faraday effect in plasmas reported that circularly polarized light is able to create an axial magnetic field of tens of kilogauss [32].

Only very recently an ultrafast inverse Faraday effect in a magnetically ordered material was demonstrated for the first time [36]. The measurements were performed at a photon energy of 1.55 eV using amplified 100 fs pulses from a Ti:sapphire laser at a repetition rate of 1 kHz (Fig. 3). Using a pump and probe configuration, a laser pulse is split into two parts. The most intensive pulse is used as a pump, while the less intensive one is used as a probe. Both pulses follow different optical paths and are focused to the same spot on the sample. Due to the direct magneto-optical Faraday effect, the magnetic changes induced by the pump can be monitored via measurements of the angle of rotation of the polarization plane of the probe α_F. For the detection of the Faraday rotation a balanced two-photodiode detector was used. Varying the time delay between pump and probe pulses one can monitor the temporal evolution of the magnetic changes in the medium induced by the pump, with a resolution limited by the pulse width. Making use of the fact that the speed of light is finite but high, the delay between pump

and probe pulses can be sensitively varied by changing the length of the pump or probe paths: a difference of 1 μm corresponds to a time difference of about 3 fs. The latter was accomplished with the help of a mechanical delay-line that moved a retro reflector in the probe path. In order to improve the signal-to-noise ratio, a combination of gate integration and lock-in technique was used. In particular, a chopper modulated the pump beam at the frequency of 500 Hz. This chopper was synchronized with the laser source such that it chopped out every second laser pulse. The signal from the detector of the probe beam was sent to the gated integrator. The latter was synchronized with the laser and set to integrate the signal from the detector in a limited time-window. Since the photo-detector was rather slow, this time-window was set to a few microseconds. The output signal from the integrator was fed to the lock-in amplifier. The latter used the 500 Hz modulation frequency of the pump as a reference for the phase sensitive detection of the signal. A detailed description of a similar detection technique has been also described elsewhere [37]. The intensity ratio between the pump and probe pulses was about a 100. Both beams were focused on the sample to a spot diameter of about 200 μm for the pump and somewhat smaller for the probe beam. The pump fluence on the sample was around 10 mJ cm^{-2}. The measurements were done in a cold finger cryostat where the temperature could be stabilized in the range 15-300 K with a precision better than 0.5 K.

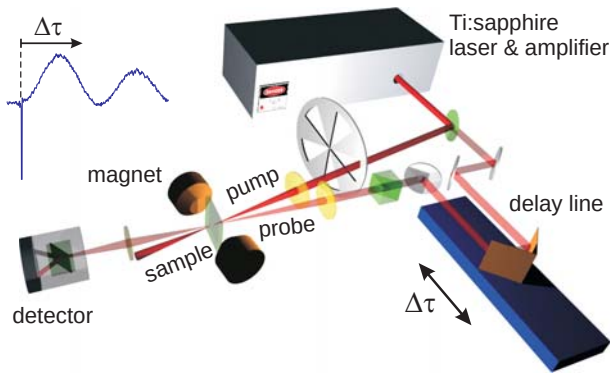


Fig. 3: *The scheme of the experimental setup for for studies of laser-induced spin excitations with sub-picosecond temporal resolution.*

In Ref. [36], the first observation of an inverse Faraday effect in a magnetically ordered structure, the material under study was $DyFeO_3$. The rare-earth orthoferrites $RFeO_3$ are a well-studied family of magnetic compounds [38]. These materials crystallize in an orthorhombic perovskite-type structure with four molecular units per unit cell, with space-group symmetry (Pbnm). The spins of the Fe^{3+} ions ($3d^5$, ground state $^6A_{1g}$, S=5/2) are coupled antiferromagnetically by isotropic exchange. The Dzyaloshinskii-Moriya interaction [39, 40] leads to a slight canting of opposite spins over an angle of about 0.5°, giving rise to a spontaneous magnetization $4\pi M_s$=100 G at room temperature. Despite the fact that the magnetization in the dysprosium orthoferrite is small, this material exhibits a large Faraday rotation owing to their strong spin-orbit interaction [27, 29]. Thus the inverse Faraday effect is expected to be large in this compound as well.

Figure 4 shows the temporal response of the Faraday rotation in a $DyFeO_3$ sample cut perpendicularly to the [001] crystallographic axis for two circularly polarized pump pulses of opposite helicities. At zero time delay instantaneous changes of the Faraday rotation are observed. These ultrafast changes of the magneto-optical signal can be explained in terms of nonlinear optical polarizations induced during virtual and real optical transitions in the Fe^{3+} ions. The instanta-

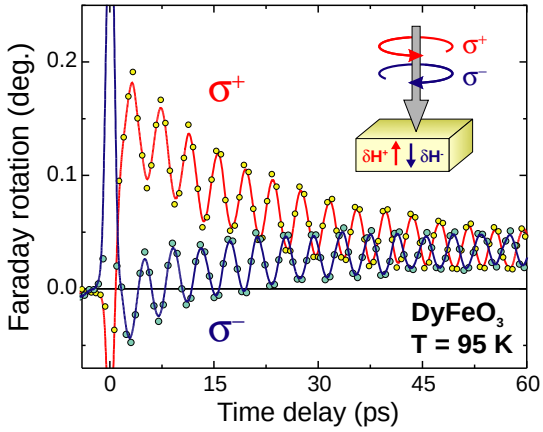


Fig. 4: *Magnetic excitations in $DyFeO_3$ probed by the magneto-optical Faraday effect. Two processes can be distinguished: (1) instantaneous changes of the Faraday effect due to the photo excitation of Fe ions and relaxation back to the high spin ground state S=5/2; (2) oscillations of the Fe spins around their equilibrium direction with an approximately 5 ps period. The circularly polarized pumps of opposite helicities excite oscillations of opposite phase. Inset shows the geometry of the experiment. Vectors δH^+ and δH^- represent the effective magnetic fields induced by right-handed σ^+ and left-handed σ^+ circularly polarized pumps, respectively. [36]*

neous changes of the Faraday rotation are followed by oscillations with a frequency of about 300 GHz. The frequency of these oscillations is in excellent agreement with the frequency of the antiferromagnetic resonance in $DyFeO_3$ (see below inset in Fig. 5). Thus the oscillations of the Faraday rotation can be clearly assigned to spin oscillations. It is seen from the figure that the phase of the laser-triggered spin oscillations depends on the helicity of the pump pulse. Right-handed and left- handed circularly polarized pulses excite spin oscillations of opposite phase. Therefore, this experiment clearly indicates an ultrafast and efficient coupling between angular momentum of photons and spins.

Sending laser pulses along different crystallographic axes of $DyFeO_3$, different modes of spin wave precession could be excited, which gives us further insight into the excitation mechanism. Fig. 5 shows the temperature dependence of the frequencies of the spin oscillations excited by circularly polarized laser pulses. In particular, filled and open circles show the frequencies of the oscillations excited by laser pulses propagating along the [001] and [100] crystallographic axes, respectively. The lines show the temperature behavior of the frequencies of the quasi-antiferromagnetic and quasi-ferromagnetic modes of magnetic resonance in $DyFeO_3$. The figure clearly shows that laser pulses propagating along the [001] crystallographic axes trigger the quasi-antiferromagnetic mode of the magnetic resonance. Note that this mode can only be triggered if light acts on the spins as an effective magnetic field directed along the [001] axis. Similarly, laser pulses propagating along the [100] crystallographic axis trigger the quasi-ferromagnetic mode of the magnetic resonance, indicating that such pulses act on the spins as an effective magnetic field directed along the [100] axis. Above T=55 K the magnetic anisotropy favors orientation of the spins along the x axis with the weak magnetic moment aligned along the z axis. The action of a strong magnetic field pulse will deviate the spins from their equilibrium. Analyzing the torques experienced by the spins S_1 and S_2 due to the action of the magnetic field pulse δH directed along the z or x axis (see Eq. 2) and taking into account the antiferromagnetic coupling between the spins, one finds that the magnetic field will trigger spin

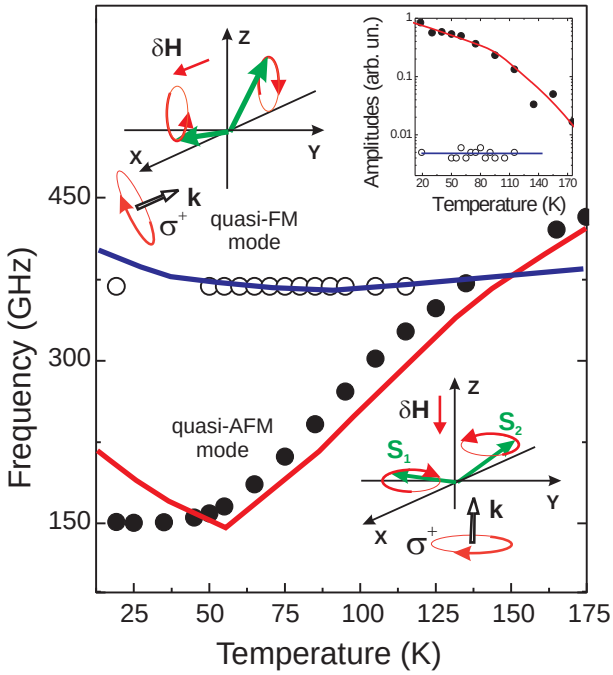


Fig. 5: *Temperature dependence of the frequencies of the observed spin oscillations. Filled and open circles show the frequencies of the excited oscillations for laser pulses propagating along [001] and [100] crystallographic axes, respectively. Red and blue lines show the frequency of the quasi-antiferromagnetic (quasi-AFM) and the quasi-ferromagnetic (quasi-FM) resonance modes from Refs. [38, 43]. Top right inset shows the temperature dependence of the oscillation amplitudes. Top left and bottom right insets are respectively schematic representations of the quasi-FM and quasi-AFM modes of the spin resonance. Vectors δH show the directions of the instantaneous magnetic field that is equivalent to the laser excitation. [36]*

oscillations around the equilibrium orientation and along the elliptical trajectories as shown on the insets to Fig. 5. This has been also confirmed by a detailed theoretical investigation of the laser-induced antiferromagnetic resonance in $DyFeO_3$ [41]. All these studies unambiguously show that a circularly polarized laser pulse acts on the spins as a short effective magnetic field pulse directed along the wave-vector of light. A rough estimate shows that spin oscillations with an amplitude as observed in the experiment can be triggered if a 100 fs laser pulse acts on the spins as an equally short pulse of an effective magnetic field up to 1 T [42].

In addition it was verified experimentally, see Fig. 6(a) [36], that the amplitude of the effect is linear-proportional to the incident laser intensity. This is in full agreement with Eq. (8) that shows the quadratic dependence of the effective field on the incoming electromagnetic field amplitudes, that is, linear on the intensity. Note that the extrapolation of the intensity dependence shows that the photoinduced effect on the magnetization would reach the saturation value of M_s at a pump fluence of about 500 mJ/cm^2. The effect of such a 100 fs laser pulse on the magnetic system would be equivalent to the application of a magnetic field pulse of about 10 T. According to our measurements, the absorption in $DyFeO_3$ in the near-infrared spectral range is of the order of 100-200 cm^{-1}. Given this low value of the absorption, a photoexcitation of 500 mJ/cm^2 is still below the damage threshold of $DyFeO_3$ and is thus quite feasible, given a sample of high optical quality.

Fig. 6(b) shows the behavior of the excitation amplitude for the two different magnetic reso-

nance modes the frequency of which is plotted in Fig. 5. To explain these dependencies, one would have to evoke the microscopic probability of light scattering on particular spin-wave modes in a particular geometry. Indeed, the process of stimulated Raman scattering as sketched in Fig. 2, involves these probabilities. Generally speaking, they define the interaction of light with magnetic materials, i.e. the coefficient α_{ijk} in the Eqs. (6-8).

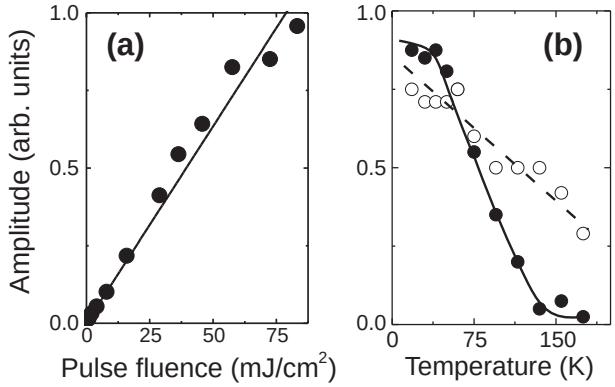


Fig. 6: *(a) Amplitude of the excited precession as a function of the incident light intensity; (b) The dependence of the excited spin precession amplitude on the sample temperature for the two different spin wave modes of Fig. 5. [36]*

4 Coherent control of spins

As shown above, a subpicosecond circularly polarized laser pulses affect the spins in magnetically ordered materials as equally short pulses of effective magnetic field with strengths up to 1 T. In this section we will show that using such opto-magnetic excitation one may coherently control the spin precession in magnetically ordered materials. As a consequence, one may amplify or stop the collective precession of spins on a femtosecond time scale.

4.1 Double pump excitation

Ultrafast coherent control of spin precession can be achieved by using multiple laser pulses in rapid succession.

In Fig. 7 the results of such coherent control experiments are shown for a magnetic garnet. Initially, for $t < 0$ the spins are aligned along the total effective magnetic field $\mathbf{H}^{eff}$ that is of the order of 0.05 T. A pump pulse of helicity σ^+ arriving at $t = 0$ acts as a strong pulse of magnetic field $H^F \gg H^{eff}$ and thus it triggers precession of the magnetization, as explained in the previous sections. A second pump pulse of helicity σ^- arriving after an odd number of half precessional periods rotates the magnetization further away from $\mathbf{H}_{\text{eff}}$, causing the subsequent precession to have almost twice the amplitude (upper graph). If, however, this second pump pulse arrives after an integer number of full periods, the magnetization is rotated back into its original equilibrium orientation along $\mathbf{H}_{\text{eff}}$ and no further precession takes place (lower graph). Fig. 8 gives a pictorial illustration of these two situations. Similarly, circularly polarized light can control the precession of antiferromagnetic spins in the THz domain (see Fig. 9).

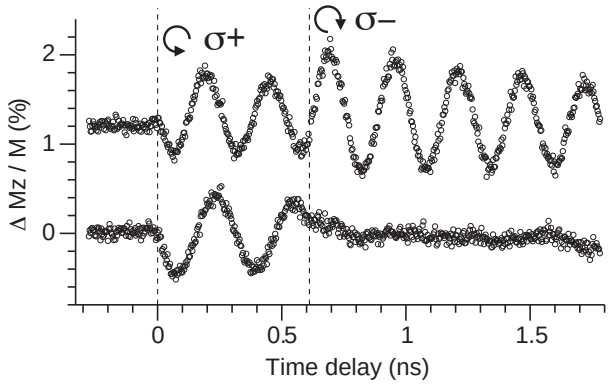


Fig. 7: *Double pump experiment in magnetic garnet with circularly polarized laser pulses of opposite helicity and* $15\ \mu\mathrm{J}$ *pulse power. The upper panel shows the pump-induced change of the sample transmittivity due to the photoexcitation of impurities. The lower panel shows how amplification and complete stopping of the magnetization precession can be achieved depending on the phase of the precession when the second laser pulse arrives. The time delay between the two pump pulses is fixed at approximately 0.6 ns, and the precession frequency is controlled by varying the external field. [45]*

These experiments clearly demonstrate that femtosecond optical pulses can be used to directly and coherently control spin motions. Depending on the phase of the precession when the second pulse arrives, energy is either transferred from the laser pulse to the magnetic system (amplification of the precession) or from the magnetic excitation to the optical pulse (stopping of the precession). In view of the low intrinsic damping in the orthoferrites and garnets, and therefore the long lifetime of their magnetic excitations, it is remarkable how ultrashort laser pulses can completely and instantaneously stop the long period coherent precession of spins. This process of transferring the energy back into the optical pulse can also be viewed as coherent laser cooling of magnons.

The complex spin oscillations in orthoferrites triggered by a train of laser pulses has recently been studied theoretically using nonlinear Landau-Lifshitz-Gilbert equations. It was demonstrated that such a periodical excitation of spins results in various patterns of spin oscillations, which depend on intensity and periodicity of the laser pulses [51].

It should be pointed out that the present double pump experiments, which demonstrate control of the magnetization in ferrimagnetic garnets and antiferromagnetic orthoferrites, are considerably different from those previously reported in diamagnetic and paramagnetic materials. During the past two decades a great number of publications has been devoted to the photoexcitation of a nonequilibrium spin polarization in direct band gap semiconductors through the phenomena of optical orientation [52, 53, 54]. In these materials, absorption of circularly polarized photons may lead to a nonequilibrium population of spin polarized electrons and holes in the conduction band and valence band, respectively. In paramagnetic semiconductors these spin polarized carriers can cause partial alignment of the moments of magnetic ions due to a sp-d exchange interaction, and thereby also affect their precession in a magnetic field [55]. Using this phenomenon of optical orientation, Akimoto *et al.* [56] have demonstrated control of the precession of Mn^{2+} moments in $\mathrm{CdTe/Cd}_{1-x}\mathrm{Mn}_x\mathrm{Te}$ quantum wells. Note that this approach, in contrast to our experiments, is based on the absorption of photons. A nonabsorptive mechanism for manipulation of spins in $\mathrm{Zn}_{1-x}\mathrm{Cd}_x\mathrm{Se}$ quantum well structures was reported by Gupta

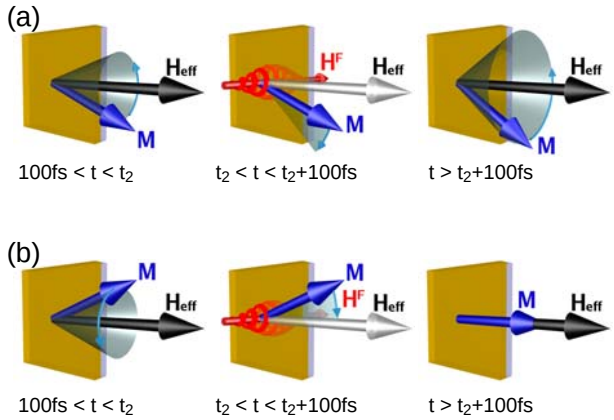


Fig. 8: *Illustration of the double pump experiment for circularly polarized pump pulses of opposite helicity arriving at an (a) odd number of half precessional periods and (b) an integer number of full precessional periods. The magnetization is either rotated further away from the effective field direction causing subsequent precession to take place with almost twice the original amplitude, or the magnetization is rotated back into the effective field direction and no further precession takes place.*

et al. [57], who used below band gap optical pulses to control the spin precession of photoexcited electrons in the conduction band via the optical Stark effect. However, these experiments were performed on paramagnetic materials, where coupling between the spins of magnetic ions is small and the spins oscillate independently. Therefore, in a double pump experiment with paramagnets, the first and second laser pulses can simply excite different spins so that the integrated signal will show either amplification or quenching of the oscillations. However, the amplification and quenching of the oscillations in such an experiment would only mean that the spins excited by the first and second pump pulses oscillate in-phase or out-of-phase, respectively. In magnetically ordered materials, discussed in this review, spins are strongly coupled by the exchange interaction and spin excitations are delocalized. Therefore, in contrast to paramagnets, laser control of spins in magnetically ordered materials indeed means control of the collective motion of spins. Additionally, control of the spin precession in paramagnetic semiconductors requires very low temperatures, typically below 10 K, and strong magnetic fields of several Teslas. In contrast, the optical control of magnetization reported here can be done at room temperature and in applied static magnetic fields well below 0.1 T.

4.2 Pulse shaping

As was shown above, the optical control of magnetism can be achieved via a two-photon process, similar to stimulated Raman scattering. Generally, multi-photon processes can be reached by many routes through a continuum of virtual levels. Therefore, the probability of a multi-photon process can be controlled by the spectral phase distribution of a femtosecond laser pulse, so that several paths leading to the final state interfere in a constructive or destructive way. This means that the efficiency of the opto-magnetic interaction must be a function of the shape of the laser pulse. By tuning the pulse shape we should thus be able to optimize opto-magnetic interactions and achieve highly efficient laser-control of magnetism. Therefore, pulse-shaping appears to be another effective tool to achieve an effective control over the spin motion at ultrafast time-scales.

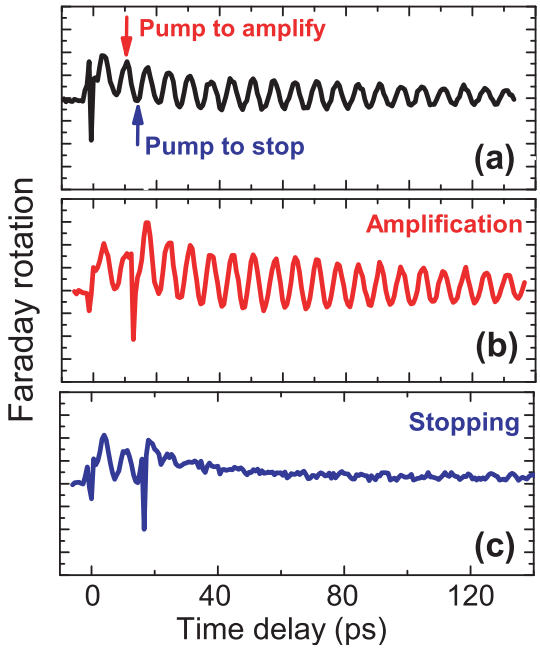


Fig. 9: *Coherent control of spins in $DyFeO_3$ with two circularly polarized laser pulses. a) precession triggered by the first laser pulse; b) amplification of spin precession by the second laser pulse that comes after an even number of full periods; c) stopping of the spin oscillations by the second pump that comes after an odd number of half periods.*

Pulse-shaping for control of quantum interference effects has attracted intense research interest in molecular physics, bio-physics and chemistry as a mean to influence the outcome of a quantum-mechanical interaction [58, 59, 60, 61, 62]. This technique, which is called coherent quantum control, is so effective that specially tailored laser pulses may steer chemical reactions [58] and control the energy flow in biological objects [59]. It has been recently demonstrated theoretically [63] as well as experimentally [64] that adaptive shaping of the amplitude, phase and polarization of femtosecond laser pulses can also be an effective tool for nanoscale localization of ultrafast optical excitations. Here we show that the technique of coherent quantum control and pulse shaping can also be used to control the energy flow from photons to spins.

To shape ultrashort laser pulses and thus to control the energy flow from photons to spins we have built a programmable pulse shaper similar to the one described earlier [65]. More particularly, the shaper was composed of a pair of diffraction gratings with 1200 lines mm^{-1}, and a pair of cylindrical mirrors that played the role of achromatic lenses with a focal length of F=300 mm (see Fig. 10). The first mirror and grating spatially map the complex spectrum of the input pulse at the Fourier plane, where a spatial modulator is inserted. The second mirror and grating reassemble the spectral components to form a modified time-shaped pulse. A one dimensional programmable liquid-crystal spatial light-modulator array, composed of 320 computer-controlled discrete phase elements was placed at the Fourier plane of the shaper, and was used for spectral phase manipulation of the pulses. The shaper was incorporated in the pump-probe setup described above and the shaped pulses were used to excite spin oscillations.

To demonstrate coherent control of collective spin oscillations, or magnons, we have chosen $HoFeO_3$. An excitation of the crystal by an ultrashort Fourier limited circularly polarized laser pulse propagating along [001] results in an effective excitation of collective spin oscillations. In particular, at T=62 K the frequency of the resonance of Fe-spins in this compound is about

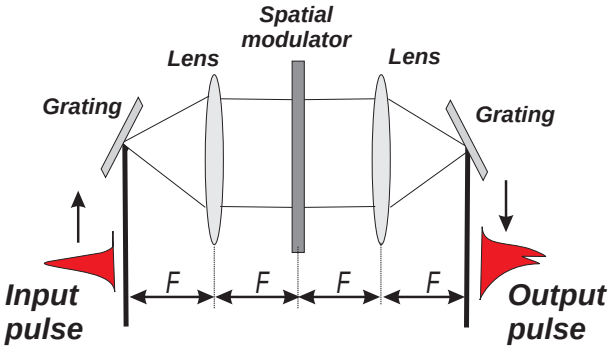


Fig. 10: *Principle scheme of the shaper for ultrashort laser pulses. F is the focal distance of the lenses. In the real schemes the achromatic lenses are often substituted by cylindrical mirrors [58, 59, 60, 61, 62, 65]*

.

f_s=280 GHz. In order to demonstrate the efficiency of the pulse-shaping we have performed a series of experiments with the shaped pulses. We have used a periodic modulation of the spectral phase of the ultrashort laser pulses for different frequencies of the modulation f_m (see Fig. 11a). From Fig. 11 one can see that the efficiency of the laser excitation of magnons is indeed a function of the pulse shape. A transform-limited pulse (f_m=0) gives the most effective excitation. Most spectral phase perturbations will attenuate the efficiency of the laser excitation of the spin oscillations, however periodic spectral phases with periodicity f_s=f_m will not. Such periodic phase structures split an initial transform-limited pulse into a pulse sequence with a repetition rate of f_m. As it has been noticed earlier [60], such a pulse sequence can be used to narrow down the response of the impulsive Raman excitation, i.e., to cancel all but the desired spin-flip transition. It should also be noted that the effectiveness of pulse sequences exciting vibrational levels was demonstrated by Weiner et al. [66]. In this subsection we have demonstrated that the same principle also holds for the laser excitation of magnons.

5 All optical magnetic recording

As was demonstrated above, using femtosecond circularly polarized laser pulses one may trigger spin oscillations and have a control over the spin motion at ultrafast time scales. However, magnetization reversal induced by a subpicosecond stimulus, i.e. a true 180° switching of the magnetization into a stable and oppositely magnetized state, has remained an important fundamental and technological challenge. Further adding to this challenge is the fact that for direct transfer of angular momentum from the light to the magnetic system, the number of photons available in optical experiments is far from enough [11]. Despite the predicted speed limit [2] and shortage of photons, we here demonstrate that a single 40 femtosecond circularly polarized laser pulse can cause well controlled permanent magnetization reversal in materials typically used for data storage. The optical excitation causes an ultrafast heating of the magnetic system which makes it highly susceptible to the magnetic field simultaneously generated by the circularly polarized light pulse. The combination of theses two effects leads to magnetization reversal in a reproducible way. No external magnetic field is required for this opto-magnetic switching, and the stable final state of the magnetization is unambiguously determined by the helicity of the laser pulse.

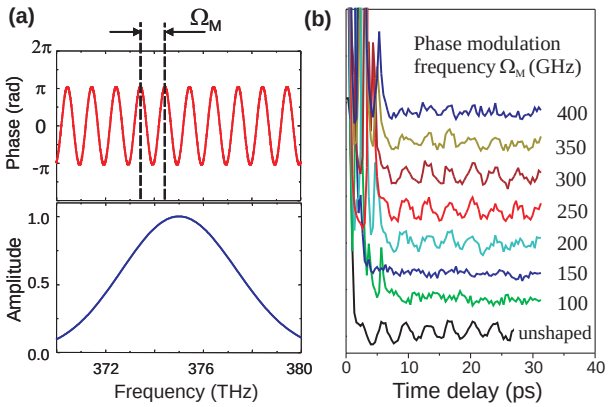


Fig. 11: *a) A periodic modulation of the spectral phase with the frequency Ω_m. The bottom panel shows the spectral profile of the pulse intensity. The phases of these spectral components get a modulation sketched at the top panel. b) Excitation of magnons with the shaped laser pulses. A transform-limited pulse (Ω_m=0) gives the most effective excitation of the spin oscillations at the frequency Ω_s=0.28 THz. Most spectral phase perturbations attenuate the efficiency of the laser excitation of the spin oscillations, however the periodic change of spectral phases with the periodicity Ω_s=Ω_m does not.*

The switching experiments were done on GdFeCo, an amorphous ferrimagnetic alloy that is widely used in magneto-optical recording [67] and known for its strong magneto-optical effects [68]. Thin films of this alloy usually exhibit strong perpendicular magnetic anisotropy, a square hysteresis loop and large magnetic domains which are easily observable in a polarizing microscope. Our samples had a typical composition of $Gd_{22}Fe_{74.6}Co_{3.4}$, a saturation magnetization of about 4πM = 1000 G at room temperature, and a Curie point T_C = 500 K. They were grown by magnetron sputtering in a multilayer structure: glass/AlTi(10 nm)/SiN(5 nm)/GdFeCo(20 nm)/SiN(60 nm), where the AlTi layer served as a heat sink and SiN was used as a buffer and capping layer. The experiments were performed by placing a sample under a polarizing microscope, where domains with magnetization “up” and “down” could be observed as white and black regions, respectively. To excite the material we used regeneratively amplified pulses from a Ti:sapphire laser at a wavelength of λ=800 nm and a repetition rate of 1 kHz. Each pulse had a Gaussian intensity profile, with a width at half-maximum of 40 fs. The laser pulses were incident normal to the sample surface, so that an effective optically generated magnetic field would be directed along the magnetization. The beam was focused down to a 100 μm spot and the laser-induced magnetic changes were studied by recording magneto-optical images of the domain patterns before, during and after the laser excitation. This is a well known technique for studying the magnetic material response to ultrashort field pulses [69, 70, 2]. The experiments were performed at room temperature in air.

The effect of polarized laser pulses on the magnetization is most readily demonstrated by slowly sweeping a laser beam across the surface of the sample. Figure 12 shows how three different sweeps with the laser beam right-handed circularly polarized (σ^+), linearly polarized (L), and left-handed (σ^-) circularly polarized affect the initial domain pattern in dramatically different ways. The region exposed to linearly polarized light is turned into small domains randomly oriented up or down. In striking contrast, the magnetization of the regions exposed to circularly polarized light is completely switched into an “up”- or “down”-state (white or black) determined only by the helicity of the optical excitation.

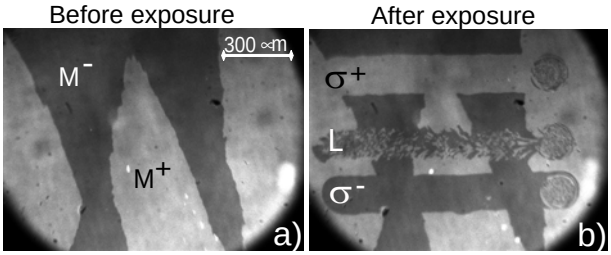


Fig. 12: *The effect of ultrashort polarized laser pulses on magnetic domains in $Gd_{22}Fe_{74.6}Co_{3.4}$. (a) Magneto-optical image of the initial magnetic state of the sample before laser exposure. White and black areas correspond to "up" (M^+) and "down" (M^-) magnetic domains, respectively. (b) Domain pattern obtained by sweeping at low speed (30 m/s) linear (L), right-handed (σ^+) and left-handed (σ^-) circularly polarized beams across the surface of the sample, with a laser fluence of about 11.4 mJ/cm^2. The central area of the remaining spots at the end of each scan line consists of small magnetic domains, where the ratio of "up" to "down" magnetic domains is close to 1 [50].*

In order to unambiguously determine whether excitation by a single 40 fs laser pulse is sufficient to reverse the magnetization, the laser beam was swept at high speed across the sample, so that each pulse landed at a different spot. The results of this experiment are shown in Figure 13 for the two helicities of the laser excitation. One can see that each of the σ^+ pulses reverses the magnetization in the black domain, but does not affect the magnetization of the white domain. The opposite situation is observed when the sample is exposed to σ^- pulses. Thus, during the presence of a single 40 fs laser pulse, information about the photons' angular momentum is transferred to the magnetic medium, and subsequently switching occurs when the optically generated field is in the opposite direction of the initial magnetization. These experiments unambiguously demonstrate that all-optical magnetization reversal can be achieved by single 40 femtosecond circularly polarized laser pulses without the aid of an external magnetic field. The actual speed of the magnetization reversal is still an intriguing question and subject to current experimental investigations. However, it can be estimated quite accurately by considering the coherence time of optical excitations in metals. During the 40 fs interaction of the optical pulse with the sample, information about the angular momentum of the photons is transferred to the electrons via electric-dipole transitions. However, the electrons loose this information very rapidly, typically within 200 fs, due to decoherence processes such as electron-electron and electron-phonon collisions [71, 72]. Moreover, the optical spectrum of $Gd_{22}Fe_{74.6}Co_{3.4}$ does not show any narrow lines, thus excluding the possibility of any long-lived electronic states acting as a reservoir of angular momentum. It follows that the actual magnetization reversal must take place on a sub-picosecond time scale, as this is the only way for the helicity information to survive. The complete recovery of the magnetization in the new state will be a much slower process determined mainly by the speed of heat diffusion in the sample. As a simple illustration of opto-magnetic recording it is shown in Figure 13(b) how optically written bits can be overlapped and made much smaller than the beam waist by modulating the polarization between σ^+ and σ^- while the laser beam is swept across the sample. High density recording may also be achieved by employing especially designed near-field antenna structures [73] such as those currently being developed for HAMR. With the recent development of compact ultrafast laser systems [74] and the successful incorporation of lasers in magnetic storage devices [73], the present demonstration of ultrafast and all-optical magnetization reversal might spur

the realization of a new generation of all-optical magnetic recording devices.

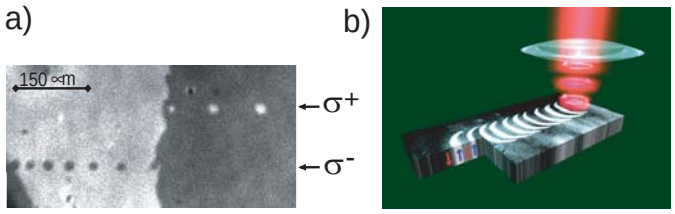


Fig. 13: *All-optical magnetic recording by femtosecond laser pulses. (a) The effect of single 40-fs circular polarized laser pulses on the magnetic domains in* $Gd_{22}Fe_{74.6}Co_{3.4}$*. The domain pattern was obtained by sweeping at high-speed (50 mm/s) circularly polarized beams across the surface so that every single laser pulse landed at a different spot. The laser fluence was about 2.9 mJ/cm*^{2}*. The small size variation of the written domains is caused by the pulse-to-pulse fluctuation of the laser intensity. (b) Demonstration of compact all-optical recording of magnetic bits. This was achieved by scanning a circularly polarized laser beam across the sample and simultaneously modulating the polarization of the beam between left- and right circular [50].*

6 Conclusions

We have summarized the recent progress in the study of ultrafast laser manipulation of spins in magnetic materials. It was shown that in addition to the well-known magneto-optical Faraday effect, where the polarization of light is affected by magnetization, the same magneto-optical susceptibility determines the inverse, opto-magnetic phenomenon: circularly polarized light affects the magnetization via the inverse Faraday effect. Due to this opto-magnetic phenomenon an intense 100 fs circularly polarized laser pulse acts on the spins similar to an equivalently short effective magnetic field pulse up to 1 T. Using this opto-magnetic effect one may selectively excite different modes of magnetic resonance, realize quantum control of magnons and trigger magnetic phase transitions on a subpicosecond time-scale. Finally, it has been demonstrated that with a single 40 fs circularly polarized laser pulse the magnetization can be reversed in a controllable way without the aid of any applied magnetic field. The direction of this switching is determined only by the helicity of the light pulse. We are strongly convinced that all these findings open new insights into the understanding of ultrafast magnetic excitation and, regarding recent progress in the development of compact ultrafast lasers [74], may provide new prospects for applications of ultrafast opto-magnetic phenomena in magnetic storage and information processing technology.

Acknowledgement

This chapter has only been possible through dedicated work of our co-workers C. D. Stanciu, A. M. Kalashnikova, F. Hansteen, A. Tsukamoto, A. Itoh, A. M. Balbashov, B. A. Ivanov, P. A. Usachev, V. N. Gridnev, and R. V. Pisarev. This work was supported by the European RTN network

DYNAMICS, INTAS (project 1000008-8112), Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) and Stichting voor Fundamenteel Onderzoek der Materie (FOM) and the Dutch NanoNed programme.

References

[1] J. Stöhr and H. C. Siegmann, Magnetism: From Fundamentals to Nanoscale Dynamics (Springer Verlag, Berlin, 2006).

[2] I. Tudosa, C. Stamm, A. B. Kashuba, F. King, H. C. Siegmann, J. Stöhr, G. Ju, B. Lu, and D. Weller, Nature **428**, 831 (2004).

[3] B. Hillebrands and K. Ounadjela (eds.), Spin dynamics in confined magnetic structures I, Topics in applied physics (Springer, Berlin, 2002).

[4] L. D. Landau and E. M. Lifshitz, Electrodynamics of Continuous Media, Theoretical physics, Vol. 8 (Pergamon Press, Oxford, 1984).

[5] E. Beaurepaire, J. C. Merle, A. Daunois, and J. Y. Bigot, Phys. Rev. Lett. **76**, 4250 (1996).

[6] J. Y. Bigot, L. Guidoni, E. Beaurepaire, and P. N. Saeta, Phys. Rev. Lett. **93**, 077401 (2004).

[7] J. Hohlfeld, E. Matthias, R. Knorren, and K. H. Bennemann, Phys. Rev. Lett. **78**, 4861 (1997).

[8] J. Güdde, U. Conrad, V. Jähnke, J. Hohlfeld, and E. Matthias, Phys. Rev. B **59**, R6608 (1999).

[9] A. Scholl, L. Baumgarten, R. Jacquemin, and W. Eberhardt, Phys. Rev. Lett. **79**, 5146 (1997).

[10] G. Ju, A. V. Nurmikko, R. F. C. Farrow, R. F. Marks, M. J. Carey, and B. A. Gurney, Phys. Rev. Lett. **82**, 3705 (1999).

[11] B. Koopmans, M. van Kampen, J. T. Kohlhepp, and W. J. M. de Jonge, Phys. Rev. Lett. **85**, 844 (2000).

[12] H. S. Rhie, H. A. Dürr, and W. Eberhardt, Phys. Rev. Lett. **90**, 247201 (2003).

[13] T. Ogasawara, K. Ohgushi, Y. Tomioka, K. S. Takahashi, H. Okamoto, M. Kawasaki, and Y. Tokura, Phys. Rev. Lett. **94**, 087202 (2005).

[14] J. Y. Bigot, C. R. Acad. Sci. Ser. IV (Paris) **2**, 1483 (2001).

[15] R. J. Hicken, Phil. Trans. R. Soc. Lond. A **361**, 2827 (2003).

[16] A. V. Kimel, A. Kirilyuk, A. Tsvetkov, R. V. Pisarev, and Th. Rasing, Nature **429**, 850 (2004).

[17] M. Vomir, L. H. F. Andrade, L. Guidoni, E. Beaurepaire, and J. Y. Bigot, Phys. Rev. Lett. **94**, 237601 (2005).

[18] G. Ju, J. Hohlfeld, B. Bergman, R. J. M. van de Veerdonk, O. N. Mryasov, J. Y. Kim, X. Wu, D. Weller, and B. Koopmans, Phys. Rev. Lett. **93**, 197403 (2004).

[19] J. U. Thiele, M. Buess, and C. H. Back, Appl. Phys. Lett. **85**, 2857 (2004).

[20] J. Hohlfeld, Th. Gerrits, M. Bilderbeek, Th. Rasing, H. Awano, and N. Ohta, Phys. Rev. B **65**, 012413 (2001).

[21] L. Landau and E. Lifshitz, Phys. Z. Union. **8**, 153 (1935).

[22] P. S. Pershan, Phys. Rev. **130**, 919 (1963).

[23] L. P. Pitaevskii, Sov. Phys. JETP **12**, 1008 (1961).

[24] J. P. van der Ziel, P. S. Pershan, and L. D. Malmstrom, Phys. Rev. Lett. **15**, 190 (1965).

[25] P. S. Pershan, J. P. van der Ziel, and L. D. Malmstrom, Phys. Rev. **143**, 574 (1966).

[26] A. F. Kabychenkov, Zh. Eksp. Teor. Fiz. **100**, 1219 (1991), [JETP **73**, 672 (1991)].

[27] A. K. Zvezdin and V. A. Kotov, Modern Magnetooptics and Magnetooptical Materials (IOP, Bristol, 1997).

[28] M. Faraday, Phil. Trans. R. Soc. Lond. **136**(1), 104–123 (1846).

[29] F. J. Kahn, P. S. Pershan, and J. P. Remeika, Phys. Rev. B **186**, 891 (1969).

[30] Y. R. Shen and N. Bloembergen, Phys. Rev. **143**(2), 372 (1966).

[31] J. Deschamps, M. Fitaire, and M. Lagoutte, Phys. Rev. Lett. **25**, 001330 (1970).

[32] Y. Horovitz, S. Eliezer, A. Ludmirsky, Z. Henis, E. Moshe, R. Shpitalnik, and B. Arad, Phys. Rev. Lett. **78**, 001707 (1997).

[33] G. Ju, A. Vertikov, A. V. Nurmikko, C. Canady, G. Xiao, R. F. C. Farrow, and A. Cebollada, Phys. Rev. B **57**, R700 (1998).

[34] P. J. Bennet, V. Albanis, Y. P. Svirko, and N. I. Zheludev, Opt. Lett. **24**, 1373 (1999).

[35] R. Wilks, R. J. Kicken, M. Ali, B. J. Hickey, J. D. R. Buchanan, A. T. G. Pym, and B. K. Tanner, J. Appl. Phys. **95**, 7441 (2004).

[36] A. V. Kimel, A. Kirilyuk, P. A. Usachev, R. V. Pisarev, A. M. Balbashov, and Th. Rasing, Nature **435**, 655 (2005).

[37] J. Wang, G. A. Khodaparast, J. Kono, A. Oiwa, and H. Munekata, J. Modern Optics **51**, 2771–2780 (2004).

[38] H. P. J. Wijn, Numerical Data and Functional Relationships, Landolt-Bornstein New Series Group III, Vol. 27f (Springer, Berlin, 1994).

[39] I. E. Dzyaloshinskii, Sov. Phys. JETP **5**, 1259 (1957).

[40] T. Moriya, Phys. Rev. **120**, 91 (1960).

[41] C. A. Perroni and A. Liebsch, Phys. Rev. B **74**, 134430 (2006).

[42] A theoretical analysis shows that the stregth of the effective magnetic generated via the inverse Faraday effect could be underestimated – see S. R. Woodford, A. Bringer, and S. Blügel, J. Appl. Phys. **101**, 053912 (2007).

[43] N. Koshizuka and K. Hayashi, J. Phys. Soc. Jpn. **57**, 4418 (1988).

[44] F. Hansteen, A. V. Kimel, A. Kirilyuk, and Th. Rasing, Phys. Rev. Lett. **95**, 047402 (2005).

[45] F. Hansteen, A. V. Kimel, A. Kirilyuk, and Th. Rasing, Phys. Rev. B **73**, 014421 (2006).

[46] A. V. Kimel, C. D. Stanciu, P. A. Usachev, R. V. Pisarev, V. N. Gridnev, A. Kirilyuk, and T. Rasing, Phys. Rev. B **74**, 060403R (2006).

[47] A. M. Kalashnikova, A. V. Kimel, R. V. Pisarev, V. N. Gridnev, A. Kirilyuk, and Th. Rasing, Phys. Rev. Lett. *in press* (2007).

[48] A. V. Kimel, B. A. Ivanov, R. V. Pisarev, P. A. Usachev, A. Kirilyuk, and Th. Rasing, *submitted*.

[49] C. D. Stanciu, F. Hansteen, A. V. Kimel, A. Tsukamoto, A. Itoh, A. Kirilyuk, and T. Rasing, Phys. Rev. Lett. **98**, 207401 (2007).

[50] C. D. Stanciu, F. Hansteen, A. V. Kimel, A. Tsukamoto, A. Itoh, A. Kirilyuk, and T. Rasing, Phys. Rev. Lett. **94**, 237601 (2007).

[51] C. A. Perroni and A. Liebsch, J. Phys.: Condens. Matter **18**, 7063 (2006).

[52] F. Meier and B. P. Zakharchenya (eds.), Optical orientation, Modern Problems in Condensed Matter Sciences, Vol. 8 (North-Holland, Amsterdam, 1984).

[53] D. D. Awschalom, J. Warnock, and S. von Molnár, Phys. Rev. Lett. **58**, 812 (1987).

[54] I. Žutić, J. Fabian, and S. Das Sarma, Rev. Mod. Phys. **76**, 323 (2004).

[55] J. K. Furdyna and J. Kossut (eds.), Diluted Magnetic Semiconductors, Semiconductors and Semimetals, Vol. 25 (Academic Press, New York, 1988).

[56] R. Akimoto, K. Ando, F. Sasaki, S. Kobayashi, and T. Tani, J. Appl. Phys. **84**, 6318 (1998).

[57] J. A. Gupta, R. Knobel, N. Samarth, and D. D. Awschalom, Science **292**, 2458 (2001).

[58] E. D. Potter, J. L. Herek, S. Pedersen, Q. Liu, and A. H. Zewail, Nature **355**, 66 (1992).

[59] J. L. Herek, W. Wohlleben, R. J. Cogdell, D. Zeidler, and M. Motzkus, Nature **417**, 533 (2002).

[60] D. Meshulach and Y. Silberberg, Phys. Rev. A **60**, 1287 (1999).

[61] N. Dudovich and Y. Oron, Nature **418**, 512 (2002).

[62] R. S. Judson and H. Rabitz, Phys. Rev. Lett. **68**, 1500 (1992).

[63] M. I. Stockmann, S. V. Faleev, and D. J. Bergman, Phys. Rev. Lett. **88**, 067402 (2002).

[64] M. Aeschlimann, M. Bauer, D. Bayer, T. Brixner, F. J. G. de Abajo, W. Pfeiffer, M. Rohmer, C. Spindler, and F. Steeb, Nature **446**, 301 (2007).

[65] A. Prkelt, M. Wollenhaupt, A. Assion, C. Horn, C. Sarpe-Tudoran, M. Winter, and T. Baumert, Rev. Sci. Inst. **74**, 4950 (2003).

[66] A. M. Weiner, D. E. Leaird, G. P. Wiederrecht, and K. A. Nelson, Science **247**, 1317 (1990).

[67] H. Awano, S. Ohnuki, H. Shirai, N. Ohta, A. Yamaguchi, S. Sumi, and K. Torazawa, Appl. Phys. Lett. **69**, 4257 (1996).

[68] P. Hansen, C. Clausen, G. Much, M. Rosenkranz, and K. Witter, J. Appl. Phys. **66**, 756 (1989).

[69] C. H. Back, D. Weller, J. Heidmann, D. Mauri, D. Guarisco, E. L. Garwin, and H. C. Siegmann, Phys. Rev. Lett. **81**, 3251 (1998).

[70] C. H. Back, R. Allenspach, W. Weber, S. S. P. Parkin, D. Weller, E. L. Garwin, and H. C. Siegmann, Science **285**, 864 (1999).

[71] H. Petek and S. Ogawa, Progress in Surface Science **56**, 239 (1997).

[72] V. V. Kruglyak, R. J. Hicken, M. Ali, A. T. G. P. B. J. Hickey, and B. K. Tanner, Phys. Rev. B **71**, 233104 (2005).

[73] W. A. Challener, T. W. McDaniel, C. D. Mihalcea, K. R. Mountfield, K. Pelhos, and I. K. Sendur, Jpn. J. Appl. Phys. **42**, 981 (2003).

[74] U. Keller, Nature **424**, 831 (2003).

E 1 Fundamental Concepts of Quantum Information Processing

Thomas Schäpers
Institut für Bio- und Nanosysteme (IBN-1)
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 Quantum Bits and Quantum Registers **3**
2.1 Quantum Bit 3
2.2 Quantum Register 5
2.3 Entangled States 6

3 Quantum-Gates **7**
3.1 Single-Qubit Gate: Basics Concept 7
3.2 The Hadarmard Transformation as a Single-Qubit Gate 9
3.3 Single-Qubit Manipulation by an Oscillating Magnetic Field 10
3.4 Two-Qubit Gates 12

4 Decoherence **15**

5 The Five DiVincenzo Criteria **16**

6 Quantum Algorithms: A Brief Overview **16**

1 Introduction

Quantum information technology is a fast developing field where the unique properties of quantum mechanics are used for computation or communication. In contrast to a bit processed in a conventional computer, which can only be in the state 0 or 1, here the bit is constituted by a quantum mechanical two-level system, the so-called qubit. A well-known example of a two-level system is the two states of an electron with spin 1/2 in a static magnetic field. What make the quantum bit so powerful is that superpositions of the two quantum states are allowed. If this property, which has no counterpart in conventional computers, is cleverly used, some computational problems can be solved much faster. The reason for this is that by employing a superposition of quantum states a quantum algorithm effectively calculates very many solutions in parallel. A prominent example of this kind of algorithm is Shor's prime number factorization, where a large number is decomposed into its prime factors [1]. As illustrated in Fig. 1, Shor's algorithm can factorize a large number exponentially fast compared to a classical computers.
A quantum computer is not commercially available, yet, as the technology to build such a device is still in an early stage and up to now systems with only a few quantum bits have been realized [2]. Of course, this is not sufficient for serious applications. One of the problems scientists face is the loss of coherence of the quantum mechanical state, due to the coupling of the quantum bit to the environment. If the coherence is lost during the computational sequence, the final result is meaningless.
Since quantum mechanics is one of the very fundamental principles in physics, many different experimental realizations in various fields of physics have been proposed. Quantum computing with a few quantum bits has been demonstrated by using nuclear magnetic resonance (NMR) [3]. Another possibility is to manipulate the electronic states of ions stored in a trap [4]. Quantum bits can also be realized in solid-state systems, in particular by using quantum dots [5, 6], single impurities in a semiconductor [7] or Josephson junctions [8, 9, 10].
In this lecture we will only discuss the basic principles of quantum computation. Quantum bits and quantum registers will be introduced first. Later on it will be discussed how the state of a quantum bit can be manipulated by means of a quantum gate. We will distinguish between two different gates, the single-qubit gate, where the state of single qubit is modified, and a

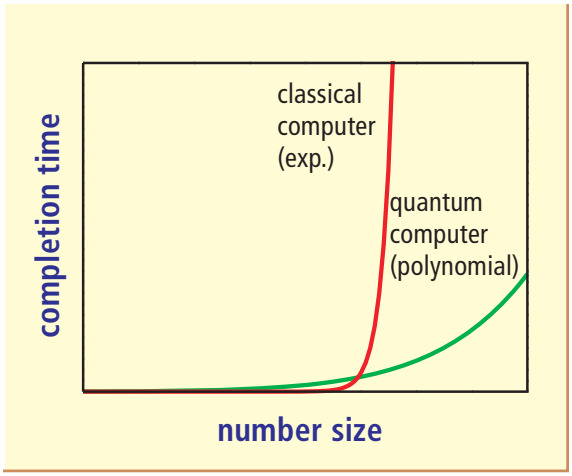


Fig. 1: *Completion time as a function of number size for a classical and a quantum factorization algorithm, respectively.*

two-qubit gate, where the first qubit controls the state of the second one. Furthermore, we will discuss what else is required to build a properly operating quantum computer. In the last part the basic ideas of the two most popular quantum algorithms, the Shor and Grover algorithm will be explained.

2 Quantum Bits and Quantum Registers

In this section the basic elements of a quantum computer, the quantum bit and the quantum register are discussed. We will start by defining the smallest unit, the quantum bit and we will see what *strange* things can happen if more quantum bits are combined.

2.1 Quantum Bit

The basic element of a quantum computer is the quantum bit, or qubit, which can be viewed as an extension of the classical notion of a bit. As illustrated in Fig. 2 (a), in a classical computer only 0 and 1 are possible values corresponding to zero and a fixed finite voltage level [11]. Similarly, a qubit consists of a quantum mechanical two-level system. However, the crucial difference is that in addition to the two eigenstates $|0\rangle$ and $|1\rangle$ a qubit can be configured in any

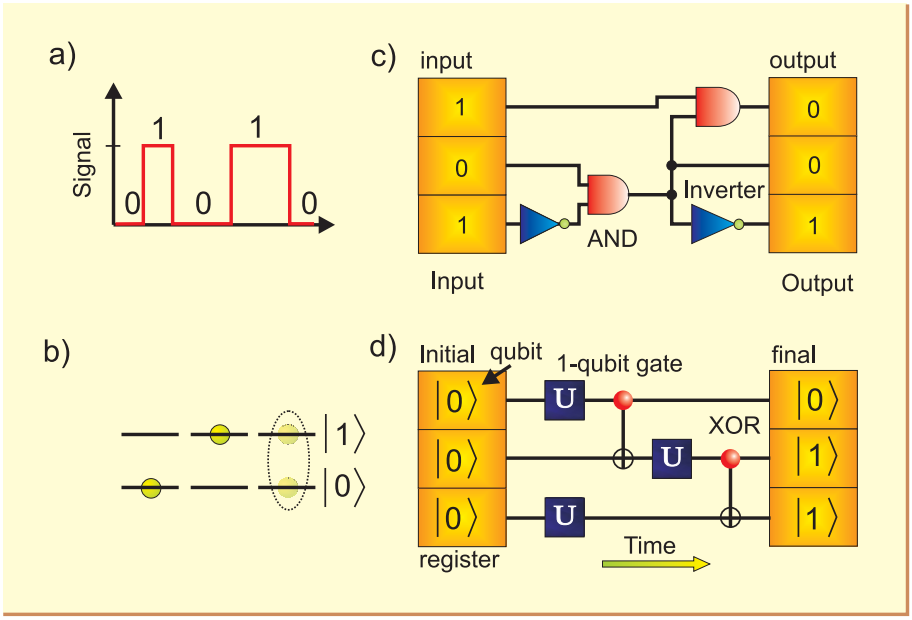


Fig. 2: *(a) Possible values in a digital computer. Here, the signal is switched between the values 0 and 1. (b) In a quantum mechanical two-level system the partical can be in the ground state $|0\rangle$ or in first excited state $|1\rangle$ but also in a superposition of both. (c) Schematic illustration of a digital computer. The calculation is performed by means of gates, e.g. inverters or NAND-gates (inverted AND). (d) Schematics of a 3-qubit quantum computer. The qubits are manipulated by single-qubit gates (U) and two-qubit gates (XOR).*

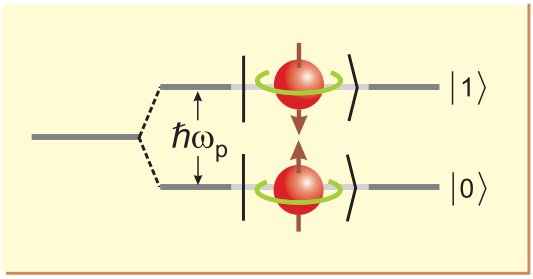


Fig. 3: *Zeeman-splitted spin up* $|0\rangle$ *and spin down* $|1\rangle$ *states representing a qubit. The energy splitting is given by* $\hbar\omega_p$*, with* ω_p *the cyclotron frequency.*

superposition of these eigenstates[1]

$$|Q\rangle = c_0|0\rangle + c_1|1\rangle \ . \tag{1}$$

In contrast to the classical bit, a qubit can be prepared in an infinitive number of appropriate quantum states by choosing different coefficients c_0 and c_1. The qubit state is normalized so that $|c_0|^2 + |c_1|^2 = 1$. The two eigenstates usually correspond to the ground and excited states of a two-level system, as depicted in Fig.2 (b). A possible realization is the Zeeman splitting of the spin-up and spin-down state of a spin-1/2 particle ($s = \pm 1/2$), i.e. an electron, in a magnetic field $\mathbf{B}$. The corresponding two basis vectors can be represented by

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ . \tag{2}$$

The qubit can then be expressed by

$$|Q\rangle = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} \ . \tag{3}$$

In case of an electron in a magnetic field $\mathbf{B}$, the two-level system is described by the Hamiltonian

$$H = \frac{1}{2} g^* \mu_B \mathbf{B} \sigma \ , \tag{4}$$

with g the effective gyromagnetic factor.[2] The Bohr magneton is defined as $\mu_B = e\hbar/2m_0$, with e the elementary charge and m_0 the free electron mass. The components of the $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli spin matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \ \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \ , \ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{5}$$

As illustrated in Fig. 3, if one assumes a constant field along the z-direction: $\mathbf{B} = (0, 0, B_z)$, the Zeeman effect leads to a splitting of the spin-up and spin-down states. The level splitting is given by $\Delta E = g\mu_B B_z$, which can be rewritten as

$$\Delta E = \hbar\omega_p \ , \tag{6}$$

[1] Here, we used the Dirac notation which allows us to describe quantum mechanical states without defining a particular set of basis vectors.

[2] In order to be consistent with the usual notation in quantum computation, we assume a negative g^*-factor so that the spin-up state is the ground state, i.e. the state with the lower energy. A negative g^*-factor is found for example in GaAs.

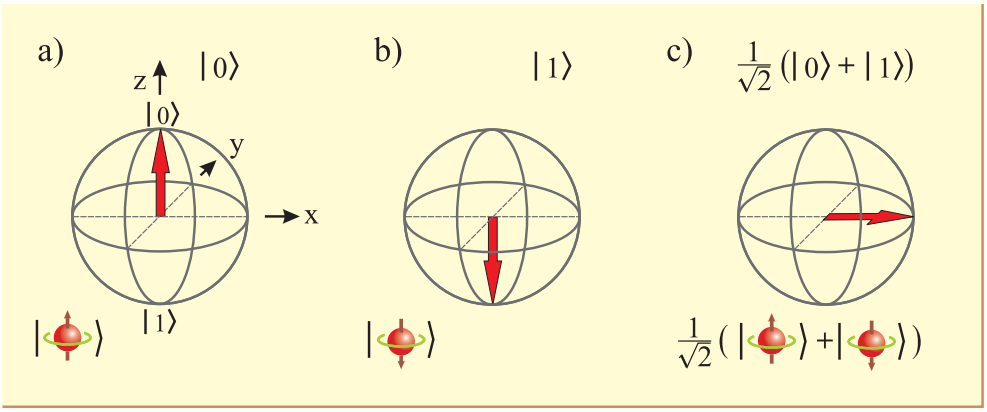


Fig. 4: *Bloch spheres illustrating different states of a spin 1/2-particle. (a) and (b) spin-up and spin-down state along the z-direction. (c) Superposition state $1/\sqrt{2}(|0\rangle + |1\rangle)$ pointing towards the x-direction. This state is an eigenstate of σ_x.*

with

$$\omega_p = |g|\mu_B B_z/\hbar \tag{7}$$

the precession frequency. We will see below, ω_p plays an important role for the manipulation of a qubit.

The state of a spin-1/2 can best be visualized by using a Bloch sphere. Here, all possible spin orientations are mapped on a sphere. Using this scheme, the spin-up and spin-down state are illustrated in Fig. 4 a) and b), where the arrows representing the spin states point towards the $\pm z$-direction, respectively. For the superposition state $1/\sqrt{2}(|0\rangle + |1\rangle)$, which is an eigenstate of σ_x, the arrow points towards the x-direction, as shown in Fig. 4 c).

Although an infinitive number of quantum states can be realized in a qubit state, a single qubit cannot be used to transmit more than one bit of information. This is due to the fact that in quantum mechanical systems only eigenstates of the quantum system are finally detected by the measurement procedure. For a superposition state, as given by Eq. (1), the final measurement will return 0 with probability $|c_0|^2$ and 1 with probability $|c_1|^2$. Thus, after the measurement the state of the qubit will either be the eigenstate $|0\rangle$ or $|1\rangle$ but definitely not in a superposition of both. Returning back to the spin 1/2-particle: If the spin orientation would be measured, e.g. by a Stern-Gerlach apparatus, the spin-up state would be measured with probability $|c_0|^2$ and the spin-down state with probability $|c_1|^2$.

Of course as a final result of a computation a superposition state is useless, because no unambiguous outcome is produced. Thus, at the end of a computation an eigenstate is required as a result. However, one should keep in mind that during calculation superposition states are very important, since they are responsible for the higher performance of quantum algorithms.

2.2 Quantum Register

A collection of qubits, which is usually called quantum register, can be used to encode more complex information. In the schematics of a quantum computer depicted in Fig. 2 (d) the quantum register contains three quantum bits. As a very simple example a two-qubit register can be composed from the following set of eigenvectors:

$$|0\rangle_2 \otimes |0\rangle_1 = |00\rangle\ , \quad |0\rangle_2 \otimes |1\rangle_1 = |01\rangle\ , \quad |1\rangle_2 \otimes |0\rangle_1 = |10\rangle\ , \quad |1\rangle_2 \otimes |1\rangle_1 = |11\rangle\ . \tag{8}$$

Here, $\otimes$ denotes the direct product of the vectors. The indices specify the first and second qubit. By using this set of basis vectors a general superposition of these state is given by

$$|\psi\rangle = c_0|00\rangle + c_1|01\rangle + c_2|10\rangle + c_3|11\rangle \ . \tag{9}$$

Instead of writing the basis vectors as given in Eq. (8), we can also use the notation

$$|00\rangle = |0\rangle \ , \quad |01\rangle = |1\rangle \ , \quad |10\rangle = |2\rangle \ , \quad |11\rangle = |3\rangle \ , \tag{10}$$

which is sometimes easier to handle if many qubits are involved. Generally, for an n-bit quantum register we can write:

$$|\psi\rangle = \sum_{i=0}^{2^n-1} c_i|i\rangle \ . \tag{11}$$

For an n–qubit state we obtain 2^n states $|i\rangle$. Please note that by preparing a superposition state all these states are simultaneously present.

2.3 Entangled States

A closer look on two or more qubit registers reveal that states can be prepared where the states of constituting qubits in a register are not independent. This situation is called entanglement. A typical two-qubit entangled state is given by:

$$\begin{aligned} |\psi\rangle &= \frac{1}{\sqrt{2}}\left(|1\rangle_2 \otimes |0\rangle_1 - |0\rangle_2 \otimes |1\rangle_1\right) \\ &= \frac{1}{\sqrt{2}}\left(|10\rangle - |01\rangle\right) \end{aligned} \tag{12}$$

Here, $|0\rangle_1$ and $|1\rangle_1$ are states constituting the first qubit, while $|0\rangle_2$ and $|1\rangle_2$ are the corresponding states of the second qubit. An entangled state cannot be factorized into an expression like:

$$|\psi\rangle = |m\rangle_2 \otimes |n\rangle_1 \ . \tag{13}$$

A typical representative of a state which can be factorized is given by

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|1\rangle_2 \otimes |0\rangle_1 - |1\rangle_2 \otimes |1\rangle_1\right) \ , \tag{14}$$

since we can simplify it to

$$|\psi\rangle = \frac{1}{\sqrt{2}}|1\rangle_2 \otimes \left(|0\rangle_1 - |1\rangle_1\right) \ , \tag{15}$$

corresponding to Eq. (13).
What is the consequence of entanglement? Let us assume that we are able to measure the state of the first qubit of the state given by Eq. (13). If the result would be $|0\rangle_1$, it directly implies that the second qubit is in the state $|1\rangle_2$. This behavior is of special importance, if the two qubits are represented by two polarized photons emitted in two opposite directions. A measurement of the polarization state of the first photon directly implies the polarization state of the second photon, although the detectors of the photons can be very far apart. Fundamental questions, how the second photon *knows* about the state of the first one are addressed in the famous article by Einstein-Podolsky-Rosen (EPR) [12]. Entanglement is also of outmost importance for quantum key distribution. Here, a secure key is generated by transmitting entangled photon between two parties. In the second step this key is used to encode information. The securities of this scheme is ensured by the fact that a possible eavesdropper can be discovered.

3 Quantum-Gates

The state of a qubit or of a set of qubits is controlled by quantum-gates. This is in analogy to boolean gates, e.g. to an inverter or AND gate in classical digital computers, where the corresponding truth tables are given in Table 1.

a	*b*
0	1
1	0

a	*b*	*c*
0	0	0
0	1	0
1	0	0
1	1	1

Table 1: *Truth table of a NOT and AND gate. For the NOT gate a is the input parameter and b the output parameter. For the AND gate a and b are the input parameters and c the output.*

Interestingly, by using an inverter (NOT gate) and an AND gate any boolean expression can be constructed. If we combine an AND and a NOT gate to form a NAND gate, even a single gate is sufficient to synthesize any boolean operation. In quantum computing the gates are represented by unitary transformations. Similar to conventional computers, one can define gates which are applied to a single qubit or gates, which connect two or more qubits. Furthermore, it was found that a set of two quantum gates is sufficient to implement all quantum computer operations.

3.1 Single-Qubit Gate: Basics Concept

By performing quantum computational operations, the qubit states are changed in course of time. The eigenstates $|0\rangle$ and $|1\rangle$ themselve are fixed, only the probabilities for their occupation, expressed by the coefficients $c_0(t)$ and $c_1(t)$ are explicitly time-dependent:

$$|Q(t)\rangle = c_0(t)|0\rangle + c_1(t)|1\rangle. \tag{16}$$

In quantum gate operations the coefficients $c_1(t)$ and $c_2(t)$ are changed in a defined manner, e.g. by applying an external magnetic field for a well-defined period of time.
In order to express the following single-qubit operations more compactly, the qubit state needs to be described a little bit more formally

$$|Q(t)\rangle = \sum_{i=0,1} c_i(t)|i\rangle \ . \tag{17}$$

The basis vectors defining the qubit are orthonormal which implies that they fulfill

$$\langle i|j\rangle = \delta_{ij} \ , \quad i,j = 0,1 \ . \tag{18}$$

Here, $\langle i|$ are the vectors of the dual space while $\langle \ldots | \ldots \rangle$ denotes the inner product. For example, by using the definition of the inner product the coefficients $c_i(t)$ can be extracted by:

$$\langle i|Q(t)\rangle = \langle i| \sum_{j=0,1} c_j(t)|j\rangle = c_i(t) \ . \tag{19}$$

In the framework of quantum mechanics the gate operation on a qubit corresponds to a transformation of the state in the course of time, which can be described by means of the Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|Q(t)\rangle = H(t)|Q(t)\rangle \ , \tag{20}$$

where $H(t)$ is the Hamilton operator. Inserting the definition of $|Q(t)\rangle$, as given by Eq. (17), into the Schrödinger equation and multiplying by $\langle i|$, we obtain a set of two equations for the coefficients $c_0(t)$ and $c_1(t)$:

$$\begin{aligned} \langle i|\mathrm{i}\hbar\frac{\partial}{\partial t}|Q(t)\rangle &= \langle i|H(t)|Q(t)\rangle \\ \langle i|\mathrm{i}\hbar\sum_{j=1,2}\frac{\partial}{\partial t}c_j(t)|j\rangle &= \langle i|H(t)\sum_{j=1,2}c_j(t)|j\rangle \qquad (21) \\ \mathrm{i}\hbar\frac{\partial c_i(t)}{\partial t} &= \sum_{j=1,2}H_{ij}(t)c_j(t) \ . \qquad (22) \end{aligned}$$

Here, $H_{ij}(t) = \langle i|H(t)|j\rangle$ are the elements of the Hamilton matrix. In order to obtain the values of the coefficients $c_0(t_1)$ and $c_1(t_1)$ at time $t = t_1$, the two equations have to be integrated in the interval $[t_0, t_1]$, where t_0 is the initial time.
A single-qubit gate operation is realized by *switching on* a Hamiltonian for a period of time $\Delta t = t_1 - t_0$ and thus modifying the coefficients $c_1(t)$ and $c_2(t)$. As we will see below, for a spin-1/2 particle this can be realized by applying an oscillating magnetic field for a certain period of time. In this case the gate operation relies on the coupling of the magnetic moment of the particle to the field.
In order to clarify how the qubit is changed after a short period of time Δt we write the Schrödinger equation as follows

$$\frac{c_i(t_0+\Delta t) - c_i(t_0)}{\Delta t} = -\frac{i}{\hbar}\sum_j H_{ij}(t_0)c_j(t_0) \ . \tag{23}$$

By regrouping this equation we get for the final state of the qubit at time t_1:

$$c_i(t_1) = c_i(t_0+\Delta t) = \sum_j \left[\delta_{ij} - \frac{i}{\hbar}H_{ij}(t_0)\Delta t\right] c_j(t_0) \tag{24}$$

The matrix on the right side is a unitary matrix and can be summarized by defining:

$$U_{ij}(t_0+\Delta t, t_0) = \delta_{ij} - \frac{i}{\hbar}H_{ij}(t_0)\Delta t \tag{25}$$

Thus, the change of the qubit state between t_0 and t_1 calculated by integrating the Schrödinger equation what can be expressed by a unitary transformation U. For a qubit represented by a two component vector defined by Eq. (3), U can be described by a unitary 2×2 transformation matrix (Fig. 5):[3]

$$\begin{pmatrix} c_0(t_1) \\ c_1(t_1) \end{pmatrix} = U \begin{pmatrix} c_0(t_0) \\ c_1(t_0) \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} c_0(t_0) \\ c_1(t_0) \end{pmatrix} \ . \tag{26}$$

[3] An unitary matrix has the property: $U^\dagger U = 1$, with $U^\dagger$ the adjoint matrix defined by $U^\dagger = (U^*)^T$.

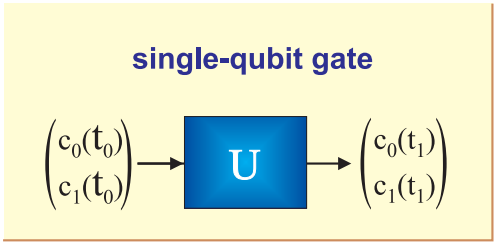


Fig. 5: *Quantum circuit diagram of an single-qubit gate. The gate operation is described by a unitary transformation.*

This unitary matrix U represents the single-qubit gate operation. In the following two sections we will specify U in more detail by discussing two possible realizations of a single-qubit gate, i.e. the Hadamard transformation and the manipulation of a spin-1/2 quantum bit by an oscillating field.

3.2 The Hadarmard Transformation as a Single-Qubit Gate

A well-known representative of a single-qubit gate is the Hadamard transformation defined by the following unitary matrix

$$U_H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} . \tag{27}$$

The Hadamard transformation can be used to conveniently generate a superposition state out of an eigenstate. This operation is often the first processing step in quantum algorithms, e.g. the Deutsch-Josza algorithm [15]. Let us assume our qubit is in the ground state $|0\rangle$. Applying a Hadamard transformation results in

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} . \tag{28}$$

Thus, as illustrated in Fig. 6, we end up with the superposition state $1/\sqrt{2}(|1\rangle + |0\rangle)$. One can

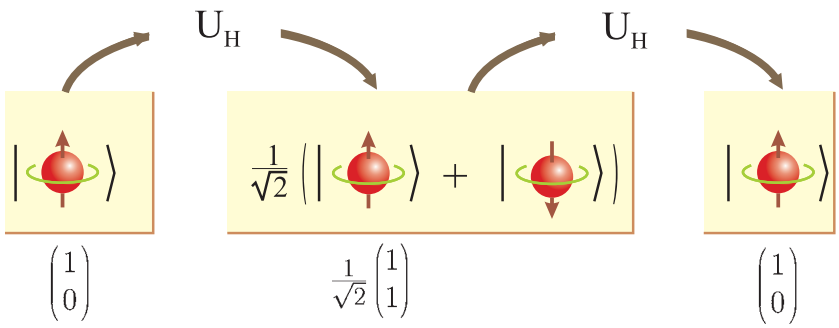


Fig. 6: *Hadamard transformation (single-qubit operation) on a spin-1/2 particle in the ground state. By applying the Hadamard transformation once again the initial spin orientation is recovered.*

easily verify that by applying a Hadamard transformation once again the superposition state returns back to the ground state $|0\rangle$ (see Fig. 6).
Of course Hadamard single-qubit gates can also be applied to qubit registers. For simplicity, let us consider a two-qubit register which is initially in the ground state $|00\rangle$. If we apply the transformations $U_{H,1}$ and $U_{H,2}$ to the first and to the second qubit, respectively, we end up with

$$U_{H,2}\, U_{H,1}\, |00\rangle = \frac{1}{2}\left(|00\rangle + |10\rangle + |01\rangle + |11\rangle\right) \ . \tag{29}$$

Thus by performing two simple one-qubit gate operation the ground state of a quantum register $|00\rangle$ is transformed to a superposition of all four basis states. As mentioned above, to be able to generate a superposition state is of outmost importance for quantum computation, since the strength of its concept lies in the parallel processing of all states.

3.3 Single-Qubit Manipulation by an Oscillating Magnetic Field

When we discussed the Hadamard transformation as a representative of a single-qubit gate we did not care how such a gate can actually be realized. Now we will explain how a single-qubit gate can be put into action by *switching on* a Hamiltonian. Let us return to our prototype qubit, i.e. the single electron with spin 1/2. By coupling the magnetic moment of the electron to an external magnetic field $\mathbf{B}$ the spin state can be modified. The process is illustrated in Fig. 7. The Hamiltonian describing this coupling was already given by Eq. (4). In contrast to the previous case with a constant magnetic field B_z, now a magnetic field rotating in the xy-plane with frequency ω_p and amplitude B_0 is applied for a certain period of time. Thus, in total the magnetic field is given by:

$$\mathbf{B} = \begin{pmatrix} B_0 \cos \omega_p t \\ B_0 \sin \omega_p t \\ B_z \end{pmatrix} \tag{30}$$

By inserting $\mathbf{B}$ into Eq. (4) one arrives at the following explicit form of the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \begin{pmatrix} c_0(t) \\ c_1(t) \end{pmatrix} = \frac{1}{2} g \mu_B \begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix} \begin{pmatrix} c_0(t) \\ c_1(t) \end{pmatrix} \ . \tag{31}$$

Assuming the ground state $|Q(t_0)\rangle = (1, 0)$ to be the initial state at $t = t_0$ one finds the following solution for the quantum bit after the oscillating field is applied for a period of time

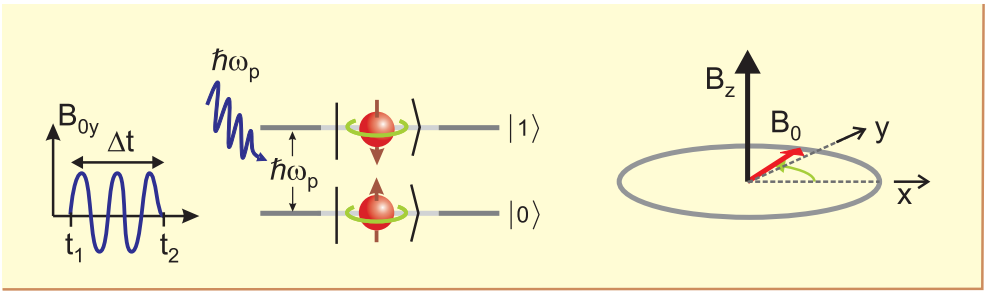


Fig. 7: *Coherent transitions of Zeeman-splitted electrons by applying a magnetic field rotating in the xy-plane with frequency ω_p. The level splitting is achieved by applying a constant field B_z along the z-direction.*

$\Delta t = t_1 - t_0$:

$$|Q(t_1)\rangle = \begin{pmatrix} \cos(\Omega\Delta t/2)\exp(+i\omega_p\Delta t/2) \\ i\sin(\Omega\Delta t/2)\exp(-i\omega_p\Delta t/2) \end{pmatrix} , \tag{32}$$

with the characteristic frequency Ω defined by

$$\Omega = |g|\mu_B B_0/\hbar \, . \tag{33}$$

In order to get some insight what happens to the qubit after the period of time Δt, one can calculate the expectation values of the spin in the different directions:

$$\langle s_x \rangle = \frac{\hbar}{2}\langle Q(t_2)|\sigma_x|Q(t_2)\rangle = \frac{\hbar}{2}\sin(\Omega\Delta t)\sin(\omega_p\Delta t) \tag{34}$$

$$\langle s_y \rangle = \frac{\hbar}{2}\langle Q(t_2)|\sigma_y|Q(t_2)\rangle = \frac{\hbar}{2}\sin(\Omega\Delta t)\cos(\omega_p\Delta t) \tag{35}$$

$$\langle s_z \rangle = \frac{\hbar}{2}\langle Q(t_2)|\sigma_z|Q(t_2)\rangle = \frac{\hbar}{2}\cos(\Omega\Delta t) \tag{36}$$

One finds by looking at $\langle s_z \rangle$ that in the course of time the spin oscillates with frequency Ω between the ground state $|0\rangle$ (spin-up) and the excited state $|1\rangle$ (spin-down), e.g. after applying the oscillation field for a period $\Delta t = \pi/\Omega$ the electron is transferred from the ground state $|0\rangle$ into the excited state $|1\rangle$. As can be inferred from Eq. (33), the time required to flip the spin is determined by the amplitude B_0 of the oscillation field.

The effect of the oscillation field on the qubit can best be visualized by making use of the Bloch sphere. For the case where the oscillation field is applied for a period $\Delta t = \pi/\Omega$ the spin is flipped from the top to the bottom pole, as illustrated in Fig. 8 a). However, the spin is not simply flipped but rather precesses about the z-axis while turning downwards. The spin precession is due to the second factors in Eqs. (34+35). The precession frequency is ω_p.

If the oscillation field is only applied for a period $\Delta t = \pi/2\Omega$ one can infer from Eq. (32) that a superposition state is obtained. The spin is now located in the xy-plane. If the amplitude B_0 of the oscillating field is properly adjusted one can for example transfer the $|0\rangle$ state to the $1/\sqrt{2}(|0\rangle + |1\rangle$ state, as illustrated in Fig. 8 b). Note, that after performing this transition and switching off the oscillating field the spin precesses about the z-axis with the precession

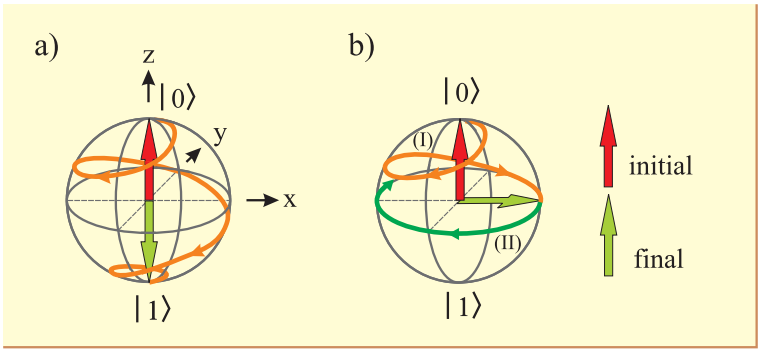


Fig. 8: *(a) Flip of the spin-up state $|0\rangle$ (red arrow) to the spin-down state $|1\rangle$ (green arrow). The trajectory of the spin is illustrated by the orange line. (b) Transition from the $|0\rangle$ state (red arrow) to the superposition state $1/\sqrt{2}(|0\rangle + |1\rangle)$ (green arrow). If the field is switched off after the transition, the spin precesses clockwise in the xy-plane as illustrated by the green line.*

frequency ω_p, thus the qubit does not keep its state in the course of time. Technically this is not a big problem. For example, one can compensate this effect by working in a reference frame rotating at with ω_p [16].
So far, we only considered the ground state as the initial state. One can generalize the special case discussed above by allowing any superposition state for the qubit. For this case the qubit manipulation can be expressed as

$$\begin{pmatrix} c_0(t_1) \\ c_1(t_1) \end{pmatrix} = U_R \begin{pmatrix} c_0(t_0) \\ c_1(t_0) \end{pmatrix} \tag{37}$$

with the corresponding unitary matrix given by

$$U_R = \begin{pmatrix} \cos(\Omega\Delta t/2)\exp(i\omega_p\Delta t/2) & i\sin(\Omega\Delta t/2)\exp(-i\omega_p\Delta t/2) \\ i\sin(\Omega\Delta t/2)\exp(i\omega_p\Delta t/2) & \cos(\Omega\Delta t/2)\exp(-i\omega_p\Delta t/2) \end{pmatrix} . \tag{38}$$

Interestingly, the matrix U_R can be decomposed in a product of two matrices

$$\begin{aligned} U_R &= \begin{pmatrix} \cos(\Omega\Delta t/2) & i\sin(\Omega\Delta t/2) \\ i\sin(\Omega\Delta t/2) & \cos(\Omega\Delta t/2) \end{pmatrix} \begin{pmatrix} \exp(i\omega_p\Delta t/2) & 0 \\ 0 & \exp(-i\omega_p\Delta t/2) \end{pmatrix} \\ &= R_x(\Omega\Delta t)R_z(\omega_p\Delta t) . \end{aligned}$$

We can interpret this outcome as a combination of two rotation of the spin, i.e. a rotation R_x with the angle $\Omega\Delta t$ about the x-axis and a rotation R_z with the angle $\omega_p\Delta t$ about the z-axis. After the period Δt the qubit is transferred to

$$|Q(t_1)\rangle = R_x(\Omega\Delta t)R_z(\omega_p\Delta t)|Q(t_0)\rangle .$$

In summary, spin rotation matrices can be used to generate any desired single-qubit gate. The spin rotation is determined by the time and amplitude of the applied oscillating field.

3.4 Two-Qubit Gates

We now turn to quantum gates, which are applied to two quantum bits. A typical two-qubit gate is the controlled NOT or exclusive OR (XOR) gate. In boolean computers the output is set to 1 if either the first or the second input is 1. If both are 1 the output is 0 again. The corresponding circuit diagram for qubits can be found in Fig. 9. In an XOR gate, a target qubit $|b\rangle$ is flipped if

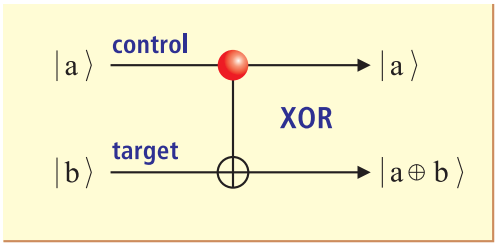


Fig. 9: *Quantum circuit diagram of an XOR gate. The lower bit $|b\rangle$ (target qubit) is flipped if $|a\rangle$ (source qubit) is set.*

a	b	$a \oplus b$		$\|a,b\rangle$	$\|a, a \oplus b\rangle$
0	0	0		$\|00\rangle$	$\|00\rangle$
0	1	1	$\Longleftrightarrow$	$\|01\rangle$	$\|01\rangle$
1	0	1		$\|10\rangle$	$\|11\rangle$
1	1	0		$\|11\rangle$	$\|10\rangle$

Table 2: *Truth table of a XOR (CNOT) gate for a boolean and for a quantum computer.*

a control qubit $|a\rangle$ is in the $|1\rangle$ state. As long as $|a\rangle$ is in the state $|0\rangle$ no action is taken on $|b\rangle$. The corresponding truth table is given in Table 2. An important conceptual difference between the the XOR gate in boolean and quantum computers is that in quantum gates, the number of inputs and outputs are always identical. The quantum bits are both still present after the gate operation.
The XOR gate is an unitary transformation on states spanned by a set of four basis vectors. If we define the following set of basis vectors

$$|00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \tag{39}$$

we can express the XOR gate by the unitary 4×4 matrix

$$U_{XOR} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{40}$$

The XOR gate can be used to produce an entangled state out of two formerly non-entangled states. For this purpose we set the control qubit $|a\rangle$ (source-qubit) to a superposition state $1/\sqrt{2}(|0\rangle - |1\rangle)$ and the target qubit to $|b\rangle = |1\rangle$. The input state is thus given by $1/\sqrt{2}\,(|01\rangle - |11\rangle)$, which is non-entangled, since it can be factorized in $1/\sqrt{2}(|0\rangle - |1\rangle) \otimes |1\rangle$. Applying an XOR gate we obtain

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}. \tag{41}$$

Thus the result is an entangled state given by $1/\sqrt{2}\,(|01\rangle - |10\rangle)$.
The XOR gate can also be used as a measurement gate to obtain the output value of a calculation, since it can reproduce the result of one input state on both output gates [17]. For this purpose we set the target gate $|b\rangle$ to $|0\rangle$. As can be seen from the truth table of the XOR gate, both lines at the output reproduce the state of the source gate $|a\rangle$. The output is either $|00\rangle$ or $|11\rangle$. This process can be used to perform a non-demolition measurement of a qubit by generating a second one in the same state, which is used for the measurement process. Furthermore, this

configuration can be used to produce fanout, i.e. coupling the output to more than one inputs, if a coupling to two gate inputs in the following calculation step is required.
The XOR gate is only one out of many possible two-qubit-gates. Another possible two-qubit gate is the so-called controlled rotation (CROT) [18], which is described by the matrix

$$U_{CROT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{42}$$

Here, a rotation on the second qubit is performed if the control qubit is set to $|1\rangle$. The name controlled rotation originates from the fact that in case the control qubit is set, the σ_z Pauli spin matrix which describes a spin rotation about the z direction is applied to the target qubit.
One can show that any two-qubit gate can be transferred to the XOR gate by performing a sequence of single-qubit operation. In the case of a CROT gate this can be achieved by performing single-qubit rotations on the target qubit $|b\rangle$ about the y-axis:

$$R_{y,|b\rangle}(\theta) = \begin{pmatrix} \cos\theta/2 & \sin\theta/2 & 0 & 0 \\ -\sin\theta/2 & \cos\theta/2 & 0 & 0 \\ 0 & 0 & \cos\theta/2 & \sin\theta/2 \\ 0 & 0 & -\sin\theta/2 & \cos\theta/2 \end{pmatrix} \tag{43}$$

As can be easily verified, performing a $\theta = \pi/2$ rotation before the CROT operation and finally a rotation with $\theta = \pi/2$ the XOR gate is recovered:

$$\begin{aligned} R_{y,|b\rangle}(-\frac{\pi}{2}) U_{CROT} R_{y,|b\rangle}(\frac{\pi}{2}) &= \frac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = U_{XOR} \end{aligned} \tag{44}$$

It was shown that any possible quantum algorithm can be realized in the basis of only a single-qubit gate and a XOR gate. However, for some physical realizations of a quantum computer it is not always straightforward to realize a XOR gate, i.e. in case of a quantum computer based on trapped ions a CROT gate is easier to implement. In order to prove that a certain type of realization of a quantum computer can be used universally, it is sufficient to find a transformation of an implemented two-qubit gate to a XOR gate. As shown above, for the CROT gate this can be achieved by performing single qubit rotations.
If we take our spin-1/2 particles for the realization of a qubit, the exchange interaction between two particles can be used to implement a two-qubit gate. The corresponding Hamiltonian for two coupled electron spins can be expressed as:

$$H = J\sigma_1 \cdot \sigma_2 , \tag{45}$$

where J is the exchange coupling parameter and σ_1, σ_2 are the spins operators of particle 1 and 2, respectively. In the famous proposal of Loss and Divincenzo [5] the single electrons

Quantum system	t_{switch} [s]	τ_{dec} [s]	Ratio
Electrons GaAs	10^{-13}	10^{-10}	10^{3}
Electrons Au	10^{-14}	10^{-8}	10^{6}
Trapped ions: In	10^{-14}	10^{-1}	10^{13}
Optical microcavities	10^{-14}	10^{-5}	10^{9}
Electron spin	10^{-7}	10^{-3}	10^{4}
Electron quantum dot	10^{-6}	10^{-3}	10^{3}
Nuclear spin	10^{-3}	10^{4}	10^{7}

Table 3: *Typical times for quantum mechanical two-level systems, which are possible candidates for a realization of a qubit [19].*

representing the qubit are confined in a split-gate quantum dot structure. The two-qubit gate operation is put into action by lowering the electrostatic barrier between both quantum dots. The onset of electron tunneling between both quantum dots results in the exchange interaction expressed by Eq. (45). Due to the coupling between both spins one qubit is changed as a function of the state of the other one.

4 Decoherence

So far, we discussed a rather ideal quantum computer composed of qubit registers and a set of quantum gates. However, reality is not that perfect. One important aspect, which has not been addressed yet, is the decoherence of quantum systems. Decoherence occurs due to the fact that a quantum system is not completely isolated from its environment. The quantum dynamics of the surrounding setup couple to a certain extent to the states of a quantum computer. Decoherence is a very serious problem, since the computational pathways which are separated at the beginning of the calculation are only recombined at the very end. Thus, if something gets wrong in between, the final result is completely meaningless.

The decoherence is characterized by the decoherence time τ_{dec}. Its inverse is a measure of the coupling of a single qubit with its environment. In order to perform a successful computation with a quantum computer, the decoherence time must be much longer than the expected operating time of the computation. The latter is determined by the number of computational switchings and the time required to perform each of these steps. Typical values of the minimum switching time t_{switch}, the decoherence time τ_{dec} and the resulting maximum number of steps given by the ratio τ_{dec}/t_{switch} are summarized in Table 3 for various two-level systems [19].

Qubit states destroyed by decoherence can be recovered if error correction schemes are applied, as was first found out by Shor [20]. The major ingredient is the introduction of redundancy. The error correction relaxes somewhat the requirements resulting from the numbers given in Table 3. However, the circuit itself becomes more complex, since additional qubits have to be introduced.

5 The Five DiVincenzo Criteria

In the previous sections different parts of a quantum computer were discussed in detail. However, the fact that all components in itself are working does not mean that the device made out of these components can actually be used as a quantum computer. Exactly this issue was addressed by DiVincenzo [21, 22], who set up a list of five criteria, the so-called *DiVincenzo criteria*. The criteria are as follows:

- ⋆ Existence of a well-defined extendible qubit array.
- ⋆ It must be possible to define an initial state $|00\rangle$.
- ⋆ A sufficiently long decoherence time is required ($\tau_{dec}/t_{switch} > 10^4$).
- ⋆ An universal set of quantum gates exist.
- ⋆ A read-out of the qubit state must be possible.

Only if all criteria can in principle be fulfilled for an envisioned realization it makes sense to follow this approach. All criteria given above have been discussed in detail in the previous sections, except for the last one which is concerned with the read-out of the final state. For the spin-1/2 particle a direct read-out of the spin state by detecting its magnetic moment is difficult. For the quantum dot system, as proposed by Loss and DiVincenzo [5], a viable way is to transfer the spin degree of freedom into a charge degree of freedom. The latter can be accessed by electronic means.

6 Quantum Algorithms: A Brief Overview

In this section two different quantum algorithms will be discussed. Since most of the quantum algorithms are quite complex, we will omit a detailed treatment of the computational steps.
The probably most-cited quantum algorithm is the factorization algorithm invented by Peter Shor at IBM [1]. By using this algorithm it is possible to factorize large numbers with only a polynomial increase of computational time with the number of digits (see Fig. 1). In contrast, for digital computers the computational time increases exponentially with number size, so that it is literally impossible to factorize large numbers, i.e. numbers of more than 100 digits. This inability of digital computers is the reason why modern encryption schemes like the Rivest, Shamir, and Adleman (RSA) system [13] are based on factorizing large numbers. The fact that by employing Shor's algorithm large numbers can in principle be factorized is therefore a real threat for contemporary encryption methods.
In Shor's factorization algorithm a method is used where a large number N is factorized by finding a period of a sequence $f(x) = a^x \mod N$, where a is a randomly chosen small number with no factors in common with N. From the period of this series the prime factors of N can be extracted. However, for an ordinary computer it is as difficult to find the period of the series as finding the prime factors directly. The basic trick of Shor's algorithm is to find the period by performing a discrete Fourier transformation on a quantum mechanical superposition state. In practice only small numbers (number: 15) have been factorized so far, owing to the problem to build large-size quantum computer systems [2].
Another very prominent quantum algorithm is the search algorithm of Grover [14]. Due to the effective parallel computation in a quantum computer, Grover's algorithm can search for an

item in a data base by only a single query, whereas multiple queries are required in a classical search algorithm. Once again, the speed of this algorithm is based on generating a superposition state given by Eq. (11) to address all entries of the data base at the same time.

References

[1] P. W. Shor. In S. Goldwasser, editor, *Proc. 35th Annual Symposium on the foundations of Computer Science*, pages 124–134. IEEE Compter Society Press, Los Alamitos, CA, 1994.

[2] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, Nature **414**, 883 (2001).

[3] N. A. Gershenfeld and I. L. Chuang. Science **275**, 350 (1997).

[4] C. Monroe, D. M. Meekhof, B.E. King, W.M. Itano, and D.J.Wineland. Phys. Rev. Lett. **75**, 4714 (1995).

[5] D. Loss and D. P. DiVincenzo. Physical Review A **57**, 120 (1998).

[6] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. K. Vandersypen, Reviews of Modern Physics **79**, 1217 (2007).

[7] B. E. Kane, Nature **393**, 133 (1998).

[8] Y. Nakamura, Yu. A. Pashkin, and J. S. Tsai. Nature **398**, 786 (1999).

[9] J.E. Mooij, T. P. Orlando, L. Levitov, Lin Tian, C. H. van der Wal, , and S. Lloyd, Science **285**, 1036 (1999).

[10] T. P. Orlando, J. E. Mooij, Lin Tian, C. H. van der Wal, L. S. Levitov, S. Lloyd, and J. J. Mazo, Phys. Rev. B **60**, 15398 (1999).

[11] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[12] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[13] R. Rivest, A. Shamir, and L. Adleman, Communications ACM **21** 120 (1978).

[14] L. K. Grover, Phys. Rev. Lett. **79**, 4709 (1997).

[15] D. Deutsch, Proc. R. Soc. Lond. A, **400**, 97 (1985).

[16] Vandersypen, L. Experimental quantum computation with nuclear spins in liquid solutions, Ph.D. Thesis, Stanford University, 2001 (quant-ph/0205193)

[17] D. Deutsch, Proc. R. Soc. Lond. A, **425**, 73 (1989).

[18] A. Steane, quant-ph/9608011, 1996, 1-36.

[19] D. P. DiVincenzo, Phys. Rev. A **51**, 1015 (1995).

[20] P. W. Shor, Phys. Rev. A **52**, 2493 (1995).

[21] D. P. DiVincenzo, Science 270, 255 (1995).

[22] D. P. DiVincenzo, Fortschr. Phys. **48**, 771 (2000).

E 2 Quantum Computing with Semiconductor Quantum Dots

Carola Meyer

Institut für Festkörperforschung (IFF-9)

Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**

2 The ”Loss-DiVincenzo” proposal **2**

3 Read-out of a single electron spin **3**
3.1 Single shot read-out 4
3.2 Singlet-Triplet read-out 7

4 Manipulation of electron spins **8**
4.1 Single spin rotation 9
4.2 The $\sqrt{SWAP}$ operation 10

5 Relaxation mechanisms **14**
5.1 Spin-energy relaxation 15
5.2 Dephasing and decoherence 18

6 Summary and outlook **20**

1 Introduction

Quantum dots can be used to confine single electrons as discussed by M. Wegewijs in the lecture "Spin and Transport in Quantum Dots". The quantum computing concepts based on quantum dots can be subdivided in two main branches: optical concepts and electrical concepts. In most of the optical concepts, the two level system representing the quantum bit (qubit) consists of exciton states. These are manipulated using polarized light. In electrical concepts, the spin states of electrons are used as qubit and manipulation can be done all-electrically.
This contribution will concentrate on spin states of electrons for quantum information focusing on the most important electrical concept known as "Loss-DiVincenzo proposal" [1]. It has been shown experimentally for this proposal that all of the "DiVincenzo criteria" (for a general introduction into Quantum Computing see lecture "Fundamental Concepts of Quantum Information Processing" by T. Schäpers) can be met as we shall see in the following.

2 The "Loss-DiVincenzo" proposal

A few years after the first implementation of the CNOT quantum gate using hyperfine and vibrational states of a $^9Be^+$ ion in an ion trap as qubits [2], a row of proposals for a solid state quantum computer appeared, based on cooper pairs [3], nuclear spins in silicon [4], and last but not least electron spins in GaAs quantum dots [1]. Daniel Loss and David DiVincenzo proposed a quantum computer based upon existing semiconductor technology.

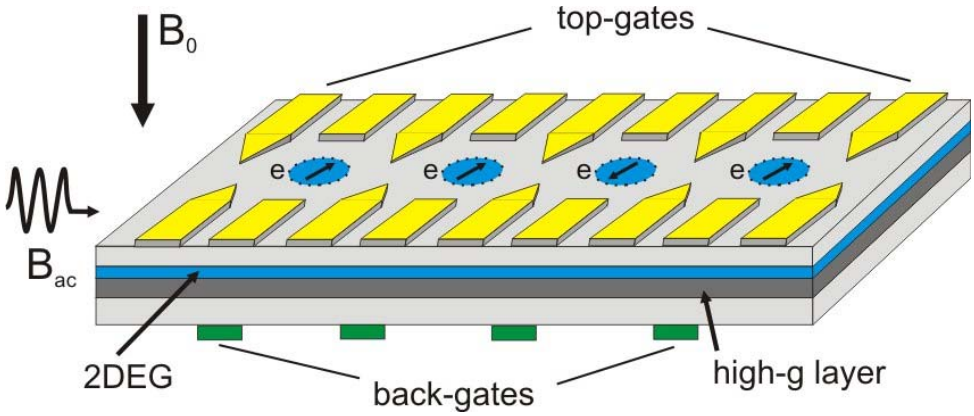


Fig. 1: *Scheme of the Loss-DiVincenzo proposal.The top gates are used to form quantum dots as well as to tune the interaction between them. An AC magnetic field is used to manipulate the electron spins. Back gates can draw the electrons into a layer with different g-factor, thus changing their resonance frequency.*

The scheme of this proposal is depicted in Figure 1. A two dimensional electron gas (2DEG) is formed by a GaAs/GaAlAs heterostructure. Voltages applied to electric top-gates are used to deplete certain regions of the 2DEG in such a way that a quantum dot with only a single electron inside remains. In a magnetic field $\boldsymbol{B_0} = (0, 0, B_z)$ the otherwise degenerate Zeeman states $|\uparrow\rangle$, $|\downarrow\rangle$ split up with energy difference $E_{Zee} = g\mu_B\hbar B_0$, with Landé factor $g = -0.44$ for GaAs and μ_B the Bohr magneton, and form the two level system used as a qubit. Initialization can be achieved by allowing the electron spins to reach their thermodynamic ground state at low temperature T, with $|E_{Zee}| \gg k_BT$ (with Boltzmann constant k_B). However, this is a very slow

process, because the relaxation rate from an exited spin state to the ground state has to be small in order not to loose the information of the qubit. We will see later that also a scheme for fast initialization exists.
The qubit states can be manipulated with an ac magnetic field applied perpendicular to B_0 just as in electron spin resonance (ESR). This ac magnetic field can be generated by passing an ac current through a wire close to the quantum dots. In order to be able to carry out single qubit rotations, the resonance frequency of the manipulated spin needs to differ from the resonance frequency of the spins in the other quantum dots. This can be achieved by a B_0 gradient along the chain of quantum dots or by g factor engineering. For the latter, the electron is pulled into a layer with a high g-factor by applying a voltage on a local back-gate. Thus, the energy splitting between the spin states and therefore the resonance condition is changed.
The Hamiltonian used for gate operations in a system with N qubits is

$$H(t) = \sum_{i}^{N} g_i(t) B_i(t) S_i + \sum_{i<j}^{N} J_{ij}(t) S_i(t) S_j, \tag{1}$$

with qubit sites i, j. The first term describes the single qubit gates as discussed above with $B(t) = B_0(t) + B_{ac}$. The second term describes two qubit gates, with the exchange interaction J_{ij} used for the qubit coupling. Only adjacent qubits need to be coupled, since information can be passed through the qubit chain with the SWAP gate. The coupling between two neighboring qubits, i.e. the potential barrier between two adjacent quantum dots, can be controlled by voltages applied to the top-gates. Therefore, the "Loss-DiVincenzo" proposal is in principle scalable. Since GaAs quantum dots have been extensively studied and the spins can be initialized in their ground state, the first two DiVincenzo criteria are fulfilled. In this lecture we will see that the other criteria, namely the qubit read-out, a universal set of quantum gates and long decoherence times are met as well.

3 Read-out of a single electron spin

In this section we will see how the electron spin state in a quantum dot can be measured. Two read-out schemes exist, one for a single quantum dot with $|\uparrow\rangle$, $|\downarrow\rangle$ as qubit states, and one with the singlet $|S\rangle$ and the triplet $|T_0\rangle$ state of a *two-electron* quantum dot as qubit. Both schemes have in common that the spin state is first converted into a charge state, which is then detected by the current through an adjacent quantum point contact (QPC). In this way, the measurement is decoupled from the qubit system and the back action of the read-out on the qubit state is minimized.
Before we look at the two schemes in more detail, we will briefly discuss the QPC detection. A QPC is a one-dimensional constriction in the 2DEG formed by top-gates (see inset Fig.2). Top-gate voltages or other potentials close by define how many electrons can pass the constriction at the same time, i.e. the number of available transport channels.
The conductance of a QPC shows a step-like behavior depending on the voltage applied to the top-gates as shown in Fig.2. Transport channels are opened one by one, while the applied gate voltage becomes more positive. Without external magnetic field the step hight is $2e^2/h$, since the two spin states of an electron are degenerate. If this degeneracy is lifted by applying an external magnetic field, additional steps appear at multiples of e^2/h [5].
In close proximity to a quantum dot, a QPC can be used as noninvasive voltage probe [6] that detects the number of electrons on the quantum dot. The QPC is operated in the middle between

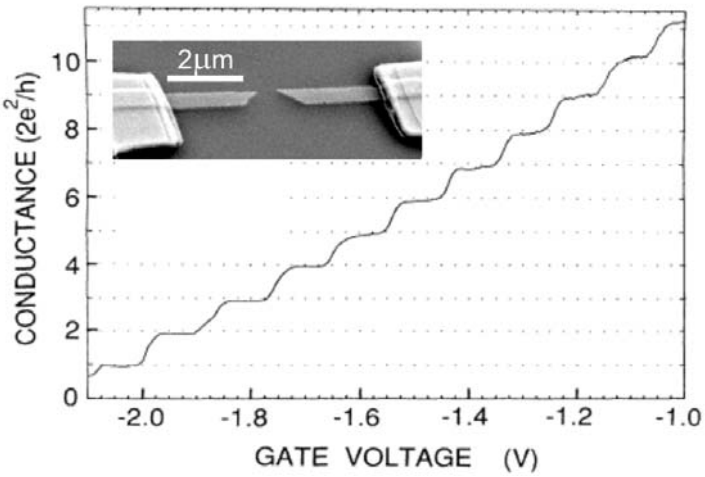


Fig. 2: *Stepwise increase of the QPC conductance at T = 0.6 K with changing top-gate voltage (from reference [5]). Inset: An example for a quantum point contact structure (adapted from http://pages.unibas.ch/phys-meso/Pictures/pictures.html).*

two current plateaus in order to obtain maximum sensitivity towards adding an electron to the quantum dot or removing it. Today this technique has been extended on double quantum dots measuring small signals of photon-assisted tunneling [7] and spin blockade [8].

3.1 Single shot read-out

In order to demonstrate the single shot read-out of a single electron spin, a quantum dot with a QPC next to it was fabricated as shown in Fig.3a. It is important that the gate R is closed completely, so that the current to the drain of the QPC is not influenced by a current through the dot. The QPC is adjusted to its working point with the gate Q. Tunneling events occur between the reservoir and the dot with rate Γ depending on the tunneling barrier influenced by gate L.

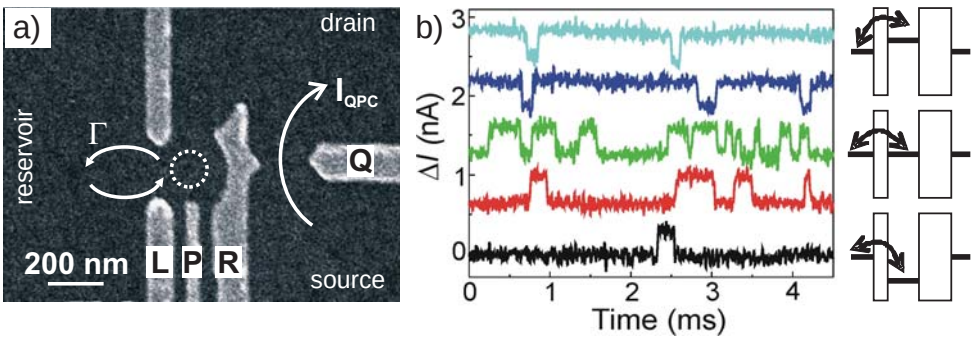


Fig. 3: *(a) Gate structure for a single quantum dot formed by gates R and L with adjacent QPC between Q and R. The potential barrier on the right is very high and tunneling between the dot and the reservoir occurs through the left barrier with rate Γ. (b) Tunneling events of a quantum dot measured trough the current of a QPC for different potentials on P. The dot is empty (high current) most of the time for the top trace while it is occupied (low current) most of the time for the bottom trace. When the electrochemical potential of the dot is aligned with the Fermi level of the reservoir, the electron tunnels back an forth. All images adapted from ref. [9].*

Since the read-out of a spin state is done via charge detection, we should first know how fast the charge state can be measured. This has been shown in ref. [9]. There, the quantum dot is set near to its $N = 0$ to $N = 1$ transition using gate P to tune the dot potential. The electron can then spontaneously tunnel back and forth between the dot and the reservoir, and the QPC current should exhibit a random telegraph signal (RTS) as shown in Fig.3b. The time the electron spends in the dot, i.e. when ΔI_{QPC} is in the low state, strongly depends on the position of the dot potential relative to the Fermi level of the leads. The current through the QPC was $I_{QPC} \approx 30$ nA with a bias voltage of $V_{bias} = 1$ mV, in agreement with the conductance of the QPC at its working point $G_{QPC} = e^2/h \approx (30\text{k}\Omega)^{-1}$. The shortest steps that clearly reached above the noise level were about 8μs long. Tunnel events occuring on a shorter timescale will be lost in the current noise of the QPC. Therefore, the spin-energy relaxation time T_1, i.e. the time after which a spin has flipped from its exited $|\downarrow\rangle$ state back to the ground state $|\uparrow\rangle$, of the spin in the quantum dot has to be much longer than 8μs. Otherwise the information stored in the qubit would be lost before it was even measured.
The single-spin-single-shot read-out was first demonstrated in the group of L. Kouwenhoven at TU Delft [10]. To detect the spin state of an electron, first a magnetic field B_0 has to be applied so that the degeneracy of the Zeeman states is lifted. In order to tune the dot potential quickly, voltage pulses with lengths of a few 100 ns are applied to gate P (Fig.3a). Figure 4 shows the pulse scheme used for the single spin read-out as well as the response of the QPC.

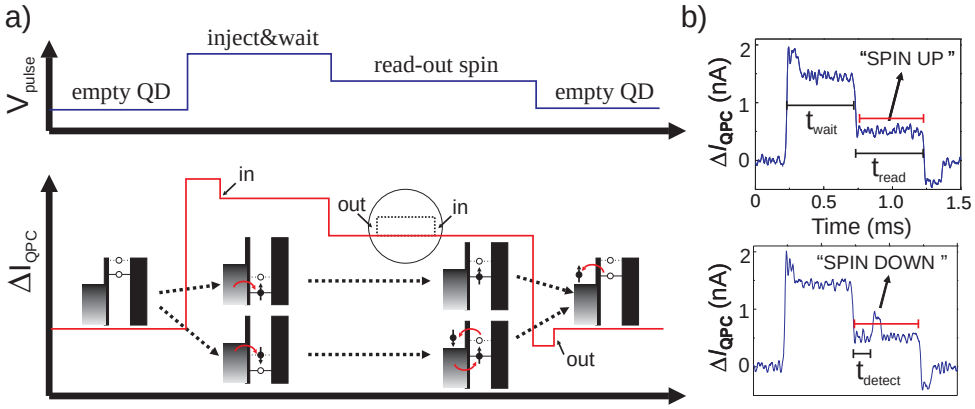


Fig. 4: *(a) Scheme of the single shot read-out. On the top the voltage levels applied as pulses on gate P (Fig.3a) are shown. The difference in the QPC current during the different stages is shown on the bottom along with the tunnel events. The signal during the read-out depends on the spin state (circle). In the case of "spin down", to additional tunnel events take place and the signal follow the dotted line. (b) Single shot measurements of a spin state. The top graph shows the trace of the QPC current for the "spin up" situation, where no tunneling events are measured during the read-out time t_{read}. On the bottom, the "spin down" case is depicted. During t_{read}, the threshold value of the QPC current (red line) is crossed indicating two additional tunneling events. The time t_{detect} is the time it takes for a "spin down" electron to tunnel out of the dot and thus related to the rate $\Gamma_\downarrow$ All figures adapted from [10].*

At the beginning the quantum dot potential is set to a low value, so that any remaining electron is pushed out of the dot. Then, a positive voltage pulse is applied to put both spin states below the Fermi level of the lead. The current of the QPC is changing as well, since it couples capacitively to the gate P as well. As soon as either a spin-up or a spin-down electron from the reservoir tunnels into the dot, the current of the QPC drops due the extra charge in the vicinity. The time

one has to wait for an electron to enter is directly connected to the tunneling rate $\Gamma = \Gamma_\downarrow + \Gamma_\uparrow$, which can be influenced by gate L (Fig.3b).
The spin to charge conversion is done in the third part of the pulse pattern. The potential of the dot is changed such that the spin-up ground state remains below the Fermi level of the lead, while the excited spin-down state lies above it. No tunneling events will happen in the first case (see Fig.4b, top), because the dot is in coulomb blockade. However, in the latter case, first the spin-down electron will tunnel out before the ground state is filled again with a spin-up electron from the lead. Therefore, two tunneling events will occur during the read-out time t_{read} (Fig.4b, bottom). Before a new cycle can be started, the potential of the dot is tuned so that both spin states are above the Fermi level and held there until the spin-up electron now occupying the dot has tunneled out.
In order to measure the relaxation time T_1, the spin-down fraction is recorded for different waiting times t_{wait}. During this time, a spin-down electron can relax to the ground state. The longer this time, the smaller the spin-down fraction will be, following an exponential decay as shown in Fig.5a. Fitting the data to $\alpha + C\exp(-t_{wait}/T_1)$ decay, a relaxation time of $T_1 \approx 0.55$ ms is obtained at $B_0 = 10$ T. This is almost two orders of magnitude longer than the time needed for the fast detection and the response of the QPC is thus quick enough.
Nevertheless, there is a finite probability α that a signal is measured during t_{read} although a spin-up electron was in the dot, for instance due to thermally activated tunneling or electrical noise (”dark counts”). This probability can be extracted directly from the T_1 measurement. It is simply the saturation value of the exponential decay. Unfortunately, a similar evaluation is not possible for the opposite case that occurs with probability β; the QPC current stays below the threshold although a spin-down electron was in the dot. The correlation between these probabilities is shown in the inset of Fig.5a.

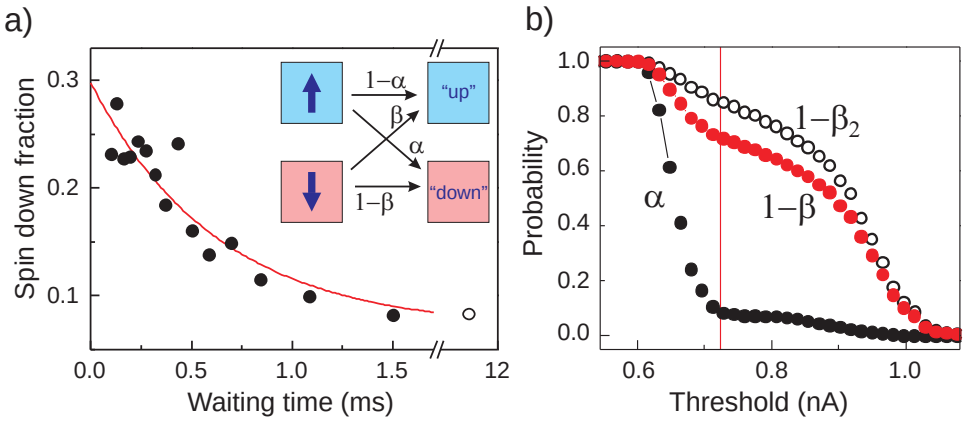


Fig. 5: *(a) T_1 relaxation measured with the single shot read-out. The probabilities for measuring a spin-up as a spin-down and vice versa are depicted in the inset. (b) Their values depend on the threshold set in the measurement (see Fig.4). The vertical red line marks the threshold value with the highest visibility. Adapted from [10].*

Two processes contribute to β which can be analyzed separately. First, a spin-down electron can relax to the spin-up state before the electron tunnels out with probability $\beta_1 = 1/(1 + T_1\Gamma_\downarrow)$. $\Gamma_\downarrow$ can be obtained from a histogram of the detection time t_{detect} (see Fig.4b for definition). In ref. [10] its value was found to be $\Gamma_\downarrow^{-1} \approx 0.11$ ms yielding $\beta_1 \approx 0.17$. Second, if the spin-down electron is replaced within 8μs with a spin-up electron the resulting QPC step may be too small to detect. The probability β_2 of this event depends on the value of the threshold

(red line in Fig.4b). It can be measured reversing the pulse sequence [10]. The empty levels are tuned to the read-out postition (4a). At the beginning of this pulse, a $|\uparrow\rangle$ should tunnel into the dot raising the QPC current above the threshold. The probability β_2 is obtained from the fraction of traces where this step is missed. The result is shown as $1-\beta_2$ in Fig.5 as well as the threshold dependence of α and $1-\beta$, the total spin-down fidelity is given by $1-\beta \approx (1-\beta_1)(1-\beta_2)+(\alpha\beta_1)$.
The so-called visibility is a very important number for quantum computing, since it is a measure for the probability of a correct qubit measurement. For the single spin read-out discussed here, the visibility is

$$V = 1-\alpha-\beta. \tag{2}$$

The red line in Fig.5b marks the threshold value at which this expression has its maximum ($\alpha \approx 0.07$, $\beta_1 \approx 0.17$ and $\beta_2 \approx 0.15$). Therefore, the fidelity for the spin-down and the spin-up state is $(1-\beta) \sim 0.72$ and $(1-\alpha) \sim 0.93$, respectively [10]. The visibility of the single shot measurement, however, is only 65%, i.e. the chance to get a wrong result is 35%. Of course, this would be inacceptable for a computer, but for a proof of concept this is a good result, especially when compared to other implementations. Repeating the same calculation several times can already improve the accuracy. Lowering the electron temperature (smaller α) and a faster QPC measurement (smaller β) will increase the visibility as well.
However, this read-out method suffers from other disadvantages. It is very sensitive to fluctuations of the electrostatic potential, the Zeeman splitting has to be much larger than the thermal energy, and high frequency noise can spoil the read-out due to photon-assisted tunneling, i.e. when the ground state electron absorbs a microwave photon and gains enough energy to tunnel out of the dot into the reservoir.

3.2 Singlet-Triplet read-out

This method circumvents the problems of the single shot read-out described before and is described in ref. [11]. It discriminates between singlet $|S\rangle$ and triplet $|T\rangle$ states of a quantum dot and is therefore used as read-out for a *two-electron* quantum dot. Thus, the quantum dot is tuned near to its $N=1$ to $N=2$ transition. The device geometry is similar to the structure in Fig.3a.
The pulse sequence used for the read-out and relaxation time measurement is shown in Fig.6a. First, the dot potential is tuned, so that the $N=1$ to $N=2$ transition is above the Fermi level of the reservoir for both, the ground state $|S\rangle$ as well as the excited state $|T\rangle$. The quantum dot now contains one electron. Then, a pulse is applied and both states are pulled below the Fermi level. After some time, an electron tunnels into the dot with Γ_T for the triplet state and Γ_S for the singlet state. The electron tunnels out in the last step again with the rate corresponding to its state.
For the spin to charge conversion, which is implemented with this step, it is required that the tunneling rate of the triplet is much larger than the rate of the singlet ($\Gamma_T \gg \Gamma_S$). The tunneling of an electron from the singlet state with $\Gamma_S = 2.5$ kHz is slow enough to be measured. As long as the dot remains occupied with two electrons, the current of the QPC will be below the starting value. Only after one electron has left, the level will be at the value corresponding to $N=1$ electrons in the dot. The tunneling of the triplet state, however, happens too fast to be detected ($\Gamma_T \sim 100$ kHz) and the current of the QPC current reaches the original value right after the end of the voltage pulse. A low pass filter of 20 kHz added to the electronic measurement assures that the tunneling from the triplet state is not detected.

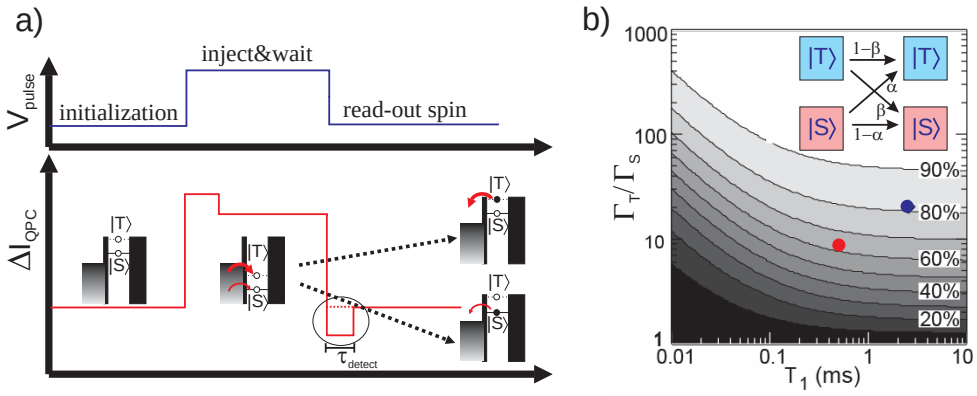


Fig. 6: *(a) Pulse sequence for the Singlet-Triplet read-out. The thicknes of an arrow depicts the tunnel rate. During the detection time τ_{detect} the QPC current drops only, if the state is a singlet. (b) Visibility depending on the ratio of the tunneling rates and the relaxation time for $\Gamma_S = 2.5$ kHz. Adapted from ref. [11].*

The visibility V as defined in equation (2) of this read-out depends on the tunneling rates Γ_T and Γ_S, the relaxation rate T_1, the time τ at which the number of electrons is measured. The probabilities α and β (for definition see Fig.6) are

$$\alpha = 1 - e^{-\Gamma_S \cdot \tau} \tag{3}$$

$$\beta = \frac{(1/T_1)e^{-\Gamma_S \cdot \tau} + (\Gamma_S - \Gamma_T)e^{-(\Gamma_S + 1/T_1)\cdot \tau}}{\Gamma_S + 1/T_1 - \Gamma_T}. \tag{4}$$

With (3) and (4) inserted in (2), the visibility depending on the ratio of the tunnel rates and the relaxation rate is shown in Fig.6b. For values of the visibility $V = 65\%$ and of the relaxation time $T_1 = 0.5$ ms as from the experiment in the previous section the ratio of the tunnel rates needed is $\Gamma_T/\Gamma_S = 10$ (marked by the red dot in Fig.6b).
The relaxation time can be obtained by measuring the triplet fraction for different waiting times as done in ref. [11]. The parameters α and β can be extracted from the same measurement (see Fig.7). The maximum visibility is 81% for optimized threshold ($\Delta I_{QPC} = -0.4$ nA) and time $\tau_{detect} = 70\mu$s (blue dot in Fig.6). The relaxation time obtained in this experiment was $T_1 = 2.58$ ms for $B = 0.02$ T. This is much longer than the relaxation time measured before at $B = 10$ T and a first indication that T_1 depends on the magnetic field, which we will discuss in more detail later.
The visibility reached with the read-out methods presented here might seem to be low. For a working quantum computer this is true, but still there are ways for improvement, e. g., lowering the electron temperature will reduce the ”dark counts” α and a faster charge detection will reduce β [10]. A higher Γ_T/Γ_S ratio will yield a larger visibility for the singlet-triplet read-out. The visibility reached so far, however, is already sufficient for first demonstrations of qubit gates and for a proof of concept we can assume the read-out DiVincezo criterium to be fulfilled.

4 Manipulation of electron spins

After learning that gate pulses can be used to quickly tune the states of a quantum dot, it is easy to understand how a fast initialization can be done. A magnetic field is applied, so that the

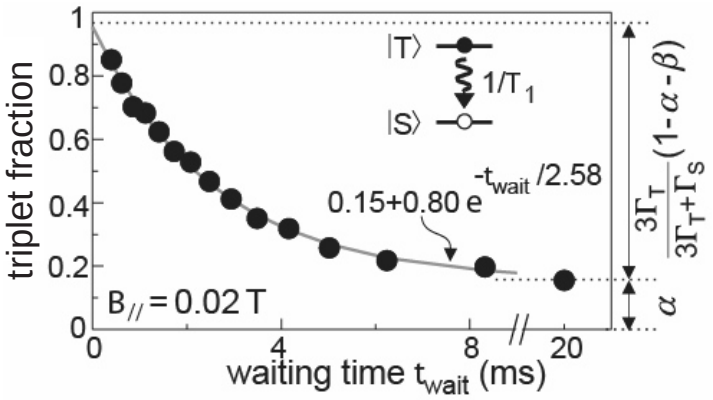


Fig. 7: *Measurement of the $|T\rangle \rightarrow |S\rangle$ relaxation time [11]. The probabilities α and β as defined in Fig.6b can be obtained as shown on the right.*

spin states are split by the Zeeman energy. First, both levels are pulsed above the Fermi level of the leads and the dot is emptied. Then, the levels are pulled down so that the spin-up level is below the Fermi level but the spin-down state is still above the Fermi level. After a time τ related to the tunneling rate, the spin-up level will be filled. The number of electrons in the dot is measured with a QPC. Now we know our initial state to be $|\uparrow\rangle$ and we can start to manipulate the spin state either by single qubit operations, i.e. single spin rotation using the first term of the Hamiltonian in equation (1), or by interaction between two qubits, using the exchange coupling J(t) of the second term, thus implementing a two-qubit gate like the $\sqrt{SWAP}$.

4.1 Single spin rotation

The state of an electron spin can be manipulated by electron spin resonance (ESR). If the spin is irradiated with an AC magnetic field B_1 with the same frequency as the Larmor frequency of the spin, i. e. the frequency of the Zeeman splitting, the spin will rotate. The angle of the rotation depends on the amplitude and duration of the B_1 pulse. This angle determines what kind of single spin gate is done, e.g., π (or 180°) corresponds to a spin-flip if the input was an eigenstate or, more generally speaking, it is a NOT gate. For more details about spin resonance, see the lecture "Donors for Quantum Information Processing" of M. Brandt or as an example for a textbook ref. [12].

In order to manipulate the electron spin in a quantum dot, an AC magnetic field of at least about 1 mT has to be coupled locally to the dot. This is much more easily said than done, since the electron temperature has to be kept very low (~ 100 mK) and high frequency irradiation always leads to dissipation of energy. The AC magnetic field is created by an AC current through a wire close to the quantum dot (see Fig.8a), with a dissipation of 10μW for $B_1 = 1$ mT and 250μW for $B_1 = 5$ mT, respectively. This requires a cooling power for the dilution refrigerator of about 300μW at 100 mK.

An ESR experiment could be done as follows. The spin is initialized in its ground state $|\uparrow\rangle$ in coulomb blockade while the level for the excited spin state $|\downarrow\rangle$ is split off by the Zeeman energy E_{Zee} and aligned between the Fermi levels of the leads (Fig.8b). In a second step, the AC magnetic field is applied, changing the spin state. Thus, the coulomb blockade is lifted and an additional current peak appears at higher gate voltage (Fig.8c,d). However, many other processes can lift the coulomb blockade as well. A current will flow independently of the ro-

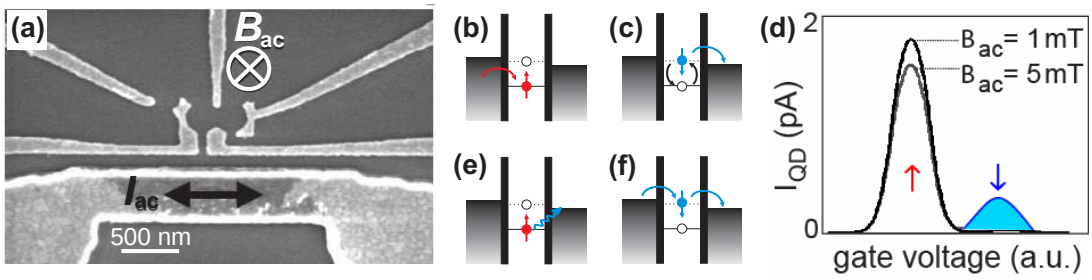


Fig. 8: *(a) Quantum dot structure with a strip line close by that creates an AC magnetic field. (b)-(d) Scheme for an ESR experiment. (e)-(f)The current due to ESR can be completely covered by photon-assisted tunneling.*

tation of the spin in the quantum dot, if the spins of the electrons in the leads have the same resonance frequency as the spin of the electron in the dot, or if heat dissipation smears out the state occupation at the Fermi level of the leads. Photon-assisted tunneling is another process that can totally mask the desired signal, which is due to ESR. In this process, the electron in coulomb blockade absorbs a photon and can tunnel directly to the drain (Fig.8e), thus lifting the coulomb blockade for transport through the excited spin state (Fig.8f). This is due to high frequency electric fields which cannot be totally suppressed. The influence of all these processes can be cancelled or at least reduced if both spin levels are pulled deep into the coulomb blockade regime by a voltage pulse. The Zeeman splitting has to be much smaller than the energy difference between the upper spin level and the Fermi level of the leads. The spin is manipulated and afterwards the electrochemical potential of the dot is pulsed back to its original position and the spin orientation is detected by either of the methods described in section 3.

The same concept can be used in a double quantum dot system with one electron in each dot (see Fig.9a). Since the exchange coupling J is very small in this configuration, the electrons can be treated as if they were separated. In this case, spin blockade as described in the lecture "Spins and Transport through quantum dots" by M. Wegewijs can be used for initialization and read-out of the system. The double dot is prepared in spin blockade, i.e. the spins in the two dots are parallel. Then, the electrochemical potential of the left dot is tuned to be deep below the transport window. An AC magnetic field rotates the spin and the electrochemical potential is raised to its former level. If the spin state has been rotated to form a singlet with the electron in the right dot, the spin blockade is lifted and a current flows. This sequence has to be repeated many times to get enough statistics. The Rabi oscillation of this experiment by Koppens et al. [13] is shown in Fig.9b. They could be observed up to pulse lengths of $1\,\mu$s, giving a lower bound for the decoherence time T_2 in this system.

One should note that the read-out scheme applied in this experiment is only sensitive to parity (parallel or antiparallel spin) and not a singlet-triplet read-out. Due to the nuclear field in GaAs, the triplet $|T_0\rangle$ and the singlet $|S\rangle$ are mixed and a $|T_0\rangle$ state will be transformed into $|S\rangle$ lifting the spin blockade. Without external magnetic field, $|T_+\rangle$ and $|T_-\rangle$ are also mixed, and no spin-blockade can be measured.

4.2 The $\sqrt{SWAP}$ operation

With regard to the requirement of a universal set of quantum gates for a quantum computer, we have seen that single qubit rotations can be done. In addition to the single spin rotations only the

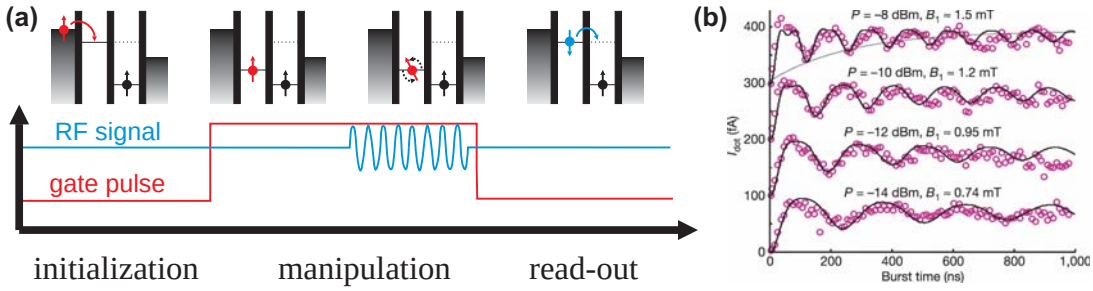


Fig. 9: *(a) Scheme of the pulse sequence for the manipulation and read-out in a double quantum dot. (b) Rabi oscillations observed experimentally (markers) and calculated (solid lines) for different magnetic fields B_1. The stronger the field, the faster is the spin rotation. Taken from ref. [13].*

CNOT gate is needed to form such a universal set. This was shown in ref. [14] and is discussed in more detail in the lecture "Fundamental Concepts of Quantum Information Processing" by T. Schäpers. On the other hand, as shown in ref. [1], the CNOT gate itself can be constructed from single spin rotations and the $\sqrt{SWAP}$ operation with

$$U_{CNOT} = e^{i(\pi/2)S_z^1} e^{-i(\pi/2)S_z^2} \sqrt{U_{SWAP}} e^{i(\pi)S_z^1} \sqrt{U_{SWAP}}. \tag{5}$$

Be aware that the operations have to be applied from right to left and that they do not necessarily commute. The single spin rotations of the two spins $i = 1, 2$ by an angle θ about the axis $a = x, y, z$ are realized by $e^{i(\theta)S_a^i}$, with the Pauli spin matrices S_a. The $SWAP$ operation exchanges the information between two qubits, i.e. $|\uparrow\downarrow\rangle$ is converted into $|\downarrow\uparrow\rangle$ while $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ do not change. With the basis

$$\begin{pmatrix} |\uparrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\uparrow\rangle \\ |\downarrow\downarrow\rangle \end{pmatrix} \widehat{=} \begin{pmatrix} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{pmatrix} \quad \text{and} \quad U_{SWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6}$$

$$\sqrt{U_{SWAP}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5+0.5i & 0.5-0.5i & 0 \\ 0 & 0.5-0.5i & 0.5+0.5i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7}$$

Starting in the product base, i. e. exchange coupling $J \to 0$, U_{SWAP} should exchange the spin information between the two qubits ($|\uparrow\downarrow\rangle \to |\downarrow\uparrow\rangle$). The product base can be expressed as coherent superposition of $|S\rangle$ and $|T_0\rangle$:

$$|\uparrow\downarrow\rangle = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/2 = (|S\rangle + |T_0\rangle)/\sqrt{2} \tag{8}$$

Now the exchange coupling J is switched on for a time t_{swap} and with

$$\int_0^{t_{swap}} J(t)/\hbar \, dt = \pi \tag{9}$$

equation (8) is transformed into

$$(|S\rangle + e^{-i\pi}|T_0\rangle)/\sqrt{2} = (|S\rangle - |T_0\rangle)/\sqrt{2} = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle - |\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/2 = -|\downarrow\uparrow\rangle \quad (10)$$

This is the state that was supposed to be reached, and the exchange coupling is switched of again. Note that there the final state has the wrong sign, but this corresponds to a "global phase" factor ($\phi = \pi$), which can be ignored [15]. The beauty of this approach is that in order to implement a $\sqrt{SWAP}$, the exchange coupling is simply turned off after the time $t_{swap}/2$ [16]. This procedure has been successfully implemented by Petta et. al [17], and in the following we shall see how it has been done.

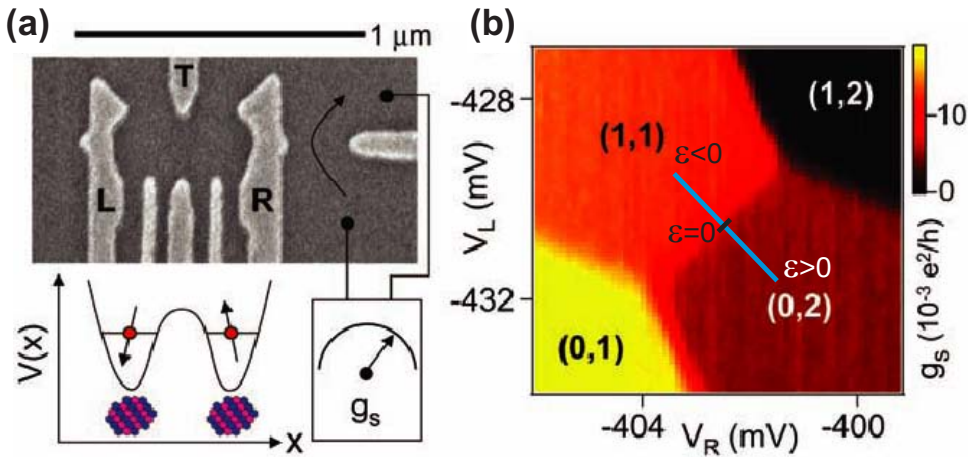


Fig. 10: *(a) Double quantum dot structure with a QPC next to the right gate. The state of the double quantum dot is detected by the current through the QPC (b). The occupation of the dot is denoted by (m,n), with m (n) the number of electrons in left (right) dot. It can be tuned by voltages V_L, V_R applied to the gates L and R. The figures are adapted from [17].*

Since a two qubit gate is to be done, a double quantum dot system as in Fig.10a has to be used. The occupation of the double dot is controlled by the voltages on the left (L) gate V_L and right (R) gate V_R, respectively, with the so-called "detuning" $\epsilon \propto (V_R - V_L)$. The gate T, which tunes the tunnel barrier between the two dots, is set to a value that gives a very weak the tunnel coupling. Therefore, the exchange interaction is very small ($J \to 0$) if the double dot is deep in the regime where each dot is occupied with one electron (1,1). A QPC next to the right dot serves as charge detector. It is tuned to be most sensitive in the regime, where either two electrons are in the right dot and the left dot is empty (0,2) for positive detuning, or where one electron occupies each dot (1,1) for negative detuning (see Fig.10b). The exchange coupling J is tuned with ϵ along the line in Fig.10b and is negligibly small for $\epsilon < -2$ mV.
Before the $SWAP$ operation can be done, the two qubit system has to be initialized in the $|\uparrow\downarrow\rangle$ state. This is done in three steps as depicted in Fig.11a-c. The system is prepared in the (0,2) singlet state $|S\rangle$ (Fig.11a). It cannot be in a $|T_0\rangle$, since this state is split off by the exchange coupling, which is large for positive detuning ϵ. Now, ϵ is changed to a negative value, thus separating the two electrons. They still form a singlet state, since they were in an eigenstate before. If there was no other interaction present, the electrons would remain in this state forever. However, besides the external magnetic field $B_0 = 100$ mT, which is the same for both quantum dots, a nuclear magnetic field B_N is present as well. This field mixes the $|S\rangle$ and the $|T_0\rangle$ state. This mixing is different for the two spins since B_N is different for the two

dots. Since they are no longer coupled to each other, the spins dephase on a time scale of about $\tau_{mix} \approx 20$ ns (Fig.11b) [17].

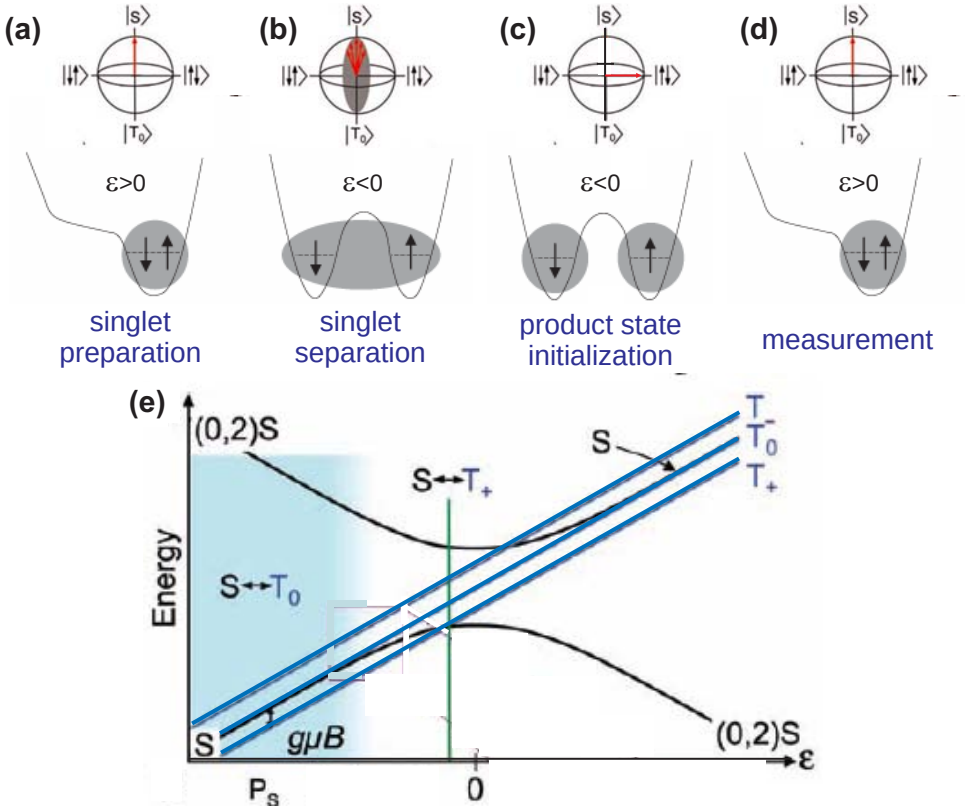


Fig. 11: *(a) Preparation of the double quantum dot in the (0,2) singlet state. (b) When the singlet is separated swiftly, the $|S\rangle$ state dephases. (c) If the separation is done slowly compared to the nuclear mixing time, the system is initialized in a product state. (d) The qubit state is measured by projection into the $|S\rangle$ - $|T_0\rangle$ base of the system. The (0,2) occupation can be reached only if the electrons form a singlet. The qubit state before the measurement can be deduced from the singlet probability. (e) Level scheme close to the (1,1)-to(0,2) transition depending on the detuning. For large negative detuning, the $|S\rangle$ and $|T_0\rangle$ states mix (blue background). At detuning of about $\epsilon \approx -1.2$ mV the $|T_0\rangle$ starts to split of from the $|S\rangle$ state due to finite exchange coupling. The $|S\rangle$ mixes with $|T_+\rangle$ at about $\epsilon \approx 0.5$ mV indicated by the green line. All triplet states are much higher in energy than the singlet (0,2). The figures are adapted from [17].*

If the transition towards negative detuning is done on a much larger timescale ($\tau_A \approx 1\,\mu$s) than this nuclear mixing time the spins still interact during the transition. This is called "adiabatic passage" and leads to a state with maximum mixing between $|S\rangle$ and $|T_0\rangle$ (both have the same probability amplitude). The phase is fixed and the spins form a product state as in eq. (8) and in Fig.11c. After some time the state is projected by tuning back to $\epsilon > 0$. If the state did not develop, it will be projected back to $|S\rangle$ Fig.11d. However, if it evolved to $|\downarrow\uparrow\rangle$ the system will now form a $|T_0\rangle$ state. Then the electron of the left dot cannot tunnel onto the right dot, because the $|T_0\rangle$ for the (0,2) configuration is too high in energy (Fig.11e).
The implementation of the $SWAP$ gate is shown in Fig.12a. The two outer Bloch spheres show the preparation and measurement of the spin states at positive detuning ϵ. Equations (8)-(10)

are represented by the three central Bloch spheres. In order to initialize the two qubit system in the $|\uparrow\downarrow\rangle$, ϵ is quickly tuned below -0.5 mV to prevent mixing between the $|S\rangle$ and $|T_+\rangle$ state due to the nuclear magnetic field (green lines in Fig.11e and 12a). Then ϵ is slowly ramped down further to provide the adiabatic passage necessary for the initialization.

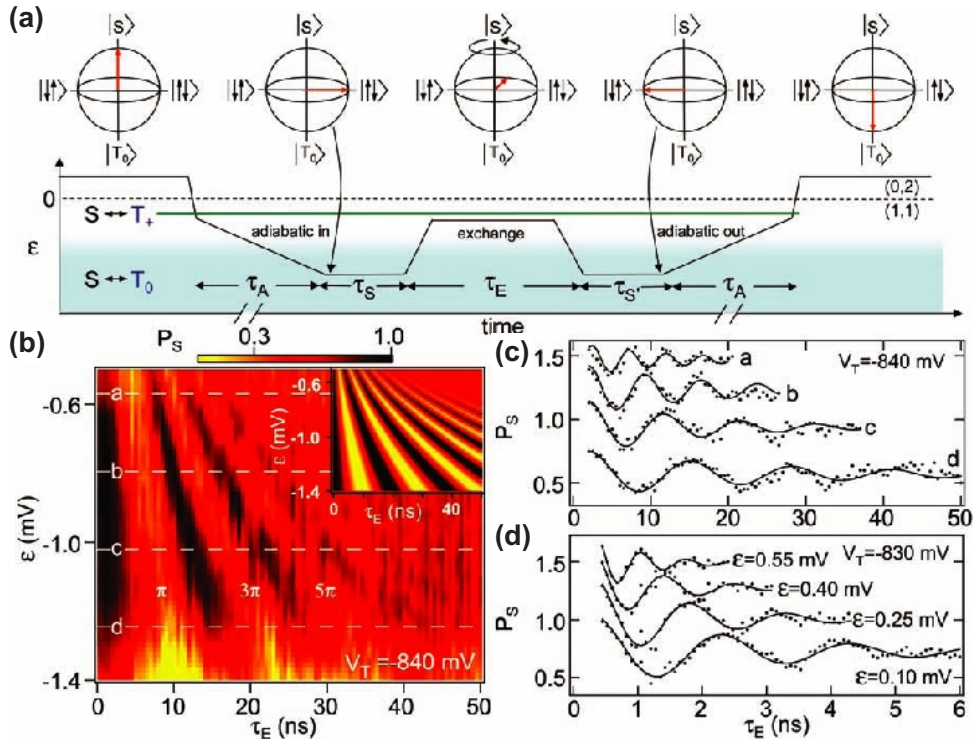


Fig. 12: *(a) Scheme for the $SWAP$ gate (b) Singlet probability P_S for different detunings during the exchange coupling and for different interaction times τ_E (c) Oscillations of the spin system during exchange coupling at different detuning marked as dashed lines in (b). (d) The oscillations are faster for weaker tunnel barrier (less negative voltage applied on gate T). The figures are adapted from [17].*

The detuning is then set to a level where the exchange coupling is larger or at least of the order of the nuclear field strength. Depending on ϵ and on the exchange time τ_E the system is rotated by an angle of $\theta = J(\epsilon)\tau_E/\hbar$. The angle of rotation θ is measured by the singlet probability (see Fig.12 b-d). A full $SWAP$ is applied for $\theta = \pi, 3\pi, 5\pi \ldots$ and the singlet probability reaches a minimum. The oscillations show that also rotations of $\theta = \frac{1}{2}\pi, \frac{3}{2}\pi, \frac{5}{2}\pi \ldots$ can be done which execute a $\sqrt{SWAP}$.
Combined with the single qubit rotations described in the previous section, a universal set of quantum gates is available for the quantum dot implementation of a quantum computer. Note that using ESR the single qubit phase gates in equation (5) cannot be carried out directly but have to be constructed form qubit rotations about the x-axis and y-axis [18].

5 Relaxation mechanisms

The fastest $\sqrt{SWAP}$ that could be done in [17] took $t = 180$ ps. This seems to be quite fast, but is it fast enough to fulfill the last DiVincenzo criterion on our list? The time it takes for a

gate has to be much shorter than the decoherence time T_2. In order to be able to apply error correction, at least 10^4 operations have to be done within T_2. The timescales and origins of spin relaxation in GaAs quantum dots will be discussed in this section.

5.1 Spin-energy relaxation

The flip of an exited spin state back to its ground state ($|\downarrow\rangle \rightarrow |\uparrow\rangle$) due to coupling with the phonon bath is called spin-energy relaxation or longitudinal relaxation and usually labeled T_1. For a spin qubit the result of such a process is a complete loss of information. It can be caused by modulation of the g-factor anisotropy due to vibrations of the crystal lattice or by relativistic coupling between the electron spin and the electric field of an emitted phonon. However, it turns out that the contributions of these direct processes to the spin energy relaxation are much smaller compared to the relaxation caused by the mixing of spin and orbital states due to spin-orbit (SO) interaction [19, 20].

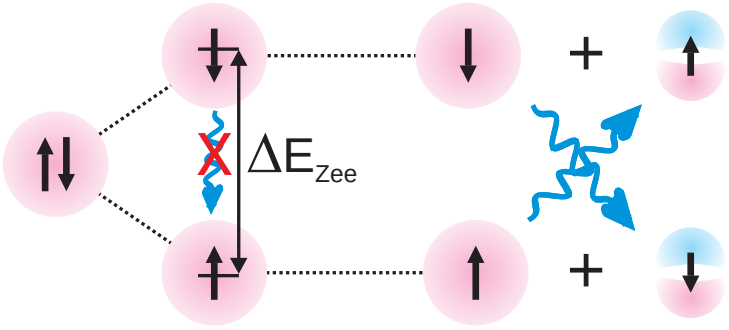


Fig. 13: *Without external magnetic field and in the absence of SO coupling, the spin states up and down are degenerate. They split by the Zeeman energy if a magnetic field is applied. Relaxation is not possible, because the direct contributions are very small and phonon coupling is prohibited. A small admixture of different spin and orbital states due to SO interaction allows phonon coupling.*

The SO Hamiltonian H_{SO} can be derived from the Dirac equation (see lecture "Electronic states in solids" by G. Bihlmayer). It consists of terms of the form $p_{x,y}\sigma_{x,y}$. Since the stationary states in a quantum dot are bound states with $\langle p_x\rangle = \langle p_y\rangle = 0$ due to the strong confinement in z, H_{SO} cannot couple different spin states of the same orbital d of the dot and

$$\langle d\downarrow|H_{SO}|d\uparrow\rangle \propto \langle d|p_x,y|d\rangle\langle\downarrow|\sigma_{x,y}|\uparrow\rangle = 0. \tag{11}$$

However, states that differ in both, the spin part as well as the orbital part, can be coupled [19]. If the Zeeman splitting is much smaller than the orbital splitting, the new eigenstates can be obtained from perturbation theory [21]

$$|d\uparrow\rangle^* = |d\uparrow\rangle + \sum_{d'\neq d}\frac{\langle d'\downarrow|H_{SO}|d\uparrow\rangle}{E_d - E_{d'} - \Delta E_{Zee}}|d'\downarrow\rangle \tag{12}$$

$$|d\downarrow\rangle^* = |d\downarrow\rangle + \sum_{d'\neq d}\frac{\langle d'\uparrow|H_{SO}|d\downarrow\rangle}{E_d - E_{d'} + \Delta E_{Zee}}|d'\uparrow\rangle \tag{13}$$

These new eigenstates (shown in Fig.13) can couple to electric fields. This leads to spin relaxation, but it also enables manipulation of the spin states by high frequency electric fields [22].
Relaxation between these new eigenstates can be of extrinsic origin, e.g., due to fluctuations of gate potentials or background charges. These and the influence of other noise sources can be kept small with a careful design of the device and turn out to be much less important compared to electric field fluctuations due to phonons. These can have two different origins. First, inhomogeneous deformations of the crystal lattice alter the band gap in space causing fluctuations of the electric field. Second, in polar crystals such as GaAs they can be caused by homogeneous strain due to the piezoelectric effect. It has been shown experimentally by studying spontaneous phonon emission that 2D and 3D piezoelectric phonons play an important role in GaAs double quantum dots [23].
The transition (relaxation) rate between the states $|d\downarrow\rangle^*$ and $|d\uparrow\rangle^*$ is given by Fermi's golden rule

$$\frac{1}{T_1} = \Gamma = \frac{2\pi}{\hbar}\sum_d |^*\langle d\uparrow|H_{e,ph}|d\downarrow\rangle^*|^2 D(\Delta E^*_{Zee}) \tag{14}$$

with the renormalized Zeeman splitting ΔE^*_{Zee}, the phonon density of states $D(E)$ at energy E and the electron-phonon coupling Hamiltonian $H_{e,ph}$ (ref. [21])

$$H^{\vec{q}j}_{e,ph} = M_{\vec{q}j} e^{i\vec{q}\vec{r}} (b^\dagger_{\vec{q}j} + b_{\vec{q}j}) \tag{15}$$

with electric field strength $M_{\vec{q}j}$ of phonon branch j (one longitudinal acoustic, two transversal acoustic) and with wave vector $\vec{q}$ at position $\vec{r}$ of the electron. $b^\dagger_{\vec{q}j}$ and $b_{\vec{q}j}$ are the phonon creation and annihilation operators. In the following we discuss the energy dependence of Γ and therefore the influence of an external magnetic field.
(i) First of all, we have to consider the phonon density of states in eq. (14). Spin-flip energies are much smaller than the energies of optical phonons and only (bulk) acoustic phonons are considered. Since they follow a linear dispersion relation, the phonon density of states increases quadratically with energy:

$$D(\Delta E_{Zee}) \propto \Delta E^2_{Zee} \tag{16}$$

(ii) The electric field strength of a phonon $M_{\vec{q}j}$ scales as $1/\sqrt{q}$ for piezoelectric phonons and as $\sqrt{q}$ for deformation potential phonons with wavenumber q. In GaAs, the effect of piezoelectric phonons dominates at energies below ≈ 0.6 meV [21]. At sufficiently small energies

$$M_{\vec{q}j} \propto 1/\sqrt{q} \propto 1/\sqrt{\Delta E_{Zee}} \tag{17}$$

Since (15) enters (14) quadratically, this adds as a factor of $1/\Delta E_{Zee}$.
(iii) Substituting eqs. (12), (13) and (15) into eq. (14), a matrix element $\langle d\uparrow|e^{i\vec{q}\vec{r}}|d'\uparrow\rangle$ is obtained describing how efficiently different orbitals are coupled by phonons. This matrix element vanishes for phonon wavelengths much shorter than the dot size l_{dot}, because the electron-phonon interaction is averaged out. The spin relaxation is fastest when the phonon wavelength is comparable to l_{dot}. For phonon wavelengths much larger than l_{dot}, the dot potential shifts uniformly up and down and different orbitals are no longer coupled efficiently. The phonon wavelength is hc_{ph}/E_{ph} and with the speed of sound in GaAs $c_{ph} \sim 4000$ m/s this yields a phonon wavelength $\lambda_{ph} \approx 16$ nm for a phonon energy $E_{ph} = 1$ meV. The Zeeman splitting and

therefore the phonon energy contributing to relaxation stays below $\Delta E_{Zee} < 200\,\mu$eV up to a magnetic field of $B_0 = 8$ T. Thus, $\lambda_{ph} \gg l_{dot}$ and the matrix element scales with

$$\langle d\uparrow|e^{i\vec{qr}}|d'\uparrow\rangle \propto q \propto \Delta E_{Zee} \tag{18}$$

This enters eq. (14) quadratically, adding a factor of ΔE_{Zee}^2.
(iv) Without finite Zeeman splitting, the various terms obtained by expanding eq. (14) using eqs. (12) and (13) cancel out [20], which is known as "van Vleck"-cancellation. It is due to the fact that the spin-orbit interaction obeys time-reversal symmetry. The SO induced rotation during half a cycle of the electric field oscillation is reversed in the second half. Thus, no net rotation takes place. Applying an external field B_0 breaks the time-reversal symmetry, because the SO interaction is of the same direction as B_0 for one half of the cycle, while it is opposite for the other half. This leads to a B_0^2 dependence of the relaxation rate [20] and

$$\Gamma_{Zee} \propto \Delta E_{Zee}^2. \tag{19}$$

Taking the contributions of eqs. (16), (17), (18) and (19) together with (15) and (14), the relaxation rate $1/T_1$ is proportional to B_0^5 since $\Gamma \propto \Delta E_{Zee}^2 \cdot \Delta E_{Zee}^{-1} \cdot \Delta E_{Zee}^2 \cdot \Delta E_{Zee}^2 = \Delta E_{Zee}^5$. For temperatures $T \gg g\mu_B B_0/k_B$ the finite phonon occupation N_{ph} leads to stimulated emission. It is accounted for by multiplying (14) with a factor $1 + N_{ph}$. The phonon occupation is given by the Bose-Einstein distribution. Therefore, $N_{ph} \propto k_B T/\Delta E_{Zee}$, and the relaxation rate is expected to follow a B_0^4 dependence. This has been observed experimentally in [24] where relaxation times up to $T_1 = 1$ s have been observed as shown in Fig.14. The same publication demonstrates the influence of the confinement on the SO interaction and thus on the relaxation time by changing the size of the quantum dot.

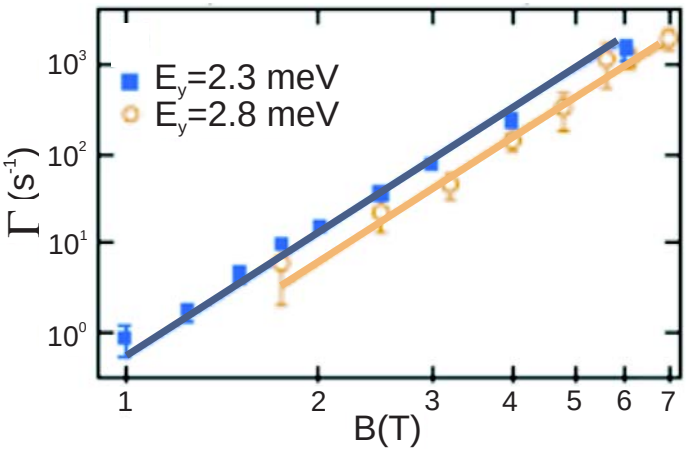


Fig. 14: *Spin-energy relaxation rates for two different confinement potentials. The markers are data points, the solid lines a fit with theory showing the expected B_0^4 dependence. The data set for weaker confinement (yellow) and thus smaller SO interaction shows smaller rates. Adapted from ref. [24].*

Besides the SO coupling there is another mechanism leading to spin-energy relaxation. Near zero field the electron spins and nuclear spins can flip-flop due to the hyperfine interaction. The electron spin evolves about the nuclear field but the nuclei also evolve around the electron spin. The field experienced by the nuclei leads to a shift of their resonance frequencies in nuclear magnetic resonance (NMR), the so-called Knight shift [25]. Since it is averaged over many

nuclei it can be taken as scalar and its strength is $A_k \approx 10\,\mu\text{s}^{-1}$ [21]. The Hamiltonian of the hyperfine interaction is

$$H_{HF} = \sum_k^N I_k \overleftrightarrow{A_k} S = \sum_k^N A_k I_k S = \sum_k^N A_k (I_k^+ S^- + I_k^- S^+ 2 I_k^z S_z)/2 \tag{20}$$

for N nuclei in the quantum dot, with $S^\pm$ and $I^\pm$ the raising and lowering operators of the electron spin and the nuclear spin, respectively. Typically, $N \approx 10^6$ in a GaAs lateral quantum dot. This leads to electron nuclear spin flip-flops and thus electron spin relaxation on a timescale of $10\,\mu$s. The energy difference between nuclear spins and electron spins grows rapidly with B_0 and flip-flops are prohibited. The SO interaction remains as the only active spin-energy relaxation mechanism at high magnetic fields.

5.2 Dephasing and decoherence

The spin-energy relaxation is due to SO coupling alone for $B_0 > 0$. If this were true as well for the phase relaxation or decoherence time T_2 (also called "transversal" spin relaxation), then we would have $T_2 = 2T_1$ [26]. Unfortunately, this is not the case. Phase relaxation does not necessarily depend on energy and fluctuations of the nuclear spins lead to decoherence of the electron spin via the hyperfine coupling. In the following we will analyze this process in more detail.

Electron spins experience a magnetic field due to the hyperfine coupling, which is called the Overhauser field. With eq. (20) and $(\sum_k^N A_k \vec{I}_k)\vec{S} = g\mu_B \vec{B}_N \vec{S}$ it is

$$\vec{B}_N = \sum_k^N A_k \vec{I}_k / g\mu_B \tag{21}$$

and of random, unknown value. Thus, the electron spin evolves in an unknown way. For fully polarized nuclear spins $B_{N,\text{max}} = 5$ T in GaAs [27]. Under experimental conditions only a small average polarization with Boltzmann statistics adds to the external field. Statistical fluctuations of the $N \approx 10^6$ nuclei of the quantum dot around this average, for spin $1/2$ similar to N coin tosses, lead to a root mean square value of the magnetic field of $B_{rms} = B_{N,\text{max}}/\sqrt{N} \approx 5$ mT, which has been confirmed experimentally [28].

The electron spin precesses about a magnetic field given by $\vec{B}_{tot} = \vec{B}_0 + \vec{B}_N$. The z-component of $\vec{B}_N$ changes the precession frequency. For $B_N^z = 1$ mT the precession rate is increased by $\Delta\nu = g\mu_B B_N^z/h = 6$ MHz and the electron spin picks up an extra phase of $180°$ within 83 ns [21]. The influence of the other components $B_N^{x,y}$ depends on their strength compared to B_0. The precession axis will be close to the x,y-plane for $B_N^{x,y} \gg B_0$. In an experiment typical values are $B_0 = 1$ T and $B_N^x \sim 1$ mT and thus, $B_N^{x,y} \ll B_0$. The precession frequency changes by $\Delta\nu \approx g\mu_B B_N^2/2B_0 = 3$ kHz causing an extra phase of $180°$ after 166 ms. The precession axis is changed by $arctan\,(B_N/B_0)$ and therefore tilted by $\approx 0.06°$. In most of the experiments $B_0 \geq 100$ mT and only B_N^z is of relevance.

If B_N^z were constant and known, its influence would not be a source of decoherence. However, B_N is fluctuating, for instance due to dynamic nuclear polarization or flip-flops of two nuclear spins with different hyperfine coupling A_k. The electron spin will pick up a random phase depending on the value of the nuclear field. For a nuclear field that is randomly drawn from a Gaussian distribution of nuclear fields with the standard deviation of $\sigma = \sqrt{\langle (B_N^z)^2 \rangle}$ (see Fig.15a), the decay of the coherence will take the form $\exp\left[-t^2/(T_2^*)^2\right]$ with (after ref. [29])

$$T_2^* = \frac{\hbar\sqrt{2}}{g\mu_B\sqrt{\langle (B_N^z)^2\rangle}} \quad . \tag{22}$$

The dephasing time T_2^* will be 37 ns for a nuclear field of $B_N^z = 1$ mT. In the experiment reported in ref. [17] and shown schematically in Fig.15b, $T_2^* = 10$ ns has been measured for the dephasing between a separated $|S\rangle$ and a $|T_0\rangle$ state corresponding to a field of $B_N^z = 2.3$ mT. Note that the dephasing time T_2^* can be much shorter than the decoherence time T_2. The effect of the nuclear field can be compensated if it assumes an unknown but constant value during the experiment, i.e. if the timescale of the fluctuations is very long compared to the timescale of the experiment (see p. 5.2).

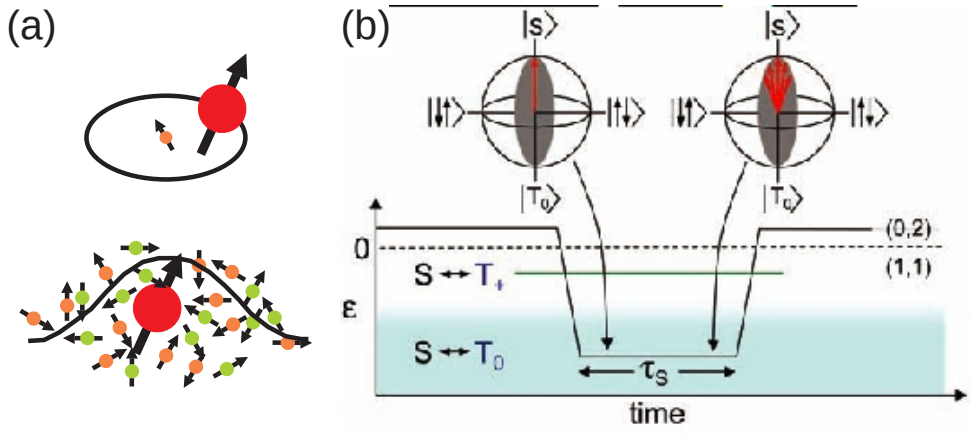


Fig. 15: *(a) One electron interacts with a single nuclear spin (top) or many nuclear spins with a Gaussian field distribution (bottom). (b) Dephasing between a separated $|S\rangle$ and a $|T_0\rangle$ state due to different random value of B_N^z at the site of each electron. Taken from ref. [17].*

The timescales of the nuclear field fluctuations depend on the interactions of the nuclei. The two most important mechanisms in this respect are the electron-nuclear hyperfine interaction [30] and the magnetic dipole interaction between the nuclei [31]. The first we already discussed in connection with the spin-energy relaxation of the electrons. Eq. (20) is only effective at $B_0 = 0$ for the electron spins. It is also most effective for nuclear spins under this condition. But the hyperfine interaction can affect B_N^z indirectly via virtual nuclear electron flip-flops between one nucleus and the electron and the electron and another nucleus. This does not affect the electron spin but leads to a flip-flop between two nuclei m and n, which changes B_N^z if $A_m \neq A_n$. As discussed in section 5.1 for the electrons, the nuclei change on a $10\,\mu$s timescale due to the Knight shift. At large magnetic fields B_0 this process will be suppressed.
In a strong external magnetic field, only the secular part of the magnetic dipole interaction Hamiltonian H_D has to be considered and

$$H_D \propto \vec{I}_m \cdot \vec{I}_n - 3I_m^z I_n^z = I_m^x I_n^x + I_m^y I_n^y - 2I_m^z I_n^z = (I_m^+ I_n^- + I_m^- I_n^+ - 4I_m^z I_n^z)/2 \quad . \tag{23}$$

The terms with the nuclear spin ladder operators $I^\pm$ vanish for coupling between different isotopes at high fields. Since the effective magnetic dipole interaction between neighboring nuclei in GaAs is about $(100\,\mu\text{s}^{-1})$ [32], $B_N^{x,y}$ change on the same timescale given by $I_m^z I_n^z$ in eq. (23). The flip-flop terms affect B_N^z but they can be strongly suppressed if $|A_m - A_n|$ is larger than

the coupling between two nuclei. Thus, B_N^z may evolve more slowly compared to the $100\,\mu$s timescale for the dipolar interaction alone.
All in all, the relevant interactions lead to moderate time scales of $t_{\text{nuc}} = 10 - 100\,\mu$s for the fluctuations of the nuclear magnetic field B_N. At high B_0, the timescale for fluctuations of B_N^z is expected to be much longer. However, this has not yet been confirmed experimentally.
For spin evolution times smaller than t_{nuc}, the influence of the fluctuations can be refocused in a Hahn echo experiment as shown in Fig.16. After a dephasing time τ_S, a rotation by $\theta = J(\epsilon)\tau_E/\hbar = \pi$ around the z-axis of the Bloch sphere is carried out. Then the spins keep evolving in the same direction as they did before the π-pulse, so that now they evolve back towards $|S\rangle$. They reach their starting state after $\tau_{S'} = \tau_S$. If there is a loss of signal, it is due to random fluctuations during $\tau_{S'} + \tau_S + \tau_E$. The spin coherence in such an experiment T_echo decays with $\exp(-t^3/t_{\text{nuc}}T_2^{*2})$ [33]. Taking $T_2^* = 10$ ns and $t_{\text{nuc}} = 10\,\mu$s, this leads to $T_echo = 1\,\mu$s. This is indeed the timescale obtained from the experiment in Fig.16.

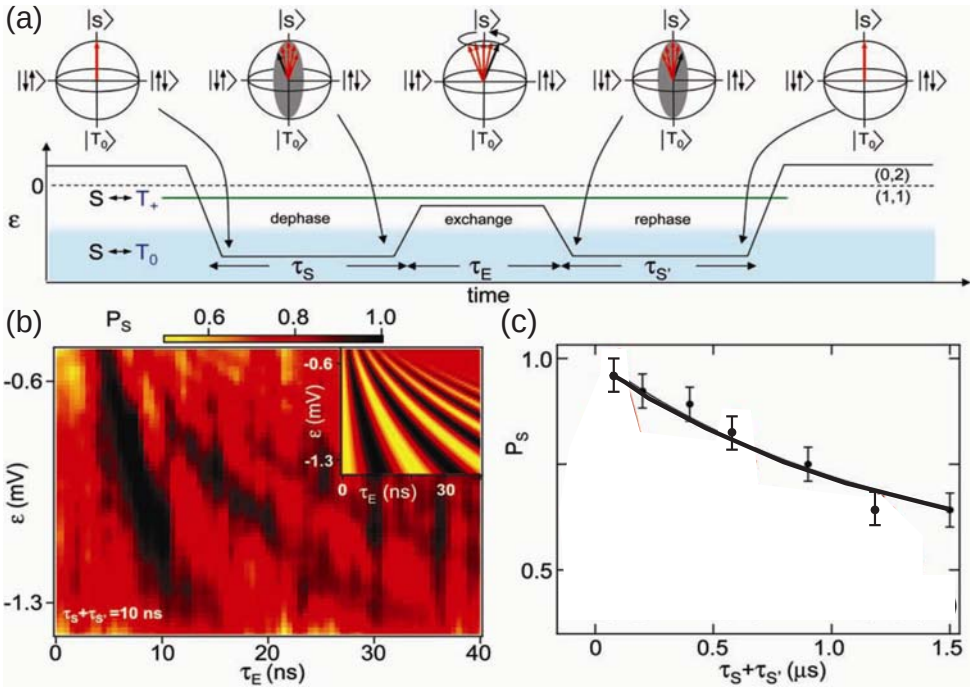


Fig. 16: *(a) The Hahn echo pulse sequence as described in the text. All transistions are done with rapid adiabatic passage so that the qubit stays all the time in the singlet-triplet base and does not change to the product base. (b) Singlet probability P_S as a function of detuning and interaction time τ_E at fixed dephasing and rephasing time. The rotation angle around the z-axis leads to an oscillation of P_S with $\theta = J(\epsilon)\tau_E/\hbar$. (c) Decay of the echo amplitude. All figures adapted from ref. [17].*

6 Summary and outlook

In this lecture we have given an introduction to quantum computing with electron spins in semiconductor quantum dots as qubits. We have shown that the qubit state can be measured with an accuracy up to 81%. Single qubit rotations can be carried out using ESR locally coupled to the quantum dot. Two qubit gates can be performed by tuning the exchange interaction between

two electrons in neighboring quantum dots. In particular, the $\sqrt{SWAP}$ operation has been introduced as a universal quantum gate. Last but not least, the origins and timescales of spin-energy relaxation and spin decoherence were discussed. The spin-energy relaxation time can be as long as $T_1 = 1$ s, depending strongly on the magnetic field B_0 and on the spin-orbit interaction. Decoherence occurs due to fluctuations of the nuclear magnetic field. Its lower bound is found so far to be $T_2 \approx 1.2\,\mu$s. Within this time, a fast $\sqrt{SWAP}$ of $180ps$ can be carried out almost 7000 times. Thus, the decoherence time seems to be sufficiently long and all DiVincenzo criteria are fulfilled.

Why then do we not already have a quantum computer? First of all, the universal gate is the $CNOT$ and composed of two $\sqrt{SWAP}$ and three single spin rotations (see eq. 5). With today's technique the latter alone take about 600 ns. They could be performed faster using a stronger B_1 field for the manipulation. However, this also increases the coupling of the electric field eventually masking the ESR effect. Materials with a larger g-factor would provide better coupling to the magnetic field and for $g \sim 2$ also the SO interaction would be small. The latter would improve the spin-energy relaxation time as well. The dephasing time itself should increase significantly in materials with less or without nuclear spins. Currently investigated as alternatives which could provide these properties are for instance quantum dots in SiGe 2DEGs or in carbon nanotubes.

Compared to the yet too short decoherence time, other limitations seem to be minor challenges. The visibility of the read-out still needs to be improved and also a gate geometry which would be scalable to hundreds or thousands of qubits needs to be developed. Although the semiconductor quantum dots remain a promising implementation for a solid state quantum computer, still a lot of work is to be done.

References

[1] D. Loss, D. P. DiVincenzo, Phys. Ref. A **57**, 120 (1998)

[2] C. Monrow, D. M. Meekhof, B. E. King, W. M. Itano, D. J. Wineland, Phys. Rev. Lett. **75**, 4714 (1995)

[3] A. Shnirman, G. Schön, Z. Hermon, Phys. Rev. Lett. **79**, 2371 (1997)

[4] B. E. Kane, Nature **393**, 133 (1998)

[5] B. van Wees, L. P. Kouwenhoven, E. M. M. Willems, C. J. P. M. Harmans, et al., Phys. Ref. B **43**, 12431 (1991)

[6] M. Field, C. G. Smith, M. Pepper, D. A. Ritchie, J. E. F. Frost, G. A. C. Jones, D. G. Hasko, Phys. Rev. Lett. **70**, 1311 (1993)

[7] J. M. Elzerman, R. Hanson, J. S. Greidanus, L. H. Willems van Beveren et al., Phys. Rev. B **67**, 161308 (2003)

[8] A. C. Johnson, J. R. Petta, C. M. Marcus, M. P. Hanson, A. C. Gossard, Phys. Rev. B **72**, 165308 (2005)

[9] L. M. K. Vandersypen, J. M. Elzerman, R. N. Schouten, L. H. Willems van Beveren et al., Appl. Phys. Lett. **85**, 4394 (2004)

[10] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp et al., Nature **430**, 431 (2004)

[11] R. Hanson, L. H. Willems van Beveren, I. T. Vink, J. M. Elzerman et al., Phys. Rev. Lett. **94**, 196802 (2005)

[12] C. P. Slichter, *Principles of Magnetic Resonance* (Springer, Berlin, 1990)

[13] F. H. L. Koppens, C. Buizert, K. J. Tielrooij, I. T. Vink et al., Nature **442** , 766 (2006)

[14] A. Barenco et al., Phys. Rev. A **52**, 3457 (1995)

[15] M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000)

[16] V. Cerletti, W. A. Coish, O. Gywat, D. Loss, Nanotechnology **16**, R27 (2005)

[17] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird et al., Science **309**, 2180 (2005)

[18] J. Stolze, D. Suter, *Quantum Computing* (Wiley-VCH, Berlin, 2004)

[19] A. V. Khaetskii, Yu. V. Nazarov, Phys. Rev. B **61**, 12639 (2000)

[20] A. V. Khaetskii, Yu. V. Nazarov, Phys. Rev. B **64**, 125316 (2001)

[21] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, L. M. K. Vandersypen, Rev. Mod. Phys. **79**, 1217 (2007)

[22] K. C. Nowack, F. H. L. Koppens, Yu. V. Nazarov, L. M. K. Vandersypen, Science **318**, 1430 (2007)

[23] T. Fujisawa, T. H. Oosterkamp,W. G. van der Wiel, B. W. Broer et al., Science **282**, 932 (1998)

[24] S. Amasha, K. MacLean, I. Radu, D. M. Zumbühl et al., Phys. Rev. Lett. **100**, 046803 (2008)

[25] W. D. Knight, Phys. Rev. **76**, 1259 (1949)

[26] V. N. Golovach, A. Khaetskii, D. Loss, Phys. Rev. Lett. **93**, 016601 (2004)

[27] D. Paget, G. Lampel, B. Sapoval, V. I. Safarov, Phys. Rev. B **15**, 5780 (1977)

[28] F. H. L. Koppens, J. A. Folk, J. M. Elzerman, R. Hanson et al., Science **309**, 1346 (2005)

[29] I. A. Merkulov, Al. L. Efros, M. Rosen, Phys. Rev. B **65**, 205309 (2002)

[30] A. V. Khaetskii, D. Loss, L. Glazman, Phys. Rev. Lett. **88**, 186802 (2002)

[31] R. de Sousa, S. Das Sarma, Phys. Rev. B **67**, 033301 (2003)

[32] R. G. Shulman, B. J. Wyluda, H. J. Hrostowski, Phys. Rev. **109**, 808 (1958)

[33] B. Herzog, E. L. Hahn, Phys. Rev. **103**, 148 (1956)

E 3 Donors for Quantum Information Processing

M. S. Brandt
Walter Schottky Institut
Technische Universität München

Contents

1 The case for silicon 2

2 Coupling of electron spins and nuclear spins 4
2.1 The spin Hamiltonian 4
2.2 Matrix representation 5
2.3 Diagonalization procedures 6

3 Device concepts 8

4 Fabrication of single-donor devices 10

5 Read-out of electron spin states 13
5.1 Silicon rf-SETs 13
5.2 Spin-to-charge conversion at the Si/SiO_2-interface 13
5.3 Decoherence 16
5.4 Outlook 18

References 19

1 The case for silicon

The quest to realize hardware for quantum information processing, and in particular for quantum computation, has motivated many research groups to assess the usability of the physical objects they study for this purpose. This has brought together experimentalists working on rather diverse topics such as ion traps, quantum optics, superconducting electronics and semiconductor physics with theoreticians developing quantum algorithms and studying decoherence processes.

It is beyond the scope of this brief introduction into donor-based qubits to fully compare the benefits and challenges of the different quantum bits or qubits investigated so far. DiVincenzo has compiled a set of requirements which can be used as a guideline for such a comparison [1]. In particular, a scalable physical system with well characterized qubits is needed, with the ability to initialize and to measure them. A universal set of quantum gates, allowing the manipulation of a single qubit as well as the controlled interaction of two qubits is required. In addition, the time needed to perform gate operations has to be much smaller than the time during which the qubits lose their coherence. And finally, one should be able to transmit qubits between distant locations. So far, nuclear spins have been used most successfully to demonstrate that quantum algorithms really work [2]. However, these groundbreaking studies have been limited to small molecules with little more than a handful of spins and qubits. Magnetic moments, both associated with nuclei and electrons, are also present in semiconductors. It is therefore an interesting question whether scalable spin-based qubit systems can be realized in semiconductors, where in conventional microelectronics memory chips with more than 10^9 classical bits can be fabricated.

Several different approaches exist to form electron spin qubits in semiconductors, such as the spins of electrons in electrostatically-defined or self-organized quantum dots, the spins of electrons bound to donors or the spins of electrons bound to defects or defect complexes. Useful nuclear spin qubits in semiconductors are in particular the nuclears spins of the donor atoms or of the atoms involved in defects and, if the semiconductor host material can be isotopically engineered, also the nuclear spins of the host atoms. To somewhat summarily go through the DiVincenzo criteria for spins in semiconductors, all qubit concepts mentioned above have been realized, and single qubit read-out, single qubit manipulation via magnetic resonance [3] as well as controlled coupling of qubits has been demonstrated for many of the concepts, but certainly not all. A general challenge of all solid state-based qubits including spins is the transfer of qubit states over long distances and it has to be noted that in this respect all solid-state qubit systems have not yet lived up to their originally presumed promise.

Which solid-state spin qubit system should therefore be pursued? As has been noted by DiVincenzo, it may still be counterproductive to even ask this question at the present stage of research into semiconductor qubits [1]. To identify possible strengths of particular spin systems, let us however compare the times T_1 and T_2 of different spin systems in semiconductors which characterize the lifetime and the coherence time of qubits, respectively. T_1 is used in the Bloch equations to describe the relaxation of a magnetization back to the equilibrium. During this relaxation, the energy of the spin system changes and therefore this relaxation is also called spin-lattice relaxation. In contrast, T_2 accounts for processes which do not change the energy of the spin system such as spin-spin scattering.

For electron spins in electrostatically-defined quantum dots fabricated in III-V compound semiconductor heterostructures, a T_1 of 0.85 ms [4] and values for T_2 longer than 1 μs [5] have been reported at temperatures of about 100 mK. For electrons in self-organized III-V quantum dots, values for T_1 up to 20 ms have been found at 1 K [6]. This comparison has to be taken with a grain of salt, since both T_1 and T_2 very sensitively depend on the temperature and the magnetic field.

The most systematic investigation of the spin properties of donors has been performed in silicon. For the electron spin of phosphorus donors at 1.25 K, a T_1 of 3000 s has been reported [7]. As in the case of quantum dots, a strong dependence on the temperature T is observed and T_1 is found to vary proportional to T^{-7} between 2.5 and 4.2 K. As expected for spin-spin interaction, T_2 is found to depend on the P concentration and T_2 times as long as 4 ms have been observed at 7 K [8]. For the corresponding nuclear spin I=1/2 of the ^{31}P nucleus, a T_1 exceeding 10 hours was reported at 1.25 K [7]. More recently, values of T_2 for phosphorus nuclear spins of nearly 2 s have been measured at 7 K [9].

Although the wave functions of electrons in quantum dots and at donors differ dramatically and therefore a direct comparison of the relevant relaxation and decoherence mechanisms is difficult, there is one key difference in the host materials: As shown in Tab. 1, all stable isotopes of the elements of groups III and V of the periodic table have a nuclear spin. In contrast, the elements of group IV have stable isotopes with no nuclear spin. In fact, isotopes of group-IV elements with a nuclear magnetic moment have a low abundance, so that Si crystals with a natural isotope composition contain only about 4.7 % of ^{29}Si with I=1/2. Therefore, cross relaxation processes between the electron and the nuclear spin systems are less important in C, Si and Ge and can in particular be engineered by artificially changing the isotope composition, e.g. by either reducing or enhancing the ^{29}Si concentration in Si crystals. Detailed experiments as a function of the isotope composition indeed show that the presence of ^{29}Si decreases both the effective T_1 and T_2 of donors in silicon [8][10].

Group III			Group IV			Group V		
^{10}B	3	19.9 %	^{12}C	0	98.9 %	^{14}N	1	99.6 %
^{11}B	3/2	80.1 %	^{13}C	1/2	1.1 %	^{15}N	1/2	0.4 %
^{27}Al	5/2	100 %	^{28}Si	0	92.2 %	^{31}P	1/2	100 %
			^{29}Si	1/2	4.7 %			
			^{30}Si	0	3.1 %			
^{69}Ga	3/2	60.1%	^{70}Ge	0	21.2 %	^{75}As	3/2	100 %
^{71}Ga	3/2	39.9 %	^{72}Ge	0	27.7 %			
			^{73}Ge	9/2	7.7 %			
			^{74}Ge	0	35.9 %			
			^{76}Ge	0	7.4 %			
^{113}In	9/2	4.3 %	^{119}Sn	1/2	8.6 %	^{121}Sb	5/2	57.2 %
^{115}In	9/2	95.7 %	^{120}Sn	0	32.6 %	^{123}Sb	7/2	42.8 %
			and eight other isotopes					

Tab. 1: *Stable isotopes of the elements of the main groups III, IV and V, their nuclear spin and natural abundance.*

The above discussion, to my opinion, is a clear case for using group-IV elements to construct the host material for spin qubits. Indeed, first experiments are under way to transfer the

investigation of spins in electrostatically-defined quantum dots to Si/SiGe heterostructures [11] with the ultimate aim of nuclear-spin free heterostructures such as ^{28}Si/^{28}Si^{70}Ge. However, a particular benefit of group-V donors as qubits in contrast to electrons in quantum dots is the nuclear spin of the donor atom, which due to the very long T_1 and T_2 times given above might very conveniently be used as qubit memory.

Why silicon, and not diamond, where the so-called NV-center, a defect complex formed by a N donor and a vacancy, shows T_2 times of up to 240 μs at room temperature [12][13] and single spin read-out has been demonstrated? This question is more difficult to answer. Read-out of the NV-center is performed optically. In contrast, most read-out concepts investigated for donors in silicon are based on electrical measurements which in principle are more easily integrated into existing microelectronics, incidentally also based on Si, rather than optical detection schemes. The comparatively shallow donor level of P in Si of 45 meV below the conduction band, on the other hand, is easily thermally ionized, so that in bulk Si the use of P donors for quantum information processing will be limited to temperatures below 30 K. In nanostructures, however, the donor becomes deeper and could therefore be used at higher temperatures as well. Another possibility to increase the accessible temperature range could be double donors such as Se, which have a much deeper donor level.

In conclusion, the possibility to isotopically engineer group-IV materials and the long spin lifetimes and coherence times of nuclear spins in these materials render donor states or donor-defect complexes in C, Si, Ge, their alloys and heterostructures well worth being studied as qubits. After a brief summary of the theoretical description of the interaction of electron spins with nuclear spins, we will discuss two concepts for donor-based quantum logic devices and methods used to position single donor atoms. Finally, we will look into experiments on how to read-out the spin state of phosphorus donors via spin-to-charge conversion at the Si/SiO_2-interface.

2 Coupling of electron spins and nuclear spins

2.1 The spin Hamiltonian

The analysis of the complete Hamiltonian of donor or defect states in a semiconductor matrix is very complex. For a description of the spin properties only, this difficulty is circumvented by the concept of a spin Hamiltonian $\boldsymbol{H}$, which explicitly includes only spin states and operators, e.g.

$$\boldsymbol{H} = \mu_B \vec{B}\hat{g}\vec{\boldsymbol{S}} + \vec{\boldsymbol{S}}\hat{D}\vec{\boldsymbol{S}} + \vec{\boldsymbol{S}}\hat{A}\vec{\boldsymbol{I}}, \qquad (1)$$

with the Bohr magneton μ_B = 9.274015×10^{-28} J/G, the magnetic field $\vec{B}$, and the electronic and nuclear spin operators $\vec{\boldsymbol{S}}$ and $\vec{\boldsymbol{I}}$. The parameters $\hat{g}$, $\hat{D}$, and $\hat{A}$ are in general spatially anisotropic matrices with at least the symmetry properties of the orbital wave function of the paramagnetic state. By convention, $\hat{g}$ represents the Zeeman interaction of the electronic spin $\vec{\boldsymbol{S}}$ with the magnetic field, $\hat{D}$ the fine structure interaction, and $\hat{A}$ the hyperfine interaction between the electronic and a nuclear spin.

To separate the spin and real space operators in such a way, it must be assumed that a ground state wave function can be factorized as the product of a spin state and some non-degenerate many-particle orbital wave function. The time-independent Schrödinger equation $\boldsymbol{H}|\psi\rangle = E_\psi|\psi\rangle$ for the spin states $|\psi\rangle$ and their eigenenergies E_ψ can often be solved

analytically. In more complex spin systems, they must be calculated numerically or via perturbation theory. The orbital parts of the wave function and spatial operators like crystal field and spin-orbit interaction are integrated out and appear only implicitly in the spin Hamiltonian via effective numerical parameters such as the effective g-tensor $\hat{g}$, whose deviation from the free electron's g-factor g_0=2.002319 stems from spin-orbit coupling $\boldsymbol{H}_{so} = \lambda \vec{\boldsymbol{L}} \vec{\boldsymbol{S}}$ treated by second order perturbation theory [14]. The form of the spin Hamiltonian is found either ad hoc, i.e. phenomenologically, motivated by symmetry considerations, or by a perturbative expansion of the full Hamiltonian [15].

2.2 Matrix representation

Spin operators $\vec{\boldsymbol{S}}$, which are used in various combinations to form the spin Hamiltonian, have the general quantum-mechanical properties of an angular momentum [16]

$$\vec{\boldsymbol{S}}^2 |m\rangle = S(S+1)|m\rangle \tag{2}$$

$$\boldsymbol{S}_z |m\rangle = m|m\rangle \tag{3}$$

$$\boldsymbol{S}_\pm |m\rangle = (\boldsymbol{S}_x \pm i\boldsymbol{S}_y)|m\rangle = \sqrt{S(S+1) - m(m\pm 1)}\,|m\pm 1\rangle \quad . \tag{4}$$

These equations are valid for spin S = 1/2 electrons as well as for many-electron systems with higher values of S. A set of basis vectors for the 2S+1-dimensional spin space is defined by unit vectors for the orthonormal states $|m\rangle$ with the eigenvalues m=−S, … , S of the operator $\boldsymbol{S}_z$. The spin operators $\boldsymbol{S}_{x,y,z}$ take the form of matrices with dimension (2S + 1) × (2S + 1). For a spin-1/2 system, these are the well-known Pauli matrices, which may be combined in a vector notation of operators to $\vec{\boldsymbol{S}} = (\boldsymbol{S}_x, \boldsymbol{S}_y, \boldsymbol{S}_z)$ to describe the anisotropic properties of spins. The matrix forms of the spin operators for all higher values of S are defined by Eqs. (2) to (4). In short form, all non-zero matrix elements $\langle n|\boldsymbol{S}_{x,y,z}|m\rangle$ of the operators $\boldsymbol{S}_{x,y,z}$ can be summarized as

$$\langle n|\boldsymbol{S}_x|m\rangle = \delta_{n,m\pm 1}\frac{1}{2}\sqrt{S(S+1) - m(m\pm 1)} \tag{5}$$

$$\langle n|\boldsymbol{S}_y|m\rangle = \mp\delta_{n,m\pm 1}\frac{i}{2}\sqrt{S(S+1) - m(m\pm 1)} \tag{6}$$

$$\langle n|\boldsymbol{S}_z|m\rangle = \delta_{n,m} m \quad . \tag{7}$$

With a spin Hamiltonian built from linear combinations of such matrix operators, the Schrödinger equation becomes a simple matrix equation, whose eigenvalues can be calculated from the characteristic polynomial $\det(\boldsymbol{H}\text{-}\boldsymbol{1}E_i) = 0$, where $\boldsymbol{1}$ is the unity matrix, and $E_i(\vec{B})$ is one of the eigenvalues of the matrix equation. The corresponding eigenvector $|\psi_i\rangle$ defines the spin eigenstate for E_i as a linear combination $|\psi_i\rangle = \sum_m \alpha_{im}|m\rangle$ of the pure basis vectors of $\boldsymbol{S}_z$. These basis vectors $|m\rangle$ of $\boldsymbol{S}_z$ are usually the eigenstates of the spin Hamiltonian in the limit of very high magnetic fields along the z-axis, where all other perturbations can be neglected.

A system of several spins, e.g. one electronic spin S and one nuclear spin I, requires a total of (2S + 1)⊗(2I + 1) orthogonal eigen-vectors. The orthonormal basis vectors $|m_s, m_i\rangle$ of the product space may again be organized as unit vectors, now with a combined index (m_s,m_i),

which takes all values (−*S*,−*I*), (−*S*,−*I*+1), … , (*S*,*I*). The matrix elements $\langle n_s, n_i | \boldsymbol{S}_{x,y,z}, \boldsymbol{I}_{x,y,z} | m_s, m_i \rangle$ of the electronic and nuclear spin operators are defined according to Eqs. (5) to (7) also in this product space. The operators $\boldsymbol{S}_{x,y,z}$ leave the nuclear spin states unaffected and are represented by the unity matrix elements $\delta_{ni,mi}$ for the nuclear spin, and vice versa $\boldsymbol{I}_{x,y,z}$ leave the electronic spin states unaffected and are represented by the unity matrix elements $\delta_{ns,ms}$ for the electronic spin, i.e.

$$\langle n_s, n_i | \boldsymbol{S}_{x,y,z} | m_s, m_i \rangle = \delta_{ni,mi} \langle n_s | \boldsymbol{S}_{x,y,z} | m_s \rangle \tag{8}$$

$$\langle n_s, n_i | \boldsymbol{I}_{x,y,z} | m_s, m_i \rangle = \delta_{ns,ms} \langle n_i | \boldsymbol{I}_{x,y,z} | m_i \rangle \tag{9}$$

with the operators on the right-hand side as defined in Eqs. (5) to (7).

2.3 Diagonalization procedures

The spin Hamiltonian matrix needs to be diagonalized in order to obtain its eigenenergies E_i and eigenvectors $|\psi_i\rangle$ [17]. In cases where it is not possible to solve the matrix Schrödinger equation analytically, approximate solutions from perturbation theory, or exact numerical solutions for one set of matrix elements at a time must be obtained. These calculations are somewhat simplified, if the magnetic field is oriented along the *z*-axis of the spin Hamiltonian, as typically the Zeeman term dominates the eigenenergies of the spin. For other orientations of the magnetic field, this can be achieved by a rotation of the coordinate system of the spin operators $\vec{\boldsymbol{S}}' = \hat{R}\vec{\boldsymbol{S}}$ [17]. In the new coordinate system, the other possible interactions then introduce only small off-diagonal elements to the spin Hamiltonian.

The case of a phosphorus donor with *S*=1/2 and *I*=1/2, $\boldsymbol{H}$ can be solved analytically by inspection. Neglecting the weak nuclear Zeeman interaction, the spin Hamiltonian is

$$\boldsymbol{H} = \mu_B \vec{B} \hat{g} \vec{\boldsymbol{S}} + \vec{\boldsymbol{S}} \hat{A} \vec{\boldsymbol{I}} = \mu_B \vec{B} \hat{g} \vec{\boldsymbol{S}} + A\left(\boldsymbol{S}_x \boldsymbol{I}_x + \boldsymbol{S}_y \boldsymbol{I}_y + \boldsymbol{S}_z \boldsymbol{I}_z\right) \tag{10}$$

with the external magnetic field $\vec{B} = (0,0,B)$ along the *z*-direction, the isotropic g-factor $\hat{g} = \boldsymbol{1}g$, and the isotropic hyperfine interaction $\hat{A} = \boldsymbol{1}A$. The basis vectors $|m_s, m_i\rangle$ of the 4-dimensional product basis of $\boldsymbol{S}_z$ and $\boldsymbol{I}_z$ in this example are

$$\left|\frac{1}{2},\frac{1}{2}\right\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad \left|\frac{1}{2},-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \quad \left|-\frac{1}{2},\frac{1}{2}\right\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad \left|-\frac{1}{2},-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix} . \tag{11}$$

The matrix form of this spin Hamiltonian is given by

$$\boldsymbol{H} = \begin{pmatrix} +\frac{\mu_B B g}{2} + \frac{A}{4} & 0 & 0 & 0 \\ 0 & +\frac{\mu_B B g}{2} - \frac{A}{4} & \frac{A}{2} & 0 \\ 0 & \frac{A}{2} & -\frac{\mu_B B g}{2} - \frac{A}{4} & 0 \\ 0 & 0 & 0 & -\frac{\mu_B B g}{2} + \frac{A}{4} \end{pmatrix} . \tag{12}$$

The off-diagonal elements $A/2 << \mu_B B g$ can be neglected for small hyperfine couplings compared to the Zeeman interaction. In this case, Eq. (12) is already approximately diagonal,

which means that the diagonal elements and the basis vectors of Eq. (11) are a first-order approximation to the eigenenergies and eigenvectors of the system. In this regime, both *S* and *I* are quantized along the external magnetic field. These first-order eigenenergies are shown as a function of *B* in the energy diagram of Fig. 1a).

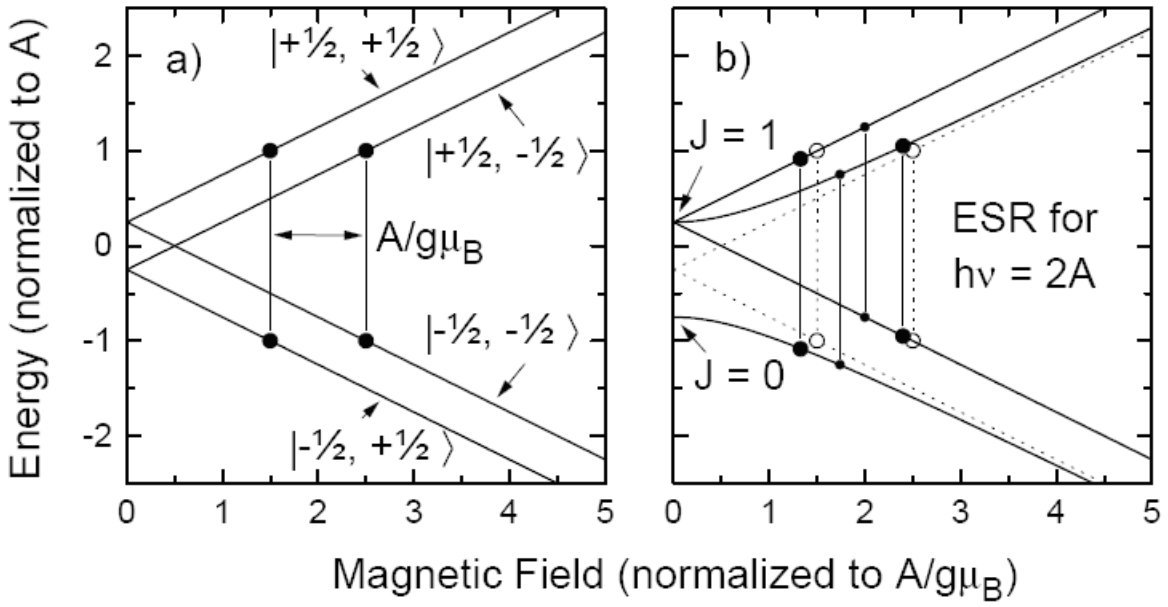


Fig. 1: *Breit-Rabi diagrams a) of the first-order solutions and b) of the exact solutions of the spin Hamiltonian of Eq. (12) with the electronic spin S = 1/2, the nuclear spin I = 1/2, and the hyperfine interaction A. For better comparison, the first-order solutions are included in b) with dashed lines. The energy levels are labelled according to the basis states* $|m_s, m_i\rangle$ *which are eigenstates of* $\boldsymbol{H}$ *at* $\mu_B Bg >> A$*. For* $\mu_B Bg << A$*, I tends to be coupled with S to an effective angular momentum J = S − I, … , S + I. The vertical lines indicate the two strongly allowed ESR transitions for a microwave energy* $h\nu = 2A$*. In (b), also the weakly allowed transitions are indicated.*

According to the dipole selection rules, the allowed electron spin resonance (ESR) transitions are those with $\left|+\frac{1}{2},\pm\frac{1}{2}\right\rangle \longleftrightarrow \left|-\frac{1}{2},\pm\frac{1}{2}\right\rangle$ with $\Delta m_s = \pm 1$ and $\Delta m_i = 0$. For a given transition energy $\Delta E = h\nu$, where h is the Planck constant, these two transitions appear at the magnetic fields $B = \left(h\nu \pm \frac{A}{2}\right)/\mu_B g$ in the first-order approximation. With the characteristic polynomial of the inner 2 × 2 block matrix, the exact energy eigenvalues $E_1, \ldots, E_4$ can be calculated analytically without approximations over the complete magnetic field range

$$E_{1,4} = +\frac{1}{4}A \pm \frac{1}{2}\mu_B Bg \tag{13}$$

$$E_{2,3} = -\frac{1}{4}A \pm \frac{1}{2}\sqrt{(\mu_B Bg)^2 + A^2} \quad . \tag{14}$$

The characteristic anti-crossing between the $\left|\pm\frac{1}{2},\mp\frac{1}{2}\right\rangle$ levels for small external fields is shown in Fig. 1 b). In the low-field regime, the eigen-vectors $\left|E_{2,3}\right\rangle$ are symmetric and antisymmetric combinations of the $\left|\frac{1}{2},-\frac{1}{2}\right\rangle$ and $\left|-\frac{1}{2},\frac{1}{2}\right\rangle$ basis vectors. As a consequence, the $\left|+\frac{1}{2},\pm\frac{1}{2}\right\rangle \longleftrightarrow \left|-\frac{1}{2},\mp\frac{1}{2}\right\rangle$ transitions are not completely "forbidden" in the low-field regime, as indicated with the smaller dots in Fig. 1 b). Physically, the system is then best described via a coupled angular momentum $J = S - I, \ldots, S + I$ for $\mu_B Bg << A$. For $A > 0$, the $J = 0$ singlet state with opposite nuclear and electronic spin orientation has lower energy than the "ferromagnetically" coupled $J = 1$ triplet state.

3 Device concepts

The original concept to use the electron spin and the nuclear spin of phosphorus donors in silicon as qubits goes back to Kane [18][19]. A device similar to the one envisaged by him is shown in Fig. 2. It consists of two phosphorus donors placed in isotopically pure ^{28}Si to suppress decoherence by the ^{29}Si nuclear spins present in natural silicon. Each donor is positioned below a separate gate, denoted A gate. In between the two donors and their respective A gates, a second gate denoted J is placed. The device is expected to operate at a temperature T of about 100 mK and at a magnetic field higher than 2 T, so that the electron spins will be fully polarized.

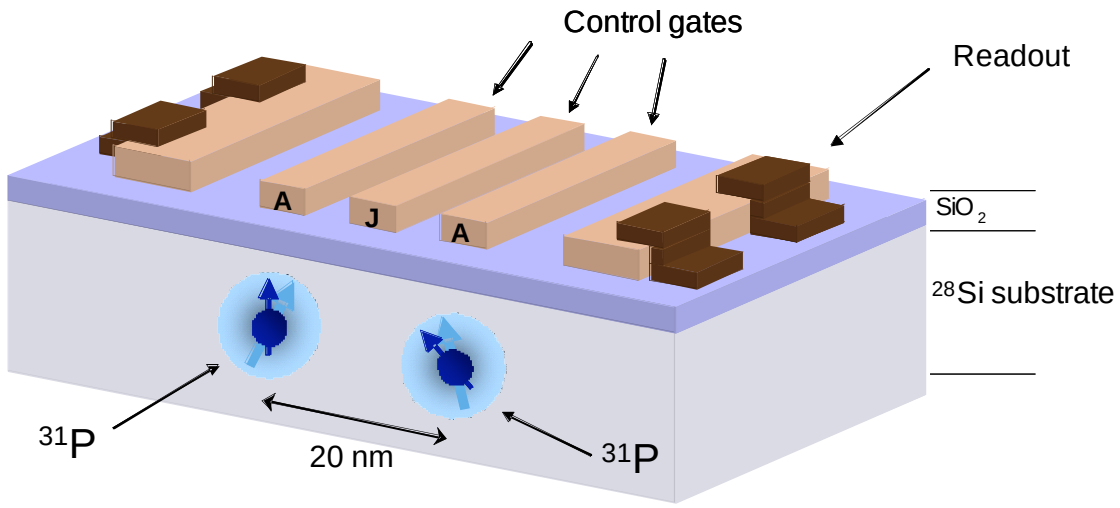


Fig. 2: *Kane concept for a quantum logic element based on the nuclear spins of ^{31}P donors in isotopically pure ^{28}Si.*

With the help of the A gate, the wave function of the donor electron can be manipulated, pulling it towards the Si/SiO_2-interface or pushing it away from it. The hyperfine interaction A between the two spins present in a single ^{31}P donor, the electron spin S and the nuclear spin I, is given by the so-called Fermi contact hyperfine interaction $A = \frac{2}{3}\mu_0 g \mu_B g_n \mu_n \left|\psi(0)\right|^2$,

where g_n is the nuclear g-factor of ^{31}P, μ_n the nuclear magneton and $|\psi(0)|^2$ the probability density of the electron donor wave function at the position of the nucleus [20][21]. Changes of the wave function due to a Stark effect should therefore lead to a variation of the hyperfine interaction. If electron spin resonance with $\Delta m_s=\pm 1$ is excited at the position of one of the hyperfine-split resonances, say $m_i=+1/2$, for a certain bias voltage at the A gate, changes of that bias voltage will lead to that particular electron spin transition going off resonance. Exciting the corresponding nuclear magnetic resonance (NMR) with $\Delta m_i = \pm 1$ effectively given by the hyperfine interaction A, a change of the bias voltage will similarly switch on or off the nuclear spin transition. Thereby, while exciting ESR or NMR globally, via the application of gate voltages either the electron or the nuclear spin can be addressed locally. This allows the realization of single qubit operations or gates. In his original proposal, Kane envisaged the use of the nuclear spins a qubits. A coupling of two nuclear spins at neighboring donors and therefore a two-qubit gate is possible via the electron spins at the respective donors. Their exchange interaction and thereby also the interaction between the nuclear spins are controlled in Fig. 1 by the coupling gate J influencing the overlap of the electronic wave functions. Since the exchange coupling, which dominates the spin-spin interaction of neighboring ^{31}P electrons, depends exponentially on the distance of the donor atoms, the accurate placement of the donors is quite crucial in this respect. To be able to effectively influence the coupling by a J gate, distances of about 5 to 10 Bohr radii are expected to be necessary.

The read-out of the electron spin state can similarly be discussed with the help of Fig. 2. For this, the gate voltages are adjusted such that the electron would be transferred from one donor to the other. Let us assume that we want to read-out the spin on the left and that the spin on the right is in a well known state such as spin down $m_s=-1/2$. The final state will be a double occupied donor state $^{31}P^-$, which according to the Pauli principle has to be in a total $S=0$ singlet state. The transfer of the electron between the donors can therefore only take place when the left electron is in a spin up state initially, or, more correctly, if the two electron spins also initially form a singlet. The Pauli principle therefore governs the charge transfer rate, so that the spin information is transferred into a charge information. Adiabatically changing the coupling of the two electron spins furthermore allows to convert the nuclear spin state into the spin symmetry of the electron spin pair, thereby facilitating the read-out of the nuclear spin. Finally, Kane suggests to use sensitive capacitive techniques such as radio-frequency single-electron transistors (rf-SETs) to determine the charge state of the read-out donor, the right donor in our case.

The effect of electric fields E on the hyperfine interaction of donors has been investigated in detail using ^{121}Sb [22]. Relative changes of the g-factor of $\Delta g/g=-1\text{x}10^{-5}\ \mu\text{m}^2/\text{V}^2\ E^2$ and of the hyperfine interaction of $\Delta A/A=-3.7\text{x}10^{-3}\ \mu\text{m}^2/\text{V}^2\ E^2$ have been found The application of high electric fields is limited by field ionization, so that only about 3 kV/cm can be used. This restricts the relative change in hyperfine interaction to about $3\text{x}10^{-4}$. The vicinity of the Si/SiO_2-interface will increase the applicable field somewhat, but still the Stark effect on the hyperfine interaction is limited.

The main advantage of the Kane concept based on the nuclear spins of the ^{31}P is their long coherence. The nutation rate with which spins can be flipped is given by $\nu_{nut} = g\mu_B B_1 / h$ or $\nu_{nut} = g_n\mu_n B_1 / h$, where B_1 is the strength of the microwave or radio-frequency magnetic field

used to drive the ESR or NMR, respectively. For the same B_1 fields, electron spins can be manipulated much faster than nuclear spins, allowing a higher clock speed in a possible quantum processor and simultaneously tolerating smaller T_2.

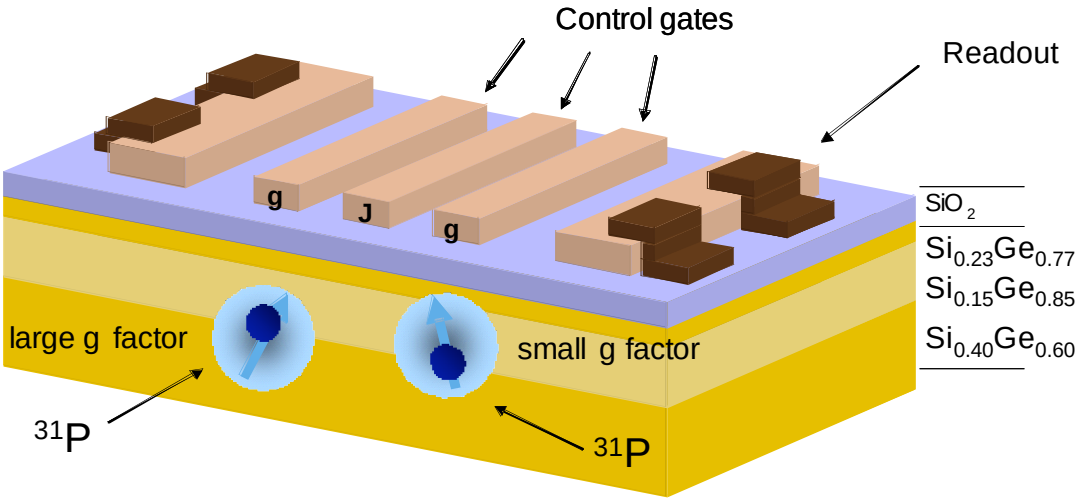


Fig. 3: *Proposal by Vrijen and coworkers on a quantum logic element based on donors placed at the interface of two epitaxial layers of SiGe with different alloy composition.*

Vrijen and coworkers suggested an alternative concept for a donor-based quantum logic element using the electron spins of donors placed at the interface of two different semiconductor materials as qubits (Fig. 3) [23]. The full miscibility of Si and Ge as well as the existence of stable nuclear spin-free Si and Ge isotopes renders the SiGe alloy system particularly interesting for this approach. Indeed, the g-factor of donors in Si is about 1.998, while in Ge a g of 1.563 is found. Therefore, placing a donor at the interface of two isotopically pure ^{28}Si and ^{70}Ge layers, or rather two isotopically pure $Si_{1-x}Ge_x$ alloys with alloy compositions such that the donor level is at the same energy in both layers, one might influence the g-factor via pushing or pulling the donor wave function from one material to the other with a gate now labeled g to a larger degree than by the Stark effect discussed for donors in pure Si above. However, it must be noted that the electrons will reside in different valleys in the two layers, since the conduction band minima of Si are near the X point in the Brillouin zone, while the band minima in Ge are at the L point. The effect of an effective intervalley scattering performed by the application of an electric field in the Vrijen concept in particular on T_2 remains to be investigated. Nevertheless, in addition to a possibly much higher change in the g-factor, the more shallow donors in Ge-rich SiGe alloys also have larger Bohr radii, so that the qubits have to be farther apart than in the pure Si design by Kane and the requirements for the accuracy of the donor placement can be less stringent.

4 Fabrication of single-donor devices

One of the main challenges in the realization of donor-based quantum logic devices is the accurate placement of single donors. As outlined above, the distance of the donors both to the Si/SiO_2-interface and to the neighboring donors should be typically of the order of 20 nm. The

main approaches to realize such a position of single donor atoms are implantation and lithography.

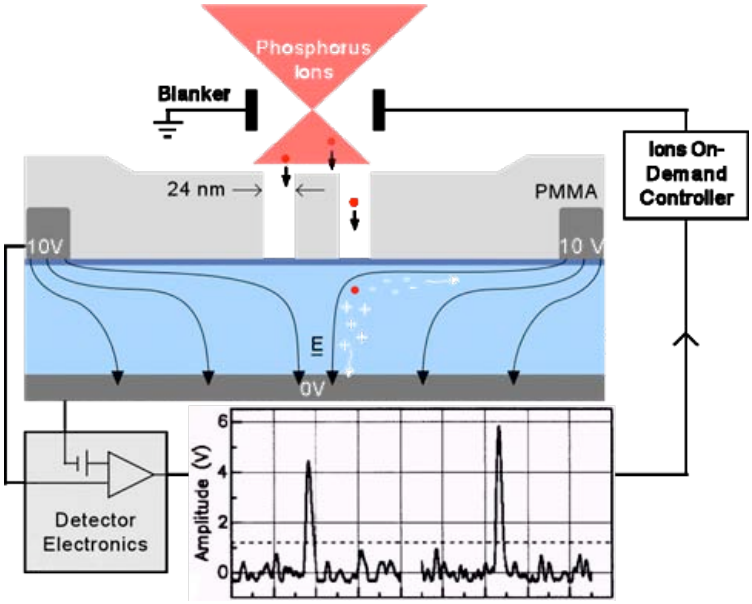


Fig. 4: *Implantation of single donors. Adapted from [24].*

A schematic diagram of a single-dopant implantation set-up is shown in Fig. 4 [24]. The silicon substrate itself is used as a particle detector. Phosphorus ions being implanted will generate electron-hole pairs, which are separated in the applied electric field. Using appropriate electronics, the current pulse thus generated by a single ion implantation event can be detected. If a low flux of incident ions is used, a beam blanker can be activated quickly enough after the observation of an implantation event to prevent further ions from reaching the substrate, allowing single ion implantations. As sketched in Fig. 4, lithographically-defined masks can define the location of the implantation. When pairs of dopants are to be formed, the approach based on two neighboring openings in the mask shown in Fig. 4 will provide such dopant pair structures with a probability of 50%. Alternatively, atomic force microscopy (AFM) tips, which have a channel in the tip through which the ions can be implanted, have been used as moveable implantation masks [25], reducing the risk of double implantation through a single opening in the lithographically-defined mask of Fig. 4.

Shallow implantations with a typical depth of 20 nm can be obtained with phosphorus ion energies of the order of 10 to 20 keV. However, the scattering between the incident ion and the atoms of the Si substrate lead to straggle. At a primary energy of 14 keV for singly negatively charged P ions, the exact position of the P atom in the Si substrate after implantation varies by about 11 nm in the distance to the substrate surface and by about 8 nm in the directions perpendicular to that. Together with the diameter of the mask openings, this straggle leads to a variation of the distance between P atoms e.g. forming a quantum logic element as sketched in Fig. 2 by several Bohr radii.

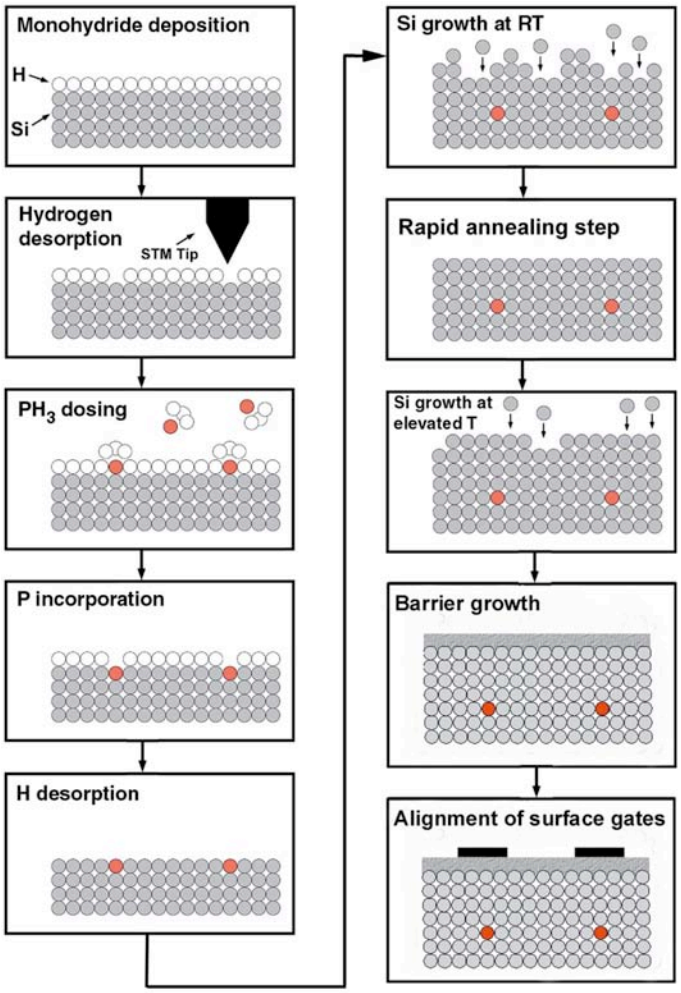


Fig. 5: *Lithography of single donors by local desorption of hydrogen atoms induced by scanning tunneling microscopy. Adapted from [26].*

An alternative, but experimentally even more challenging approach is the use of lithography based on scanning tunneling microscopy (STM) sketched in Fig. 5 [26][27][28]. The fundamental idea is to fabricate a hydrogen-terminated surface and, with the help of an STM tip, to locally remove this atomic analog to a photoresist. The surface Si atoms not terminated by hydrogen are very reactive, and PH_3 molecules brought onto the surface will bind to them, which allows to position single phosphorus atoms with the atomic precision of STM. It only remains to form an epitaxial Si cap layer, followed by growth of a SiO_2 barrier and the fabrication of the gate contacts. However, Si tends to move with the growth front at the temperatures usually used for high quality epitaxial growth, which would significantly reduce the accuracy of the P placement in the final device structure. Using an intermediate growth step at room temperature, followed by a rapid thermal anneal and a high temperature overgrowth, the "straggle" in the P position can be reduced to below 1 nm, a value smaller than the Bohr radius of effective mass donors in Si. However, it should be noted that also in the case of STM lithography, every P donor pair formed will have such a different overlap of the wave functions that the relevant voltages at the J gates in Fig. 2 for the switching of the electronic spin-spin interaction will have to be adjusted separately for each pair. Apart from devices for quantum information processing, the STM-based lithography summarized also allows the fabrication of other interesting and novel devices for "atomic electronics" in general.

5 Read-out of electron spin states

The second main challenge is the read-out of single donor spins. While this has been demonstrated for electrostatically-defined [4] and self-organized quantum dots [29] as well as for the NV-center in diamond [30], the successful determination of both the electron and the nuclear spin state of single donors is subject to intense work by several groups.

5.1 Silicon rf-SETs

In parallel to the fabrication of single dopant devices, significant progress is being made in the development of techniques for the sensing of single charges in Si devices, an integral step in the measurement of spin states. As an example, Fig. 6 a) and b) show a radio-frequency single electron transistor (rf-SET) formed at the Si/SiO_2-interface of a nearly undoped Si substrate [31]. Between the source and drain contacts, an n-type channel with a low electron density is formed. Via two gate electrodes B1 and B2, an island can be controllably created in the middle of the channel. The radio-frequency reflectance of the device, a measure of the source-drain conductance, shows clear Coulomb diamonds in Fig. 6 c), demonstrating the possibility to measure the change of the number of charges on the islands on a single-electron level. The particular detection via rf reflectance allows charge measurements with a bandwidth as high as 2 MHz, comparable to that of mature aluminium rf-SETs. These results suggest that the quantum logic element as well as the charge detector required for the spin read-out could be fabricated as an integrated device from Si.

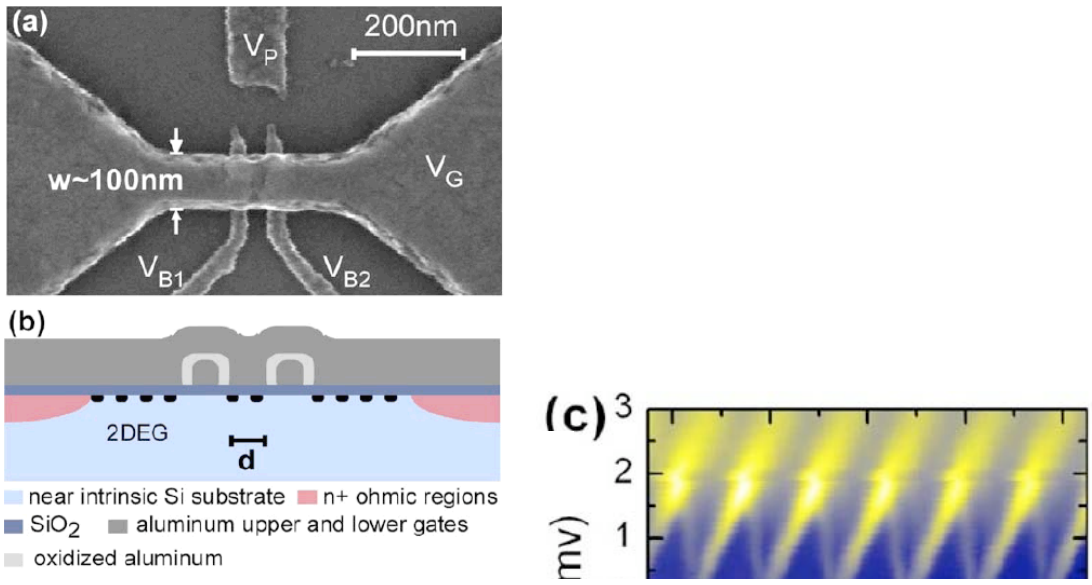


Fig. 6: *a) Scanning electron microscopy top view and b) schematic side view of a Si rf-SET. c) Coulomb diamonds demonstrating the single electron sensitivity of the device. Adapted from [31].*

5.2 Spin-to-charge conversion at the Si/SiO_2-interface

So far, we have discussed the read-out of the electron spin state via a spin-to-charge conversion process involving two neighboring ^{31}P donor states. However, any paramagnetic state can be used as a partner to read-out the donor spin state. To be able to manipulate the qubits, the donors have to be near to the gate electrodes, which are insulated from the silicon

substrate via an oxide. At the interface of the silicon and this oxide, defects such as unsaturated Si "dangling" bonds occur naturally in concentrations of typically between 10^{11} and 10^{13} cm^{-2}, depending on the exact oxide growth conditions. By passivation with hydrogen in a forming gas anneal or via compensation, the density of these defects can be reduced. However, they can also be used very conveniently as the partner required for spin-to-charge conversion [32].

Figure 7 shows a spin-dependent recombination process involving a ^{31}P donor electron and the dominant Si/SiO_2-interface defect named P_{b0} [33]. As in the case of the charge transfer between neighboring P donors discussed above, the recombination step can only proceed if the ^{31}P-P_{b0} pair is initially in a singlet spin configuration. In this case, a negatively charged P_{b0}^- state is formed. If mobile charge carriers are present such as electrons and holes generated by illumination, an electron will be trapped by the positively charged ^{31}P and a hole by the negatively charged P_{b0} center, leading to a reduction of the carrier densities and to a reduction of the conductivity. The symmetry of the spin pair and therefore also the electron spin state of the donor can thus be detected by changes in e.g. the photoconductivity.

Fig. 7: *Spin-to-charge conversion at the Si/SiO_2-interface.*

To demonstrate the feasibility of this electrical read-out scheme of the electron spin state, let us look at an ensemble of ^{31}P-P_{b0} pairs. Due to the spin-allowed recombination, most ^{31}P-P_{b0} singlet pairs will have recombined in thermal equilibrium, most ^{31}P-P_{b0} pairs remaining will be in a triplet configuration. Turning triplets into singlets via electron spin resonance of either the ^{31}P or the P_{b0} (which is possible when the ^{31}P-P_{b0} coupling is small), pair recombination is increased [3]. Continuous resonance excitation will lead to a continuous oscillation of the ensemble being dominantly in the triplet or singlet configuration and, simultaneously, an oscillation of the ^{31}P-P_{b0} recombination rate. The observation of these Rabi oscillations in the conductivity will be a clear demonstration of this spin read-out concept. However, the *RC*-timeconstants of the samples typically studied do not allow the direct monitoring of these oscillations. To overcome this limitation, two technical tricks summarized in Fig. 8 are currently used [34].

Figure 8 a) shows the oscillation of the recombination during the application of a microwave pulse exciting electron spin resonance of either partner in the spin pair. After the pulse, the

spin system is most-likely in a non-equilibrium state determined by the symmetry of the ^{31}P-P_{b0} pairs and will relax back into equilibrium. Governed by different time constants including the singlet- and triplet recombination times and the *RC*-timeconstant of the detection system, a current transient will be observed during this relaxation. The amplitude of this current transient, however, is proportional to the deviation of the singlet/triplet content of the spin pair ensemble at the end of the microwave pulse from thermal equilibrium (Fig. 8 b), so that measurements of this transient long after the excitation of the spin resonance has finished provide the wanted information on the spin orientation [34]. To further improve the signal-to-noise ratio of these experiments, the current transient is integrated over a certain time window, yielding a charge *Q* as the primary physical quantity measured (Fig. 8 c).

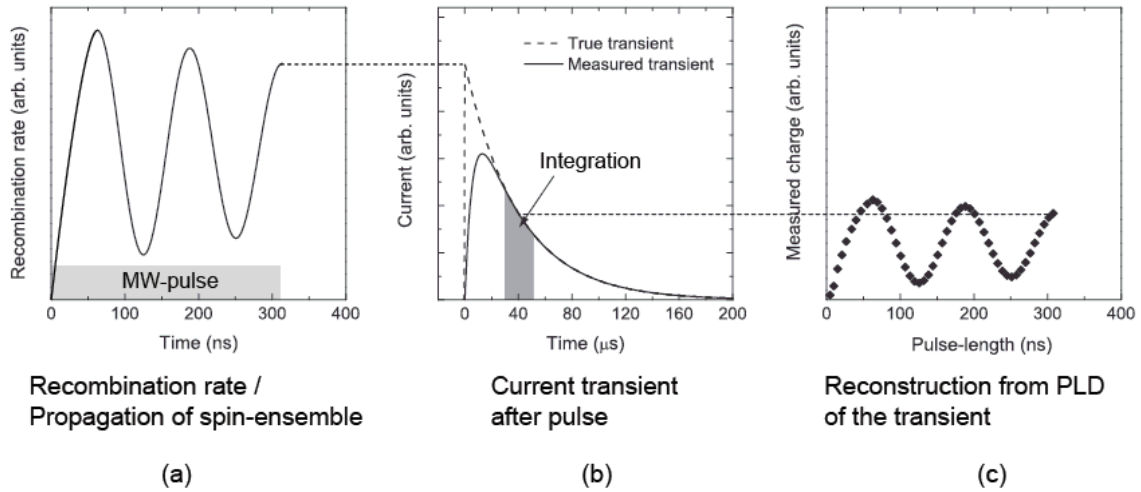


Fig. 8: *Measurement of the recombination rate at the end of the microwave pulse exciting the electron spin resonance of ^{31}P-P_{b0} pairs.*

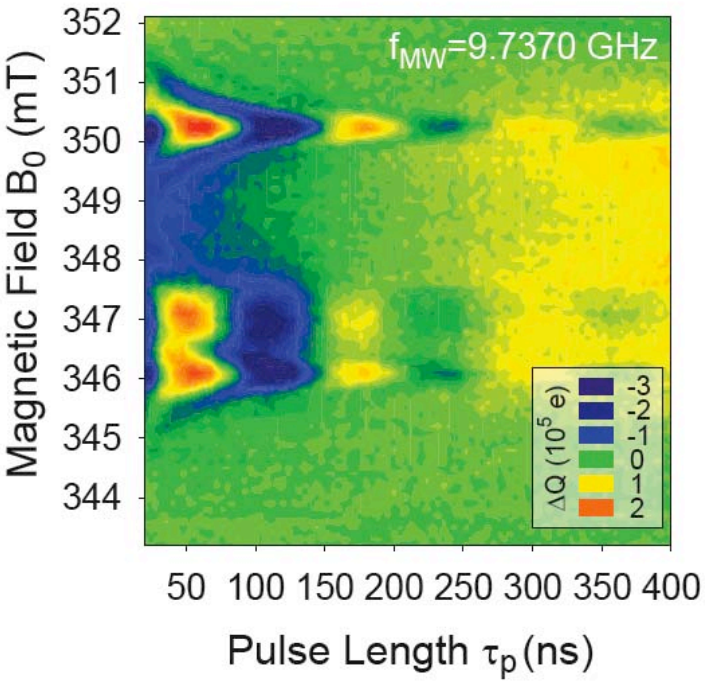


Fig. 9: *Rabi oscillations of the ^{31}P-P_{b0} recombination at the Si/SiO_2-interface observed by electrically detected magnetic resonance (EDMR).*

The result of such pulsed electrically detected magnetic resonance (EDMR) experiments are shown in Fig. 9 as a function of the length of the microwave pulse τ_p for a sample containing about 10^{10} P donors in a 15 nm thin Si layer capped with a native SiO_2, corresponding to a phosphorous concentration of about 10^{17} cm^{-3}. At three different magnetic fields, clear oscillations of *Q* are found [32]. The resonances at 346.1 and 350.3 mT are the hyperfine split signature of the ^{31}P donor. The difference in the magnetic fields of the two resonance positions is $A/\mu_B g$ as shown in Fig. 1. The resonance at 347 mT is caused by the P_{b0} centers. Final proof that indeed Rabi oscillations induced by magnetic resonance are observed comes from the reduction in the Rabi oscillation frequency $\nu_{nut} = g\mu_B B_1 / h$ upon lowering of the microwave power $P_{\mu w} \propto B_1^2$ used (Fig. 10).

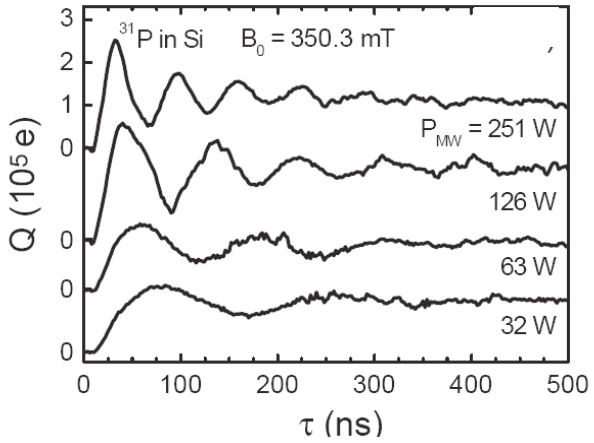


Fig. 10: *Dependence of the Rabi oscillation frequency on the microwave power used.*

5.3 Decoherence

Spin coherence can be studied with a variety of different techniques, including the measurement of the spin resonance lineshape and intensity as a function of microwave power and dynamical variables such as the rate with which the magnetic field is changed. These approaches can be summarized by the term “passage effects”. Alternatively, so-called echo techniques can be used [35]. The coherence time T_2 can e.g. be determined via a Carr-Purcell echo experiment, which consists of a pulse sequence denoted by $\pi/2$-τ_1-π-τ_2, where $\pi/2$ and π denote the rotation angle of the spin system induced by resonant microwave pulses and τ_1 and τ_2 are the free evolution periods between the pulses [36]. Such a Carr-Purcell echo is shown in Fig. 11 a) for an ensemble of identical spins (e.g., the spins of phosphorus donors) plotted in a Bloch sphere, starting, e.g., with the spin ensemble in the down eigenstate m_s=-1/2. The microwave pulses are assumed to rotate the spins around the *x*-axis of the Bloch sphere. The echo develops in the *x*-*y* plane of the Bloch sphere, giving rise to a pulse in the transverse magnetization at $\tau_1=\tau_2$, which is easily detectable in conventional ESR. However, the ^{31}P-P_{b0} spin-to-charge conversion process is sensitive to the singlet-triplet symmetry of a spin pair, which is not changed by the formation of an echo in the transverse magnetization. A successful detection of such echoes via charge transport therefore requires so-called echo tomography [37], where after the second free evolution period τ_2 a final $\pi/2$ pulse rotates the spin system back into singlet or triplet eigenstates of the pair, shown in Fig. 11 c) in more detail. For $\tau_2<\tau_1$ and $\tau_2>\tau_1$, no echo has developed in the *x*-*y* plane so that after the final $\pi/2$ pulse, the spins of the ensemble point to all directions in the *x*-*z* plane of the Bloch sphere.

Both triplet and singlet configurations will therefore be found in ensembles of the ^{31}P-P_{b0} spin pairs under these conditions. However, for $\tau_2=\tau_1$, an echo has developed, so that after the final $\pi/2$ pulse, the ^{31}P spin ensemble is in the original m_s=-1/2 down eigenstate again. If the P_{b0} partner in the spin state is in its m_s=-1/2 state, we find only the triplet configuration for the ^{31}P-P_{b0} spin pairs. Therefore, echoes can be formed also in the singlet-triplet symmetry of spin pairs and are accessible to purely electrical detection.

We can predict the experimental signature of the echoes. The pulse sequence $\pi/2$-τ_1-π-τ_2-$\pi/2$ contains microwave pulses with a total length of 2π. Ideally, we therefore expect a value of Q after a Carr-Purcell echo sequence with $\tau_2=\tau_1$ equal to the Q found after a rotation by 2π in a Rabi-flop experiment shown in Fig. 11 b). For $\tau_2<\tau_1$ and $\tau_2>\tau_1$, the spin-pairs are not in the steady-state configurations. Rather, the ensemble of spin-pairs will contain singlet and triplet configuration in about equal contributions and therefore a larger Q corresponding to Rabi oscillations by $3\pi/2$ and $5\pi/2$ is expected as the result of such Carr-Purcell sequences. A quantitative comparison of Q observed on the high-field ^{31}P resonance at 350.3 mT during the echo in Fig. 11 d) to the Q observed in the Rabi oscillation in Fig. 11 b) demonstrates that the echo amplitude ΔQ is indeed as large as expected.

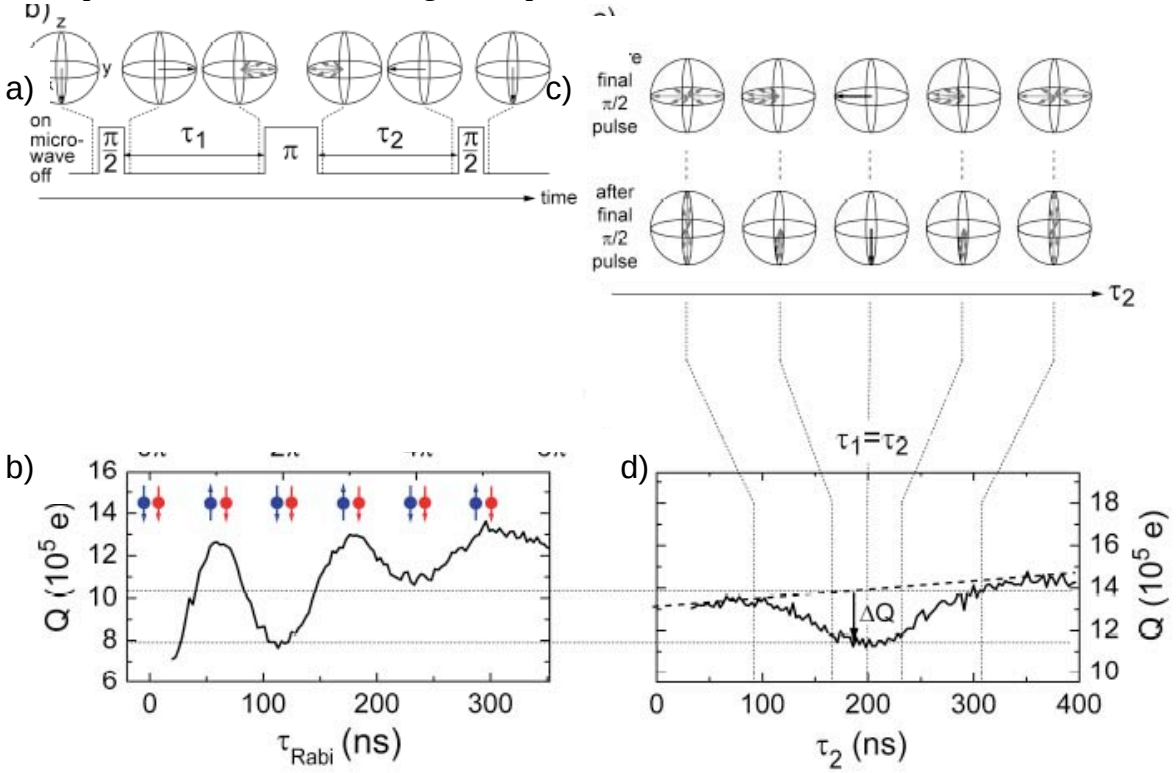


Fig. 11: *Application of echo tomography to the Carr-Purcell method to measure the coherence of ^{31}P-P_{b0} spin pairs via charge transport. b) and d) show that the echo amplitude ΔQ can be understood quantitatively from the amplitude of the Rabi oscillations.*

To determine the echo decay time and therefore the effective coherence time, the echo sequence is measured as a function of τ_1 and τ_2. In all cases, the echo is observed at $\tau_2=\tau_1$ and

its intensity decreases monoexponentially for longer values of τ with a characteristic time constant of about 1.7 μs both for echos detected on the ^{31}P and the P_{b0} resonance [37]. We have already mentioned the much longer values for T_2 observed in bulk crystals of isotopically pure ^{28}Si with low ^{31}P concentrations [8]. In bulk ^{nat}Si with a donor concentration of 10^{17} cm^{-3} as in our samples, the T_2 determined by magnetization echo experiments is already reduced to about 10 μs [38]. The presence of the Si/SiO_2-interface is also expected to lead to a reduction of the coherence [39], as has been shown experimentally for implanted Sb donors [40]. However, EDMR experiments using pulses at different microwave frequencies to induce spin flips of both partners in the ^{31}P-P_{b0} pair are able to measure the singlet recombination time. A quantitative comparison of the singlet recombination time and the effective echo decay time shows that the coherence of the ^{31}P-P_{b0} is limited by the lifetime of the spin pair, rather than by spin-spin scattering. The observation of spin echos in the charge transport opens the possibility to apply pulse sequences such as DEER and ESEEM including free evolution times to study spin-spin interactions in these devices, allowing the determination of the coupling between the electron spins at ^{31}P and P_{b0} or between the donor electron spin and the nuclear spins of ^{29}Si, respectively, and ultimately the realization of entanglement between these spins.

5.4 Outlook

The results discussed above show that, at least for ensembles, the read-out of the electron spin state via ^{31}P-P_{b0} pairs is feasible. This observation opens up a wealth of opportunities: Using the A gates in Fig. 2, it can be envisaged that by changing the gate voltages the coupling between the donor wave function and the read-out spin at the Si/SiO_2-interface can be varied, which would allow the selective addressing and reading of the single ^{31}P spins [41]. Electrical detection of spin resonance, but not yet actual spin read-out has already been achieved on samples containing as few as 50 P donors [42]. Furthermore, several different approaches for the electrical read-out of the nuclear spin state are being discussed. Irrespective of the possible use of this particular read-out scheme or even the use of donors for quantum information processing, these studies allow a more detailed understanding of the complex charge carrier and spin dynamics in semiconductor nanostructures.

Acknowledgements

I would like thank my current and former students H. Huebl, A. Stegner, F. Hoehne, J. Lu, B. Grolik, B. Galler, L. Dreher and C. Bihler for discussions and their dedicated work in the lab. In fact, most should be coauthors of this contribution. This holds especially for T. Graf, who originally wrote Sec. 2 for his PhD thesis [43]. However, all errors are mine. The initial investigations on the read-out of donor electron spin states were performed in collaboration with K. Lips and C. Böhme. Discussions with S. Angus, R. Clark, A. Dzurak, A. Ferguson, W. Hutchison, K. Itoh, D. Jamieson, D. McCamey, S. McKibbin, S. Rogge, T. Schenkel and M. Simmons are gratefully acknowledged. The research into donor spin physics at the Schottky Institut is funded by Deutsche Forschungsgemeinschaft through SFB 631.

References

[1] D. P. DiVincenzo, Fortschr. Phys. **48**, 771 (2000)
[2] L. M. K. Vandersypen and I. L. Chuang, Rev. Mod. Phys. **76**, 1037 (2004)
[3] A modern introduction into magnetic resonance is e.g. M. H. Levitt, *Spin Dynamics* (Wiley, Chichester, 2001)
[4] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp, L. M. K. Vandersypen and L. P. Kouwenhoven, Nature **430**, 431 (2004)
[5] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson and A. C. Gossard, Science **309**, 2180 (2005)
[6] M. Kroutvar, Y. Ducommun, D. Heiss, M. Bichler, D. Schuh, G. Abstreiter and J. J. Finley, Nature **432**, 81 (2004)
[7] G. Feher and E. A. Gere, Phys. Rev. **114**, 1245 (1959)
[8] A. M. Tyryshkin, J. J. L. Morton, S. C. Benjamin, A. Ardavan, G. A. D. Briggs, J. W. Ager and S. A. Lyon, J. Phys. Condens. Matter **18**, S783 (2006)
[9] J. J. L. Morton, A. M. Tyryshkin, R. M. Brown, S. Shankar, B. W. Lovett, A. Ardavan, T. Schenkel, E. E. Haller, J. W. Ager and S. A. Lyon, Nature **455**, 1085 (2008)
[10] E. Abe, K. M. Itoh, J. Isoya and S. Yamasaki, Phys. Rev. B **70**, 033204 (2004)
[11] N. Shaji, C. B. Simmons, M. Thalakulam, L. J. Klein, H. Qin, H. Luo, D. E. Savage, M. G. Lagally, A. J. Rimberg, R. Joynt, M. Friesen, R. H. Blick, S. N. Coppersmith, M. A. Eriksson, Nature Phys. **4**, 540 (2008)
[12] L. Childress, M. V. Gurudev Dutt, J. M. Taylor, A. S. Zibrov, F. Jelezko, J. Wrachtrup, P. R. Hemmer and M. D. Lukin, Science **314**, 281 (2006)
[13] http://lukin.physics.harvard.edu/Theses/LilyThesis.pdf
[14] J. A. Weil, J. R. Bolton and J. E. Wertz, *Electron paramagnetic resonance* (Wiley, New York, 1994)
[15] C. Rudowicz, Magn. Reson. Rev. **13**, 1 (1987)
[16] J. J. Sakurei, *Modern quantum mechanics* (Addison-Wesley, New York, 1994)
[17] J. W. Orton, *Electron paramagnetic resonance* (Iliffe Books, London, 1968)
[18] B. E. Kane, Nature **393**, 133 (1998)
[19] B. E. Kane, Fortschr. Phys. **48**, 1023 (2000)
[20] F. Feher, Phys. Rev. **114**, 1219 (1959)
[21] H. Huebl, A. R. Stegner, M. Stutzmann, M. S. Brandt, G. Vogg, F. Bensch, E. Rauls, and U. Gerstmann, Phys. Rev. Lett. **97**, 166402 (2006)
[22] F. R. Bradbury, A. M. Tyryshkin, G. Sabouret, J. Bokor, T. Schenkel and S. A. Lyon, Phys. Rev. Lett. **97**, 176404 (2006)
[23] R. Vrijen, E. Yablonovitch, K. Wang, H. W. Jiang, A. Balandin, V. Roychowdhury, T. Mor and D. DiVincenzo, Phys. Rev. A **62**, 012306 (2000)
[24] D. N. Jamieson, C. Yang, T. Hopf, S. M. Hearne, C. I. Pakes, S. Prawer, M. Mitic, E. Gauja, S. E. Andresen, F. E. Hudson, A. S. Dzurak and R. G. Clark, Appl. Phys. Lett. **86**, 202101 (2005)
[25] A. Persaud, F. I. Allen, F. Gicquel, S. J. Park, J. A. Liddle, T. Schenkel, Tzv. Ivanov, K. Ivanova, I. W. Rangelow and J. Bokor, J. Vac. Sci. Technol. B **22**, 2992 (2004)
[26] S. R. Schofield, N. J. Curson, M. Y. Simmons, F. J. Rueß, T. Hallam, L. Oberbeck, and R. G. Clark, Phys. Rev. Lett. **91**, 136104 (2003)
[27] K. E. J. Goh, L. Oberbeck, M. Y. Simmons, A. R. Hamilton and R. G. Clark, Appl. Phys. Lett. **85**, 4953 (2004)
[28] T. Hallam, T. C. G. Reusch, L. Oberbeck, N. J. Curson and M. Y. Simmons, J. Appl.

Phys. **101**, 034305 (2007)
[29] J. Berezovsky, M. H. Mikkelsen, O. Gywat, N. G. Stoltz, L. A. Coldren and D. D. Awschalom, Science **314**, 1916 (2006)
[30] F. Jelezko, T. Gaebel, I. Popa, A. Gruber and J. Wrachtrup, Phys. Rev. Lett. **92**, 076401 (2004)
[31] S. J. Angus, A. J. Ferguson, A. S. Dzurak and R. G. Clark, Appl. Phys. Lett. **92**, 112103 (2008)
[32] A. R. Stegner, C. Boehme, H. Huebl, M. Stutzmann, K. Lips and M. S. Brandt, Nature Phys. **2**, 835 (2006)
[33] E. H. Poindexter, P. J. Caplan, B. E. Deal and R. R. Razouk, J. Appl. Phys. **52**, 879 (1981)
[34] C. Boehme and K. Lips, Phys. Rev. B **28**, 245105 (2003)
[35] A good introduction into pulse sequences can be found in R. Freeman, *Spin Choreography* (Oxford Univ. Press, Oxford, 1998)
[36] H.Y. Carr and E. M. Purcell, Phys. Rev. **94**, 630 (1954)
[37] H. Huebl, F. Hoehne, B. Grolik, A. R. Stegner, M. Stutzmann and M. S. Brandt, Phys. Rev. Lett. **100**, 177602 (2008)
[38] M. Chiba and A. Hirai, J. Phys. Soc. Jpn. **33**, 730 (1972)
[39] R. de Sousa, Phys. Rev. B **76**, 245306 (2007)
[40] T. Schenkel, J. A. Liddle, A. Persaud, A. M. Tyryshkin, S. A. Lyon, R. de Sousa and K. B. Whaley, Appl. Phys. Lett. **88**, 112101 (2006)
[41] A. R. Stegner, C. Boehme, H. Huebl, M. Stutzmann, K. Lips and M. S. Brandt, arXiv:quant-ph/0607178 (2006)
[42] D. R. McCamey, H. Huebl, M. S. Brandt, W. D. Hutchison, J. C. McCallum, R. G. Clark, and A. R. Hamilton, Appl. Phys. Lett. **89**, 182115 (2006)
[43] http://mediatum2.ub.tum.de/download/602968/602968.pdf

E 4 Quantum algorithms

M. Richter
Institute for Advanced Simulation
Forschungszentrum Jülich GmbH

Contents

1 Introduction **2**
1.1 Why quantum computation? 2
1.2 Complexity classes 2

2 Deutsch's algorithm: constant or balanced function? **3**

3 Grover's algorithm: how to seek a needle in a haystack? **4**
3.1 Geometric visualization 5

4 Shor's algorithm: factoring of numbers **7**
4.1 An illustrative example 7
4.2 The factorization algorithm 9
4.3 *Step 2:* Period finding with a quantum computer 10

1 Introduction

These lecture notes cover the topic quantum algorithms on an introductory level. We assume that the reader is familiar with the basic concepts of quantum mechanics (see for instance [8]). Instead of getting lost into details we tried to present the underlying ideas in a comprehensible manner. The only exception is Grover's algorithm which has been described more thorough in order to achieve a deeper understanding of the workflow of a typical quantum algorithm. Suggestions for further readings are given in the references.
Throughout the concept of these lecture notes especially [4, 5] have been heavily used.

1.1 Why quantum computation?

According to Moore's law [1] the number of transistors of an integrated circuit, with respect to minimum component costs, doubles approximately every 18 – 24 months. Up to now, this exponential growth has not saturated and by simply extrapolating this behavior the space for storing a single bit of information will scale down to the atomic size around 2020. At that point, quantum effects will become unavoidably dominant and instead of pushing the silicon-based transistor to its physical limits it might be more reasonable to exploit the *principles of quantum mechanics* in an intrinsic way.
The power of quantum computation is due to typical quantum phenomena, such as the *superposition* of quantum states and *entanglement*. There is an inherent quantum parallelism associated with the superposition principle. In simple terms, a quantum computer can process a large number of classical inputs in a single run. On the other hand, this would lead to a large number of possible outputs. It is the task of quantum algorithms to amplify the desired output by interference of all states. To be useful, quantum computers require the development of appropriate quantum algorithms. In specific cases — like the factoring of numbers (Shor's algorithm) or the searching of an unstructured data base (Grover's algorithm) — a remarkable speedup in comparison with classical solutions can be achieved. We discuss the most prominent examples of such algorithms in the following chapters.

1.2 Complexity classes

Complexity theory addresses the question how difficult it is to solve a given mathematical problem. Problems are classified according to the increase of time, a classical computer would need to solve the task, in dependence of the "size" of the task. For example, the size can be the number of bits that define the problem.
We say that a problem belongs to the computational class **P** if it can be *solved* in polynomial time, i.e. in a number of steps that is polynomial in the input size. Instead, the computational class **NP** is the class of problems whose solution can be *verified* in polynomial time. It is clear that **P** is a subset of **NP**. Here we find all the problems whose solution can be easily verified and that are also easy to solve. It is an open problem whether $\mathbf{P} \neq \mathbf{NP}$. If this were the case, there would be problems hard to solve but whose solution could be easily checked. For instance, the integer factoring problem belongs to the class **NP**, since it is easy to check if a number m is a prime factor of an integer N, but no algorithm is known that allows efficiently to compute the prime factors of N on a classical computer. Therefore, it has been conjectured, though not proven, that the integer-factoring problem does not belong to the class **P**.

2 Deutsch's algorithm: constant or balanced function?

Deutsch's problem illustrates the computational power of quantum *interference*. We consider a black box called the *oracle* which evaluates the function $f : \{0,1\} \to \{0,1\}$. There are four of such functions which are listed in the following table:

x	f_0	f_1	f_2	f_3
0	0	0	1	1
1	0	1	0	1

Those functions can be classified according to a *global* property: two of them are *constant* (f_0 and f_3) and two *balanced* (f_1 and f_2). *The problem is to decide whether a given function is constant or balanced*. On a classical computer this task requires two queries of the oracle. A quantum computer can solve the same problem with only one oracle query:
We need one ancillary qubit $|y\rangle$. On a quantum level the oracle corresponds to a unitary transformation U_f

$$U_f \, |x\rangle|y\rangle = |x\rangle \, |y \oplus f(x)\rangle \tag{1}$$

where $\oplus$ denotes addition modulo 2. That is, the second qubit is flipped if and only if $f(x) = 1$. In the case that the second qubit is in a superposition we obtain

$$\begin{aligned} \tfrac{1}{\sqrt{2}} \, |x\rangle \, (|0\rangle - |1\rangle) \quad &\xrightarrow{U_f} \quad \tfrac{1}{\sqrt{2}} \, |x\rangle \, (|0 \oplus f(x)\rangle - |1 \oplus f(x)\rangle) \\ &= \quad \begin{cases} \tfrac{1}{\sqrt{2}} \, |x\rangle \, (|0\rangle - |1\rangle) & \text{if } \; f(x) = 0 \\ \tfrac{1}{\sqrt{2}} \, |x\rangle \, (|1\rangle - |0\rangle) & \text{if } \; f(x) = 1 \end{cases} \\ &= \quad (-1)^{f(x)} \tfrac{1}{\sqrt{2}} \, |x\rangle \, (|0\rangle - |1\rangle) \; . \end{aligned} \tag{2}$$

Both qubits remain in their primary state with $(-1)^{f(x)}$ acting as a global phase factor. For a superposition of both qubits we obtain

$$\begin{aligned} \tfrac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle) \quad &\xrightarrow{U_f} \quad \tfrac{1}{2} \left((-1)^{f(0)} \, |0\rangle + (-1)^{f(1)} \, |1\rangle \right) (|0\rangle - |1\rangle) \\ &= \quad \tfrac{(-1)^{f(0)}}{2} \left(|0\rangle + (-1)^{f(0) \oplus f(1)} \, |1\rangle \right) (|0\rangle - |1\rangle) \; . \end{aligned} \tag{3}$$

The relevant information is now coded in the relative phase of the superposed states of the first register. By applying a Hadamard operation we get

$$\tfrac{(-1)^{f(0)}}{2} \left(|0\rangle + (-1)^{f(0) \oplus f(1)} \, |1\rangle \right) (|0\rangle - |1\rangle) \xrightarrow{H} (-1)^{f(0)} \, |f(0) \oplus f(1)\rangle |1\rangle \tag{4}$$

which corresponds to $|0\rangle|1\rangle$ if f is constant and $|1\rangle|1\rangle$ in the case that f is balanced. Therefore, a global property of the function $f(x)$ has been encoded in a single qubit after a single call of f. This is because a quantum computer can evaluate both $f(0)$ and $f(1)$ simultaneously. The main point is that these two alternative "paths" are combined by the final Hadamard gate, giving the desired interference pattern. The interference is constructive for the outcome $f(0) \oplus f(1)$ and destructive for the alternative outcome.

3 Grover's algorithm: how to seek a needle in a haystack?

Imagine that you have a telephone number of a person who is living in your town and you want to find out whose number it is by just using your telephone book. The best you can do classically is to go through the names one by one until you find the corresponding number. If the phone book contains N entries, you would have to check $N/2$ numbers in the mean. Fortunately, a *quantum search* can do better: in 1996 Lov Grover could show [10] that only $\sqrt{N}$ queries are needed. Furthermore, Grover's algorithm is known to be *optimal* [11], i.e. no classical or quantum algorithm can solve the problem (of searching an unstructured database) faster. Since the quantum algorithm does not lie in a different complexity class than the best classical one[1], the speedup is still quadratic.

The underlying idea of Grover's algorithm is to start with a superposition of *all* states and amplify the amplitude of the state searched for step by step by repeatedly applying a certain sequence $\mathcal{G}$ of operations. After a fixed number k of iterations the amplitude of this state has gained a value close to 1, which means a measurement yields the searched element with a high probability.

For simplicity[2] let us assume that we have an unstructured database that contains $N = 2^n$ different elements. We label the items as $\{0, 1, \ldots, N-1\}$ and x_0 is the element searched for. The result of the quantum search process is stored in a register $|x\rangle$ which yields the index x_0 with a high probability when measured. In addition a single ancillary qubit $|y\rangle$ is needed to store the result of the oracle query. The oracle $\mathcal{O}$ computes the n-bit binary function

$$f \,:\, \{0,1\}^n \to \{0,1\}\,, \tag{5}$$

defined as

$$f(x) = \begin{cases} 1 & \text{if } x = x_0\,, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Grover's algorithm in detail:

- Start with the state
$$|x\rangle|y\rangle = |00\ldots0\rangle|1\rangle \ . \tag{7}$$
- Apply $H^{\otimes n+1}$ ($n+1$ Hadamard gates) $\Longrightarrow$ equal superposition of all basis states:
$$|00\ldots0\rangle|1\rangle \to \tfrac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle \, \tfrac{1}{\sqrt{2}} \left(|0\rangle - |1\rangle\right) \tag{8}$$
- Evaluate the oracle function[3] $|x\rangle|y\rangle \xrightarrow{\mathcal{O}} |x\rangle|y \oplus f(x)\rangle$.
This flips the sign of the $|x_0\rangle$ amplitude:
$$\tfrac{1}{\sqrt{2^{n+1}}} \sum_{x=0}^{2^n-1} (-1)^{f(x)} \, |x\rangle \left(|0\rangle - |1\rangle\right) \tag{9}$$

[1] Both grow polynomial in time with the number of database elements.

[2] Grover's algorithm does also work with partially identical elements, but this issue is a little bit more complicated (see e.g. [2]). In the case $2^{n-1} < N < 2^n$ the database can be filled up with distinguishable items so that $\tilde{N} = 2^n$ holds.

[3] $\oplus$ means addition modulo 2 which corresponds to a XOR operation.

- Let us *define* the Grover iteration $\mathcal{G}$, with $\mathcal{G} = \mathcal{I}_M \mathcal{O}$, where $\mathcal{O}$ denotes the oracle query and

$$\mathcal{I}_M = -H^{\otimes n}(\mathbf{1} - 2\,|0\rangle\langle 0|)H^{\otimes n} = -(\underbrace{\mathbf{1} - 2\,|S\rangle\langle S|}_{\mathcal{R}_{|S\rangle}}) \tag{10}$$

  is often referred to as the "inversion about the mean[4]". The uniform superposition $|S\rangle$ is given by $|S\rangle \equiv H^{\otimes n}\,|0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle$ and the operator[5] $\mathcal{R}_{|S\rangle}$ mirrors a given state at the hyperplane perpendicular to $|S\rangle$ (cf. section 3.1).

- Now apply $\mathcal{I}_M$ (remember: we have already used $\mathcal{O}$, so one Grover iteration $\mathcal{G}$ is now completed).

- (For $N > 4$):[6] Apply $\mathcal{G}$ several times until a measurement of $|x\rangle$ gives x_0 with maximal probability[7].

- Perform a measurement of the first register in the computational basis, giving outcome $x = \bar{x}$. If $f(\bar{x}) = 1$ the search was successful, otherwise repeat the algorithm.

3.1 Geometric visualization

The underlying idea of Grover's algorithm gets much clearer in terms of a geometric interpretation. Since this visualization goes back to the fact that geometrically a reflection operator induces a mirroring, let us start with a short warm-up.

Preliminary consideration: reflection operator and mirroring Consider a bidimensional space spanned by the vectors $\{|x_0\rangle, |x_0^\perp\rangle\}$ and a generic vector $|\psi\rangle = \alpha\,|x_0\rangle + \beta\,|x_0^\perp\rangle$. The action of the reflection operator $\mathcal{R}_{|x_0\rangle} = \mathbf{1} - 2\,|x_0\rangle\langle x_0|$ on $|\psi\rangle$ is $\mathcal{R}_{|x_0\rangle}\,|\psi\rangle = -\alpha\,|x_0\rangle + \beta\,|x_0^\perp\rangle$. Therefore, $\mathcal{R}_{|x_0\rangle}$ changes the sign of the $|x_0\rangle$ amplitude. Geometrically this corresponds to a mirroring at the axis $|x_0^\perp\rangle$, that is, at the hyperplane perpendicular to $|x_0\rangle$ (see Fig. 1).

Next, let us prove that $-\mathcal{R}_{|S\rangle} = \mathcal{R}_{|S^\perp\rangle}$ holds. We consider a generic vector $|u\rangle = \mu\,|S\rangle + \nu\,|S^\perp\rangle$. Application of $\mathcal{R}_{|S\rangle}$ yields $-\mu\,|S\rangle + \nu\,|S^\perp\rangle$ while $\mathcal{R}_{|S^\perp\rangle}\,|u\rangle = \mu\,|S\rangle - \nu\,|S^\perp\rangle = -\mathcal{R}_{|S\rangle}\,|u\rangle$. As a result, the "inversion about the mean" operator $\mathcal{I}_M$ in eqn. (10) can be written as

$$\mathcal{I}_M = \mathcal{R}_{|S^\perp\rangle}\,. \tag{11}$$

Grover's algorithm begins with the uniform superposition state (8)

[4]The operation $\mathcal{I}_M$ applied to a general state $\sum_x \alpha_x\,|x\rangle$ yields $\sum_x [-\alpha_x + 2\,\langle\alpha\rangle]\,|x\rangle$, where $\langle\alpha\rangle \equiv \sum_x \alpha_x/N$ is the mean value of the amplitudes α_x.

[5]A projection Operator $\mathcal{P} = |a\rangle\langle a|$ satisfies $\mathcal{P}^2 = \mathcal{P}$. Since $(\mathbf{1} - \mathcal{P})^2 = \mathbf{1} - \mathcal{P}$, the right side of this equation is also a projection operator. In contrast, $\mathbf{1} - 2\mathcal{P}$ is a *reflection* operator, $(\mathbf{1} - 2\mathcal{P})^2 = \mathbf{1}$, which changes the sign of the projection onto $|a\rangle$.

[6]In the case $N = 4$ we are already done. This means that quantum mechanically we can search an unsorted data base containing 4 different elements with a single query. Classically we can handle only 2 different items with one question (e.g. by asking: "Is this the element that I want?"). As a remarkable coincidence nature uses also *four* different nucleotide bases to code the genetic information in DNA. This gave rise to the speculation that quantum search processes might be involved on a genetic level [12].

[7]The exact number of iterations k will be derived in section 3.1.

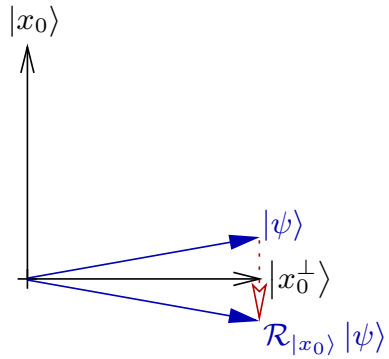


Fig. 1: *The reflection operator $\mathcal{R}_{|x_0\rangle}$ flips the sign of the $|x_0\rangle$ amplitude: $\mathcal{R}_{|x_0\rangle}(\alpha\,|x_0\rangle + \beta\,|x_0^\perp\rangle) = -\alpha\,|x_0\rangle + \beta\,|x_0^\perp\rangle$. This mirrors $|\psi\rangle$ at the $|x_0^\perp\rangle$ axis.*

$$|\psi_0\rangle \equiv |S\rangle = \tfrac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle \tag{12}$$

(for simplicity, we drop the second register whose value will not change during the rest of the algorithm). Since the plane spanned by $\{|S\rangle\,, |x_0\rangle\}$ can also be generated by[8] $\{|x_0\rangle\,, |x_0^\perp\rangle\}$, we have

$$|\psi_0\rangle \equiv |S\rangle = \sin\theta\,|x_0\rangle + \cos\theta\,|x_0^\perp\rangle \tag{13}$$

where θ denotes the angle between the vectors $|x_0^\perp\rangle$ and $|S\rangle$. As illustrated in Fig. 2, the oracle $\mathcal{O}$ mirrors $|\psi\rangle$ at $|x_0^\perp\rangle$ and afterwards the result $\mathcal{O}\,|\psi\rangle$ is mirrored at $|S\rangle$ by virtue of $\mathcal{I}_M$. Since both mirrorings take place in the same plane the result of the whole operation is a rotation. Fig. 2 demonstrates that the Grover iteration $\mathcal{G}$ rotates a generic vector $|\psi\rangle$ by an angle of 2θ towards the searched element $|x_0\rangle$. After j steps of Grover's iteration the n-qubit state is given by

$$|\psi_j\rangle \equiv \mathcal{G}^j\,|\psi_0\rangle = \sin((2j+1)\theta)\,|x_0\rangle + \cos((2j+1)\theta)\,|x_0^\perp\rangle \tag{14}$$

since all rotations take place in the primary plane. The process must stop after k steps, where k is such that $|\psi_k\rangle$ is very close to the marked state $|x_0\rangle$. This is the case when $\sin((2k+1)\theta) \approx 1$. The smallest integer k that fulfills this condition is determined by

$$(2k+1)\theta \approx \frac{\pi}{2}\,, \tag{15}$$

which implies

$$k = \text{round}\left(\frac{\pi}{4\theta} - \frac{1}{2}\right)\,, \tag{16}$$

where round specifies the nearest integer. Since we started from the uniform superposition state (12) we have

$$\sin\theta = \langle x_0|\psi_0\rangle = \frac{1}{\sqrt{N}}\,, \tag{17}$$

which leads to the following relation between the number of Grover iterations k and the volume of the database N:

$$\boxed{k = \text{round}\left(\frac{\pi}{4\arcsin(1/\sqrt{N})} - \frac{1}{2}\right)} \tag{18}$$

[8] In fact, this condition defines $|x_0^\perp\rangle$.

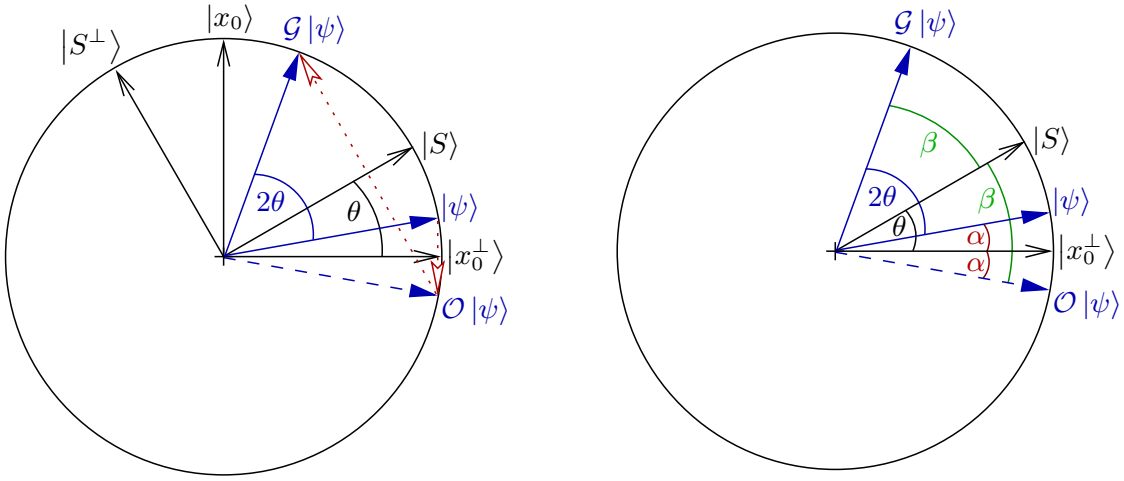


Fig. 2: Left figure: *Geometric visualization of the Grover iteration $\mathcal{G}$. The oracle query $\mathcal{O}$ mirrors $|\psi\rangle$ at $|x_0^\perp\rangle$ (which is achieved by $\mathcal{R}_{|x_0\rangle}$). The "inversion about the mean" $\mathcal{I}_M$ mirrors $\mathcal{O}|\psi\rangle$ at the uniform superposition $|S\rangle$, by virtue of $\mathcal{R}_{|S^\perp\rangle}$ or the application of $-\mathcal{R}_{|S\rangle}$ (cf. eqn. (10)).* Right panel: *Since $\theta = \beta - \alpha$, we have $2\theta = \beta + \theta - \alpha$. Thus, $\sphericalangle(|\psi\rangle, \mathcal{G}|\psi\rangle) = 2\theta$ so that the Grover iteration $\mathcal{G}$ rotates $|\psi\rangle$ by 2θ towards $|x_0\rangle$.*

For large N the approximative behavior $\arcsin(1/\sqrt{N}) \approx 1/\sqrt{N}$ holds and we can demonstrate the quadratic speedup of Grover's algorithm

$$k = \text{round}\left(\frac{\pi}{4}\sqrt{N} - \frac{1}{2}\right) = \mathcal{O}\left(\sqrt{N}\right) \tag{19}$$

in contrast to classical algorithms in which the number of database queries grows like $\mathcal{O}(N)$.

4 Shor's algorithm: factoring of numbers

In 1994 Peter Shor — employed by the US company AT&T Labs Research — published a quantum algorithm which allows to compute the prime factors of a given number with an exponential speed up [13]. The time needed by a classical computer to solve this problem grows exponentially with the number of digits. For instance, to factorize a number with 130 digits about 10^{18} operations are needed. If we assume 10^{12} floating point operations per second (1 Tflops) this would take 42 days. In the case we would double the number of digits to 260 the calculation would consist of about 10^{25} operations which would last 1 million years. In practice it is not possible to solve the task. On the other hand, the verification of the result is trivial, we just have to multiply the numbers. This asymmetry is the crucial point that makes to RSA encryption scheme so successful.

4.1 An illustrative example

Let us examine an illustrative example: 15 is the product of the prime numbers 3 and 5. In this case it is possible to find the factors by trial and error: first we try 2 as a factor which fails. Next we try 3 and succeed. If we denote the number we want to factorize by N, it takes $\sqrt{N}$ trials to

find the factors in the worst case. According to the binary representation of the number, which we assume to have L digits, this corresponds to $2^{L/2}$ attempts which means that the time for this simple algorithm grows exponentially with the number of digits. The best algorithm known grows like $e^{L/3}$. In contrast, Shor's algorithm grows polynomial in time like L^3.
It has been known that the factorization of a number is connected to the period of a certain function of that number. Unfortunately, the period finding of of a given function is also exponentially time consuming. However, quantum mechanically this can be done in polynomial time.
In the following we will focus on the underlying idea of Shor's algorithm. For further details see eg. [2]. The factorization of N is equivalent to the period finding of

$$f(n) = y^n \bmod N \ , \tag{20}$$

where y is an arbitrary fixed number $y < N$ which does not divide N and $n \in \mathbf{N}$. Let's reconsider the example $N = 15$. We choose $y = 7$ which yields $f(n) = 7^n \bmod 15$. The following table shows $f(n)$ in dependence on n:

n	1	2	3	4	5	6	...
$f(n)$	7	4	13	1	7	4	...

In this case we have the period $r = 4$. Once the period of $f(n)$ has been found, we can calculate factors of $N = p \cdot q$ according to

$$\begin{aligned} p &= \text{gcd}\left(y^{r/2} + 1, N\right) \\ q &= \text{gcd}\left(y^{r/2} - 1, N\right) \ , \end{aligned} \tag{21}$$

where gcd denotes the *greatest common divisor*. In case of our example we get

$$\begin{aligned} p &= \text{gcd}\left(7^{4/2} + 1, N\right) = \text{gcd}\left(50, 15\right) = 5 \\ q &= \text{gcd}\left(7^{4/2} - 1, N\right) = \text{gcd}\left(48, 15\right) = 3 \end{aligned} \tag{22}$$

and thus $15 = 5 \cdot 3$. Finally, we remark that the computation of the gcd is only of polynomial complexity (already 300 B.C. Euclid found an algorithm for this task).
Classically, the period finding of $f(n)$ is as hard as the factoring of N itself. Fortunately, quantum mechanics helps in this case. As in Grover's algorithm we need a second register. We start with the uniform superposition

$$|0\rangle|0\rangle \rightarrow \tfrac{1}{\sqrt{2^L}} \sum_{x=0}^{2^L-1} |x\rangle|0\rangle \ . \tag{23}$$

In the next step the value of the function f is calculated and stored in the second register. Since we started from an uniform superposition state all function values are calculated in a single step

$$\tfrac{1}{\sqrt{2^L}} \sum_{x=0}^{2^L-1} |x\rangle|0\rangle \rightarrow \tfrac{1}{\sqrt{2^L}} \sum_{x=0}^{2^L-1} |x\rangle|f(x)\rangle \ . \tag{24}$$

In a third step we carry out a measurement on the second register. As a result of this measurement the first register will be in a superposition of states with the periodicity searched for. The period can be extracted with a special technique called *fast quantum Fourier transformation*.

In the case $N = 15$ and $a = 7$ we have

$$|0\rangle\,|0\rangle \rightarrow |1\rangle\,|0\rangle + |2\rangle\,|0\rangle + |3\rangle\,|0\rangle + \ldots . \tag{25}$$

due to the initialization process (for simplicity we omit the normalization factor). In the next step the value of the function $f(x)$ is written in the second register

$$\rightarrow |1\rangle\,|7\rangle + |2\rangle\,|4\rangle + |3\rangle\,|13\rangle + |4\rangle\,|1\rangle + |5\rangle\,|7\rangle \ldots . \tag{26}$$

Now we measure the second register. Since all of the amplitudes occur with equal probability we get 7, 4, 13 or 1. Let us assume that the result is 7. In this case the first register corresponds to

$$|\psi\rangle \propto |1\rangle + |5\rangle + |9\rangle + \ldots . \tag{27}$$

Those states differ by 4 which is the period we are looking for. The period can be extracted by the fast quantum Fourier transformation.
The crucial point is that the period manifests itself as a *global* property of the wave function. This global feature can efficiently be read out. This leads to the remarkable exponential speed up in comparison to classical algorithms.

4.2 The factorization algorithm

Let p and q be prime numbers and let $N = pq$. We want to factor N into a product of p and q.

1. Take a positive integer y less than N randomly. Calculate the greatest common divisor $\gcd(y, N)$ by the Euclidean algorithm. If $\gcd(y, N) \neq 1$, we are extremely lucky: y is either p or q, and we are done. Suppose $\gcd(y, N) = 1$.

2. Define $f_N : \mathbf{N} \rightarrow \mathbf{N}$ by $n \mapsto y^n \bmod N$. Find the smallest $r \in \mathbf{N}$, such that $y^r = 1 \bmod N$. The number r is called the *order* or *period*. It is known that this takes exponentially large steps in any classical algorithm, but is takes only polynomial steps in Shor's algorithm. *A quantum computer is required only for this task,* and the rest can be executed on a classical computer.

3. If r is odd, it cannot be used in the following steps. Go back to step 1 and repeat the above steps with different y until an even r is obtained. If r is even, proceed with step 4.

4. Since r is even, it holds that

$$(y^{r/2} - 1)(y^{r/2} + 1) = y^r - 1 = 0 \bmod N. \tag{28}$$

   If $y^{r/2}+1 = 0 \bmod N$, then $\gcd(y^{r/2}-1, N) = 1$; go back to step 1 and try with different y. If $y^{r/2} + 1 \neq 0 \bmod N$, $y^{r/2} - 1$ contains either p or q, and we proceed with step 5. Note that the number $y^{r/2} - 1$ cannot be a multiple of N in the latter case. If this would be the case, it would lead to $y^{r/2} = 1 \bmod N$, which contradicts the assumption that r is the smallest number which satisfies $y^r = 1 \bmod N$.

5. The number

$$d = \gcd(y^{r/2} - 1, N) \tag{29}$$

   is either p or q, and factorization is done.

4.3 *Step 2:* Period finding with a quantum computer

Given N, choose $q = 2^L$ between N^2 and $2N^2$. Choose a random $y < N$ and begin with an L qubit register in the state $|0\rangle$. A Hadamard operation on this register yields

$$\frac{1}{\sqrt{q}} \sum_{a=0}^{q-1} |a\rangle \, .$$

Compute $y^a \bmod N$ and store the result in a second register

$$\frac{1}{\sqrt{q}} \sum_{a=0}^{q-1} |a\rangle \, |y^a \bmod N\rangle \, .$$

This can be done efficiently eg. by repeatedly squaring mod N to get y^{2^i} and multiplying selected ones corresponding to the binary expansion of a. Now perform a measurement on the second register; suppose the result is z

$$z = y^l \bmod N \qquad \text{for some least } l$$

also

$$y^l = y^{jr+l} \bmod N \qquad \forall j$$

Thus, the measurement will select values

$$a = l, l + r, l + 2r, \ldots, l + Ar$$

where A is the greatest integer less than $(q-l)/r$. Notice that $l \leq r$ and $q \sim O(N^2) \Rightarrow A \sim q/r$. After the measurement we find the first register in the state

$$|\phi_l\rangle = \frac{1}{\sqrt{A+1}} \sum_{j=0}^{A} |jr + l\rangle \ .$$

We consider now the simple case in which r *devides* q *exactly*[9] so that $A = q/r - 1$ and

$$|\phi_l\rangle = \sqrt{\frac{r}{q}} \sum_{j=0}^{\frac{q}{r}-1} |jr + l\rangle = \sum_{a=0}^{q-1} f(a) \, |a\rangle$$

with

$$f(a) = \begin{cases} \sqrt{r/q} & (a-l) \text{ is a multiple of } r \\ 0 & \text{otherwise} \end{cases} \ .$$

After application of the quantum Fourier transform we obtain

$$|\phi_l\rangle \mapsto \sum_c \tilde{f}(c) \, |c\rangle$$

[9] The general case is technically more involved but it is based on the same ideas.

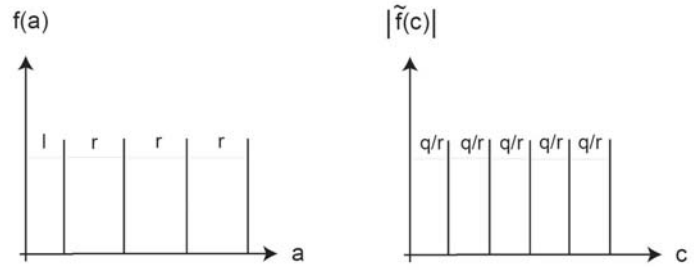


Fig. 3: *The figure shows the amplitudes $f(a)$ and the Fourier transformed ones. Note that the initial shift l disappears due to the translational invariance property of the QFT.*

where $\tilde{f}(c)$ is the discrete Fourier transform of $f(a)$

$$\begin{aligned}
\tilde{f}(c) &= \frac{\sqrt{r}}{q}\sum_{j=0}^{q/r-1}\exp\left(\frac{2\pi i(jr+l)c}{q}\right) \\
&= \frac{\sqrt{r}}{q}\underbrace{\left[\sum_{j=0}^{q/r-1}\exp\left(2\pi i\frac{jrc}{q}\right)\right]}_{=0 \text{ or } q/r \text{ iff } c \text{ is a multiple of } q/r}\exp\left(2\pi i\frac{lc}{q}\right) \\
\tilde{f}(c) &= \begin{cases} \frac{1}{\sqrt{r}}\exp(2\pi ilc/q) & c \text{ is a multiple of } q/r \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

With $c = jq/r$ the final state reads

$$\frac{1}{\sqrt{r}}\sum_{j=0}^{r-1}\exp\left(\frac{2\pi ilj}{r}\right)\left|j\frac{q}{r}\right\rangle \ .$$

Thus a measurement (on the first register) of the states labeled by jq/r will yield a multiple $\lambda q/r$ with $\lambda = 0, \ldots, r-1$ chosen equiprobably. The initial shift l disappears due to the translational invariance property of the QFT. After the measurement we know a value c satisfying $c/q = \lambda/r$ (q is also known). If $\gcd(\lambda, r) = 1$, we can determine r by canceling c/q down to an irreducible fraction. Since λ is chosen at random, it can be shown that

$$\text{Prob}(\gcd(\lambda, r) = 1) \geq \frac{1}{\log r}$$

holds for large r. Thus, repeating the computation $O(\log r) < O(\log N)$ times guarantees a success probability close to one and hence an efficient determination of r.

References

[1] G. Moore, *Cramming more components onto integrated circuits*, Electronics 38 (1965) 114.

[2] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).

[3] *Information, Science, and Technology in a Quantum World*, Los Alamos Science, Number 27 (2002).
`http://www.fas.org/sgp/othergov/doe/lanl/pubs/number27.htm`

[4] G. Benenti, G. Casati and G. Strini, *Principles of Quantum Computation and Information, Volume I: Basic Concepts*, World Scientific, Singapore (2004).

[5] D. Bruß, *Quanteninformationstheorie*, lecture notes WS2004/2005 (in German),
`http://www.thphy.uni-duesseldorf.de/~ls3/QI-Skriptum-050210.pdf`
D. Bruß, *Quanteninformation*, Fischer, Frankfurt (2003) (in German).

[6] A. Steane, *Quantum computing*, Rept. Prog. Phys. 61 (1998) 117, `quant-ph/9708022`.

[7] D. Aharonov, *Quantum computation*, Annual Reviews of Computational Physics, ed. Dietrich Stauffer, World Scientific, vol VI, 1998, `quant-ph/9812037`.

[8] J.J. Sakurai, *Modern Quantum Mechanics*, Addison-Wesley, Reading (1985).

[9] G. Arnold, Th. Lippert, N. Pomplun and M. Richter, *Large Scale Simulation of Ideal Quantum Computers on SMP-Clusters*, to appear in the proceedings of the PARCO05 conference.

[10] L.K. Grover, *A fast quantum mechanical algorithm for database search*, in Proc. of the 28th Annual ACM Symposium on the Theory of Computing, p. 212, ACM Press, New York (1996), `quant-ph/9605043`; L.K. Grover, *Quantum Mechanics helps in searching for a needle in a haystack*, Phys. Rev. Lett. 79 (1997) 325, `quant-ph/9706033`.

[11] C. Zalka, *Grover's quantum search algorithm is optimal*, Phys. Rev. A60 (1999) 2746, `quant-ph/9711070`.

[12] A. Patel, *Quantum Algorithms and the Genetic Code*, Pramana 56 (2001) 365, `quant-ph/0002037`; *Live Force*, New Scientist, April 15th (2000).

[13] P.W. Shor, *Algorithms for quantum computation: discrete logarithms and factoring*, Proceedings of the 35th Annual Symposium on Foundations of Computer Science, IEEE Press, Los Alamitos (1994).

F1 Technology of Hard Disk Drives

A. Berger
CIC nanoGUNE
E – 20018 Donostia – San Sebastian, Spain

Contents

1 Introduction 2
1.1 General Functionality of Hard Disk Drives 2
1.2 Longitudinal vs. Perpendicular Recording 4
1.3 State-of-the-Art Performance Levels 5

2 Magnetic Recording Media 7
2.1 Basic Properties 7
2.2 Soft-Magnetic Underlayer (SUL) for Perpendicular Recording 9
2.3 Thermal Stability 10
2.4 Modern Materials Designs 11

3 Magnetic Write Heads 12
3.1 Basic Functionality 13
3.2 Modern Designs 14
3.3 Technology Challenges 15

4 Magnetic Sensors as Read Heads 15
4.1 Basic Functionality and Requirements 16
4.2 Magneto Resistance Effects 17
4.3 GMR Sensors 18
4.4 CIP vs. CPP Geometry 19
4.5 TMR Sensors 20

References 22

1 Introduction

Hard Disk Drives (HDD) are a key component of today's technology that empowers our information age. They are a core non-volatile memory device technology that finds widespread use in computers, data centers and home electronics, with more than 500 million units produced annually. The actual non-volatile memory inside the HDD is based on the ferromagnetic hysteresis effect and thus, HDDs are a prime example of magnetics technology.

Fig. 1: *HDD and some of its core applications in computers and digital video recorders*

The present material will give a general introduction into HDDs and their magnetic core components, trying to relate fundamental aspects of magnetism to this very important real-world application. In chapter 1, the overall device function and technology is discussed, while chapters 2 – 4 are reserved for a more detailed view at the key magnetic components: Magnetic Recording Media (2), Magnetic Write Heads (3) and Magnetic Sensors as Read Heads (4).

1.1 General Functionality of Hard Disk Drives

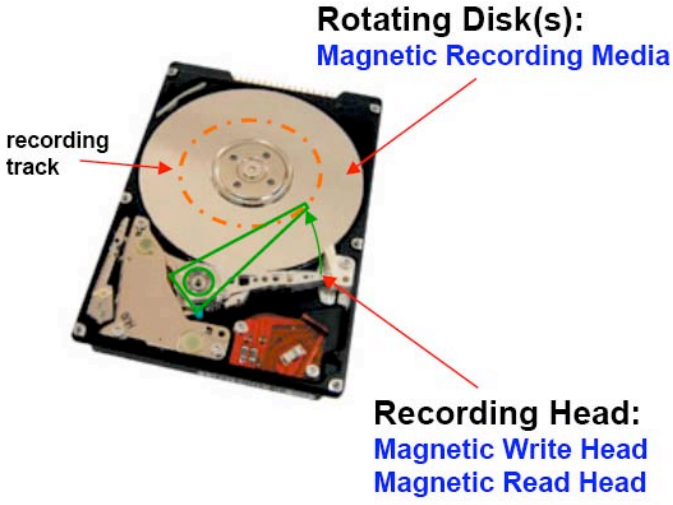


Fig. 2: *Basic HDD elements*

Figure 2 shows a picture of an (opened up) HDD. The largest part of the interior is made up by the disk (or by several of them stacked on top of each other) that gives the HDD its name. The disk itself is made of ultra-smooth glass or metal that is covered by a thin magnetic film, which is the Magnetic Recording Media layer. It is this thin magnetic overcoat to top of the disk surfaces that is the actual magnetic storage device, in which information is stored as a sequence of magnetic bits or domains. As we will see later the Magnetic Recording Media

layer is not a single film out of one magnetic material, but a very specific layer stack or sequence out of several different materials to enable the storage of magnetic bit patterns with the required density and stability.

While the long-term storage of information performed by the disk is the key functionality of HDDs, this can only be achieved if the information can be "written" into the Magnetic Recording Media in the first place and later be "read" back. Therefore, the HDD contains a Magnetic Write Head and a Read Head, which are both located on an arm element that can be moved over the disk as shown in figure 2. To achieve today's storage capacity, the information must be written (and read back) with ultra-high density, i.e. in the form of extremely small magnetic bits. It is the Magnetic Write and Read Heads that define the size of the magnetic bits, and therefore they have to be extremely small, as we will discuss later on.

The arm, on which the Recording Head sits, can move over the disk to address different parts of the disk. However, this movement is a 1-dimensional movement only. So, to make the entire 2-dimensional disk surface accessible for information storage, the disk rotates at typical speeds in between 4,200 to 15,000 revolutions per minute. Only with these rather high rotation speeds can sufficiently high linear speeds be achieved that are necessary to enable high data rates for write-in and read-out of information. The basic HDD operation mode is given by first moving the Recording Head to a fixed radial position and then writing or reading a track of magnetic bits to or from the Magnetic Recording Media layer on the disk, as shown schematically in figure 3.

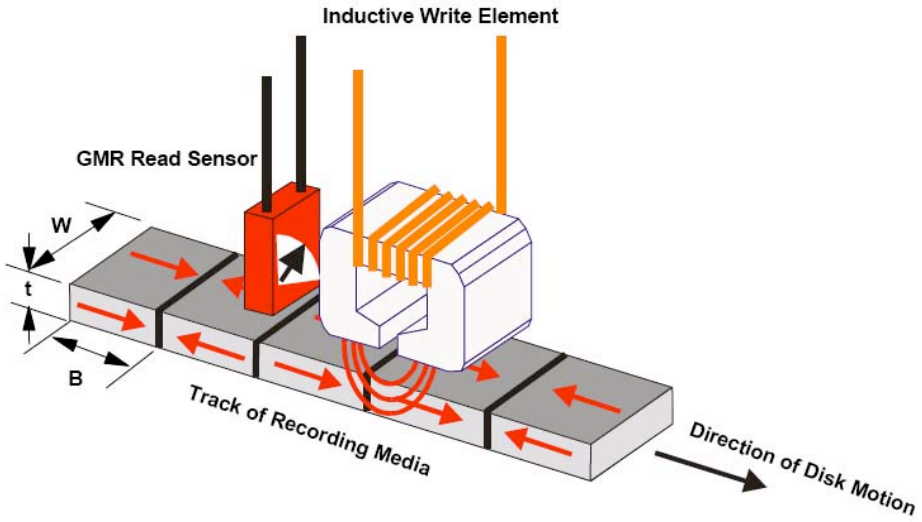


Fig. 3: *Schematic of recorded track*

Given its mechanical movements, the HDD is not a solid-state device in the traditional sense (having no mechanical movements), but rather a hybrid technology that uses solid-state physics and technology, namely ferromagnetism, within a mechanically operated device. This aspect of HDD technology may be best illustrated by the slider, the lithographically fabricated device that contains both the Magnetic Write and the Magnetic Read Head, as shown in Figure 4. In addition to the magnetic components designed onto the backend of the slider (shown on top in figure 4), it also has a lithographically defined Air Bearing Surface on the bottom (shown on the left side in figure 4), which makes it into a tiny (mechanical) airplane that flies at constant height over the disk surface, and is lifted my means of the air flow pressure due to the rotating disk. Present day technology allows products to operate at less than 10 nm flying height.

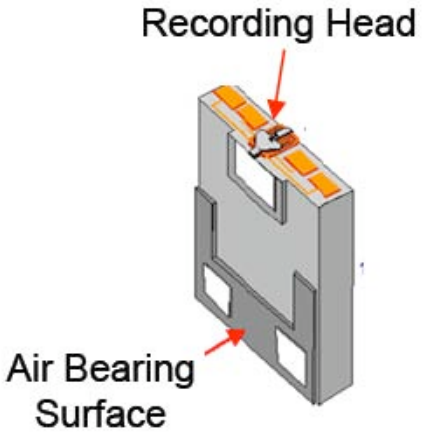


Fig. 4: *Recording Head Slider*

As mentioned already above, a HDD drive can contain more than one disk. Laptop HDDs can have 1 or 2 disks (even though there are some specialized devices, that contain 3 disks), while normal desktop drives may have from 1 to 5 disks. Generally, the disks have Magnetic Recording Media on both sides and two Recording Heads are used per disk, one for each side.

1.2 Longitudinal vs. Perpendicular Recording

For the first 50 years of their existence, i.e. from the first commercial drive build by IBM in 1956 until to about 2005/06, HDDs operated exclusively with the so-called longitudinal recording technology. The term longitudinal refers hereby to the in-plane direction of the magnetization within the Magnetic Recording Media parallel to the disk. To magnetize the Magnetic Recording Media in this direction, a magnetic Write Head is used, that basically is a magnetic ring head with a gap as shown on the left hand side of figure 5. The magnetic field that extrudes from the gap has a large longitudinal component, which makes it very suitable to switch the magnetization of the material in the disk, allowing it to “write” a longitudinal Magnetic Recording Media.

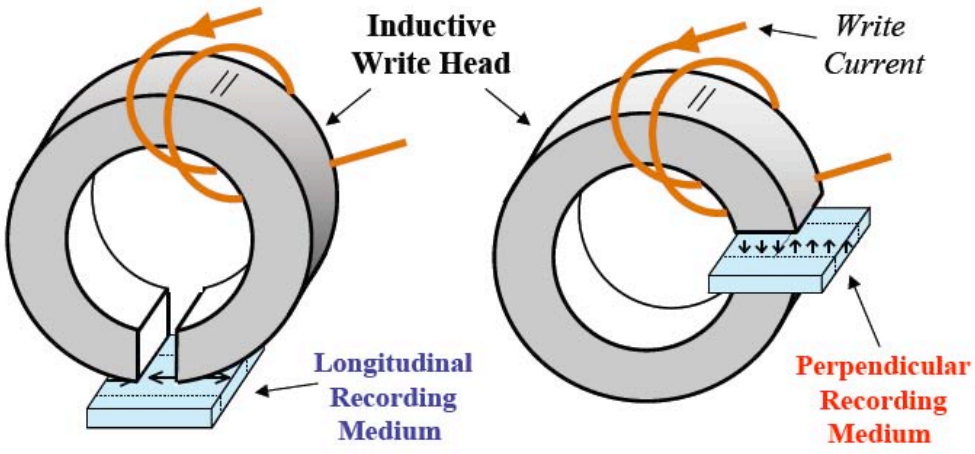


Fig. 5: *Longitudinal vs. Perpendicular Recording*

The big drawback of this longitudinal recording is the fact that only the fringe fields of the magnet gap can be utilized for magnetic recording, because the disk is located adjacent but

outside the magnet gap. The magnetic fields available in this geometry are substantially lower than the deep gap field that can be achieved inside the gap of the ring magnet itself. This is very relevant for magnetic recording because achieving high field values is important. As we will see in the following chapters, specifically in chapter 2.3., the size of the available write field sets an upper limit to the magneto-crystalline anisotropy K_U of the material that can be used for the Magnetic Recording Media. A reduction of K_U in turn will cause a reduction in stability for the magnetic state, i.e. the magnetic bit pattern, which needs to be avoided. On the other hand, only very limited improvements of the saturation magnetization for the ring head material can be expected in the future. Today's FeCo-based alloys can achieve a saturation flux density B_S in the range of 2.4 Tesla, but this appears to be somewhat of a fundamental limit as no materials with substantially higher B_S are available for recording head applications. Alternative methods to generate high magnetic fields, such as superconducting coils, etc. have not been practical (because, these Tesla-sized magnetic fields have to be generated with frequencies of several GHz to achieve the required data rates).

Given the above mentioned situation, perpendicular recording provides a fundamental advantage, as seen on the right hand side of figure 5, because the Magnetic Recording Media are moved through the gap of the Write Head, making the larger deep gap fields available for recording. The term perpendicular recording describes the fact that the recording media magnetization and the magnetic write field are oriented perpendicular to the plane of the Magnetic Recording Media and disk. While this type of magnetic recording technology has been pursued since at least the 1970s [1], a move towards it only happened in late 2005 when the first large-scale manufacturer started to move major parts of its product portfolio over to perpendicular recording. Presently, the technology shift towards perpendicular recording is almost complete and no new products using longitudinal recording have come onto the market since 2007. The reason why it took so long to implement perpendicular recording technology is two-fold. For one, until about 2003 longitudinal recording technology kept on improving so fast that no alternative technology was necessary [2] and secondly, perpendicular recording is more complex in its component design as we will see in the following chapters. In reality, the geometry shown in figure 5 for perpendicular recording is not even schematically correct, because it would not be feasible to actually move a disk through the physical gap of an electromagnet with the required speed and precision. The figure only serves as an illustration of the high magnetic fields available for magnetic recording in this perpendicular geometry.

1.3 State-of-the-Art Performance Levels

Figure 6 shows a comparison of the first commercial HDD, which IBM started to produce in 1956 in comparison to the latest class of laptop type products that were released by Hitachi GST at the end of 2008 [3]. The key driver of the technological improvement that is responsible for the wide applicability of HDDs today, is the so-called areal density, that describes how many bit can be stored per surface area. As one can see, today's products achieve 375 Gbits/in^2 (Gigabits (= billion bits) per square inch), meaning that each bit is only (41 nm)^2 in size. This number represents an improvement of about 200,000,000 times if compared to the original IBM RAMAC product. Following in the footsteps of this areal density improvement are the data rate, the size and number of components needed to reach a certain capacity point, and last but not least the price reduction, which makes the HDD a widely applicable device today.

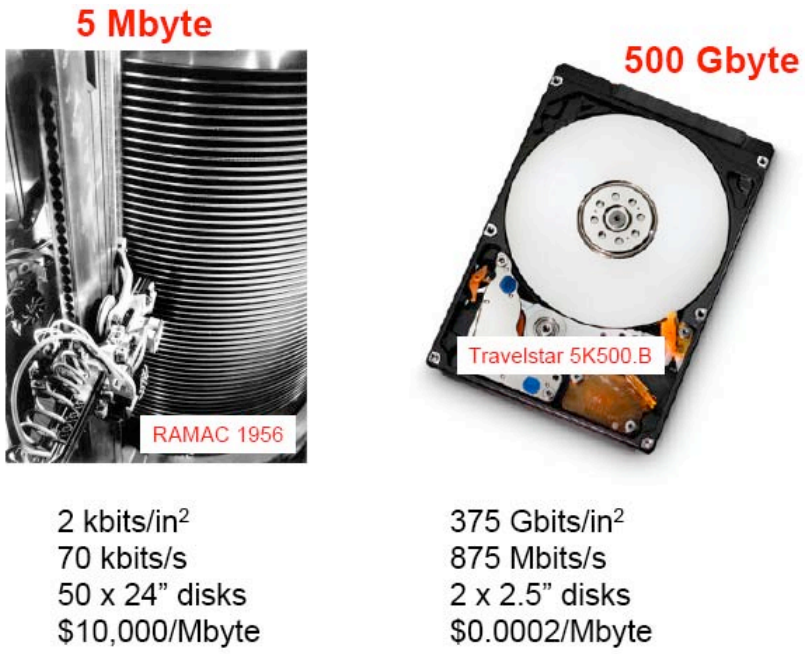


Fig. 6: *Products now and then*

As the areal density plays such a pre-eminent role for HDD technology and is the most commonly cited benchmark for technology achievements, we will briefly discuss here, how this quantity is actually determined and tested. Figure 7 shows a schematic picture of a recording media. One can see two tracks of bits that were written into this material, which extend from left to right. The region between these tracks is displayed as a gray area, which symbolizes a not yet defined demagnetized state. Hereby, it is important to mention that Magnetic Recording Media materials exhibit a fully magnetized remanent magnetization state even for perpendicular recording media, because they would otherwise not be suitable for the purpose of non-volatile data storage. In the picture shown in figure 7, each bit has a bit length B and a width W across the track that is significantly larger than B. Typical cross-track to down-track width ratios are of the order of 4 – 6 for conventional magnetic recording.

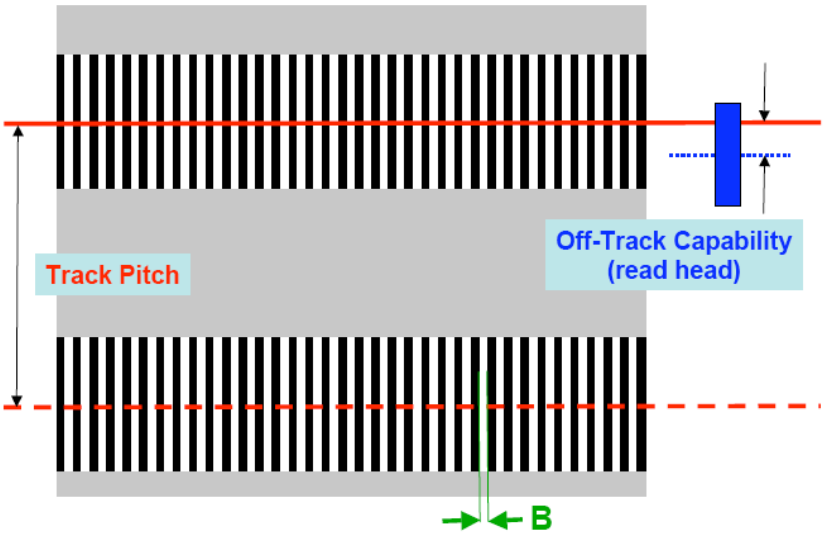


Fig. 7: *Schematic of areal density measurement*

To test the areal density, one first writes multiple tracks of a preprogrammed linear density $LD = B^{-1}$ at a fixed distance from each other, which is called the track pitch TP. Subsequently, one reads the center track back and tests if the original bit pattern can be retrieved (typically, this has to be done with less than one error in 10^5 -10^6 bits. For such "error rates" one can recapture the original information, because magnetic recording uses a bit pattern coding scheme that has sufficient redundancy to compensate such error occurrences). One also tests, what distance one can move the Read Head to the side of the track center and still recover the initial signal sequence. The maximum distance where data recovery is still possible is called Off-Track Capability. As one reduces the track pitch, adjacent patterns start to overwrite each other from the sides, so that the off-track capability is being reduced. The demonstrated track pitch is now defined as the track pitch value, at which the off-track capability becomes too small for reliable operation. The cut-off is typically defined as a fixed percentage of the track pitch itself, such as 20 - 40% for typical products and 10% for many technology demonstrations that are published in the literature [4]. Once the achievable track pitch has been measured in this way, one can determine the achieved areal density AD as

$$AD = \frac{LD}{TP} \qquad (1).$$

2 Magnetic Recording Media

As already mentioned above, the core function of the Magnetic Recording Media is the long-term preservation of information as a magnetic bit pattern. But besides this core functionality several other aspects of Magnetic Recording Media deserve mentioning. One of them is the required smoothness, mechanical hardness and chemical stability to allow for stable operation with Recording Head flying heights of less than 10 nm. Therefore, Magnetic Recording Media contain a non-magnetic overcoat layer and a lubricant. They also exhibit ultra-smooth surfaces with RMS roughness values of 1 nm and below.

In contrast to many other devices and components in the electronics world, the media disk is used as a whole and is not diced up such as silicon based electronic chips, for instance. Thus, it requires especially high uniformity and process control. The usage profile also demands very low production costs, because the disks themselves are not being miniaturized or changing in size with every new generation of HDDs, only the magnetic bits on them are.

2.1 Basic Properties

From the above discussion of the HDD, it is clear that Magnetic Recording Media have the core functions of

(i) allowing for the creation of very small magnetic bit structures
(ii) conserving these bit structures for a long time (typically 10 year stability is required)
(iii) produce a field pattern that allows for a non-destructive read-back of the bit pattern and recovery of the input information

While the down-track bit sizes are nowadays already smaller than 20 nm, one needs to realize that the required bit placement precision is actually much more stringent than the bit length itself, typically only 10% of that value, i.e. approximately 2 nm. This placement precision together with the condition that this placement has to be stable for years under normal

environmental conditions such as varying temperatures as well as small magnetic fields from the surroundings of the HDD, makes it evident that domain walls in Magnetic Recording Media will have to be very narrow and precisely pinned.

One efficient way to pin domain walls as well as make them narrow [5] is by introducing geometric constrictions into the magnetic material. This causes a pinning because the total energy contained in a domain wall is proportional to its length, so that a shorter wall has less energy than a longer wall, simply because the elevated energy density inside the domain wall only occurs within the ferromagnetic material and vanishes outside. Magnetic Recording Media take this concept to the extreme by completely decoupling the magnetic building blocks from each other. The result is a granular magnetic structure such as the one displayed in figure 8, which shows a electron microscopy picture, in which the grains appear dark and the grain boundaries light (the different gray-scales of the grains itself are caused by the crystalline grain axes distribution). Such a structure cuts the domain wall width effectively to the size of a grain boundary, which is of the order of only 1 nm. The placement precision of a bit in such a structure however, is not limited by the grain boundary width, but rather by the randomness of the grain locations. Nonetheless, due to the averaging effect of multiple grains across the track one can nowadays achieve bit pattern placement precision of better than 2 nm [6].

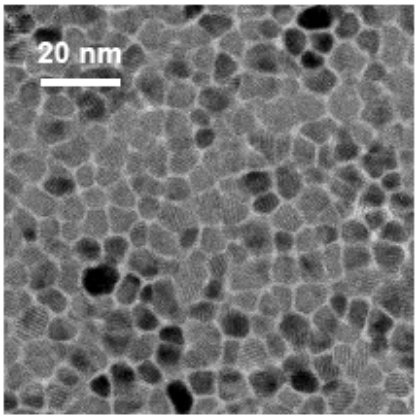


Fig. 8: *Recording Media grain structure*

Fundamentally, there are two ways to produce the required granular magnetic structure. One way is to actively structurize the media layer, so that it has a predefined pattern. This is basically the concept of patterned media, which are presently considered as the follow-on technology to conventional Magnetic Recording Media. At the present time, they are not cost competitive and will have to overcome numerous fabrication and system level design challenges, but they have a fundamental advantage related to (ii), i.e. they are very much superior in conserving the magnetic bit pattern. This will be discussed further in chapter 2.3. and 2.4.

The other pathway to reach the type of media structure shown in figure 8 is by using materials, processes, and layer stacking sequences that can be directed in such a way that a precise lateral granular structure forms itself. In particular segregation of non- or only weakly soluble materials is used to cause a segregation dynamics that guides materials into the desired granular shape. This basic process has been utilized for the fabrication of Magnetic Recording Media for the past decades and still is the leading industrial technology, as it was successfully adapted to the ever-increasing technical demands. The structure shown in figure 8 was fabricated like this.

Another aspect that is relevant for the basic functionality of magnetic materials is the crystal orientation of the grains. Magnetic Recording Media use magnetic materials that have uniaxial magnetic anisotropy with the magnetic easy axis along the crystalline c-axis, so that each grain has only two stable magnetization directions, which is best suited for the purpose of binary information storage. Correspondingly, it is important that all magnetic grains are aligned so that they behave in a similar fashion magnetically. Thus, one not only needs a well-defined uniform grain structure for Magnetic Recording Media, but HDD technology also requires the grains to be uniformly aligned, in particular have a c-axis orientation along the surface normal for today's perpendicular recording. To achieve this it is necessary to devise an adequate growth sequence of Template Layers, as shown schematically in figure 9, that provide the right base for the actual Magnetic Recording Layer to grow while forming uniformly sized, well segregated and well oriented magnetic grains. Typical grain orientation distributions for high performance recording media have a standard deviation from perfect alignment of only 2 – 4 degrees today.

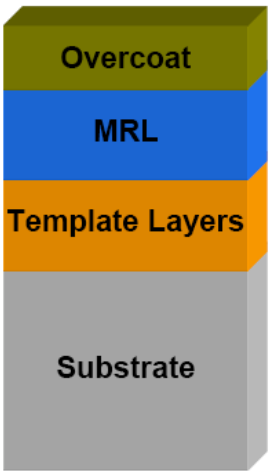


Fig. 9: *Basic Recording Media stacking sequence*

2.2 Soft-Magnetic Underlayer (SUL) for Perpendicular Recording

In chapter 1.2., it was mentioned that the geometry shown in figure 5 for perpendicular recording was not even schematically correct, because it is not possible to move a disk through the gap of a macroscopic recording head with nm flying height precision and create magnetic tracks and bits that require nm-scale positioning precision as well. Figure 10 shows what is done instead to emulate the existence of a real ring head: one places a Soft-magnetic Under Layer (SUL) below the recording layer. This SUL can now channel the magnetic flux from the main write pole back to the return pole, which leads to a concentration of flux under the main pole, i.e. the desired high magnetic field for magnetic recording. Under the return pole the magnetic flux is much less concentrated and results in perpendicular magnetic fields that are too low to the switch the Magnetic Recording Media grains or perturb the magnetic bit pattern. Thus, only under the main write pole of the recording head is the Magnetic Recording Media really written. So, overall the SUL functions as a magnetic mirror that produces a magnetic image of the real recording head and causes a magnetic flux pattern as if there really were a closed magnetic ring head with the recording layer in its gap.

From the above-described functionality, it is clear that the SUL needs to be as close to the magnetic Write Head as possible so that the magnetic flux can be channeled most effectively and the highest possible magnetic field can be reached. Thus, it has to be placed right below the Magnetic Recording Layer and therefore has to become part of the Magnetic Recording

Media itself. So, in perpendicular recording, Magnetic Recording Media have two functional magnetic layers, the actual Magnetic Recording Layer and the SUL that is part of the magnetic recording head circuit. For the SUL, one typically uses high moment amorphous alloys with a total thickness of up to several 100 nanometers.

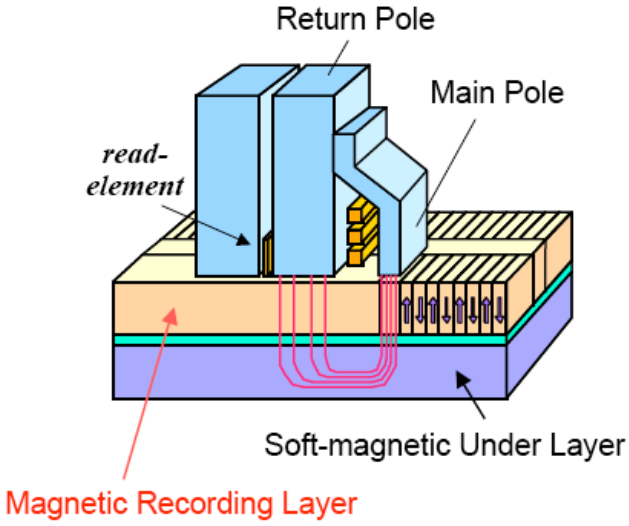


Fig. 10: *Perpendicular Recording Media with SUL*

2.3 Thermal Stability

As discussed in chapter 2.1., the grains in Magnetic Recording Media are separated from each other to allow for a precise placement of narrow bit transitions. But this individual addressability of each grain comes at a high price, namely that each grain has to be magnetically stable by itself as well. As for any finite magnetic system, the magnetization state of a recording media grain is not completely stable, but will switch after a certain amount of time. As long as the switching probability is so low that it does not affect operation during the lifetime of a HDD, this is not a problem. Today's HDD are, however, not far from the point where thermal stability is a serious problem, an effect known as the super-paramagnetic limit [7]. Super-paramagnetism is a fundamental effect of ferromagnetic grain assemblies that behave effectively like a paramagnet above a certain temperature, because their magnetic moments flip back and forth faster than the observation speed.

To estimate the relevance of this effect, one generally makes the assumption that within one grain all spins are aligned with each other. This is a sensible simplification, because the strong exchange coupling within each grain will keep all spins aligned on the typical nm length scale of Magnetic Recording Layer grains. Under the assumption of spin alignment, one can then treat the magnetic grain as a single macro-spin S, whose total energy depends only on the uniaxial anisotropy constant K_U, its magnetization orientation angle θ and the total grain volume V. Specifically one finds that the activation energy necessary for magnetization reversal without an external field applied is simply given as [8]

$$E_A = K_U \cdot V \qquad (2)$$

Correspondingly, the average lifetime τ of a predefined magnetic grain state can be estimated [9] to

$$\tau^{-1} = f \cdot \exp(-\frac{K_U \cdot V}{k_B \cdot T}) \qquad (3)$$

with f being the attempt frequency that typically is in the several to several ten GHz range, and k_B and T being the Boltzmann constant and the temperature, respectively. It is evident that the size of K_UV is a crucial material parameter, responsible for the thermal stability of magnetic bit patterns and for the non-volatility of HDD recorded information. This stability concern now clashes with (a) the ever-decreasing bit size, because smaller bits require smaller grains, i.e. smaller V. On the other hand K_U cannot be increased as already discussed in Chapter 1.2., because the magnetic switching field H_S of a grain with saturation magnetization M_S in the above macro-spin model [10] is found to be

$$H_S = 2\frac{K_U}{M_S} \qquad (4)$$

i.e. directly proportional to K_U, which means that K_U can only be increased if larger magnetic fields are made available for the write process. This need for higher magnetic fields was a key driving force behind the recent technological transition from longitudinal to perpendicular recording. Due to this transition and other technological advances, one can now achieve recording densities that are more than a factor of 10 larger than what was predicted just over ten years ago to be the “fundamental” super-paramagnetic limit of magnetic recording [11].

2.4 Modern Materials Designs

One of the recent ideas to improve HDD technology and Magnetic Recording Media beyond the above mentioned limitations are the so-called Exchange Spring media that are schematically shown in figure 11. In this structure, the single Magnetic Recording Layer is

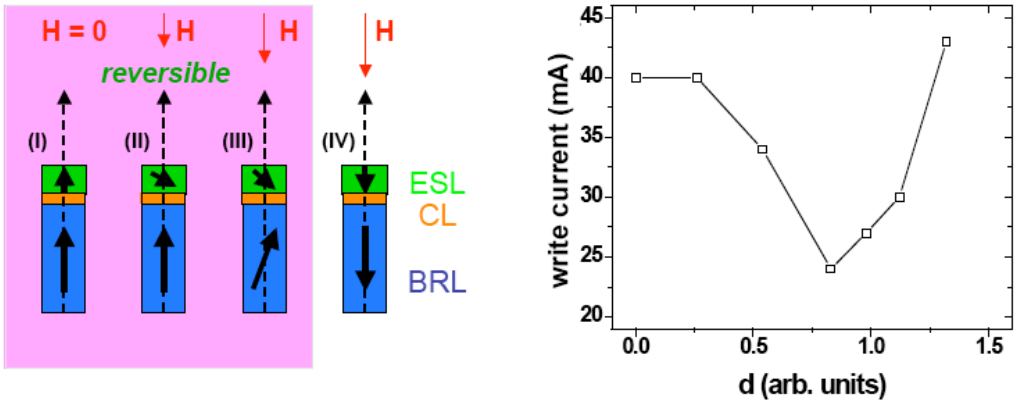


Fig. 11: *Exchange Spring Layer media*

replaced by a tri-layer stack of a Base Recording Layer (BRL) on the bottom, made from hard magnetic material, and the Exchange Spring Layer (ESL), made from a soft magnetic material, on top. In between is a Coupling Layer that allows for a tuning of the ferromagnetic coupling strength between BRL and ESL. Upon optimizing the ferromagnetic interlayer coupling strength, which can be adjusted by varying the CL thickness d, the magnetization reversal of such a grain proceeds via a non-uniform process. Such a process requires only a reduced field strength and partially circumvents the switching field considerations and

limitations discussed in 2.3. The validity of this concept has recently been demonstrated in real recording media, by achieving better writability (shown on the right hand side of figure 11) and an overall improved media performance in an optimized structure [12]. The name Exchange Spring media comes from the fact that until one switches the magnetically hard BRL, the magnetization rotation of the ESL is fully reversible and can snap back like a spring.

A much more radical deviation from conventional Magnetic Recording Media are the so-called patterned media, which are schematically shown in figure 12. Here, every pattern or quasi-grain has a predefined position, which is generally lithographically defined. This has the key advantage that every quasi-grain can carry a bit, because its position is known and can be individually addressed in the write process. About 30 - 50 grains are needed in conventional media per bit, because the grains are randomly distributed and multiple grain averaging is needed. With this alternative technology, one can in principle achieve more than 10 times higher storage densities, because every quasi-grain can be made from magnetic materials with strong lateral exchange coupling that can very easily fulfill the stability condition according to equation (2). The main drawback of patterned media is the fact that the previously unstructured and therefore cheap media disk has now to be made by means of a lithographic process with very small tolerances. Solving this problem as well as related technological issues such as the non-planar disk topography, the need for absolute write synchronization etc. are presently a major research effort in the HDD industry. It is anticipated by many that an eventual technological transition to patterned media might occur at the 1 Tbit/in^2 density point.

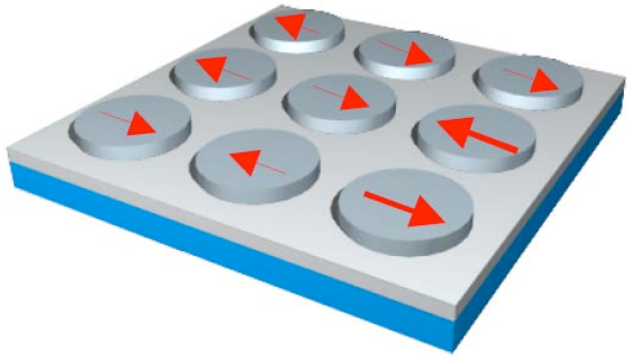


Fig. 12: *Patterned Recording Media*

An intermediate step towards patterned media technology is the so-called discrete track medium (DTM), which is a hybrid technology of conventional recording media and patterned media. Here, only the tracks are lithographically pre-defined, while the tracks themselves are made from conventional recording media. Such structures have basically the same limitations as conventional recording media because the data storage is still done by encoding magnetic bits into multiple random grains, but they do have the advantage that many serious write and read-back problems that are caused by the track edges are suppressed. In comparison to patterned media, DTM have the advantage that the lithography demands are less severe and other technology changes, such as the disk topography are easier to handle.

3 Magnetic Write Heads

Similar to Magnetic Recording Media discussed in Chapter 2, Magnetic Recording Heads have to fulfill a large number of technical specifications that are not directly related to their core magnetic functionality but are crucial for the purpose of reliably writing and reading

information. For instance, the head assembly including the air bearing surface have to be designed in such away that they not only fly precisely with less than 10 nm flying height above the disk surface, but also allow for a rapid and precise lateral shift of the entire assembly to any random track without triggering mechanical oscillations. Hereby, they have to settle with nm precision on the target track in a millisecond time frame. These aspects of the head assembly go beyond the scope on this description and will not be considered here.

3.1 Basic Functionality

The key function of the magnetic Write Head is the creation of a very strong, but locally confined magnetic field inside the Magnetic Recording Layer that allows for the magnetization reversal of the media grains. To produce a dense bit pattern along the track as the disk runs underneath it, the recording head must also be able to switch the field direction very rapidly. Typical data rates for server type HDDs are reaching several Gbit/s nowadays, which illustrates that a magnetic Write Head reversal must occur within only several hundred picoseconds.

Figure 13 shows a schematic of a Write Head. While this schematic is for a longitudinal Write Head, it is not fundamentally different for a perpendicular Write Head. We can see the two main magnetic layers, top and bottom, that are connected by a magnetic connector element, around which a lithographically defined copper coil is placed. The magnetic circuit is opened at the write pole gap to allow the magnetic flux to exit the Write Head and produce an outside field. The core difference between this picture and a perpendicular Write Head are that for perpendicular recording: (i) the write pole is thinner, which allows the flux to exit along the pole axis, i.e. right along the perpendicular magnetization axis of the disk, and (ii) a larger gap between top and bottom pole, which prohibits the direct flux from the write pole to the return pole and forces the magnetic flux through the SUL as shown in figure 10. This enhances the achievable field and properly aligns it with the preferred magnetization axis of the adjacent disk.

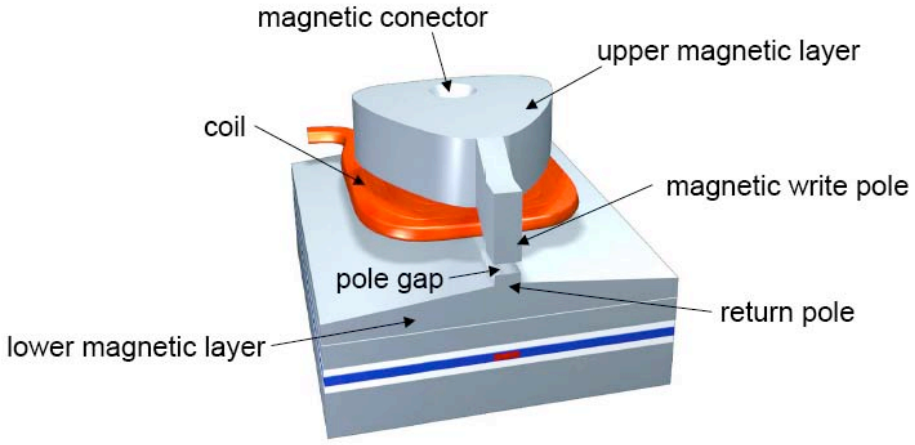


Fig. 13: *Magnetic Write Head schematic; courtesy of Jeff Childress (Hitachi GST)*

To achieve the most precise bit structure, it is actually advantageous not to define the bit in the center of the recording head where the magnetic field is the highest, but at one of the edges, the so-called write field edge, where the magnetic field has the highest gradient. Here, the lateral separation of grains that switch and grains that do not switch is most precisely defined, assuming that the total magnetic field at this write field edge is exactly tuned to match the switching field of the Magnetic Recording Layer. This is one of the core design criteria for magnetic Write Heads. Typically, magnetic Write Heads reach field gradients of

several 100 Oe/nm nowadays. Given this system design and operation, magnetic grains on the disk are actually reversed multiple times as they enter into the magnetic Write Head region, in which the Write Head field is larger than H_S. This region is called “magnetic write bubble”. The magnetic state of grains and the resulting magnetic bit pattern is defined the moment the grains leave this write bubble due to the mechanical movement of the disk.

3.2 Modern Designs

While the above discussion of the magnetic write bubble of a recording head explains the timing of the actual recording process along the track direction, one has to consider that there is also a cross-track extension and shape of this write bubble. Fundamentally, the recording head geometry ought to be designed in such a way that also this cross-track direction of the write-bubble is as sharply defined as possible, allowing for adjacent tracks to be brought together very closely and to achieve a high areal recording density.

While this is a challenge in itself even under the best of circumstances, the problem is worse if one considers that the entire disk has to be addressed by one recording head. As already shown in figure 2, the recording head assembly sits on an arm that can be moved across the disk to define or find any arbitrary track position. As one moves the head from the inside to the outside, the write pole is aligned well with the azimuthal track direction for only one specific track diameter in a simple rotating arm. For other tracks there is a misalignment skew that becomes extreme at the inner and outer track radius. This skew now causes the problem that the Write Head becomes effectively wider as can be seen in figure 14. Worse than that, substantial parts of the written track are not useable, but constitute the so-called erase band. Here, only the front part of the write bubble crosses the media, while the actual writing edge never crosses these regions. The result is that adjacent tracks will be erased (therefore the name erase track) in this range, but no useful information is written here, because of the very poorly defined field gradient, with which these disk segments leave the write field bubble.

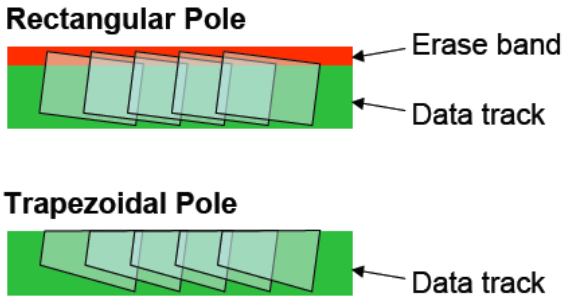


Fig. 14: *Trapezoidal Write Head*

As also shown in figure 14, this problem can be suppressed by using a trapezoidal pole shape with the trapezoidal angle being larger than the maximum skew angle that will occur for the inner or outermost track position. In this way it is insured that the trailing edge, which is designed to do the actual magnetic write process, is indeed the one writing the entire track width, hereby suppressing (all or most of) the erase band. The key drawback of the trapezoidal pole shape is that its accurate fabrication and production is very challenging, as it needs to be done with nm scale precision.

3.3 Technology Challenges

While there are many challenges related to the Write Head and its purpose, namely to produce a strongly confined high magnetic field by means of a tiny magnetic write pole, we highlight here two aspects that have a nearly fundamental character: the write-pole shape anisotropy and the saturation magnetization of the pole material.

Given its purpose the Write Head pole has a needle or tip like shape. Correspondingly, the magnetostatic self-energy is such that this tip tends to produce a magnetic state that generates a not insignificant magnetization along the pole axis even if no write-coil current is applied. This causes the Write Head to produce a substantial magnetic field, even if the head is not intended to write, i.e. during read back operation or in the idle state, which can compromise already written information directly or reduce the long-term stability as an accumulative effect. Upon further reduction of the Write Head pole size in future product generations, this effect is expected to get increasingly worse, because the build-up of a stray-field free stable magnetic configuration is becoming less likely for ever shrinking lateral dimensions, because such structures require more and more exchange energy. One possible way to circumvent this problem, or at least delay its onset is the lamination of the Write Head material; i.e. making it out of multiple segments that are not exchange coupled with each other, which allows for a reduction of the exchange energy associated with the desired flux-compensated magnetization state [13].

To achieve continued increases in areal density, one would want to further increase the magnetic field gradient of the Write Head, so that the down-track position of the magnetic bits could be better defined. As this gradient is associated with the trailing write pole edge, it is fundamentally defined by the edge geometry and the saturation magnetization of the Write Head pole material. Thus, to achieve further advances, an increase in the saturation magnetization of the pole head material would be most desirable. However, such progress appears to be most unlikely, because already today recording heads use materials that have some of the highest known saturation magnetization values and progress in terms of finding materials with higher saturation magnetizations has been virtually non-existent. So, only the geometry can be utilized to achieve higher field gradients by using write pole shields to produce artificially enhanced magnetic field gradients (a) along the track: trailing-edge shield, (b) across the track: side shields, or (c) both: wrap around shield. While these geometries do improve the overall field confinement and give a better-defined write field bubble, they do come at the cost of a reduced total field strength, because they guide flux away from the SUL pathway. They also require substantially more complex fabrication processes.

4 Magnetic Sensors as Read Heads

As their name indicates, Magnetic Read Heads are being used to read out the information that was previously written onto the Magnetic Recording Layer by the Write Head. Because the magnetization pattern is not directly accessible for external observation, the magnetic stray-field that is generated by the bit-pattern is used for this purpose. Thus, the Read Head is in essence a magnetic field sensor (in the early days of HDD technology, it was actually a flux change sensor that measured the flux changes caused by the stray fields above the rotating disk). Generally, the detection of magnetic fields or flux changes by a nearby Read Head does not perturb the magnetic bit pattern itself, so that the read process is non-destructive, which is an important fact and advantage of this non-volatile storage technology.

Over the years, the Read Head technology has undergone several major changes, each of which had a very significant impact on the overall HDD performance. This can be seen from figure 15, in which the historic development of HDD areal recording density is shown. In

particular the 1st Magneto-Resistance (MR) head, based initially on the Anisotropic Magneto-Resistance (AMR) effect, and the subsequent introduction of the Giant Magneto-Resistance (GMR) head had a major impact and led to an accelerated technology improvement with annual growth rates of up to 100% for the areal density. Interestingly, the recent transition from longitudinal to perpendicular recording did not impact the Read Head technology very much, because the field sensing requirements are very similar in both cases despite the rather different field patterns that are produced by the Magnetic Recording Media.

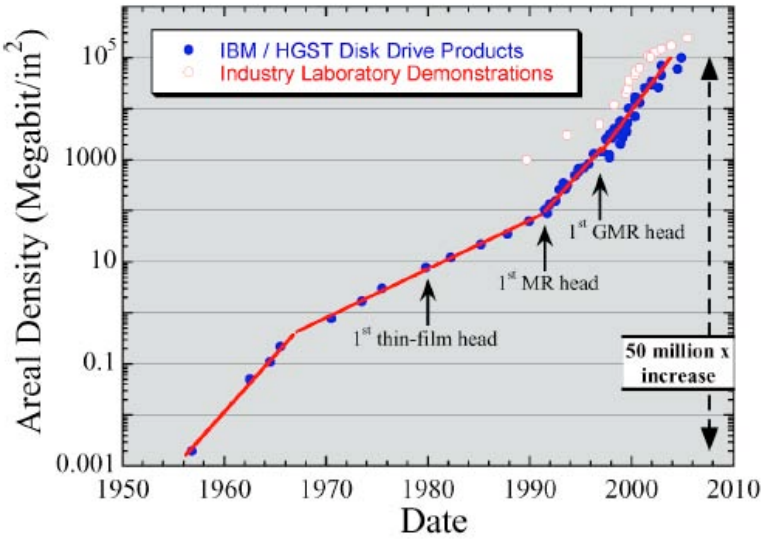


Fig. 15: *Achieved areal density vs. time; courtesy of Jeff Childress (Hitachi GST)*

4.1 Basic Functionality and Requirements

HDD Read Heads have to be very localized probes to allow for the recovery of the original information by means of measuring the bit pattern magnetic stray-fields. Otherwise, the various local field contributions would compensate each other and no relevant information could be retrieved from the spinning disk.

Figure 16 shows a schematic of the recording head and emphasizes the core Read Head components in their relation to the underlying track pattern on the disk. As one can see, the MR sensor of the Read Head is indeed very small to allow for the generation of a read-back signal that is able to identify every single bit. The cross-track width of the MR sensor is made smaller than the written track width, so that (a) a slight misalignment away form the track center still allows for a reliable operation and (b) the edge segments of the written tracks are avoided, because they generally are not straight. They are therefore not in alignment with the Read Head edges and would produce a considerable noise contribution to the Read Head signal. Typical ratios of read width (RW) to write width (WW) are 50 – 80 %.

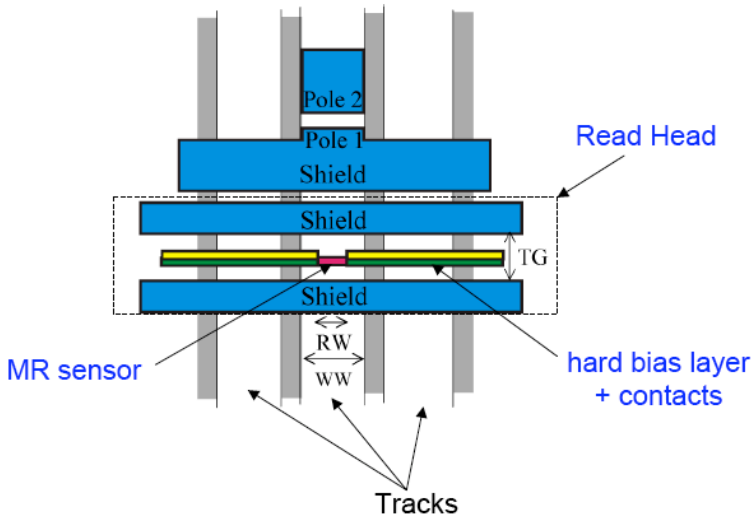


Fig. 16: *Basic Read Head geometry; courtesy of Jeff Childress (Hitachi GST)*

In the down-track direction, magnetic Read Heads need an even better resolution than in the cross-track direction because the bit length is significantly smaller than the track width. This down-track resolution is achieved by two geometric features of the Read Head. First, the active sensing layer in the MR sensor is thin and only the field that actually penetrates the sensing layer will generate a signal. However, the sensing layer also has a considerable height above the surface (perpendicular to the plane shown in figure 16), because this dimension is determined by the lithographic resolution just as RW is. Therefore, magnetic fields coming from bits farther up and down the track would penetrate the MR sensor and compromise its resolution. To limit this effect, magnetic shields are introduced as the second down-track geometric feature in magnetic Read Heads. They are actually more important for the down-track resolution than the sensing layer thickness itself. These shields are made from a soft magnetic material and absorb any magnetic flux from outside the gap area, so that only the magnetic fields generated inside the shield distance TG (see figure 16) actually reach the sensing layer. As a rule of thumb, TG should not exceed two times the bit length [2].

The fact that TG can be larger than the bit length at all is enabled by the signal processing that is used in HDDs to analyze the Read Head signal of several bits simultaneously. This allows for a de-convolution of overlapping fields from neighboring bits. For the signal processing to work reliably, however, it is important that the MR sensor is linear in its field characteristics and does not saturate during operation. While the sensor layer itself is optimized in this regard, one also uses the so-called hard bias in HDD Read Heads to enable sensor linearity and stability. This hard bias uses a fully magnetized hard magnetic material, shown in green in figure 16 that produces a magnetic field in the cross-track direction. This field gives the sensing layer a preferred magnetization direction and causes a sufficiently stiff field response of the sensor to prevent signal saturation. Also the bias field from the hard bias layer suppresses the formation of domains in the sensing layer, which would otherwise cause Read Head instabilities. The yellow segments, shown in figure 16 are the electrical leads of the sensor.

4.2 Magneto Resistance Effects

While Magneto-Resistance (MR) effects and their physical origin are discussed in great detail elsewhere in this course, they are addressed here only very briefly to ensure a consistent use of terminology. In HDDs, three MR effects are or were relevant: the anisotropic magneto-

resistance (AMR), the giant magneto-resistance (GMR), and the tunnel magneto-resistance (TMR).

The AMR describes the fact that the resistance of a ferromagnetic material depends on the angle between the direction of the magnetization and the direction of the electrical current, generally showing a sinusoidal angular dependence. The effect can be several percent in size and enabled a major simplification of HDD technology. Instead of the use of pick-up coil circuits to sense magnetic flux changes of a passing bit pattern, AMR sensors allowed the direct measurement of stray fields by a single magnetic film in a layered sensor stack. Correspondingly, technology improvements accelerated substantially with the introduction of the AMR Read Head in the early 1990s, as can be seen in figure 15.

The GMR, first discovered simultaneously, but independently by P. Grünberg and A. Fert in 1988 [14], describes the fact that the resistance of adjacent ferromagnetic films, separated by a non-magnetic conducting layer, depends strongly on the relative orientation of the magnetization directions in these layers. While the GMR relies on a more complex structure than the AMR, it is also a much larger effect that can reach several hundred percent. Therefore, GMR sensors replaced AMR sensors in the late 1990´s and started an era of tremendous advances in HDD technology with unprecedented growth rates for areal density in subsequent years.

The TMR is very similar to the GMR, except that the non-magnetic layer that separates the two adjacent ferromagnetic layers is an insulator. Thus, only tunnel currents can flow in between the two ferromagnetic layers, and their amplitude again depends on the relative orientation of the magnetizations. The effect is large and can reach several hundred percent, which makes it desirable for Read Head applications. It was introduced into products starting in 2005.

4.3 GMR Sensors

Figure 17 shows a schematic drawing of a GMR sensor stack. Center part of this stack is the ferromagnetic free layer (FL). It is this layer that responds to external magnetic fields (generated by the bit pattern) by rotating its magnetization away from its equilibrium position parallel to the cross track direction in the disk plane. Together with the metallic spacer layer (SL) and the reference magnetic layer (RL) it constitutes the basic GMR trilayer stack, that is responsible for the GMR effect in these sensors. The magnetization direction of the RL is kept fixed during operation of the Read Head, so that a change in the FL magnetization orientation will simultaneously change the magnetization angle between FL and RL and therefore cause a GMR change. Using constant current operation conditions, the magnetization rotation in the FL therefore produces a voltage change, which is the read-out signal. In this way, the stray field patterns associated with the magnetic bits in the Magnetic Recording Media are transformed into an electrical signal sequence that correspond to the originally stored data set.

As mentioned above, it is important that the voltage vs. field characteristic of the sensor is linear. Given that the GMR effect has a sinusoidal angular dependence, one therefore needs a base operation point for the GMR sensor, in which the magnetizations of the FL and the RL are perpendicular to each other. For this purpose, the RL magnetization is perpendicular to the disk plane.

We will now address the question, how the reference layer magnetization is fixed. This can be done most effectively by attaching an Anti Ferromagnetic (AFM) layer, the so-called pinning layer, to the RL. It is well-known that AFM layers in contact with ferromagnets can produce an exchange anisotropy (also called exchange bias effect), which stabilizes one particular magnetization direction, i.e. pins the adjacent ferromagnet in this very direction [15]. This is done in GMR sensors by using PtMn and IrMn as typical AFM materials.

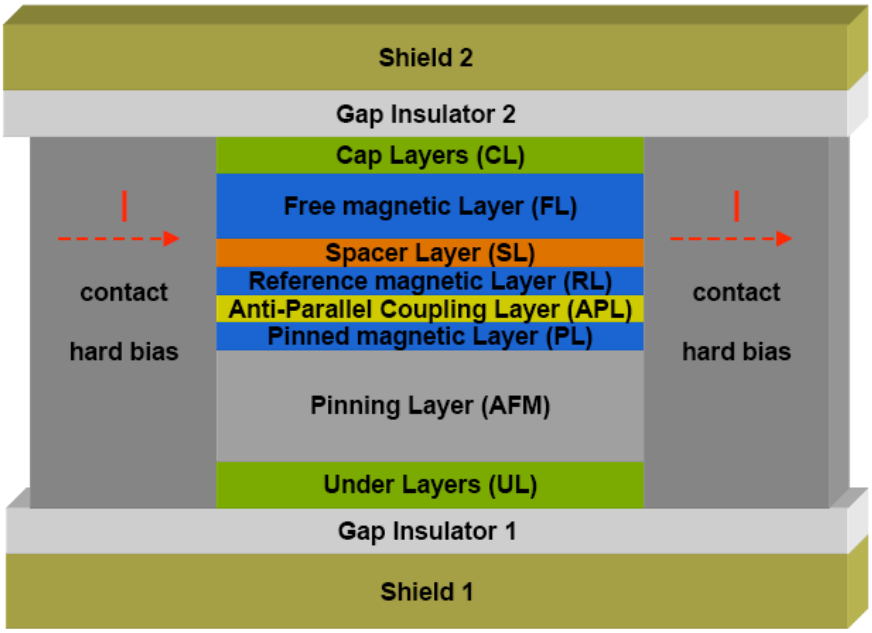


Fig. 17: *GMR sensor*

However, as one can see from figure 17, the RL is not directly attached to the AFM, but rather coupled by means of anti-ferromagnetic interlayer exchange coupling [16] to another ferromagnetic layer, the pinned magnetic layer (PL), which is the one attached to the AFM. The interlayer exchange coupling is typically mediated by means of an ultra thin Ru layer (APL) that provides sufficient anti-parallel coupling strength between the RL and the PL [17]. This more complex GMR sensor stack is utilized to suppress the magnetic fields generated by the RL. Due to its role as the GMR reference electrode, the RL needs to be in a uniform magnetization state. However, such a state also generates a substantial magnetic field and correspondingly biases the FL, which would cause a loss of dynamic range and deviations from linearity. To avoid this, one uses the RL/APL/PL tri-layer structure. Due to the anti-parallel magnetization orientation of the RL and PL, it can be designed in such a way that there is no net magnetic field bias onto the FL. The GMR effect itself, however, is dominated by the interface magnetization regions directly adjacent to the SL, which means that only the RL and not the PL magnetization contribute to the GMR sensor signal. Therefore, the magneto-electric GMR functionality is not altered by the addition of APL and PL, and only a more stable operating point is facilitated by them.

The GMR layer stack is completed by under layers (UL) on the bottom and cap layers (CL) on top, both of which have no operational function. The UL have the purpose of providing the right template for high-quality layer growth in the fabrication process and the CL allow for the protection of the FL against chemical contamination from adjacent materials. The complete GMR Read Head furthermore comprises electrical leads and the hard bias layer that connect from the side, as well as the top and bottom gap insulator layers that insolate the GMR sensor electrically from the top and bottom magnetic shields.

4.4 CIP vs. CPP Geometry

The GMR sensor discussed in the previous section and displayed in figure 17 has the so-called Current-In-Plane (CIP) geometry, in which the current flows through the plane of the sensor. This is in contrast to the so-called Current-Perpendicular-to-the-Plane (CPP)

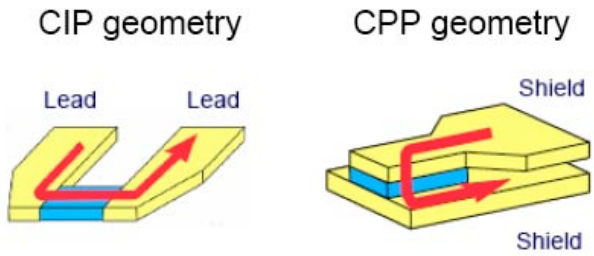


Fig. 18: *CIP vs. CPP geometry; courtesy of Jeff Childress (Hitachi GST)*

geometry, in which the current flows perpendicular across the plane of the GMR sensor. Both geometries are schematically compared in figure 18. They each have advantages and disadvantages that are widely discussed in the literature [2]. So, far only CIP-GMR sensors have been used in products. The core drawback of the CPP-GMR sensor is that for today's typical sensor dimensions it produces a rather low total resistance because its cross section (perpendicular to the current) is large and its length (along the current path) is small. Thus, parasitic resistances such as the one coming from the leads are rather large in comparison to the resistance of the active sensor element and reduce the achievable total resistance change. Also, the total resistance of such devices is presently too small to produce the required signal levels in the mV range for acceptable current densities. However, as the dimensions of Read Head sensors continue to shrink in the future, the geometry will start to favor the CPP over the CIP design for GMR sensors, so that they are expected to show up in products at about the 500 Gbit/in^2 - 1 Tbit/in^2 density point.

4.5 TMR Sensors

Despite the disadvantages discussed above for the CPP-GMR sensor, the CPP sensor geometry also has key advantages. For instance, in CPP geometry the magnetic shields can be directly utilized as electrical leads for the sensor. Correspondingly, the in CIP-geometry necessary adjacent insulation layers can be removed, which simplifies the fabrication process. Furthermore, one has to realize that in the CIP geometry that is shown in figure 17, all the layers of the GMR sensor stack carry a current in parallel. Thus, the secondary GMR sensor layers: UL, AFM, PL, APL, and CL effectively constitute a resistor, which shunts the "real" GMR sensor consisting of FL, SL, and RL. This shunting reduces the GMR effect that can be utilized as a signal. In the CPP geometry, shunting is avoided, even though the secondary GMR sensor layers will still contribute a parasitic resistance.

The TMR Read Head sensor utilizes these core CPP geometry advantages together with the large TMR effect, which made it a viable product technology for the past few years [18]. Figure 19 shows a schematic of the TMR Read Head. As one can see, the basic stack structure is identical to the CIP-GMR head shown in figure 17. The only crucial difference is the fact that the SL is now made from a non-conducting layer (typically MgO) that is thin enough to allow for a substantial tunnel current in between FL and RL. The resistance of such structures containing a tunnel barrier is now high enough to avoid the previously discussed problems of the CPP geometry at today's sensor dimensions. Therefore, TMR sensors are presently the technology of choice for commercial Read Heads in HDDs. However, it is expected that as dimensions shrink further, the sheet resistance of tunnel barriers will be too high for practical sensor operations, at which point the above mentioned product transition towards CPP-GMR sensors is anticipated.

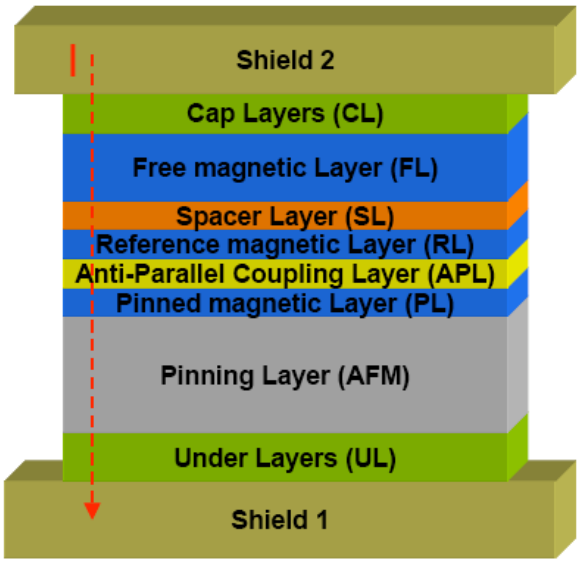


Fig. 19: *TMR sensor*

To complete the discussion of figure 19, one should mention that the TMR Read Head also contains a hard bias layer (not shown), just as the CIP-GMR head in figure 17. The difference is that this hard bias layer is now electrically isolated from the sensor stack, because it does not function as part of the electrical contact anymore. However, it fulfills the identical magnetic purpose as in the case of the CIP-GMR sensor. One should also clarify that a CIP-TMR sensor is not possible, because in the CIP geometry the insulating SL would simply isolate the parallel metallic conductivity of the FL and RL from each other, making their conductivity independent from their relative magnetization orientation. Also, the transport mechanism would be purely metallic without any tunneling effect.

References

[1] S. Iwasaki and Y. Nakamura, IEEE Trans. Magn. **13**, 1272 (1977)
[2] J.R. Childress, R.E. Fontana Jr., C. R. Physique **6**, 997 (2005)
[3] data sheet: Hitachi GST Travelstar 5K500.B
[4] M. Xiao et al., J. Appl. Phys. **99**, 08E712 (2006)
[5] A. Berger, B. Feldmann, H. Zillgen, and M. Wuttig, J. Magn. Magn. Mater. **183**, 35 (1998); P. Bruno, Phys. Rev. Lett. 83, 2425 (1999).
[6] H. Neal Bertram, *Theory of Magnetic Recording* (Cambridge University Press, 1994)
[7] E. E. Fullerton, D. T. Margulies, N. Supper, H. Do, M. Schabes, A. Berger, and A. Moser, IEEE Trans. Magn. **39**, 639 (2003)
[8] More detailed considerations will produce corrections to this rather simple physical picture, which are not so relevant from a fundamental point here, but important for an accurate description of real materials; see also [6].
[9] D. Weller and A. Moser, IEEE Trans. Magn. **35**, 4423 (1999)
[10] Giorgio Bertotti, *Hysteresis in Magnetism: for Physicists, Materials Scientists, and Engineers* (Academic Press, 1998)
[11] S. H. Charap, P. L. Lu, and Y. He, IEEE Trans. Magn. **33**, 978 (1997)
[12] A. Berger, N. Supper, Y. Ikeda, B. Lengsfield, A. Moser, and E. E. Fullerton, Appl. Phys. Lett. **93**, 122502 (2008)
[13] see, for instance: US Patent 7420780
[14] M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, Phys. Rev. Lett. **61**, 2472 (1988); G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, Phys. Rev. **B 39**, 4828 (1989)
[15] see, for instance: J. Nogues and I. K. Schuller, J. Magn. Magn. Mater. 192, 203 (1999)
[16] P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky and H. Sowers, Phys. Rev. Lett. **57**, 2442 (1986)
[17] S. S. P. Parkin, N. More, and K. P. Roche, Phys. Rev. Lett. **64**, 2304 (1990)
[18] C.Chappert, A.Fert, and F. Nguyen Van Dau, Nature Materials **6**, 813 (2007)

F 3 Magnetic Logic

G. Reiss, A. Thomas, J. Schmalhorst, , D. Meyners
Universität Bielefeld
Fakultät für Physik
V. Höink
National Institute of Standards and Technology
Gaithersburg, USA
D. Meyners
Universität Kiel
Inorganic Function Materials

Contents

1 Introduction 2
1.1 Defintion of logic elements 3
1.2 The approaches to Magnetic Logic 6

2 The Quantum Cellular Automata 7

3 The domain wall approach to Magnetic logic 8

4 Magnetic Tunnel Junction Logic 11

5 Summary 16
Acknowledgements 17

References 17

1 Introduction

Everything in the digital world is based on the binary number system. Numerically, this involves only two symbols: 0 and 1. Logically, we can use these symbols or we can equate them with others according to the needs of the moment. Thus, when dealing with digital logic, we can specify that:

0 = false = no , 1 = true = yes

Using this two-valued logic system, every statement or condition must be either "true" or "false;" it cannot be partly true and partly false. While this approach may seem limited, it actually works quite nicely, and can be expanded to express very complex relationships and interactions among any number of individual conditions. It makes it possible to design and build circuits that will remain indefinitely in one state unless and until they are deliberately switched to the other state. This makes it possible to construct a machine (the computer) which can remember sequences of events and adjust its behavior accordingly.

Digital logic may be divided into two classes: *combinational logic*, in which the logical outputs are determined by the logical function being performed and the logical input states at that particular moment; and *sequential logic*, in which the outputs also depend on the prior states of those outputs. Both classes of logic are used extensively in all digital computers.

Since both types of logic circuits begin with logic *gates* to combine logical input signals in various ways to produce the desired outputs, we will begin on the next page by seeing how the basic logic gates work: While each logical element or condition must always have a logic value of either "0" or "1", we also need to have ways to combine different logical signals or conditions to provide a logical result. For example, consider the logical statement: "If I move the switch on the wall up, the light will turn on." At first glance, this seems to be a correct statement. However, if we look at a few other factors, we realize that there's more to it than this. In this example, a more complete statement would be: "If I move the switch on the wall up *and* the light bulb is good *and* the power is on, the light will turn on." If we look at these two statements as logical expressions and use logical terminology, we can reduce the first statement to:

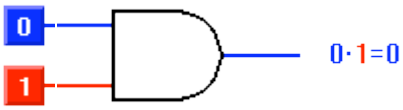
Light = Switch

This means nothing more than that the light will follow the action of the switch, so that when the switch is up/on/true/1 the light will also be on/true/1. Conversely, if the switch is down/off/false/0 the light will also be off/false/0.

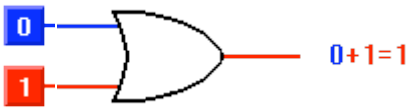
Looking at the second version of the statement, we have a slightly more complex expression:
Light = Switch *and* Bulb *and* Power

1.1 Defintion of logic elements

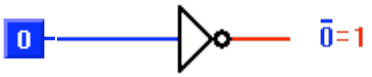
When we deal with logical circuits (as in computers), we not only need to deal with logical functions; we also need some special symbols to denote these functions in a logical diagram. There are three fundamental logical operations, from which all other functions, no matter how complex, can be derived. These functions are named *and*, *or*, and *not*. Each of these has a specific symbol and a clearly-defined behavior, as follows:



The AND Gate implements the AND function. With the gate shown to the left, both inputs must have logic 1 signals applied to them in order for the output to be a logic 1. With either input at logic 0, the output will be held to logic 0.
There is no limit to the number of inputs that may be applied to an AND function, so there is no functional limit to the number of inputs an AND gate may have. However, for practical reasons, commercial AND gates are most commonly manufactured with 2, 3, or 4 inputs



The OR Gate is sort of the reverse of the AND gate. The OR function, like its verbal counterpart, allows the output to be true (logic 1) if any one or more of its inputs are true. Verbally, we might say, "If it is raining OR if I turn on the sprinkler, the lawn will be wet." Note that the lawn will still be wet if the sprinkler is on and it is also raining. This is correctly reflected by the basic OR function.



The NOT Gate, or Inverter is a little different from AND and OR gates in that it always has exactly one input as well as one output. Whatever logical state is applied to the input, the opposite state will appear at the output.
The NOT function is necessary in many applications and highly useful in others. A practical verbal application might be: The door is NOT locked = You may enter

In terms of microelectronic devices, the simplest *realization* of logic circuits consists of diodes and resistors: The so called Diode-Logic (DL):

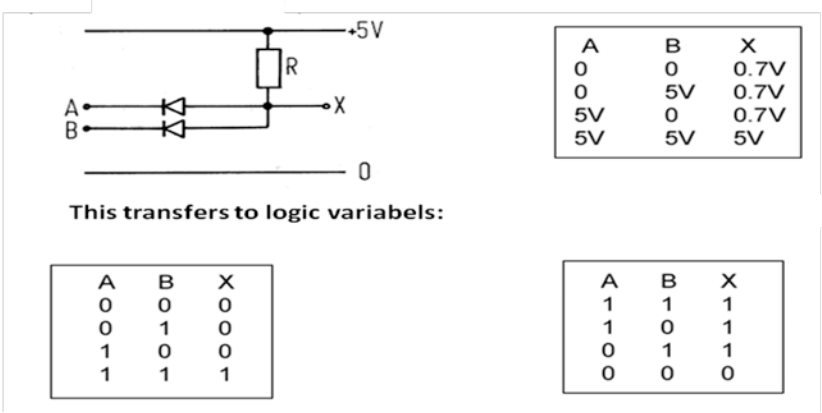


Fig. 1: *A diode based logic gate for an AND function in positive and a NAND function in negative logic. The threshold voltage for the current in forward bias is assumed to be 0.7V.*

Many more realization of logic functions with standard electronic components are on the market today, The two most common are the Transistor-Transistor-Logic (TTL) and Complementary Metal Oxide Semiconductor (CMOS) circuits:

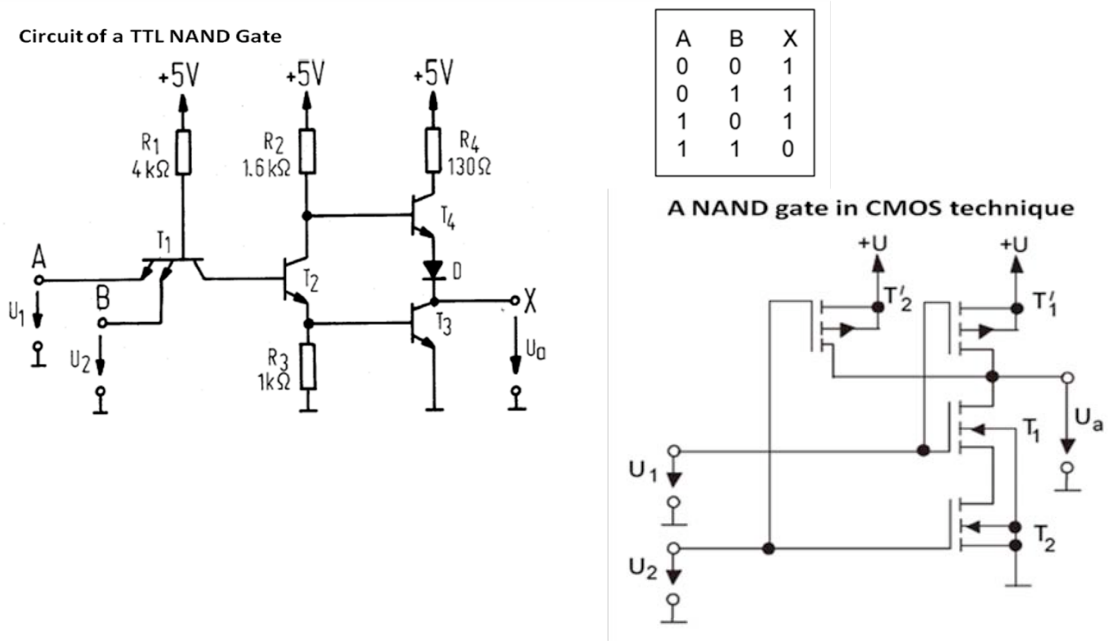


Fig. 2: *NAND Gates with transistors (TTL, left) and with complementary MOSFETs (CMOS, right).*

In terms of applications, logic circuits are nowadays not only used in the processor units of computers. Large markets are open for the integration of simple logic circuits on “smart cards”, where the magnetically stored security code is replaced by a processing unit (RISK processors, for example) with a logic function.

The next level of complexity is reached in, e.g., hand held devices, where PIN and other personal codes are used for securing only authorized use. For this, logic based on SRAM and / or FLASH is used.

The main disadvantages of these systems are either a relatively high power consumption and non volatility (CMOS) or a relatively expensive technology mix in as, e.g., the case of SRAM and FLASH.

Magnetic systems are used already for decades for data storage in hard disk drives and tapes. Here, the hysteresis of the magnetization is used for storing data:

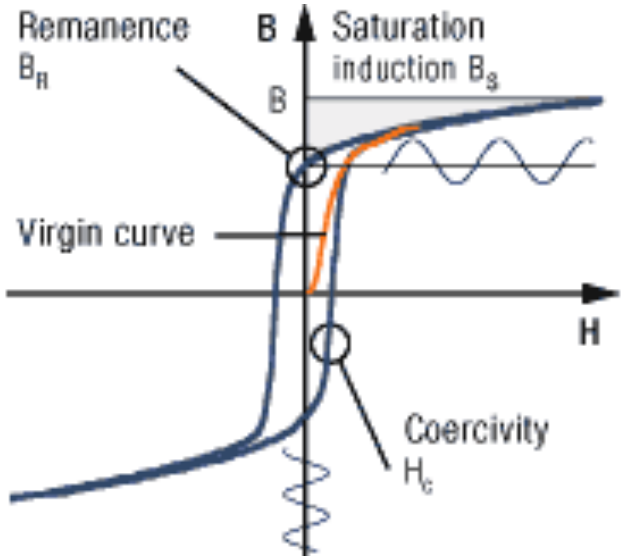


Fig. 3: *A magnetic hysteresis curve and the explanation of the related parameters.*

The two stable states at H=0 (see fig. 3) can be used to store digital information. This is realized in thin ferromagnetic films on hard disks and tapes.

The idea to also process data with magnetic systems is quite old: Back in the late 1960'ies and early 1970'ies, ideas to process information with magnetic bubbles have been discussed and realized [1]. These magnetic systems, however, failed to become a major tool for data processing due to the enormous success of the semiconductor based microelectronics.

Today, however, the integration of computers into every day's life requires smaller, more powerful but less power consuming processing units. According to the ITRS roadmap, devices based on Si-only technology will face serious obstacles hindering downscaling and power reduction within a few years. Parallel data processing, faster interactions between logic functions and memories, and re-configurable logic functions are needed in particular for applications that require image processing.

In modern Magnetic Logic the direction of the magnetisation in sub-micron patterned ferromagnets is identified with Boolean logic 1 or 0. The input changes the magnetisation of one structure and the output – i.e. the logic function – is then defined by the layout and the magnetisation states of the subsequent structures. As first demonstrated by groups in England [2] and Germany [3], Magnetic Logic promises low power non-volatile computing, "on the fly" re-programmability, massively parallel character and ultrafast operation and thus has the potential to become a key technology for logic circuits. Magnetic Logic could also bring together real time reconfiguration and universal memory - key success factors compared to other semiconductor techniques for re-configurable logic.

1.2 The approaches to Magnetic Logic

Three complementary approaches to Magnetic Logic (ML) will be discussed here:

- In Approach 1, magnetic dots are neighbouring each other closely and thus have a dipolar interaction. Thus magnetizing one of these dots will influence the neighbours. With a proper arrangement, information on the magnetic state of one dot can propagate along one or several chain(s) of these patterns and finally be detected by optical or electronic means. This approach was named “Quantum cellular Automata” [2].

- In Approach 2, driving magnetic domains in a circuit enables to convert input currents creating a magnetic field on one end of the circuit into an output represented by the magnetic state of the nanostructure at the other end. The sense of rotation - clockwise or counter-clockwise - gives control over the flow of information [3]. This scheme can be highly parallel and of low cost. Readout can be performed by optical means, or magnetic GMR- or tunnelling elements, which directly convert the magnetic state into an electrical signal.

- In the Approach 3, the logic gates are built from these magnetic tunnelling junctions that are arranged in a bridge type configuration. The logic input switches the magnetisation of one electrode. The output produced by the magnetic states of the nano-structured elements is the voltage drop across the gate which can be programmed to be logic e.g. NAND, AND, NOR by setting the magnetizations of reference cells, thereby promising the ability to perform re-configurable computing [4].

In any approach to logic operations, two routing functions are also required for complex logic circuits: a fan-out structure, which makes two identical copies of an input signal; and a cross-over structure, which allows two signals to pass over each other without interference.

2 The Quantum Cellular Automata

A Quantum Cellular Automata (QCA) is a nanostructure-compatible computation paradigm that uses arrays of quantum-dot cells to implement digital logic functions. As it was shown in [5], a configuration of single electron transistors can be QCA and shows the ability to perform logic operations. These devices, however, only work at very low temperatures.

The group of R. Cowburn and co-workers demonstrated [3] experimentally that QCA architectures using relatively large ferromagnetic dots (~100 nm) can be made to work at room temperature.
Each QCA network (see fig. 4) consists of a single large input dot followed by a chain of 69 small circular dots (diameter 110 nm), placed on a pitch of 135 nm. The dots were made by e-beam lithography from a magnetically very soft material ("Supermalloy").

Fig. 4. *Scanning electron micrographs of the left (input) region of two of the room temperature magnetic QCA networks. The large dot at the left end is the input dot.*

By definition, a logic 1 is signaled when the dot's magnetization vector points to the right, and a logic 0 when it points to the left (note, that the magnetization wants to point along the chain due to magnetostatic interactions). The classical magnetostatic interactions are then used to propagate information along the chain of dots. The system is thus binary, because only right- and left-pointing magnetization states are stable states.

The elongated shape of the input dot at the left end of the chain (fig. 4) increases its shape anisotropy and thereby the switching field. Thus the input dot is still able to influence the neighboring circular dots but remains unaffected by them.

For propagating information, the state of the input dot is set by an external magnetic field pulse and remains stable thereafter. Then, the magnetization of the neighboring circular dot will switch, influencing the next one. Thus the information will propagate along the chain until the last dot at the very right end of the chain has switched.

Probing the information can be performed by, e.g., the magneto-optical Kerr effect, where the polarization state of light gives information about the magnetization. The fact that the signal shown in fig. 5 has an amplitude of ~±60 shows that almost the entire chain (69 dots) is switching. Thus, the magnetization direction of the single input dot determines the magnetic

state of all other dots. Signal propagation along the length of the chain is thus clearly occurring.

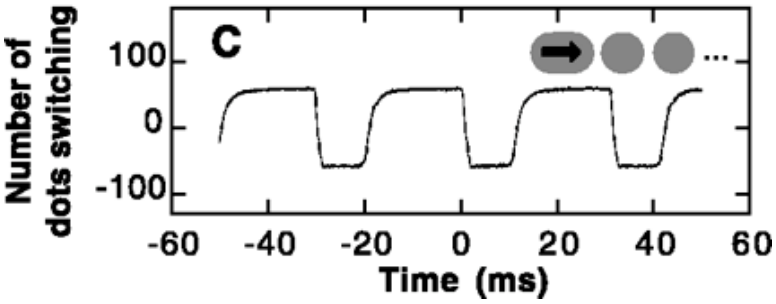


Fig. 5. *Room temperature experimental results from the magnetic QCA networks shown in fig. 4 for the input dots set to 1, calibrated in number of dots switching within one network. The width of the edge transitions is not due to the temporal response of the magnetic QCA networks but rather the distribution in switching fields, combined with the finite switching time of the applied oscillating field.*

In this approach, a very simple magnetic QCA network was chosen as proof of the principle that logic operations and information propagation can be performed at room temperature. In this particular network, the first circular dot acts as an AND gate, where the oscillating clock field is one input and the elongated dot is the other input. The circular dots 2 to 69 act as output interconnects.

There is, in principle, no physical problem in splitting the magnetic soliton running along the chain into two channels, and so fan-out of signals should also be possible. Simulations show that NOT functions, concatenated gates, and crossing chains are also possible.

3 The domain wall approach to Magnetic logic

An alternative to propagate pure spin information without moving charge was presented in [3]: A magnetic logic architecture referred to as "domain-wall logic" was developed which uses no transistors and exhibits very little heating caused by data switching. A domain wall is a mobile interface between regions of oppositely aligned magnetization. Submicrometer planar nanowires made from a soft magnetic material such as Permalloy ($Ni_{80}Fe_{20}$) are excellent conductors for domain walls [6], because the high shape anisotropy of the nanowire forces the magnetization to align with the long axis of the wire. Such domain walls can be propagated through complex networks of nanowires under the action of an externally applied magnetic field. This field rotates in the plane of the device and acts as both the clock and the power supply.

The main operating mechanism for processing the information of the spin system is the reversal of the magnetization direction in a cusp-shaped planar nanowire (see fig. 6A to E). If

a logical "1" is defined as the magnetization pointing in the direction of propagation of domain walls and a logical "0" as the opposite, the cusp in effect performs the logical NOT operation.

In order to implement any arbitrary logic function, additional basis functions must be implemented. At least function with two inputs, such as AND or OR, is required to complement the NOT function so that any computational calculation can be performed.

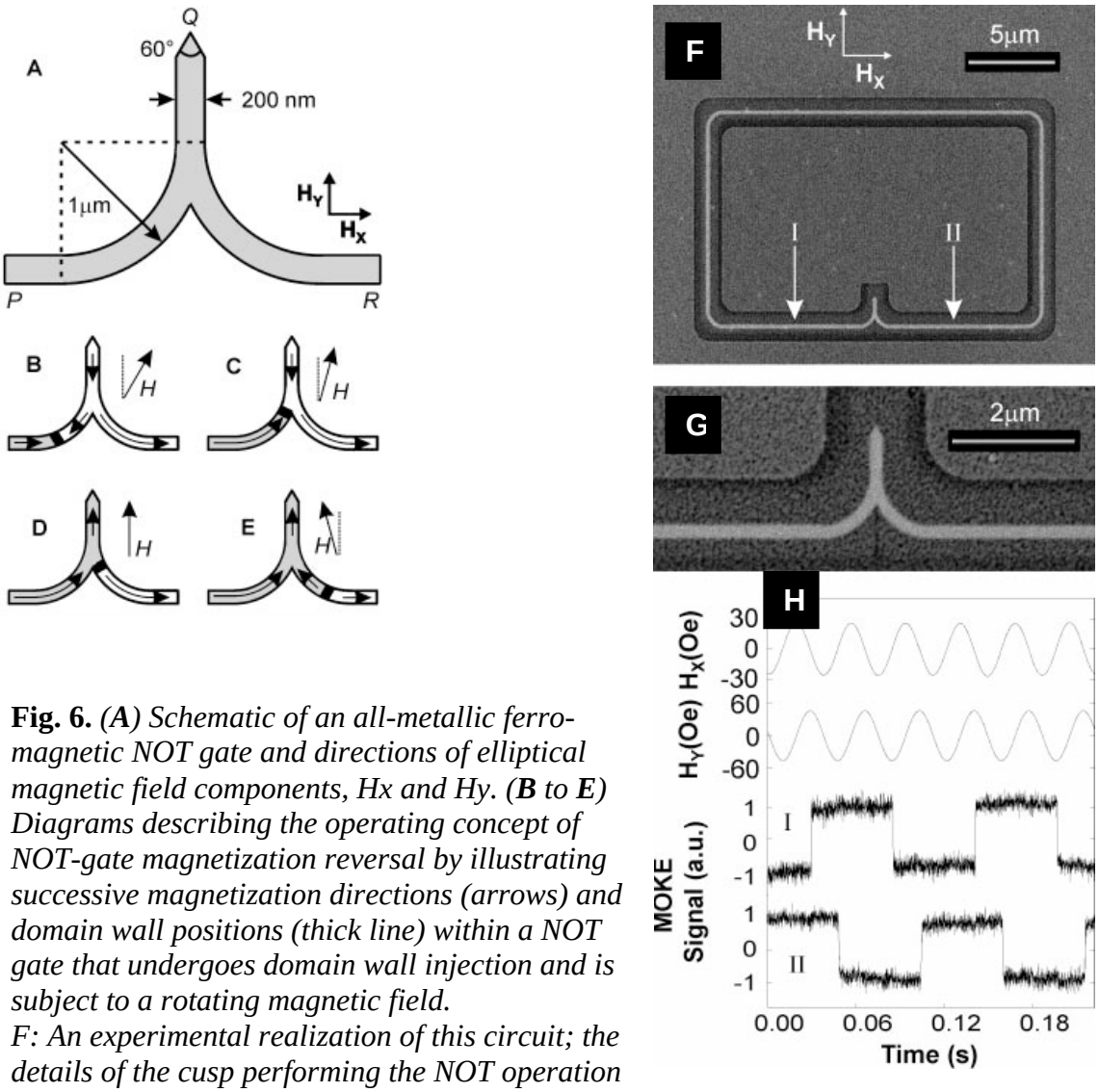


Fig. 6. *(**A**) Schematic of an all-metallic ferromagnetic NOT gate and directions of elliptical magnetic field components, Hx and Hy. (**B** to **E**) Diagrams describing the operating concept of NOT-gate magnetization reversal by illustrating successive magnetization directions (arrows) and domain wall positions (thick line) within a NOT gate that undergoes domain wall injection and is subject to a rotating magnetic field.*
F: An experimental realization of this circuit; the details of the cusp performing the NOT operation is shown in G. The graphs in H show the oscillating fields H_x and H_y and the magneto-optical signals at points I and II of the device shown in F.

In fig. 6 F-H, an experimental realization of this basic circuit for the domain wall logic is shown: A narrow Permalloy ($Ni_{80}Fe_{20}$) wire is patterned by e-beam lithography on a Si substrate to a rectangle, where in the lower rim, the cusp is integrated (fig. 6 G). The observation of the processing of the spin information can be done again by magneto-optical Kerr effect. Here, the laser was placed on two spots marked by I and II in fig. 6 F. In the results shown in fig. 6 H, a clear inverted magnetization direction is resolved when going from spot I to spot II. Thus the cusp between these two spots acts as a NOT function for the domain wall coming in from the left side.

This basic circuit thus demonstrates the capability of performing logic operations with domain walls moving around in a system of magnetic nanowires. More sophisticated circuitries where constructed (see fig. 7), showing also the crossing, fan in and fan out functionalities [3].

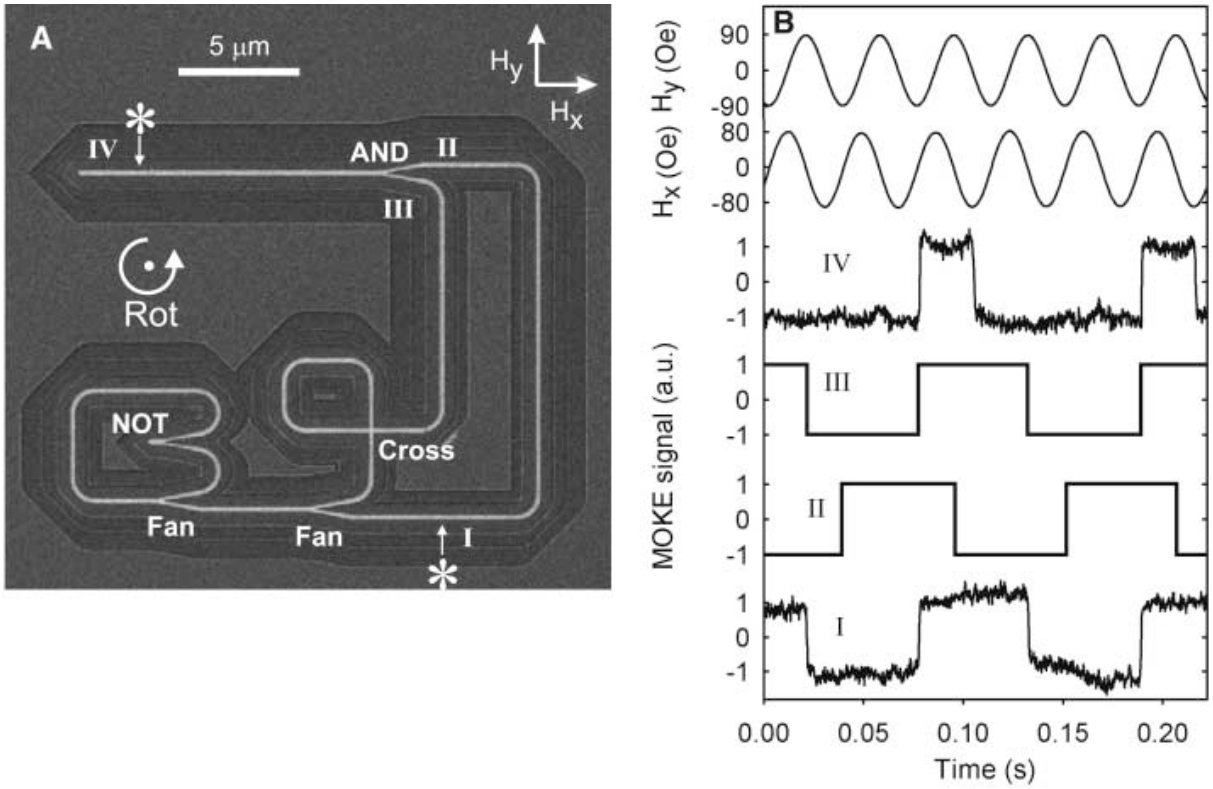


Fig. 7. *(A) Focussed Ion Beam image of a magnetic nanowire network with one NOT gate, one AND gate, two fan-out junctions, and one cross-over junction. MOKE measurements were made at positions I and IV, indicated by asterisks, and positions II and III denote the inputs to the AND gate. Also indicated are the directions of field components (Hx and Hy) and the sense of field rotation (Rot). (B) MOKE traces describing the operation of the magnetic circuit within a counterclockwise rotating field with amplitudes H_x^0= 75Oe and H_y^0= 88Oe and dc offset of H_x^{DC} = –5 Oe. Experimental MOKE measurements from positions I and IV of the circuit are shown. Traces II and III are inferred from trace I and show the magnetization state of the AND gate's input wires. (from [3]).*

Thus four basic functions (NOT, AND, fan-out, and cross-over) operating simultaneously under the action of a single global applied rotating magnetic field have been demonstrated in the domain wall approach. Using these basic functions, any logic circuit can be implemented using complex combinations of the four basis functions.

For a fast and reliable readout, however, magnetoresistive elements need to be integrated in minimum feature size (35). One additional drawback is the outer magnetic field needed for operating the logic gates. Here, the domain-wall propagation by spin transfer [7] may ultimately overcome these inefficiencies.

4 Magnetic Tunnel Junction Logic

The last approach to be discusse is the use of magnetic tunnel junctions for creating logic circuits [4]. Such junctions are already under development for several years for the Magnetic Random Access Memory (MRAM) or sensors. The realization of a logic based on the same principles and technologies as the –potentially universal- memory MRAM is of large interest, because it opens the way to a common technology platform for storing and computing data. Moreover, magnetic logic gate arrays can be field programmable, leading to field programmable logic gate arrays (FPGAs). Such FPGAs are programmable ”on the fly” and thus open also a path to fast reconfigurable computing[11].

The general concept of using MTJs for logic operations is visualized fig. 8.

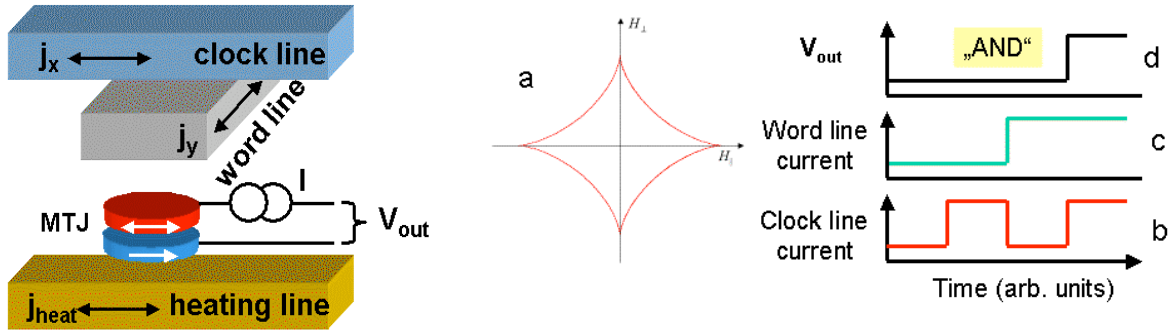


Fig. 8: *Left: The simplest logic gate consisting of only one tunneling cell (MTJ). The inputs are the clock- and the wordline and optionally a current carrying line heating the cell. The output is the voltage drop across the MTJ.*
Right: The Stoner astroid of a idealized soft layer relevant for the switching with two perpendicular field pulses from the clock- and the wordline. b) and c) The time dependence of the currents on clock- and wordline, respectively and d) the output voltage V_{out} if only the combination of clock- and wordline field can switch the magnatization.

A current flowing through one (or more) magnetic tunnel junction(s) causes a voltage drop V_{out}. Because the resistance of the MTJ depends on the relative orientations of the

magnetizations of upper and lower electrode, this V_{out} will change, when one of the electrodes switches its magnetization (indicated by the double arrow in fig. 8). Generally, the resistance is higher in the antiparallel magnetization state as compared to the parallel one. Switching of the magnetization can be accomplished by several means. The simplest is driving current through word- and clock lines close to the MTJ. These currents create a magnetic field which can switch the magnetization of the MTJs soft electrode. Other possibilities are heating the MTJ so that also the hard electrode can be switched magnetically.

A scheme of operation of this simplest logic gate with only one MTJ is shown in fig. 8 (right): Because the switching condition of the magnetization of the soft layer is given approximately by the Stoner asteroid, only the simultaneous presence of both currents on clock- and wordline causes switching if appropriate values for the currents are chosen. Thus the output voltage V_{out} will change only if both currents are present. If the MTJ is switched by this procedure to the high resistive state (antiparallel magnetizations), then V_{out} represents the logic "AND" relation between the clock- and word-current. In the reversed case of switching from the high to the low resistive state, the function is "NAND". Thus in this simplest version, already two functions can be realized and by using the clockline current, also clocked operation of this circuit is possible.

In an extension of this scheme, Ney *et al.* [8] discussed a concept, where only one tunneling junction can form a programmable logic gate with more than these two functions, if it is possible to also switch the hard (fixed) magnetic electrode. This could be accomplished by, e.g., using the heating line sketched in fig. 8.

Here, we concentrate on the first approach using a bridge type circuit consisting of four tunneling junctions (fig. 9).

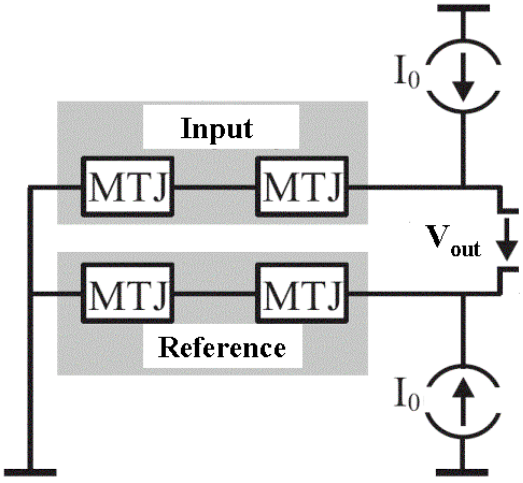


Fig. 9: *Schematic circuit of a field programmable logic gate array consisting of two input MTJs and two reference MTJs. The value of the current driven through all MTJs is I_o as shown. The output V_{out} is the voltage difference between the input and the reference chain.*

The preparation of sub-µm logic gates with four MTJs as shown in fig. 9 requires a series of relatively sophisticated e-beam processes. First, the MTJs are patterned followed by the preparation of the upper contact lines of the MTJs. Then, the clock- and word lines are prepared with insulating 100nm thick SiO_2 separating both the contact lines of the MTJs as

well as clock- and wordline. After that, the complete structure is covered by a TaO_X protection layer (5nm). The chips with four nominally identical logic gates are then cut and wire bonded in a conventional chip socket. In our layout, each MTJ can be also characterized individually in order to evaluate variations of the resistance and the TMR within one gate.

Examples for the result of this procedure are shown in fig.'s 6a and 6b: A complete logic gate consisting of four electrically connected MTJs as already shown in fig. 9 (fig. 6a) and one individual tunneling cell with a size of about 200nm x 100nm (fig. 6b). Provided all lithography steps are successfully done, the properties of the MTJs do not change with respect to their expected resistance and TMR within the margins found already on extended samples.

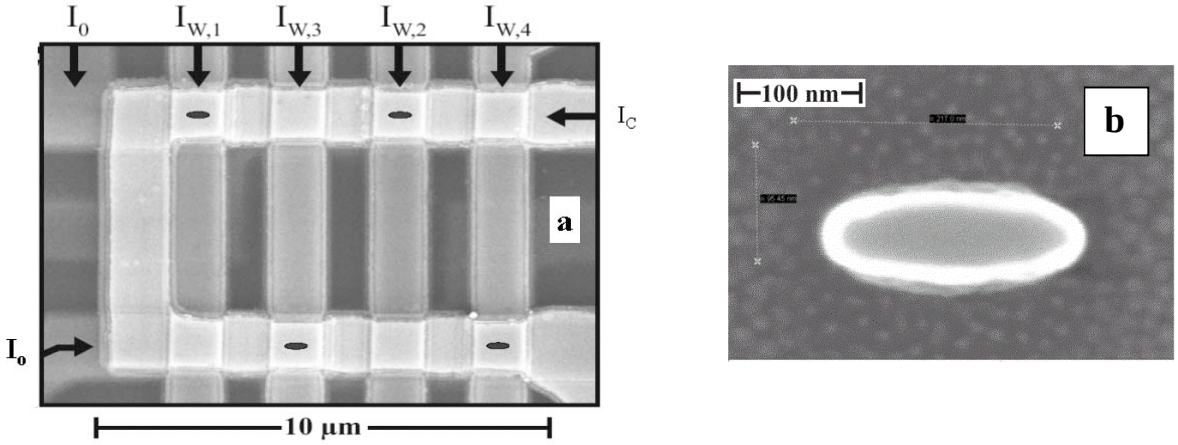


Fig. 10: a) SEM image of a field programmable logic gate made by subsequent e-beam lithography steps. The burried magnetic tunnel junctions are indicated by black ellipses. I_o is the current running through the upper (input) and the lower (reference) MTJs. I_c is the clock-current and I_{W1-4} are the word-currents for switching the MTJs. b) SEM image of the smallest tunnel junction employed in the FPGAs (elliptic, 200nm x 100nm).

For the logic operations, current pulses are passed through the wordlines on two of the MTJs, which create a magnetic field able to switch the soft electrode of two MTJs connected in series. In fig. 11, we show the resulting resistance of this chain as a function of the applied current pulses in the two wordlines (I_{W1} and I_{W2} in fig. 10a, respectively).

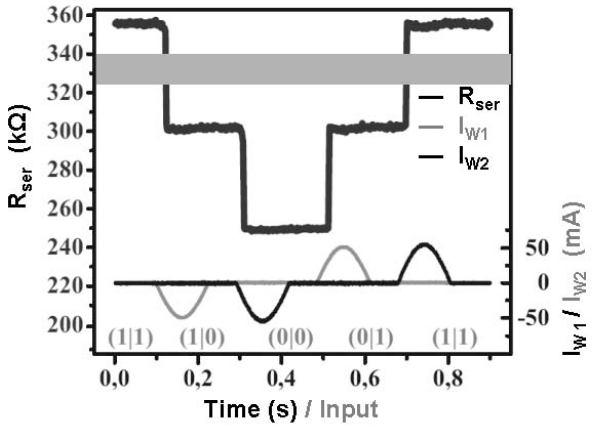


Fig. 11: *Resistance of the series of two (input) MTJs upon switching. The resistance takes the distinct values corresponding to the states R_H/R_H, R_H/R_L, R_L/R_L and R_H/R_L of the MTJs.*

The resistance changes according to the switching of the individual MTJs soft layers between a maximal value with both MTJs being in the high resistive state R_H to a mixed state (R_H / R_L), the minimum resistance value (R_L / R_L), the next mixed state (R_L / R_H) and then back to the original state (R_H / R_H). If one identifies resistance values smaller than –in this case- 320kΩ with logic 0 and larger than 340kΩ with logic 1, then this corresponds to an "AND" function relating the two (wordline-) inputs. This would be then the non-programmable operation of only two MTJs. For programming the function, the two other MTJs (connected also in series) are necessary.

This mode of operation (as sketched in fig. 9) is demonstrated in fig.12:

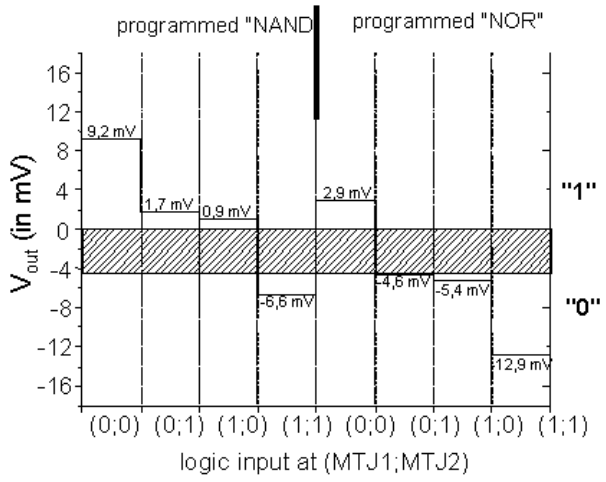


Fig. 12: *Output voltage V_{out} of a complete FPGA (compare fig. 5) upon changing the reference MTJs from the state R_L/R_L to R_L/R_H corresponding to a programmed "NAND" and "NOR", respectively.The function is veryfied by switching the input MTJs with appropriate word- and clock-currents. The shaded area marks the gap necessary between V_{out} corresponding to logic 1 and 0. Note, that in this case the gap amounts to only 5mV.*

Here two functions (NAND and NOR) have been implemented by setting the two reference-MTJs in fig. 9 to the states (R_L / R_L) for the NAND function and (R_L / R_H) for the NOR.
As can be seen from fig. 12, the programming action changes the output of the logic gate array from one Boolean function to another. Because the programming action consists of magnetically switching the soft electrode of one of the reference MTJs, this is as fast as transferring the input from the current pulses to the resistance state of the input-MTJs. Moreover, the state of the logic gate array (i.e. inputs and output) are stored in a nonvolatile manner in the MTJs.

Although these gates thus can provide very flexible field programmable logic gate arrays for reconfigurable computing, the scalability issue is critical: Simply spoken, the currents in the word- and clocklines are too large (some mA) to allow for a downscaling in the deep sub-µm range.

The solution for this obstacle could be current induced magnetization switching [9] (CIMS), where, however, the area-resistance product of the barrier must not exceed around 50Ωµm², because only then enough current can be pressed through the MTJ without destroying it. We

have therefore tried to prepare low resistive MgO based tunneling elements. In fig. 13, we show a typical minor loop for this structure, where, the area resistance product is 6.5Ωμm² allowing for current density of up to around $1.3 \cdot 10^7$A/cm². Nevertheless, the TMR is around 126%, i.e. well above the threshold of 100% for a reliable production.

For these junctions, we carried out CIMS experiments, i.e. the voltage across the MTJ was pulsed to a certain value and then the resistance was measured at a low voltage of 20mV. The result of this attempt is shown in fig. 12, where the applied voltage pulse was already replaced by the corresponding value of the current density.

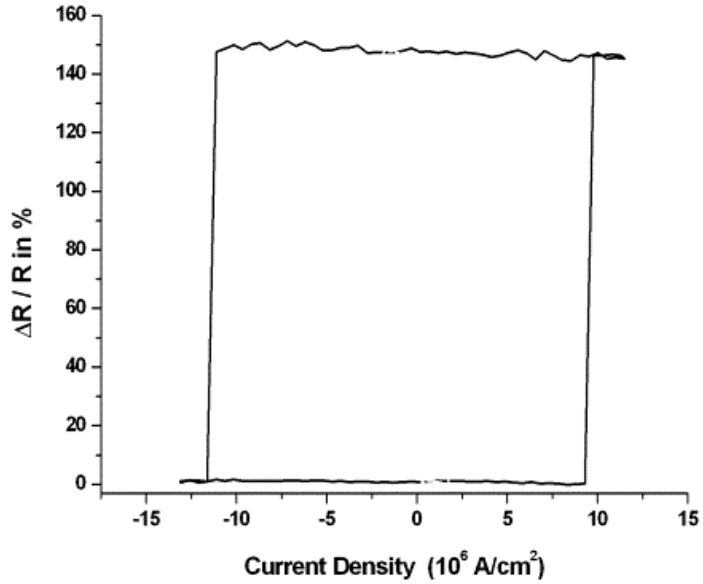


Fig. 13: *A minor loop of a tunnel junction taken by current induced switching. Here, a current pulse was applied to the junction and the resistance was subsequently determined at low bias voltage (20mV). The junction switches from R_L to R_H (antiparallel state) at $9 \cdot 10^6$A/cm² and back to R_L at $13 \cdot 10^6$A/cm².*

Pulsing a current density between $8 \cdot 10^6$A/cm² and $13 \cdot 10^6$A/cm² through these MTJs leads to switching back and forth the magnetization of the soft layer between parallel and antiparallel alignement, respectively.

Because this switching current in CIMS scales down in parallel with the area of the tunneling junctions, these MgO based tunnel junctions seem to be the ideal candidates for the realization of working and economically producable field programmable logic gate arrays.

Because these junctions also are extremely stable at high temperature, they could open a way to also switch the hard magnetic layer by thermally changing the exchange bias direction. Moreover, for Heusler materials as magnetic electrodes[19] it was found, that a change of the sign of the TMR can be found if the bias voltage is changed from low to high values [10]. Thus, using these effects as new ways for changing the functions of the FPGAs, new degrees of freedom can be introduced which can be used to further increase their variability.

5 Summary

Conventional microelectronic technology has relied on shrinking transistors to produce increasingly smaller, faster and more powerful computers. However, because the laws of physics prevent conventional devices from working below a certain size, that method is nearing its physical limits.

In this lecture, we have discussed three approaches to overcome the limits of traditional downscaling by using the spin of the electrons rather than their charge for processing information.

The **Quantum Cellular Automata** molecular chip could contain as many as 1 trillion devices, as opposed to the 6 million devices in the most advanced conventional chip. And since it does not rely on flowing electrons to transmit a signal, no electric current is produced and heat problems are avoided.

One of the most attractive features of **Domain-Wall Logic** is its great simplicity. Logical NAND is achieved in CMOS by using four transistors, whereas domain-wall logic uses two elements (NOT and AND). A logical AND function, usually requiring six CMOS transistors, can be achieved simply by bringing two magnetic nanowires together. Unlike Si CMOS architecture, a domain wall cross-over junction can be achieved in a single plane and without multilevel metallization, which means that in principle, extremely low-cost devices could be produced. Whereas most applications of magnetic logic (and indeed the wider field of spintronics) will involve a hybrid system on a chip that includes silicon-based CMOS, certain applications such as biomedical implants or wearable computing hardware would benefit from the ability to fabricate devices on, for example, flexible polyimide substrates. One could imagine nanowires constructed into three-dimensional (3D) neural networks or hugely dense 3D nonvolatile memories. The ability to supply power, clock, master reset, and serial input with a single externally applied magnetic field, is particularly attractive in the 3D case, where signal access is limited. There may also be potential for interfacing domain-wall logic with emerging dilute ferromagnetic semiconductors that allow electrical control and sensing of magnetization.

The perspective of the **Magnetic Tunnel Junction Logic** is the integration with the universal MRAM memory on one technological platform and its ability to form field programmable logic circuits with all basic logic functions and additional operations such as XOR, which today require a large number of transistors. Because the logic function can be programmed at the same speed as the logic operation, reconfigurable computing schemes could be realized in this technique. In high performance computing, such schemes could accelerate the operation of processing units by a factor of 100 if the program optimizes the functions of the processor during operation. At the same time, the power dissipation, which is one of the most critical issues in nowadays high performance computing, could be reduced by orders of magnitude enabling an even better scalability of this approach.

Acknowledgements

We thank W. Maass, B. Ocker and J. Langer from Singulus Technologies for supplying MgO based MTJs and valuable discussions. The work and results reported were obtained partly with research funding from the European Community (6. Framework, Contract Number 510993). The views expressed are solely those of the authors, and the other Contractors and/or the European Community cannot be held liable for any use that may be made of the information contained.

References

[1] See, e.g., R. C. MINNICK, P. T. BAILEY, R. M. SANDFORT, W. L. SEMON, Proceedings of the December 5-7, 1972, fall joint computer conference, Anaheim, California, part II, page 1279 ff
[2] R.P. Cowburn, M. E. Welland, Science 287, 1466 (2000)
[3] D. A. Allwood, G. Xiong, C. C. Faulkner, D. Atkinson, D. Petit, R. P. Cowburn, Science 309, 1688 (2005)
[4] R. Richter, L. Bär, J. Wecker, G. Reiss, Appl. Phys. Lett., 80, 1291 (2002)
[5] I. Amlani, A. O. Orlov, G. Toth, G.H. Bernstein, C. S. Lent, G. L. Snider, Science 284, 289 (1999)
[6] R. D. McMichael, M. J. Donahue, *IEEE Trans. Magn.* 33, 4167 (1997)
[7] M. Tsoi, R. E. Fontana, S. S. P. Parkin, Appl. Phys. Lett. 83, 2617 (2003)
[8] A. Ney, C. Pampuch, R. Koch, and K.H. Ploog, Nature 418, 509 (2002)
[9] J.C. Slonczewski, *J. Magn. Magn. Mater.* 159, L1 (1996)
[10] A. Thomas, D. Meyners, D. Ebke, N.-N. Liu, M.D. Sacher, J. Schmalhorst, G. Reiss, H. Ebert, A. Hütten, *Appl. Phys. Lett* 89, 012502 (2006)

Index

Symbole

90^o type-coupling, B1.7
2-dimensional electron gas, B3.12, B5.2
3Q-state, A3.35
3d-metals, B1.6

A

absorption of spin current, D3.5
addition energy, C6.7
adiabatic STT term, D3.26
Aharanov-Bohm measurements, C5.6
AMR, I1.4
Anderson model, C6.10
anholonomy angle, A7.2
anisotropic magneto-resistance (AMR), F1.16
anisotropic magnetoresistance, I1.4
anomalous g factor, A7.26
anomalous Hall conductivity, C2.3, C2.11, C2.22, C2.25
anomalous Hall effect, A7.17, C2.2, C2.5, C2.25
anomalous velocity, A7.20, C2.2, C2.10
anti-Stokes scattering, D2.7
antiferromagnetic giant magnetoresistance (AGMR), D3.27
antiferromagnetic interlayer exchange, B1.11
antiferromagnetic orders: AFM-A, AFM-G and AFM-E, A5.9
antiferromagnetism, A2.11
antisymmetric exchange interaction, A5.6
antivortex, D1.23
asymmetric Bloch wall, D1.24

B

backward volume (BV) geometry, D2.5
balanced function, E4.3
ballistic system, C5.12
bandstructure, A1.6
Bardeen's approach, B2.10
Bells inequality relation, B1.5
Berry connection, A7.5, C2.5, C2.7
Berry curvature, A7.5, C2.7, C2.10, C2.12, C2.14, C2.20
Berry phase, A7.4, C2.8, C2.9, C5.6
bilinear coupling, B1.7, D2.9
biquadratic coupling, B1.7, D2.9
Biquadratic exchange, A3.9
Bir-Aronov-Pikus mechanism, B6.7
Bloch functions, B5.2
Bloch point, D1.8
Bloch wall, D1.19
Bloch's theorem, A1.4
Bloch-equation, C6.16
Bohr-Sommerfeld quantization, A7.21
Bohr-van Leeuwen theorem, A2.2
Boltzmann equation, B6.20, C2.4, C2.6, C2.8, C2.10, C2.15, C2.20
Breit-Wigner resonance, C5.11
Brillouin light scattering (BLS), D2.6
Brillouin zone, B1.11, B3.6

C

C-state, D1.29
$Ca_{2-x}Sr_xRuO_4$, A4.9

caliper of the Fermi surface, B1.10
called giant magnetoresistance (GMR), B3.2
canonical quantization, A7.24
carbon, C5.2
carbon nanotubes, C5.2
Carr-Purcell, E3.16
charge and spin currents, C1.7
charge quantization, C6.3
charge transfer salts, A4.20
charge-degeneracy point, C6.7
charging energy, C6.4
chemical potential, B3.12, C1.4
Chern number, A7.6
circularly polarized photons, B3.7
cluster DMFT, A4.20
coercive field, D3.12
coherent control of spins, D4.9
coherent laser cooling of magnons, D4.10
competing interactions, D1.3
complex bandstructure, B2.15
complexity classes, E4.2
conductance quantization, C4.8
conduction band, B3.6
conduction electrons, B1.12
conductivity mismatch, B3.8, C1.15
confinement, B1.9
constrained density functional, A2.14
contact resistance, C1.6
controlled NOT gate, E1.12
controlled rotation (CROT), E1.14
Coulomb blockade, C6.4
Coulomb diamonds, C5.11
critical current, D3.8
cross-tie wall, D1.23
crystal field splitting, A4.10
crystal symmetry, A1.3
Curie temperature, A3.10
Curie-Weiss, A6.5
current density, B2.4
current polarization, D3.6
current-driven domain wall motion, D3.25
current-induced magnetization switching, D3.3
current-induced spin-density wave, D3.27
current-induced torque, D3.6, D3.28

D

D'yakonov-Perel' mechanism, B6.6, B6.19
Damon-Eshbach mode, D2.5
damping, D1.16
Datta-Das transistor, C4.2
decoherence, E1.15, E2.18
decoherence time, B3.15
demagnetization factor, D1.13
density functional theory, A1.10
Deutsch's algorithm, E4.3
DFT bandgap, A1.11
diamagnetism, A2.3
diamond structure, D1.27
diffusive transport, D3.7
dilute magnetic semiconductors, I1.12
diluted magnetic semiconductors, B3.4, B3.8
Dipole-dipole energy, A3.14
Dirac, C5.4
Dirac equation, A1.29, A2.4, E2.15
Dirac Hamiltonian, C2.23
Dirac point, C5.12
direct band gap, B3.6
displacive phase transitions, A6.7

DiVincenzo criteria, E1.16, E2.2, E3.2
Domain theory, D1.8
domain wall logic, F3.8
domain walls, D1.18
doping, A4.14
double exchange mechanism, A5.6
double pump excitation, D4.9
Dresselhaus term, B5.7
Dresselhaus-effect, C4.2
dynamical mean field theory, A4.3, A4.4
Dyson equation, A4.5
Dzyaloshinskii-Moriya interaction, A3.18, A5.6, D4.6

E

e-beam lithography, D3.5, D3.19
edge states, C4.6
effective field, D1.15, D3.10
effective magnetic field, D2.3, D4.3
effective mass equation, B5.2
effective spin polarization, B1.7
Einstein-Podolsky-Rosen (EPR) paradox, B1.2
electric current, B2.5
electric field, A2.2
electric polarization, A7.12
electrical resistivity, B1.12
electrical spin injection, B3.4, B3.8
electrically detected magneticresonance (EDMR), E3.16
electrochemical potential, C6.6
electroluminescence, B3.8
electromagnetic environment, C6.4
electron spin, E3.4
Elliott relation, B6.16
Elliott-Yafet mechanism, B6.6, B6.15
energy minimization, D1.9
entangled photons, B1.4
entangled state, E1.6
envelope-function, B5.5
epitaxial nanopillars, D3.17
equivalent circuit, C1.13
Europium chalcogenides, B4.6
exact diagonalization, A4.6
exchange bias, F1.18
exchange biasing, B1.12
exchange energy, D1.10
exchange field induced by tunneling, C6.16
exchange length, D1.22
exchange splitting, D3.28
exchange spring media, F1.11
exchange-correlation potential, A1.16
extended Gilbert equation, D3.9, D3.11

F

Fabry-Perot interferometer, D2.8
factorization, E4.9
Faraday effect, I1.4
Faraday rotation, B3.11
Fe/MgO/Fe, B2.13
femtosecond time-scale, D4.5
Fermi level, B1.6
Fermi surface, A1.9
Fermi's golden rule, E2.16
ferromagnet-semiconductor interfaces, B3.5
ferromagnetic semiconductors, B3.8
ferromagnetism, A2.11
fiber bundle, A7.6
field programmable logic gate array, F3.11
field-effect-transistors, C5.2
finite-size effects, D1.3

Flower state, D1.29
Four-spin exchange interaction, A3.9
fullerenes, C5.6

G

g-factor, E3.5
gap energy, B3.6
giant magnetoresistance, B1.11, I1.6
giant magnetoresistance effect (GMR), D3.3
Gilbert, I1.2
Gilbert damping constant, D3.10
Gilbert equation, D1.16
GMR ratio, B1.13
Goodenough-Kanamori rules, A5.6
graphene, C5.2, I1.14
Grover's algorithm, E4.4
GW approximation, A1.22
gyromagnetic ratio, D1.14
gyrotropic vortex motion, D1.33

H

Hadamard transformation, E1.9
halfmetallic ferromagnets, I1.11
Hall-bar geometry, C4.6
Hanbury-Brown-Twiss experiment, C5.15
Hanlé effect, B3.14
Hanlé polarization, B3.16
Hanle effect, B6.8
hard disk drive, F1.2
Hartree-Fock method, A1.10
head-to-head domain walls, D1.25
heavy hole states, B3.7
Heisenberg, A6.7
Heisenberg Model, D1.6
Heisenberg model, A2.15, A3.8, B4.6
hetero-structures, B5.2
heterostructures, A4.14, A4.18
Heusler alloy, I1.11
Heusler alloys, B3.8
Hg/Te, C4.4
hidden parameter, B1.5
Homochiral mesoscale spin spiral, A3.42
Hubbard bands, A4.10
Hubbard model, A4.4
Hund's rules, A5.4
hyperfine interaction, C5.6, E2.17

I

III-V semiconductors, B3.5
implantation, E3.11
improper phase transition, A6.12
inhomogeneous DMFT, A4.19
interaction, C4.2
interfaces, B1.11
interlayer exchange coupling, B1.7, D2.9, F1.19, I1.5
intrinsic mechanism, C2.5, C2.10, C2.20
inverse Faraday effect, D4.4
inverse GMR effect, B1.12
inverse SHE, C4.11
inversion about the mean, E4.5
isotopes, E3.3
itinerant picture, B1.6

J

Jullière, I1.7
Julliere model, B2.5

K

Kane concept, E3.8
Karplus and Luttinger theory, A7.17
Kerr microscope, B1.7
Kerr rotation, B3.11

kinetic equation, B6.21, C6.12
Kohn-Sham equation, A2.6
Kondo peaks, C5.8
kp-perturbation expansion, B5.4
Kramers degeneracy, B6.13

L

$La_{1-x}Sr_xTiO_3$, A4.14
Lamor precession, C5.19
Lanczos procedure, A4.9
Landau levels, C4.3
Landau structure, D1.27
Landau-Bütiker conductance, D3.27
Landau-Lifshitz equation, D4.3
Landau-Lifshitz torque equation, D2.3
Landau-Lifshitz-Gilbert equation, D1.17
Landauer formula, B2.6
Landauer-Büttiker formalism, C4.8
large-angle precessional modes, D3.13
Larmor frequency, B3.3, D3.10
laser excitation of spins, D4.4
$LaTiO_3$, A4.12
LDA+U model, A1.17
light hole states, B3.7
local charge neutrality, C1.5
local realism, B1.4
logic elements, F3.3
longitudinal recording, F1.4
Lord Kelvin, I1.4
Luttinger properties, C5.6

M

macrospin, D1.8, D3.14
magnetic anisotropy, A2.16
Magnetic anisotropy energy, A3.12
magnetic charges, D1.11
magnetic correlations, A4.22
magnetic domains, B1.7
magnetic field, A2.2
magnetic frustration, A4.20
magnetic insulator, B4.6
magnetic monopole, C2.14
magnetic random-access memory, B3.3, D3.2
magnetic recording, F1.2
magnetic recording head, F1.10
magnetic recording media, F1.5
magnetic ringing, D1.31
magnetic storage devices, D4.15
magnetic tunnel junction logic, F3.11
magnetic tunneljunction, B2.2
Magnetite, I1.2
magnetization dynamics, D3.2, D3.9
magneto-opitcal Faraday effect, D4.3
magneto-opitcal Kerr effect, B3.11
magneto-optic Kerr effect, B1.8
magneto-optical effects, D2.7
magneto-optical susceptibility, D4.3
magneto-optics, D4.3
magnetocrystalline anisotropy, A1.28, D1.12
Magnetocrystalline anisotropy energy, A3.13
magnetoresistance, B2.2, B4.10, B4.13, I1.4
magnetoresistance curves, C5.10
magnetoresistance effects, C5.8
Magnetostatic energy, D1.11
magnon, B1.7, D4.4
magntic vortex, D3.26
master equation, C6.13
Maxwell's equations, I1.2
Maxwells equations, B3.11
Mean-field approximation, A3.11
metallic antiferromagnets, D3.26
MgO barrier, D3.16

micro-BLS, D2.6
micromagnetism, D3.9
molecular-beam epitaxy (MBE), D3.18
Moore's law, E4.2
motional narrowing, B6.8, B6.19
Mott transition, A4.3
MRAM, B3.3, D3.2
multiferroics, I1.13
multilayer heterostructures, B3.2
multilayers, B1.11

N

N-V Center, C5.16
$Na_{0.3}CoO_2$, A4.16
nanopillar, D3.4, D3.8
non local method, C5.10
non-adiabatic STT term, D3.26
non-commuting variables, B1.3
non-locality, C4.8
non-uniform magnetization structures, D3.25
nonequilibrium resistance, C1.14
nonlocal spin-injection, C1.22
nonthermal optical control of magnetism, D4.2
NV-center, E3.4
Néel caps, D1.25
Néel wall, D1.20

O

Oersted field, D1.38, D3.5
one-dimensional systems, C5.6
optical pulses, D4.2
optical seletion rules, B3.7
opto-magnetic switching, D4.13
opto-magnetism, D4.3
orbital ordering, A1.19
orbital polarization, A4.12
orbital-selective Mott transition, A4.10
order-disorder transitions, A6.8
oscillation period, B1.10
oscillatory coupling, B1.9

P

parallel transport, A7.2
paramagnetism, A2.3
Peierls substitution, A7.25
period, E4.9
periodic boundary conditions, D1.5
perovskite, B4.8
perpendicular standing spin waves, D2.4
persistent charge current, A7.8
persistent spin current, A7.8
perturbation theory, B2.10
phase diagram, A4.2, A4.17, A4.20
phosphorus donors, E3.3
photon, B1.3
photon-assisted tunneling, E2.7, E2.10
point group, B5.3
Poisson equation, D1.11
polarization of light, D4.3
pole avoidance principle, D1.26
precession, D1.14
precessional switching, D1.32
proper phase transition, A6.12
pulse shaping, D4.11

Q

quality factor, D1.13
quantization of the spin wave mode, D2.13
quantized Hall conductance, C4.3
quantum algorithm, E1.16
quantum bit, E1.3
quantum cellular automata, F3.7

quantum charge pump, A7.14
quantum computation, E3.2
quantum dot, C6.5
quantum dots, C5.21
quantum gate, E1.7
quantum Hall effect, A7.14
quantum Hall effect (QHE), C4.2
quantum impurity, A4.5
quantum interference, D4.12
quantum of conductance, C6.3
quantum oracle, E4.4
quantum point contact, E2.3
quantum register, E1.5
quantum spin Hall effect, C2.24, C4.2, I1.9
quantum transport, C6.12
quantum well, B1.9, B3.8, B4.14, C4.4
quantum-well states (QWS), B1.10
quantum-wires, B5.2
quasi-antiferromagnetic resonance , D4.8
quasi-ferromagnetic resonance, D4.8
qubit, C5.21, E2.2, E3.8

R

racetrack memory, D3.26
Random phase approximation, A3.11
random telegraph signal, E2.5
rare-earth orthoferrites, D4.6
Rashba constant, B5.7
Rashba effect, A1.26, A3.19
Rashba-effect, C4.2
reentrant behavior, A4.21
resonant tunneling, B4.14
rigid band model, B1.6

S

S-state, D1.29
Savtchenko switching mode, D1.32
scanning electron microscopy with spin analysis (SEMPA), B1.9
scattering matrix, C2.16
scattering rate, C2.15
Schottky barrier, B3.4, B3.9
self-energy, A4.4
semiclassical dynamics, C2.4, C2.7, C2.8
semiconductor heterointerfaces, B3.5
semiconductor spintronics, B3.3
Shape anisotropy, A3.13
shape anisotropy, D1.13
Shor's algorithm, E4.7
side jump, C2.6, C2.7, C2.15, C2.16, C2.18, C2.19, C2.22
silicon, E3.3
Silsbee-Johnson spin-charge coupling, C1.16
single electron tunneling, C6.4
single spin precession, B3.14
single-domain ferromagnetic particle, A7.10
single-qubit gate, E1.8
skew scattering, A7.18, C2.5, C2.7, C2.21, C2.22, C2.25, C2.27
Slater determinant, A1.13
Slater-Pauling curve, A2.12, B1.6
spatial quantization, B1.2
spectrum of discrete energy levels, B1.10
speed of light, B1.3
spin accumulation, D3.7, D3.20
spin blockade, C6.18, E2.10
spin bottleneck, C1.14
spin centers, C5.13
spin coherence times, B3.5
spin coherent state, A7.10

spin dephasing, B3.15
spin extraction, C1.13
spin field-effect transistor (spin FET), B3.3
spin filter, B4.6
spin filter tunnel junction, B4.10, B4.13
spin filter tunneling, B4.3
spin filtering, D3.6
spin Hall effect, A7.26, C2.2, C2.3, I1.8
spin Hamiltonian, E3.4
spin injection, B4.16, C1.2, C4.9, C5.11, I1.8
spin injection efficiency, C1.12
spin LED, B3.5, B3.8
spin oscillations, D4.7
spin polarisation, B5.2
spin polarization, C1.2
spin polarized tunneling, B4.4
spin precession, B5.2
spin precession of a spin ensemble, B3.15
spin quantum computation, B3.5
spin quasichemical potentials, C1.26
spin relaxation, B6.2
spin relaxation length, B6.2, B6.3
spin relaxation time, B6.2
spin transfer, I1.10
spin transistor, B5.11
spin transport, B3.5, C5.8
spin valve system, B1.12
spin waves, D2.2
spin-accumulation, C6.16
spin-based electrons, C5.2
spin-charge coupling, C1.10
spin-dependent reflectivity, B1.9
spin-dependent scattering, B1.12
spin-energy relaxation, E2.15
spin-flip length, B6.3
spin-flip process, D4.4
Spin-flip scattering at impurities, B6.7
spin-flip scattering length, D3.8
spin-orbit (SO), C4.2
spin-orbit coupling, A1.24, A2.16, B3.3, B6.9, C5.6, D4.4
spin-orbit coupling constant, B3.6
spin-orbit interaction, B3.6, E2.15
spin-orbit interaction in semiconductors, B5.2
spin-polarization, B1.2, C6.14
spin-polarized currents, B3.2
spin-polarized transport, B2.5, C1.4
spin-precession, C6.16
spin-relaxation, C6.16
Spin-spiral state, A3.9
spin-split density of states, D3.6
spin-torque nano-oscillator (STNO), D3.16
spin-transfer torque, D3.3
spin-transfer torque dynamics, D3.3
spin-valve effect, C6.14
spin-valve structures, B3.2
spinel ferrite, B4.8
spinor, B5.9, D3.5
spintronics, B3.2, C5.6, D3.2
SQUID, B1.8
Sr_2RuO_4, A4.3
$SrVO_3$, A4.12
stability-diagram, C6.7
standing electron waves, B1.10
standing wave resonances, D2.13
Stark effect, C5.21
stationary vectors, B1.11
steady-state oscillatory modes, D3.12

Stern-Gerlach experiments, B1.2
STM lithography, E3.12
Stokes-process, D2.6
Stoner, A6.6
Stoner criterion, A2.10
Stoner model, A2.7, A3.4
strongly correlated electron systems, A4.2
Structure inversion asymmetry, A3.2
STT effect, D1.37
STT efficiency function, D3.11, D3.21
super-paramagnetic limit, F1.11
super-paramagnetism, F1.10
superexchange mechanism, A5.5
susceptibility, A2.10
SWASER, D3.29
synchronization of STNOs, D3.17

T

tetrahedral symmetry, B5.3
time-reversal symmetry, E2.17
TKNdN theory, A7.15
TMR, B2.2, C5.11, C6.15, D3.16
transition metal oxide, B2.19
transverse spin dephasing time, B3.15
tunnel contacts, C1.21
tunnel magneto-resistance) (TMR), F1.18
tunnelbarrier, B2.2
tunneling, B2.3
tunneling magnetoresistance, B2.2, I1.7
tunneling magnetoresistance (TMR), B3.3
two resistor model, B2.5
two-current model, B1.13, I1.4
two-qubit gate, E1.12
two-step switching process, D3.21

U

ultrafast laser pulses, B3.15
ultrafast laser sources, D4.2
ultrafast time-scales, D4.11
uncertainty principle of quantum mechanics, B1.3
unified description of GMR and STT, D3.19

V

V_2O_3, A4.15
valence band, B3.6
Van den Berg Scheme, D1.27
visibility, E2.7, E2.8
vortex core, D1.33
vortex core switching, D1.33
vortex polarization, D1.34
vortex-antivortex pair, D1.35

W

wavepacket, C2.8, C2.10, C2.17
wavepacket dynamics, A7.19, C2.10

Y

Yafet relation, B6.17

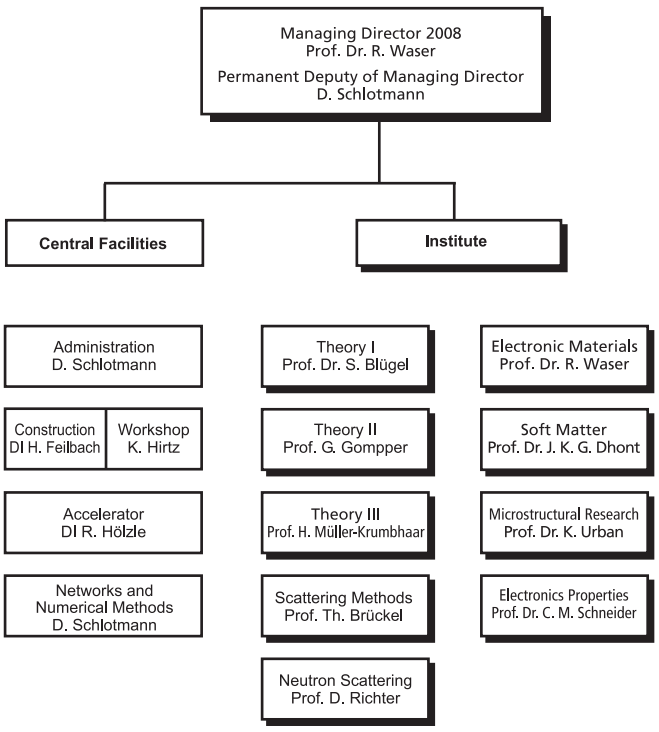
Z

Zeeman polarization, B3.10
Zeemann effect, B3.11
Zener tunnel junction, B3.8

Institute for Solid State Research (IFF)

52425 Jülich
Telefon +49 2461 61-4465
Fax +49 2461 61-2410
http://www.fz-juelich.de/iff

Organisation of the IFF



Founded in 1969, the scientific reputation of the Institute of Solid State Research (Institut für Festkörper-forschung – IFF) still owes much to the conception of its founders that new discoveries are made at the boundaries of disciplines. This is as true today as it was almost forty years ago. In this spirit, the IFF has pioneered new research fields and set trends towards multi- and cross-disciplinary activities in both fundamental research as well as technological innovations. This is highlighted by the 2007 Noble Prize in Physics, awarded to Prof. Dr. P. Grünberg, – jointly with Prof. Dr. A. Fert, Paris – for the discovery of Giant Magnetoresistance (GMR), an activity which triggered the new research field of spintronics.

Today, the IFF is engaged in investigating a multitude of condensed matter phenomena with special emphasis on three prime objectives:

- studies of fundamental physical mechanisms and phenomena of condensed matter,
- the development and improvement of experimental and theoretical analysis methods, as well as
- the elucidation and utilization of new material properties in complex systems.

The corresponding research programs exploit the full scale of analytical and numerical methods to elucidate relations between structural, electronic, magnetic and functional properties of condensed matter, and to achieve a description of the underlying physical mechanisms.

Research efforts are directed at obtaining a detailed understanding of phenomena, ranging from microscopic and atomistic interactions in the solid state to the competition of entropic and energetic forces in macromolecular fluids. Research at the IFF rests firmly on the two pillars of quantum mechanics and statistical physics. On a microscopic scale, we describe the interactions of electrons and atomic building blocks and determine how these entities respond to external influences. Particular strengths encompass the theory of electronic structures, clusters, micro-mechanics of lattice imperfections, the dynamics of structure formation and phase transitions, materials and phenomena

of magneto- and nanoelectronics, spintronics, spin dynamics, and strongly correlated electron systems. On the mesoscopic scale, we study soft matter systems, such as polymer solutions and melts, colloidal dispersions, surfactants, membranes, vesicles, and cells. Important questions concern self-assembly and structure formation, the competition of entropic and enthalpic forces, geometric and topological interactions, the complexity of multi-component systems, systems far from equilibrium, as well as the behavior under external fields. In all cases, the instrumentation of electron, neutron, and synchrotron sources and their application is essential to the study of condensed matter.

The experimental portfolio together with an acknowledged expertise enables the IFF to tackle complex problems in close cooperation with scientists and industry worldwide. Special state-of-the-art laboratories exist for thin film deposition and growth of bulk crystals, for nanostructuring as well as for the preparation of soft matter materials. In addition to standard methods for materials characterization, highly specialized techniques ranging from superconducting microscopy and spin-polarized microscopies to femto-second laser spectroscopy are available at the IFF and are being constantly improved in performance.

The Jülich Centre for Neutron Science (JCNS) operates advanced neutron scattering instruments at the worldwide most modern and highest flux neutron sources. As a complement to local research opportunities, instruments are designed and operated at external national and international neutron sources, such as the FRM-II in Munich or the neutron spallation source in Oak Ridge.

With the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons (ER-C) the IFF operates a national user facility housing several of the world's most advanced electron microscopes and tools for nanocharacterization. In-house research programs cover topical issues in

condensed matter physics and – as a matter of course – future developments of subångström and sub-electronvolt microscopy.

The newly founded Peter Grünberg Centre provides the competence and method platform for the research on fundamentals of future information technology. It is operated jointly by the institutes IFF and IBN in the framework of the Jülich-Aachen Research Alliance JARA. The Grünberg Centre houses several units. The Helmholtz Nanoelectronic Facility (HNF) provides manifold means to prepare complex material systems and nanoelectronic structures in combination with an in-depth characterization. In the Synchrotron Radiation Laboratory (SRL) a broad variety of spectroscopy, microscopy, and scattering experiments at various synchrotron radiation facilities are designed and realized. The lab also provides expertise for the development of new beamlines and experimental concepts and, thus, acts as a valuable partner for synchrotron radiation laboratories throughout the world.

The IFF puts particular emphasis on the support of young scientists and has successfully established a number of Young Investigator Groups in the fields of quantum transport, spinelectronics, thermoelectrics, complex liquids and multifunctional materials.

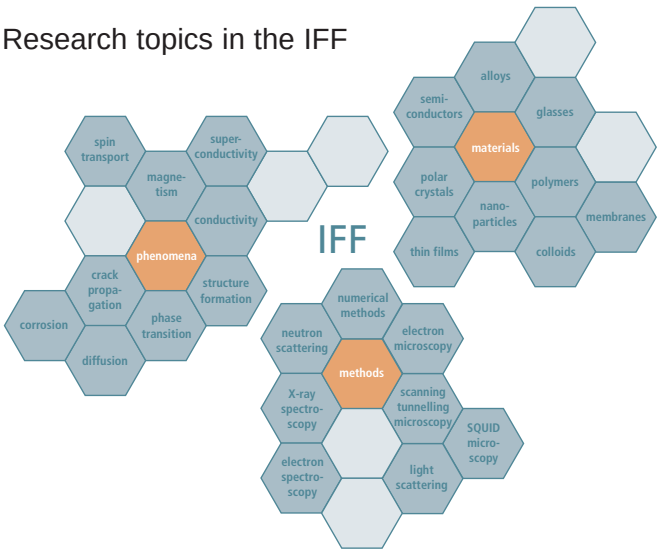
Last but not least, the IFF has a long tradition in teaching and training of students, not only through the approximately 30 IFF staff scientists steadily giving lectures at universities, but in particular through the annual IFF Spring Schools and Neutron Laboratory Courses. Dissemination of knowledge to students and scientist is provided through the organisation of workshops and conferences, in particular the yearly Jülich Soft Matter Days. In addition, the IFF is involved in the German Research School for Simulation Science (GRS) established together with the RWTH Aachen University.

The IFF is a department, which comprises six experimental and three theoretical divisions as well as joint service facilities. As part of the Research Centre Jülich – a member of the Hermann von Helmholtz Association of German Research Centres (HGF) – the IFF provides key contributions to the strategic mission of the HGF within three research programs:

- Fundamentals of Future Information Technology (FIT)
- Research with Photons, Neutrons and Ions (PNI)
- Molecular Systems and Biological Information Processing (BioSoft)

The success of the IFF rests upon the inventiveness and initiative of its more than 300 members. The IFF supports independent research by encouraging the responsibility of individual scientists – a philosophy that contributes greatly to the stimulating atmosphere in the department. In order to sustain this level on the long run, special encouragement is given to young scientists.

Research topics in the IFF



1. **Soft Matter**
   From Synthetic to Biological Materials
   Lecture manuscripts of the 39th IFF Spring School March 3 – 14, 2008
   Jülich, Germany
   edited by J.K.G. Dhont, G. Gompper, G. Nägele, D. Richter, R.G. Winkler (2008),
   c. 1000 pages
   ISBN: 978-3-89336-517-3

2. **Structural analysis of diblock copolymer nanotemplates using grazing incidence scattering**
   by D. Korolkov (2008), III, 167 pages
   ISBN: 978-3-89336-522-7

3. **Thermal Nonequilibrium**
   Thermal forces in fluid mixtures
   Lecture Notes of the 8th International Meeting on Thermodiffusion,
   9 – 13 June 2008, Bonn, Germany
   edited by S. Wiegand, W. Köhler (2008), 300 pages
   ISBN: 978-3-89336-523-4

4. **Synthesis of CMR manganites and ordering phenomena in complex transition metal oxides**
   by H. Li (2008), IV, 176 pages
   ISBN: 978-3-89336-527-2

5. **Neutron Scattering**
   Lectures of the JCNS Laboratory Course held at the Forschungszentrum Jülich and the research reactor FRM II of TU Munich
   edited by R. Zorn, Th. Brückel, D. Richter (2008), c. 500 pages
   ISBN: 978-3-89336-532-6

6. **Ultrafast Magnetization Dynamics**
   by S. Woodford (2008), 130 pages
   ISBN: 978-3-89336-536-4

7. **Role of Surface Roughness inTribology: from Atomic to Macroscopic Scale**
   by C. Yang (2008), VII, 166 pages
   ISBN: 978-3-89336-537-1

8. **Strahl- und Spindynamik von Hadronenstrahlen in Mittelenergie-Ringbeschleunigern**
   von A. Lehrach (2008), II, 171 Seiten
   ISBN: 978-3-89336-548-7

9. **Phase Behaviour of Proteins and Colloid-Polymer Mixtures**
   by C. Gögelein (2008), II, 147 pages
   ISBN: 978-3-89336-555-5

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies

10. **Spintronics – From GMR to Quantum Information**
    Lecture Notes of the 40th IFF Spring School March 9 – 20, 2009
    Jülich, Germany
    edited by St. Blügel, D. Bürgler, M. Morgenstern, C. M. Schneider,
    R. Waser (2009), c. 1000 pages
    ISBN: 978-3-89336-559-3

Vorlesungsmanuskripte des
17. IFF-Ferienkurses 1986

Dünne Schichten und Schichtsysteme

Einführung in die modernen Methoden der Herstellung, Modifikation, Mikroanalyse und technische Anwendungen dünner Schichten und Schichtsysteme.

Hauptthemen: Präparation · Haftung und innere Spannungen · Grenzflächen · Kohärente Vielfachschichten · Mikroanalytik · Schichten für Informationstechnik

744 Seiten, EUR 25,00

Vorlesungsmanuskripte des
18. IFF-Ferienkurses 1987

Synchrotronstrahlung in der Festkörperforschung

Einführung in die Anwendung der Synchrotronstrahlung zu Strukturuntersuchungen in der Festkörperforschung.

Hauptthemen: Quellen · Elektronentheorie · Strukturuntersuchungen · Absorptionsspektroskopie · Photoelektronenspektroskopie · Lithographie · Mikroanalytik

812 Seiten, EUR 25,00

Vorlesungsmanuskripte des
19. IFF-Ferienkurses 1988

Supraleitung und verwandte Quantenphänomene

Einführung in die Supraleitung und verwandte Quantenphänomene.

Hauptthemen: Supraleitung und Anwendungen · Supraflüssigkeit · Stark korrelierte Elektronen · Lokalisierung

786 Seiten, EUR 25,00

Vorlesungsmanuskripte des
20. IFF-Ferienkurses 1989

Computersimulation in der Physik

Einführung in numerische Methoden zur Simulation komplexer Systeme in der Physik.

Hauptthemen: Molekular-Dynamik-Methoden · Monte-Carlo-Methoden · Partielle Differentialgleichungen

1011 Seiten, EUR 17,90

ISBN 3-89336-013-1

Vorlesungsmanuskripte des
21. IFF-Ferienkurses 1990

Festkörperforschung für die Informationstechnik

Einführung in die materialphysikalischen Grundlagen der modernen Informationstechnik. Schwerpunkte: Physik der Halbleiter · Physik der Magnetspeichermaterialien · Funktion typischer Bauelemente · Industrielle Fertigung aktueller Bauelemente · Physikalische Aspekte zukünftiger Technik

1028 Seiten, EUR 17,90

ISBN 3-89336-033-6

Vorlesungsmanuskripte des
22. IFF-Ferienkurses 1991

Physik der Polymere

Einführung in die physikalischen Stoffeigenschaften von Polymeren und Kunststoffen.

Schwerpunkte: Struktur, Dynamik und Thermodynamik · Konformation · Kristallisation und Glasumwandlung · Gummi- und Viskoelastizität Molekulare Dynamik · Computersimulation · Polymermischungen · Werkstoffe

900 Seiten, EUR 23,00

ISBN 3-89336-055-7

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Vorlesungsmanuskripte des
23. IFF-Ferienkurses 1992

Synchrotronstrahlung zur Erforschung kondensierter Materie

Einführung in die Anwendung von Synchrotronstrahlung zur Bestimmung der geometrischen und elektronischen Struktur von Festkörpern und Grenzflächen.

Schwerpunkte: Erzeugung und Charakteristika der Synchrotronstrahlung · Instrumentierung · Wechselwirkung von Licht und Materie · Elektronische Struktur von Festkörpern, Oberflächen und Grenzflächen; Theorie und Experiment · Spinauflösende Messungen und Magnetismus · Halbleiter-Oberflächen und Schichtsysteme · Adsorbate auf Oberflächen · Streuung und Beugung von Röntgenstrahlung · Strukturanalyse mit EXAFS · Mikroskopie und Lithographie

ca. 900 Seiten, EUR 28,00

ISBN 3-89336-088-3

Vorlesungsmanuskripte des
24. IFF-Ferienkurses 1993

Magnetismus von Festkörpern und Grenzflächen

Einführung in den Magnetismus von metallischen Festkörpern und Grenzflächen.

Schwerpunkte: Konventionelle und moderne Messmethoden · Theorie und Experimente zum Band-magnetismus · Magnetische Domänen und Abbildungsverfahren · Volumen- und Grenzflächenanisotropie · Spinwellen · Phasenübergänge und kritische Exponenten · KKY-Wechselwirkung · Kopplungsphänomene und elektrischer Transport in Schichtsystemen · Werkstoffe

ca. 900 Seiten, EUR 28,00

ISBN 3-89336-110-3

Vorlesungsmanuskripte des
25. IFF-Ferienkurses 1994

Komplexe Systeme zwischen Atom und Festkörper

Einführung in die bei komplexen Systemen auftretende Phänomene und die zur Beschreibung verwendeten Konzepte und Theoretischen Methoden.

Schwerpunkte: Moleküle, Cluster, Fullerene (Dichtefunktional, Molekulardynamik, komplexe Energieflächen, Clusterspektroskopie) · Defekte in Quasikristallen, Fluiden, Mustern (Topologische Methoden, Energiefunktional) · Grenzflächen, Schichtsysteme (Adsorption, Benetzung, Aufrauhung, Reibung, STM, Kraftmikroskop, Heterostrukturen) · Komplexe Fluide (Hydrodynamik, Polymere, Polyelektrolyte, Kolloide, Membranen, Gläser) · Musterbildung, Chaos und Turbulenz (in Flüssigkeiten, chemischen Reaktionen und Biologie) · Poröse und granulare Medien (Transport, Nichtmischbare Flüssigkeiten)

1022 Seiten, EUR 30,70

ISBN 3-89336-128-6

Vorlesungsmanuskripte des
26. IFF-Ferienkurses 1995

Elektrokeramische Materialien

Grundlagen und Anwendungen

Einführung in die physikalischen und physikalisch-chemischen Grundlagen elektrochemischer Materialien, vor allem mit Perowskit- oder Perowskit-ähnlicher Struktur, die sich durch außergewöhnliche elektrische und dielektrische Eigenschaften auszeichnen.

Schwerpunkte: Kristallstruktur · Elektronenstruktur · Defekt-Physik und Chemie · Elektronische und ionische Transporteigenschaften · Mechanische Eigenschaften

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Untersuchungs- und Präparationsmethoden: Impedanz-, dielektrische und mechanische Relaxationsspektroskopie · Neutronenstreuung · Dünne Schichten · Sintern · Kristallzüchtung
Anwendungen: Dielektrische Schichten · Varistoren · PTC · Piezo- und pyroelektrische Keramik · Ferroelektrika · Optische Bauelemente · Ionenleiter und Brennstoffzellen · Oxidische Supraleiter
ca. 900 Seiten, EUR 33,20
ISBN 3-89336-146-4

Vorlesungsmanuskripte des
27. IFF-Ferienkurses 1996

Streumethoden zur Untersuchung kondensierter Materie

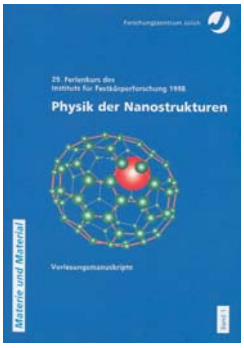
Einführung in die Wechselwirkung und Streuung von Strahlung zur Untersuchung der atomaren Eigenschaften von kondensierter Materie, insbesondere in die Streutheorie zur Interpretation der Meßdaten; Experimentelle Methoden der Diffraktometrie zur Strukturanalyse, der Kleinwinkelstreuung, der Reflektometrie und der Spektroskopie von Neutronen und Röntgenstrahlen. Strahlungsquellen und Instrumentierung. Relaxationsspektrometrie mit Licht.
Die Untersuchungsmethoden werden an den wichtigsten Stoffklassen demonstriert, beispielsweise Metalle, Legierungen, magnetische Verbindungen, Oxide, Gläser, Kunststoffe, Membranen und Proteine sowie Schichtsysteme von Polymeren und Halbleitern.
ca. 850 Seiten, EUR 35,80
ISBN 3-89336-180-4

Vorlesungsmanuskripte des
28. IFF-Ferienkurses 1997

Dynamik und Strukturbildung in kondensierter Materie

Einführung in dynamische Vorgänge und Strukturbildung in verschiedenen Formen der kondensierten Materie auf atomaren bis hin zu makroskopischen Längen- und Zeitskalen.
ca. 900 Seiten, EUR 38,40
ISBN 3-89336-204-5

Vorlesungsmanuskripte des
29. IFF-Ferienkurses 1998



Physik der Nanostrukturen

Vorstellung der Methoden, die es erlauben, die Nanostrukturen mit höchster Präzision herzustellen und zu charakterisieren. Vorstellung von Beispielen nanostrukturierter Festkörper und Oberflächen aus der aktuellen Forschung, in

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

denen es gelungen ist, neue Materialeigenschaften zu beobachten und auch teilweise schon technologisch nutzbar zu machen. Erforschung von Clustern, die entweder isoliert im Molekularstrahl oder als Cluster-Materialien vorgestellt und diskutiert werden.

ca. 1000 Seiten, EUR 40,00

ISBN 3-89336-217-7

Vorlesungsmanuskripte des
30. IFF-Ferienkurses 1999

Magnetische Schichtsysteme

Einführung in die Grundlagen der Wechselwirkungen, die zu kollektivem Magnetismus führen. Dazu gehört eine Einführung in die Dichtefunktionaltheorie und den Bandmagnetismus. Weiterhin wird eine Übersicht über verschiedene magnetische Materialklassen sowie relevante Messmethoden gegeben. Die Behandlung des Dünnfilmmagnetismus beginnt mit der Einführung der wichtigsten Präparationsmethoden und strukturellen Charakterisierungsmöglichkeiten. Daran schließt sich die Beschreibung der Eigenschaften des Magnetismus in Einzelschichten an, sowie deren Kopplungsphänomene in Multischichten. Besondere Beachtung erhalten Transportphänomene, die durch spinabhängige Elektronenstreuung und Reflektivität an den Grenzflächen oder einen spinabhängigen Tunneleffekt über isolierende Zwischenschichten verursacht werden. Auch die bereits in Einzelschichten auftretenden Anisotropien, sowie die Magnetostriktion werden theoretisch und experimentell behandelt, ebenso die Magnetooptik und der Röntgendichroismus.

Die wichtigsten Themen sind:
- Grundlagen von Magnetismus und magnetischen Materialien
- Präparation magnetischer Schichtsysteme
- Spinstrukturen in Schichtsystemen
- Zwischenschichtaustauschkopplung
- Magnetische Anisotropie und Magnetooptik
- Spinabhängiger Transport

ca. 1300 Seiten, zahlreiche z.T. farb. Abb.,
EUR 56,24

ISBN 3-89336-235-5

Vorlesungsmanuskripte des
31. IFF-Ferienkurses 2000

Femtosekunden und Nano-eV

Dynamik in kondensierter Materie

Die charakteristischen Längen- und Zeitskalen der Dynamik kondensierter Materie erstrecken sich über viele Größenordnungen: von ultraschnellen elektronischen Prozessen im Subfemtosekundenbereich über die Bewegung von Makromolekülen im Nano- bis Mikrosekundenbereich bis hin zu langsamen kollektiven Relaxationsprozessen im Stundenbereich, wie sie etwa bei Spingläsern auftreten.

Diese extreme Bandbreite, die in anderen Disziplinen ihresgleichen sucht, erfordert einen ganzen Zoo von experimentellen Untersuchungsmethoden, die auf der Wechselwirkung von elektromagnetischer Strahlung und von Teilchenstrahlung mit Materie beruhen. Durch revolutionäre Entwicklungen bei den Strahlungsquellen und neue raffinierte experimentelle Verfahren wurden in den letzten Jahren die Grenzen der Festkörperspektroskopie zu extremen Werten der Orts-, Zeit- und Energieauflösung verschoben.

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

So können heute Nanopartikel mit einer Zeitauflösung im fs-Bereich bzw. Energieauflösungen im neV-Bereich untersucht werden.

In dem IFF-Ferienkurs 2000 werden diese modernen experimentellen Verfahren vorgestellt, wobei als Sonden elektromagnetische Strahlung (von Infrarot- bis harter Röntgenstrahlung), Neutronen und Elektronen diskutiert werden. Viele Experimente wurden erst durch neue Strahlungsquellen realisierbar, wie Synchrotronstrahlungsquellen der dritten Generation oder Spallationsneutronenquellen.

Es werden die Eigenschaften dieser Quellen diskutiert, wobei Begriffe wie Kohärenz und Zeitstruktur eine zentrale Rolle spielen, und ein Ausblick auf Großprojekte im neuen Jahrtausend gegeben. Moderne experimentelle Verfahren wie fsec-Laserspektroskopie, neV-Neutronen-spinechospektroskopie, Röntgenkorellations-spektroskopie oder mehrdimensionale Kern-spinresonanz werden zusammen mit den wissenschaftlichen Anwendungsfeldern vorgestellt.

Die Dynamik der kondensierten Materie wird in ihrer ganzen Breite behandelt: von elektronischen Anregungen und Transport über die klassische Elementaranregungen wie Phononen und Magnonen hin zu diffusiven Bewegungen oder die wesentlich entropisch getriebene Dynamik weicher Materie z.B. von Polymeren, Membranen und komplexen Flüssigkeiten. Weitere Schwerpunkte bilden ultraschnelle Prozesse in Molekülen, der zeitliche Ablauf chemischer Reaktionen, metallische und magnetische Systeme und schließlich die dynamischen Prozesse an Oberflächen. Anwendungsnahe Themen wie etwa die Magneto-elektronik werden ebenfalls angesprochen.

Die wichtigsten Themen sind:

- fsec-Spektroskopie
- inelastische Steuung von elektromagnetischer Strahlung, Neutronen
- Elektronenelektronische Anregungen und Transport
- Magnetisierungsdynamik und Magnetoelektronik
- Oberflächendynamik
- chemische Reaktionen und transiente Zustände
- Diffusion, Relaxationsprozesse, Elementaranregungen.

ca. 1000 Seiten, zahlreiche z.T. farb. Abb.,
EUR 48,57 (Hardcover)

Vorlesungsmanuskripte des
32. IFF-Ferienkurses 2001

Neue Materialien für die Informationstechnik

Die Möglichkeiten der heutigen Informations- und Kommunikationstechnik werden zunehmend bestimmt durch den Einsatz neuer elektronischer Materialien - jenseits der klassischen Siliziumtechnik. Dies betrifft sowohl die Erweiterung der konventionellen Mikroelektronik durch neuartige Funktionen als auch den Einsatz in integrierten Hybridsystemen. Der IFF-Ferienkurs 2001 behandelt alle Materialklassen, die in modernen Systemen der Informationstechnik heute bereits eingesetzt werden oder ein hinreichend großes Marktpotential für zukünftige Anwendungen aufweisen.

Der Ferienkurs hat einen betont interdisziplinären Charakter. Er verbindet Aspekte der Festkörperphysik, der Kristall- und Molekülchemie sowie der Biologie mit der Mikroelektronik. Der Schwerpunkt liegt auf der Verknüpfung des mikroskopischen Verständnisses der Materialien und der Nutzung ihrer Eigenschaften in Bauelementefunktionen.

Es werden materialübergreifende Konzepte der Informationstechnik diskutiert und aufgezeigt, in wie großer Breite neue Materialien in Bauelemente und Systeme vordringen. Das Spektrum der Vorlesungen umfasst u. a. folgende Themen:

- **dielektrische und ferroelektrische Oxide** für künftige Generationen schneller, nicht flüchtiger Arbeitsspeicher und hochintegrierter Logik-ICs,

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

- **ultradünne Metallschichten** mit spin-abhängiger Leitfähigkeit für magnetische Speicher und Leseköpfe heutiger Festplattengenerationen,
- **Phasenwechsel-Materialien** für wiederbeschreibbare DVDs,
- **photoadressierbare Polymere** für holographische 3-D-Speicher,
- **akusto-optische und elektro-optische Oxide** für die optische Datenübertragung,
- abstimmbare **keramische Dielektrika** für die Mikrowellentechnik, komplexe organische Verbindungen für hochauflösende **Flüssigkristall-Displays** und die künftige **Molekularelektronik**,
- Komponenten für Plasma- und Feldemissionsdisplays,
- Materialien für die Verknüpfung der **biologischen** und der **mikroelektronischen** Informationsverarbeitung.

ca. 800 Seiten, zahlreiche z. T. farb. Abb.,
EUR 61,36 (Hardcover)

Lecture manuscripts of the
33rd IFF Spring School

Soft matter:

complex materials on mesoscopic scale

In the last years, the traditional research areas of polymers, biological macromolecules, colloids, amphiphilic systems and membranes, as well as liquid crystals have merged into a new research field - *Soft Matter*. This field combines all materials, which are characterized by structures on typical length scales between nanometers and micrometers. Due to the large structural length scale, the number density of their translational degrees of freedom is many orders of magnitude smaller than for an ordinary, molecular material. This and the weak interactions between the structural units, which is typically on the order of the thermal energy k_BT, implies that these materials are easily deformable by external forces - they are soft.

The growing together of the different, previously disjoint areas of *Soft Matter* arises on one hand from the recognition of the same under-lying mechanisms in the structure and the properties of these systems, and on the other hand from the combination of many of these components in a single material. Examples are polymer-colloid mixtures such as ink, or the cell membranes of biological cells, in which a large number of different, cooperative components are involved.

Research in the field of *Soft Matter* is an interdisciplinary enterprise. This includes chemistry, which, with the synthesis of increasingly complex molecules, provides the building blocks for new materials. Physics provides the methods to investigate the properties of soft materials, and aims at reaching a detailed understanding of the connection between the molecular units and their interaction, and the observed macroscopic properties. Some of the investigated questions are derived from, or have implications, for example, for pharmacology or cell and molecular biology. As far as applications are concerned, material science is interested in finding materials with properties, which can be custom-tailored and tuned over a wide range.

The book consists of four parts:

- Techniques, Methods and Synthesis. Scattering, NMR, dielectric spectroscopy, computer simulations and the synthesis of polymers and colloids, including templating methods, are discussed.
- Structure and Phase Behavior. After an introductory lecture on statistical mechanics, the microstructural properties and phases of many different systems are dealt with, like polymers, polyelectrolytes, films and surfaces, amphiphilic systems, membranes and colloids.
- Equilibrium and Non-equilibrium Dynamics. This part discusses dynamics and kinetics of soft matter in- and out- of equilibrium,

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

like polymer and colloid dynamics, glassy polymers and colloids, membrane and microemulsion dynamics, phase separation kinetics and rheology.
- Industrial Applications. This part discusses applications of soft matter systems in industrial processes.

c. 1000 pages, many partly coloured illustrations, EUR 70,00 (Hardcover)

Lecture manuscripts of the
34th IFF Spring School

Fundamentals of Nanoelectronics

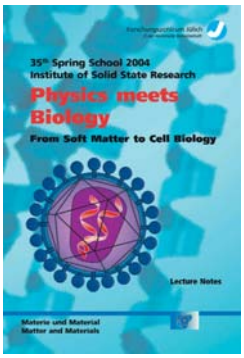
Information technology stands today at the edge of a revolutionary change: from microelectronics to nano-electronics. The characteristic structure sizes will soon fall below the 100 nm mark and will continue to shrink to approximately 20 nm around the year 2015. On this length scale quantum effects are decisive and lead to completely new possibilities and challenges.
As an alternative to the continually increasing cost of conventional semiconductor technologies for manufacturing integrated circuits, new strategies are examined in research, which are based on fundamental principles of physics and chemistry. For example, molecular self organization mechanisms are developed in order to manufacture well defined nanostructures with desired properties.

The IFF Spring School treats the basic principles of physics, chemistry and information technology that prepare the way into the new and fascinating world of nanoelectronics.
The spectrum of the lectures covers, among other subjects, the following topics:
- Fundamentals
  Theory of electron scattering
  Tunnelling processes
  Coulomb blockade effects
  Spin dependent transport
  Quantum communication/computing
- Analyses
  Scanning probe methods
  Spectroscopy with atomic resolution
  High resolution electron microscopy
- Technology
  New lithography techniques
  Atomic layer by layer deposition
  Self organization techniques
  Scanning probe manipulation
- Concepts for nanoelectronic devices
  Limits of Si technology
  Metallic nanowires
  Spintronics
  Electronics based on carbon nanotubes
  Single electron devices
  Concepts for QC devices
  Molecular electronics

c. 500 pages, many partly coloured illustrations, EUR 55,00 (Hardcover)

Lecture manuscripts of the
35th IFF Spring School

Physics meets Biology



Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

In the last two decades, we have witnessed breathtaking advances in molecular biology. In particular, the elucidation of several genomes including humans, mice, worms, flies and plants have provided the sequence of hundreds of thousands of different genes and proteins. In the years to come, it will be a formidable task to elucidate the location, structure, function, and interactions of these cellular components. Key requirements for a successful investigation and a detailed understanding of these enormously complex systems are:

- The development and application of an array of modern physical techniques to study the structure and dynamics of macromolecules and their assemblies over a wide range of length and time scales. Such techniques include scattering methods, nuclear magnetic resonance (NMR), atomic force techniques, singlemolecule fluorescence and electron tomography,
- A detailed understanding of macromolecules and their assemblies. Soft Matter physics investigates the behavior of complex mixtures consisting of several different components of polymeric, colloidal or amphiphilic cha~ racter with the aim to understand the cooperative behavior of systems with a large number of interacting degrees of freedom. A profound understanding of the fundamental mechanisms underlying biological function and self organisation may be fostered by investigations of simpler systems which are composed of synthetic soft materials.
- Analytical and synthetic organic chemistry. The principles of biological structure formation and function may be used to design and build similar synthetic molecules with tunable properties. Markers and labels have to be attached to biological macromolecules to visualize them and trace their motion. Finally, synthetic chemistry is required to provide model systems for macromolecular assemblies.

In order to cope with the enormous challenges posed by understanding the structure and function of the cellular machinery, biologists and physicists have to come together and to share their expertise. Biologists have to master the state of the art technologies required to study and understand biological processes. In a similar vein, physicists and chemists have to learn about and to be inspired by the large variety and diversity of exciting biological problems, and the immense research opportunities in the Life Sciences in general, and Cell Biology in particular.

With the IFF Spring School 2004, we want to contribute to the integration process of soft matter physics, biophysics and biology of the cell. The lectures are grouped into four topics:

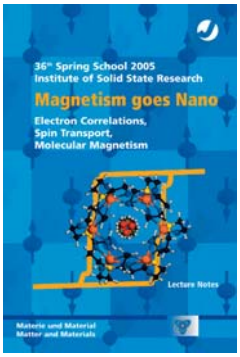
- General concepts and basic facts will provide an overview that covers statistical mechanics, Brownian motion, self assembly, and the components of biological cells.
- Techniques and methods introduces several modern experimental techniques as well as computer simulations.
- Polymers, Biopolymers, and Proteins covers a wide range of systems, including flexible and semiflexible polymers, polyelectrolytes, dynamics of proteins, protein crystallisation, and the cytoskeleton.
- The lectures on Membranes range from the statistical mechanics of fluctuating surfaces to the complexity of membrane proteins, endo and exocytosis and cell adhesion.

c. 850 pages, many partly coloured illustrations, EUR 55,00 (Hardcover)

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Lecture manuscripts of the
36th IFF Spring School

Magnetism goes Nano



Magnetism has always been a fascinating phenomenon and continues to be an exciting research field. Even today it provides stunning scientific discoveries combined with vast technological potential and economic impact. Recent years have seen the advent of magnetoelectronics and spintronics, fields in which magnetism and solid state electronics are joining to exploit spin-dependent transport processes. This creates novel electronic functionalities that in part already have entered the market, for example, in hard disk read heads and non-volatile,magnetic random access memories (MRAM). A pioneering work in magnetoelectronics was the discovery of giant magnetoresistance (GMR) by Peter Grünberg (IFF) and Albert Fert (Université Paris Sud). This discovery was honoured in 1998 by the prestigious "German Future Award – awarded by the President of the Federal Republic of Germany for a breakthrough in technology and innovation".

The continuing need to increase storage density leads to smaller and smaller magnetic entities. But the impact of these "nanomagnets" is not limited to technology. In cutting edge research several new phenomena, such as spin-torque transfer, spin-current induced magnetic switching, or spin-current induced microwave generation have been found very recently. Their discovery was only possible on account of the ability to fabricate magnetic nanostructures in the 100 nm regime. Clearly, the challenge for the future is to understand and control magnetism and magnetic phenomena on very small length scales and in reduced dimensions. The relevant physical systems range from thin films through quantum wires and quantum dots down to individual clusters, magnetic molecules, or even single magnetic atoms on a surface.

Programme

The IFF Spring School 2005 will address these new developments in magnetism on a graduate student level. The lectures will first build a basis for the understanding of the major phenomena and aspects in magnetic systems, including the theoretical framework for a quantitative description. The School will then advance to the peculiarities of magnetism in systems of reduced dimensions, covering thin films, quantum wires and dots, and magnetic clusters. Finally, it will bridge the gap to molecular magnetism and touch upon the major principles of quantum information physics.

The topics of the lectures cover:

- Theoretical Concepts in Magnetism
- Magnetism in Reduced Dimensions
- Electronic Correlations
- Spin Transport
- Magnetization Dynamics
- Novel Materials for Spintronics
- Molecular Magnets
- Quantum Information Physics
- Preparation of Nanomagnetic Systems
- Advanced Experimental Approaches

The IFF-Spring School is organized in close collaboration with universities, research institutions and industry.

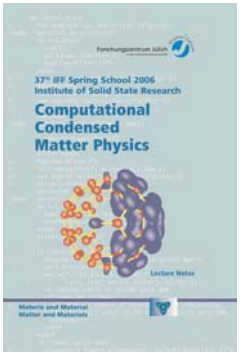
The School offers about 50 hours of lectures plus discussions, as well as the opportunity to take part in practical courses and visits to the participating institutes at the Research Centre Jülich. All lectures will be given in English. The lecturers are internationally renowned experts in their areas of research. All registered participants will be given a copy of the Lecture Notes (in English), which contains all the material presented during the school. Participants are expected to have a basic knowledge of quantum mechanics and condensed matter physics.

c. 850 pages, many partly coloured illustrations, EUR 55,00 (Hardcover)

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Lecture manuscripts of the
37th IFF Spring School

Computational Condensed Matter Physik



During the last decades we have witnessed dramatic advances in the simulation of physical systems on the computer. This partly due to an impressive growth in computer power. Equally or even more important, however, has been the outstanding progress in the development of new theoretical concepts and computational methods: In the simulation of condensed matter systems, the main challenge is to find models, which capture the essential physics of the real material, while still being susceptible to an efficient treatment an a computer.

As a result, we are now seeing more and more areas of condensed matter physics, where computer simulations achieve predictive power. Hence, they are becoming increasing important in identifying or designing new materials with fascinating and advantageous properties. Thus, computer simulations are now an essential tool in nanoscience, materials science, chemistry, and even biology. The important challenges in these Fields are:

• Many characteristic properties ot transition-metal oxides, nanostructures, and organic crystals are due to the strong repulsion between the electrons. An important focus of current research is the development of new methods for an efficient simulation of this quantum mechanical many-body problem.

• In Soft Matter Science – which studies the behavior of polymer solutions and melts, membranes, colloidal suspensions, and biological macromolecules – simulation methods have to be developed which bridge the large length – and time-scale gap between the atomistic scale of the solvent molecules and the mesoscopic scale ot the embedded macromelecules.

• A similar problem occurs in the investigation of macroscopic properties. The elementary processes often happen on the atomic scale, which is separated by many orders of magnitude from the macroscopic lengths and times of day-to day experience, as in solidification patterns of high-performance materials or earthquake rupture. Multi-scale simulation techniques have to be developed in Order to tackle this problem.

The basic idea of quantum computing is to use linear operations in Hilbert space to perform massively parallel calculations. While no quantum computer of any substantial size has yet been built, quantum computing holds the promise of a qualitatively new way of simulating physical systems.

Programme

The IFF Spring School 2006 addressed modern computational approaches to condensed matter physics at a graduate Student level. Introductory lectures will build the basis for the understanding of the major theoretical Methods and the phenomena they are meant to describe. More advanced lectures will address practical aspects of the methods and demonstrate how computer simulations contribute to our understanding of physics. Highlighting exemplary applications will lead the audience from the basic numerical methods to the frontiers of current research.

The topics of the lectures cover:

- Simulations of Quantum Systems
- Density Functional Theory
- Correlated electrons
- Quantum Computing
- Complex Materials
- Supercomputing
- Mesoscopic Hydrodynamics
- Monte Carlo Simulations
- Biophysics
- Soft Matter
- Pattern Formation
- Friction & Fracture

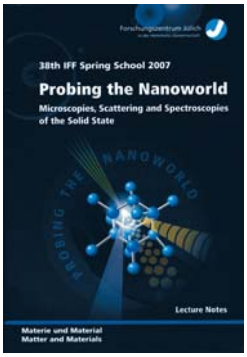
The School offerd about 50 hours of lectures plus discussions, as well as the opportunity to take part in practical courses, visits to the participating institutes and the supercomputer facilities at the Research Centre Jülich.

Source of Supply:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich · Germany
Tel. ++49(0)2461 61-5368 · Fax ++49(0)2461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Lecture Manuscripts of the
38th IFF Spring School

Probing the Nanoworld

Microscopies, Scattering and Spectroscopies of the Solid State



In recent years we have witnessed dramatic improvements in the capabilities of experimental methods involving the interaction of electrons, neutrons and Synchrotron radiation with the solid State. This is partly due to substantial advances in the Performance of sources, beamlines and imaging lenses, detectors as well as monochromators and spectrometers all at roughly the same time. Equally important, however, has been the outstanding progress in the development of vitally new analysis concepts together with concomitant numerical methods, which allow deriving a more unmitigated benefit from advanced instrumentation.

As a result, we are now seeing more and more areas of solid State research where state-of-the-art analysis techniques, e.g. aberration corrected electron microscopy, high energy or resonant Synchrotron X-ray scattering, neutron scattering with spherical polarization analysis or with neV energy resolution, as well as high-resolution electron spectroscopy and spectromicroscopy, permit conceiving macroscopic material properties epistemologically from microscopic observations. Hence, this basic approach is becoming increasingly important in identifying material imperfections and designing new materials with fascinating and likewise advantageous properties. All of the above probing techniques - irrespective of the use of electrons, neutrons or Synchrotron radiation - are nowadays essential tools to probe the nanoworld of complex materials in both, bulk and reduced dimensions.

Therefore, this is also the ideal time to appraise when, how, and why appropriate techniques should be used for specific Problems to gain deeper scientific insights to Condensed matter phenomena as well as to optimize processing Steps in solid State technology. Consistently, the 38th IFF Spring School 2007 will focus on both, essential fundamentals and the latest developments in probing the solid State by application of advanced microscopy, scattering and spectroscopic techniques. Lecturers will address late-breaking aspects of experimental analysis techniques in tandem with materials science related applications.

Programme

The IFF Spring School 2007 will address both, advanced experimental techniques and theoretical fundamentals at a graduate Student level. Introductory lectures will focus on the structural and electronic properties of the solid State in bulk and in reduced dimensions, and, thus, build the basis for the understanding of prime techniques and major methodical concepts. Supplementary lectures will address

basic excitation mechanisms and the interaction of radiation with matter.

More specialized lectures will exemplify a multitude of techniques in electron microscopy, neutron and Synchrotron scattering in tandem with solid State related applications. Highlighting selected applications will lead the audience from basic methods to the frontiers of current research.

The main topic areas to be covered are as follows:

- Electron Excitations & Interaction of Radiation with Matter
- Scanning Probe Microscopy & Spectroscopy
- Aberration Corrected Transmission Electron Microscopy
- Electron Holography & Tomography
- Electron Energy Loss Spectroscopy
- Photoemission Spectroscopy & Microscopy
- Synchrotron Radiation & Neutron Scattering
- X-ray Absorption Spectroscopy & Resonant Scattering
- Magnetic Fluctuations & Lattice Excitations
- Soft Matter Structure & Dynamics

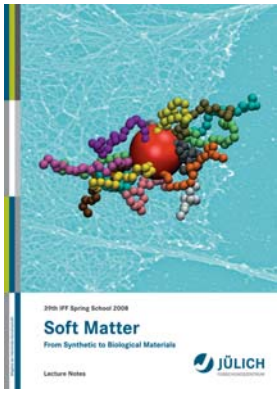
The School offered about 50 hours of lectures, discussions and a variety of laboratory courses.

Ca. 1000 pages, many partly coloured illustrations,
ISBN 978-3-89336-462-6; EUR 69,90 (Hardcover)

Bezugsadresse:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich
Tel. 0 24 61 / 61- 53 68 · Fax 0 24 61 / 61-61 03 · e-mail: zb-publikation@fz-juelich.de

Lecture Manuscripts of the 39th IFF Spring School

Soft Matter – From Synthetic to Biological Materials



Soft matter is ubiquitous in a vast range of technological applications and is of fundamental relevance in such diverse fields as chemical, environmental, and food industry as well as life sciences. Over the past years, soft matter science has been largely extended in its scope from more traditional areas such as colloids and polymers to the study of biological systems, soft nanoscale materials, and the development of novel composites and microfluidic devices. Soft and biological materials share fundamental structural and dynamical features including a rich variety of morphologies and non-equilibrium phenomena, self-organisation, an unusual friction-dominated flow dynamics, and a high sensitivity to external fields. These properties emerge from the cooperative interplay of many degrees of freedom, with spatio-temporal correlations that can span a huge range from nano- to millimetres and nanoseconds to days. The key requirements for the advancement in the field of these highly complex soft materials are:

- The development of novel experimental techniques to study properties of individual components in processes and the cooperative behavior of many interacting constituents. The synthesis of complex materials, self-organized and biomimetic systems with novel or unusual properties will broaden the spectrum of applications.
- The exploration of advanced theoretical and computer simulation methods that span the large range of time and length scales and allow to cope with an increasing complexity of molecular constituents. Existing methods need to be extended and new approaches are required to describe systems far from equilibrium, e.g., in life sciences and material processing.
- Structural and novel functional properties of soft and biological materials need to be studied invoking self-organization and hierarchical structure formation, entropic particle interactions and fluid-like aspects of biological materials such as vesicles and cells.
- The unusual dynamics of complex fluids requires special approaches to gain insight into diffusion transport properties, rheology and mesoscopic flow behavior, which are influenced by a delicate interplay of hydrodynamic interactions, thermal fluctuations, and external fields.

The IFF Spring School 2008 at the Forschungszentrum Jülich, Germany, addressed advanced experimental techniques and applications, and theoretical and computer simulation methods on an undergraduate and graduate student level. Introductory lectures provided the basis of important experimental and theoretical tools. More advanced lectures explained practical aspects of various methods and lead the participants from basic methods to the frontiers of current research.

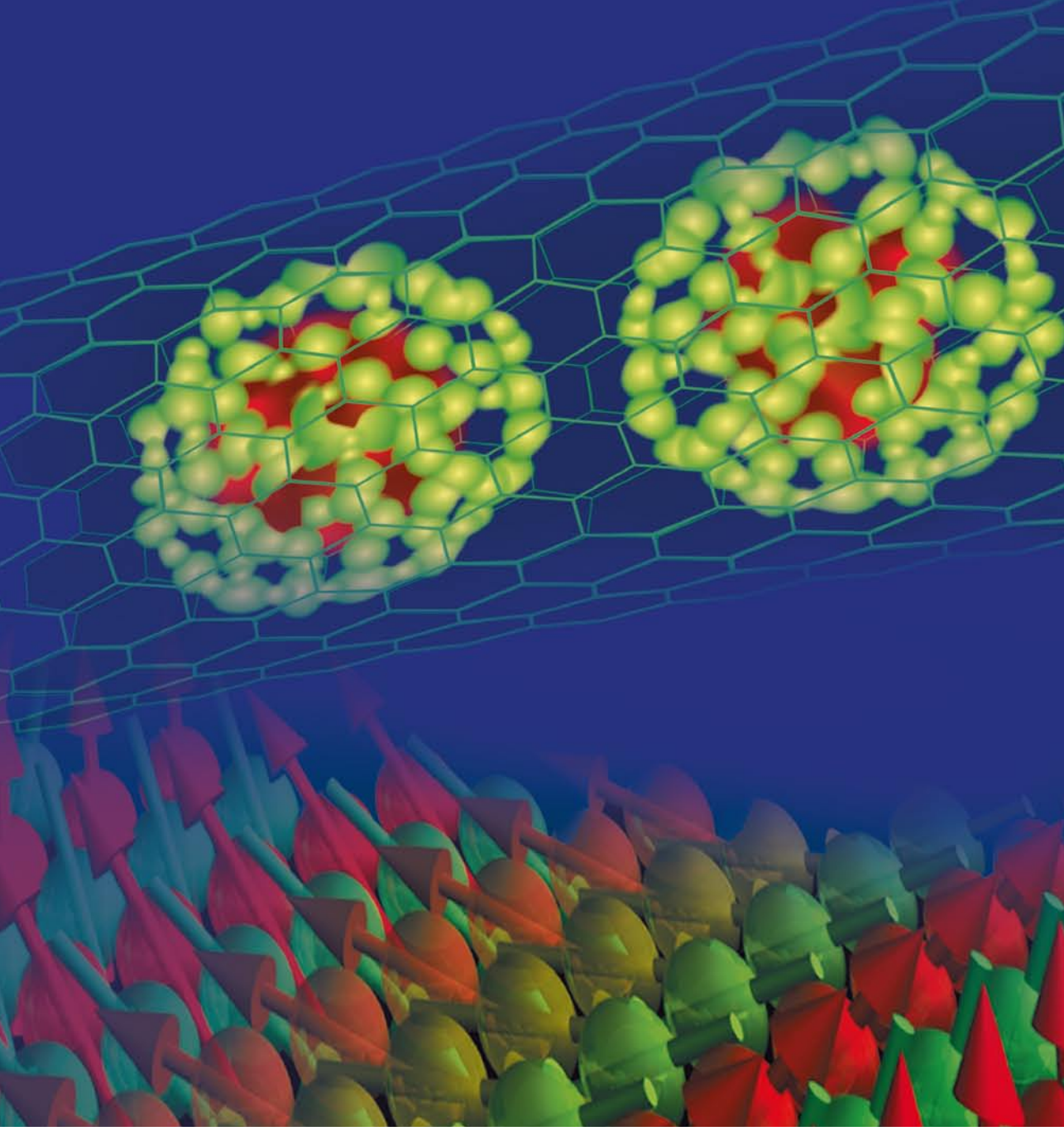
The lectures covered the following topics:

- Scattering Techniques
- Single Molecule Techniques
- Equilibrium- and Non-equilibrium Statistical Physics
- Microfluidics
- Computer Simulations
- Synthesis
- Self-Organisation
- Flow Properties and Rheology
- Biomechanics
- Macromolecules and Colloids
- Membranes and Interfaces
- Biomimetic Systems
- Glasses and Gels

The school offered about 50 hours of lectures plus discussions, as well as the opportunity to participate in practical courses and visits to the participating institutes at the Forschungszentrum Jülich.

Ca. 1000 pages, many partly coloured illustrations, ISBN 978-3-89336-517-3; EUR 69,90 Hardcover).

Bezugsadresse:
Forschungszentrum Jülich GmbH · ZENTRALBIBLIOTHEK · 52425 Jülich
Tel. 0 24 61 / 61- 53 68 · Fax 0 24 61 / 61-61 03 · e-mail: zb-publikation@fz-juelich.de



Band | Volume 10
ISBN 978-3-89336-559-3

JÜLICH
FORSCHUNGSZENTRUM