Springer Series in Optical Sciences 157

Hans-Joachim Lewerenz

Photons in Natural and Life Sciences

An Interdisciplinary Approach

Springer

Springer Series in

OPTICAL SCIENCES 157

founded by H.K.V. Lotsch

Springer Series in

OPTICAL SCIENCES

The Springer Series in Optical Sciences, under the leadership of Editor-in-Chief *William T. Rhodes*, Georgia Institute of Technology, USA, provides an expanding selection of research monographs in all major areas of optics: lasers and quantum optics, ultrafast phenomena, optical spectroscopy techniques, optoelectronics, quantum information, information optics, applied laser technology, industrial applications, and other topics of contemporary interest.
With this broad coverage of topics, the series is of use to all research scientists and engineers who need up-to-date reference books.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors. See also www.springer.com/series/624

Please view available titles in *Springer Series in Optical Sciences*
on series homepage http://www.springer.com/series/624

Hans-Joachim Lewerenz

Photons in Natural and Life Sciences

An Interdisciplinary Approach

With 218 Figures

Springer

Professor Dr. Hans-Joachim Lewerenz
Helmholtz Zentrum Berlin for Materials and Energy
Institut für Solare Brennstoffe
Hahn-Meitner Platz 1
D-14109 Berlin, Germany
E-mail: lewerenz@hmi.de

Springer Series in Optical Sciences ISSN 0342-4111 e-ISSN 1556-1534
ISBN 978-3-642-23748-5 e-ISBN 978-3-642-23749-2
DOI 10.1007/978-3-642-23749-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011944227

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to Mrs. Rodenberg, Physics teacher at the Ulrich-von-Hutten Gymnasium in Berlin Lichtenrade

Preface

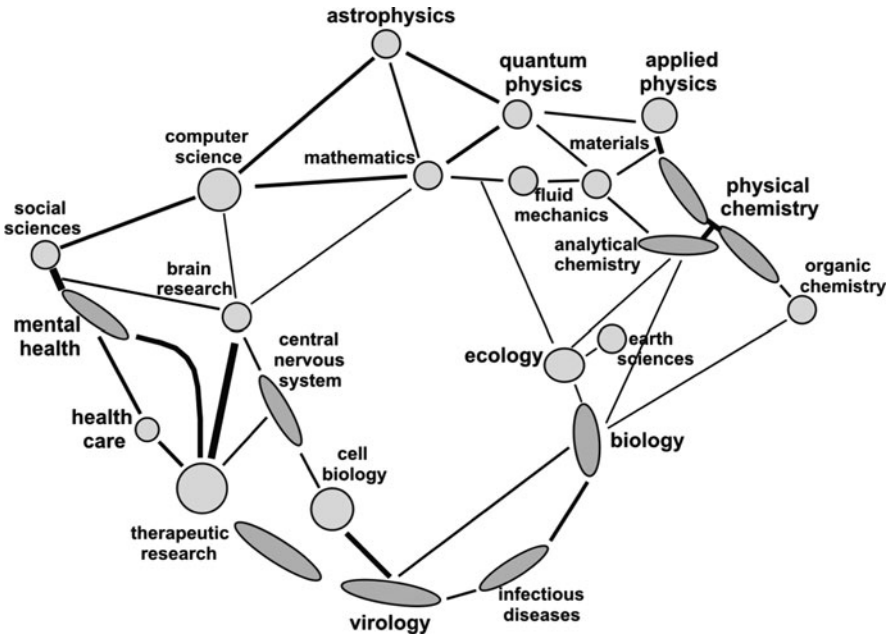
Advances in science, often curiosity driven, can also address societal needs such as energy, health and lifestyle, and, of course, the increase of fundamental knowledge. Technology, which largely defines our way of living, is strongly dependent on the underlying scientific achievements, which, in turn, are related to the advances in the fundamental and applied sciences including the life sciences.

With the advent of nanoscience, advances have become possible that are based on novel multidisciplinary approaches where connections between hitherto rather separated disciplines such as, for example, physics, biology, chemistry, mathematics, materials science and information technology have been made. In this sense, science, besides its progress in the respective fields, is presently experiencing a renaissance of bridging between disciplines similar to the situation that existed at the beginning of the age of *Enlightenment*. Accordingly, this text focusses on the interdisciplinarity that characterizes many aspects of modern nanoscience. I have selected photons as a "glue" that demonstrates the interlinking of so many apparently disparate scientific disciplines in natural and life sciences. Therefore, instead of treating scientific topics according to their traditional compartmental structure, photons and their interactions serve here as a *guide for a visit* in the respective disciplines and their sometimes unusual connections[1].

The general topics, made visible by my subjective selection of the chapters, shall emphasize the bridging character in modern science. The approach presented here is rather unique as we can see in the schematic above (see footnote 1) which demonstrates that there is still a long road to travel for interdisciplinary work in the sciences; this text is just one of the beginnings in that direction.

The interaction of matter with electromagnetic radiation involves absorption, scattering and emission of photons. The corresponding electronic and vibronic excitations are localized processes making photons a suited probe of nanoscale matter. In most cases, this probe is nondestructive except for high-intensity beams

[1] Image adapted from W.B. Paley, K. Boyack, D. Klavans, SEED Magazine, March 7 (2007), commiss. K. Börner.



where radiation chemistry can occur. The use of pulse lasers, femtosecond X-ray sources and free electron lasers allows ultrahigh time resolution with sufficient beam intensity. Besides being an analytical tool, photons can be used to change matter by localized interactions. Deep UV lithography is a well-known example of structuring matter with light. In advanced communication, based on entangled photons, the delocalized nature of the photon field is exploited, whereas in quantum computing, localized interactions, for instance, with the hyperfine structure of ion states in a Pauli trap, are exploited.

In life sciences, the use of photons ranges from the traditional use of light microscopy that has revolutionized medicine, to present day's photothermal cancer therapy, for example. Photosynthesis has produced our atmospheric oxygen and the analysis of the light-induced and consecutive processes in photosynthesis has resulted in a wealth of insights that is now used for realization of bio-inspired solar fuel generation, known as *artificial photosynthesis*. A particularly interesting phenomenon learnt from photosynthesis is the excitation energy transfer where photonic energy is transferred to catalytic sites without an actual charge transfer. The analysis of the energy states in such centres has become possible by the development of synchrotron third-generation light sources. An important future route is the steering of enzymatic reactions, e.g., bio-catalytic activity, by photonic excitation energy from neighbouring absorbers and to replace enzymes by suited more robust catalytic macromolecules.

The conceptual approach of a multidisciplinary text results from inspection of scientific achievements that were made by scientific cross-fertilization and also

by the author's work over the last three decades in various disciplines such as surface science, electrochemistry, materials science of semiconductors, solar cell development and, to a smaller extent, on plasmonics and biological systems.

Important parts of this monograph have been written during visits of Baldhead Island (North Carolina), during a sabbatical at North Carolina State University (NCSU) and during a visit at Jilin University in Changchun (China) where my host was Wenbo Song. I am indebted to Zlatko Sitar and Michael Rigsbee for hosting me for a sabbatical year at NCSU. Valuable discussions with Raoul Schlesser, Ramon Collazo and Klaus Bachmann are gratefully acknowledged. Many stimulating discussions with the members of my research team at the Helmholtz Zentrum for Materials and Energy in Berlin, Germany, are gratefully acknowledged. Last but not least, I want to express my gratitude to Claus Ascheron of Springer Heidelberg for his patience, help and continuing support of this project.

This book addresses a readership with a pronounced interest in the multidisciplinary challenges in nanoscience. The text has been written for readers with a basic knowledge in physics and mathematics and the frequent use of equations and figures was a means to condense the width of the highly interdisciplinary content into a manageable volume. The book shall provide an overview of developing areas and give insights into already existing cross-disciplinary activities and, also, hopefully lead to the stimulation of further interdisciplinary work in many areas of natural and life sciences. It has been written for interested graduate students, teachers and experts working in the respective fields who want to broaden their perspective. Part of this text has been presented as a lecture on "nanophotonics" at the Technical University Berlin (Germany) and, accordingly, this book might be of use for those lecturers who want to stimulate interdisciplinary research and education by teaching.

Berlin *Hans-Joachim Lewerenz*

Contents

Introductory Remarks

Due to the presently still rather unusual connections between separate scientific disciplines, the description of fundamental aspects has been included in this text for intelligibility of the sometimes advanced cross-disciplinarity of the content. Therefore, the reader will find some basic scientific material as well as more advanced descriptions when combinations of disciplines are treated that sometimes carry an element of surprise. Examples for the latter are the use of γ-quanta in the determination of phase diagrams of a ternary solar material, the application of Cherenkov radiation for the analysis of the energy band structure of photonic crystals, light emission in scanning tunnelling experiments and its application to submolecular spectral analysis or the use of X-ray emission techniques for analysis of metallo-centres in proteins, to name just a few topics. These and other topics have intentionally been selected to demonstrate the huge (dormant) innovation potential of interdisciplinary research and cooperation.

This text comprises basically two aspects of photonics: the generation of light in various processes, describing the "mystical" interaction of electrons and photons, and the application of electromagnetic radiation to modern scientific areas. The former is treated in the first three chapters which can be considered as the first part of this book but fundamental aspects of photons reappear in Chap. 6 where their quantum nature is described. Applications in various areas of science are found in Chaps. 4–6. The chapter on spintronics (Chap. 7) has a hybrid character as it also focusses on the behaviour of electrons in solids. It has been included because of recent developments in spin-photonics.

The book begins with basic considerations on the nature of light and its origin, including historical aspects. Photon generation by momentum changes of moving charges (electrons) is described by a comparably intelligible semiclassical concept for a wider readership. Further, the inception of quantum electrodynamics (QED), related to Fermat's and Huygen's principles transferred to non-zero mass particles, is presented and light propagation based on these concepts is described in a phenomenological manner, yielding a rather intuitive insight into Feynman's path integrals in QED. Based on a recent finding of an amino acid in the dust of the comet Wild 2 and on earlier panspermia-supporting events, the present state of discussion

on the origin of life and, in particular, its relation to photonic energy supply are included here.

Chapter 2 treats advanced light sources used for generation of high-energy photons and, in part, ultrashort photon pulses. The anticipated role of the International Thermonuclear Reactor (ITER) for provision of ultrahigh energy photons is outlined here as well. A considerable part is devoted to the use of synchrotron radiation from wigglers and undulators. The operational modes of free electron lasers are outlined and rather recent methods such as femtosecond laser slicing which relies on ponderomotive forces are included. The selection of these light sources is to a high degree commensurate with later presented applications in natural and life sciences, found, for example, in Chaps. 4 and 5.

Whereas Chap. 2 describes photons generated by kinetic electrons, Chap. 3 addresses photon generation by discrete quantum processes. Here, the energy spectrum encompasses high-energy photons from X-ray plasma lasers, for instance, then moving into the longer wavelength regime of bio- and chemiluminescence. As an interesting example of highly localized photon generation, surface plasmon decay into single particle excitations during scanning tunnelling spectroscopy (STM) is reviewed. The renewed interest in the field of plasmonics that has substantial implications in areas such as solar energy conversion and medical sciences is included here as the interaction and decay of the collective excitations are related to photon absorption and generation.

The second part of this text (Chaps. 4–7) shows the actual interrelation of various disciplines by selection of applications of photons in bridging fields of natural and life science. Applications of advanced light sources to the study of various areas of micro- and nanoscience are found in Chaps. 4 and 5. In Chap. 4, the investigation of processes at the semiconductor–electrolyte interface by *in-system* synchrotron radiation photoelectron spectroscopy and by X-ray emission spectroscopy is covered. Subsequently, a phase diagram analysis of the ternary system Cu–In–S, relevant for the development of $CuInS_2$-based photovoltaic solar cells, is presented. As nuclear probe, gamma rays are used that are emitted in the decay of radioactive ^{111}In atoms which have been incorporated into the host material during growth. Chapter 5 contains two major aspects of the management of photons, the artificial route provided by the development of photonic crystals and the route used by nature to funnel photonic energy into the reaction centres in photosynthesis. In the former, the analysis of the photonic band structure for a two-dimensional crystal by Cherenkov radiation represents another example of connectivity between apparently unrelated scientific subjects. After a note on metamaterials, the chapter ends with a scanning tunnelling microscopy investigation of enzymes that have been immobilized on semiconductor surfaces. The derived concepts of the electronic properties of semiconductor–protein junctions allow to propose future systems where charge or excitation energy transfer between semiconductors and enzymes enables designed catalytic operation. A first example on a bio-photocathode that uses an enzyme will be presented.

The effects associated with the quantum nature of photons and the related puzzling observations are outlined in Chap. 6. After an introduction to entanglement,

the Einstein–Podolsky–Rosen criticism of the Copenhagen interpretation of quantum mechanics is described and of Bell's inequality is introduced that allows experimentalists to distinguish in principle between nonlocal and local quantum theories. In addition, processes are reviewed that are based on the quantum nature of photons. They encompass quantum teleportation and applications such as quantum cryptography and quantum computing and include early realization of several quantum computers.

Finally, Chap. 7 treats the exploration of the quantum coherence of the electron spin that has resulted in the field of spintronics, a neologism of the words spin and electronics. Before entering the more recent developments of spin photonics, the chapter gives an overview of the status of the field, beginning with more traditional aspects and devices such as magnetic tunnel junctions, the giant magnetoresistive effect and later advancing to semiconductor spintronics. The more theoretical subchapter on magnetism and spin presents an excursion into related science and also gives a concise overview on the basic magnetic properties relevant for spintronics. The chapter ends with an assessment of the present state of the so-called spin photonics, thereby returning to the main theme of this book.

Chapter 1
On the Origin of Light

1.1 Introduction

This chapter treats fundamentals of photon generation, giving an overview of physical principles of photon generation. The considered energies range from ultra-hard radiation to the visible and (far) infrared terahertz regime. A rather intriguing example is the future possibility to use high-energy photons from nuclear fusion: once mastered in a sustainable way, nuclear reactions could provide energetic photons unaccounted for so far, since in stars, the comparably low-energy black body radiation of the surface temperature dominates [1]. The direct and stable high-energy photon output from a nuclear reaction might become available in a foreseeable future [2], providing possibilities for gaining new insights into photon–matter interactions and allowing corollary applications.

In view of the intended interdisciplinary bridging character of this text, theoretical concepts are, when possible, first presented in a classical or semi-classical approach to facilitate intelligibility. For example, the interrelation of electrons and photons is described here semi-classically, based on the work of Weizsäcker and Williams [3, 4] for relativistic electrons, using the picture of virtual quanta.

A substantial part of this chapter treats the various modes of Bremsstrahlung, also referring to applications. This is followed by photon generation from discrete electronic quantum transitions. Subsequently, the quantum electrodynamic concepts of light generation and propagation are presented in a concise overview, followed by an overview of what is presently known regarding the contribution of light to the origin of life (Sect. 1.5). The content of this chapter is summarized in Table 1.1.

1.2 Bremsstrahlung

The general aspects of radiation emitted from motion changes (acceleration and deceleration) of electrons are described. Although synchrotron, wiggler and undulator radiations belong to this category, they are treated separately because

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_1,

Table 1.1 Overview of the content of Chap. 1

1.1.	Introduction
1.2.	Bremsstrahlung
1.2.1.	Radiation from charges in motion
1.2.2.	Coulomb collisions
1.2.3.	Relativistic Bremsstrahlung: the Weizsäcker–Williams model
1.2.4.	Coherent Bremsstrahlung
1.2.5.	Polarization Bremsstrahlung
1.3.	Principle of light generation from quantum transitions
1.4.	Quantum electrodynamic concepts: light generation and propagation
1.4.1.	Background
1.4.2.	Light propagation concepts
1.4.3.	Photon generation for quantum information science
1.5.	On the origin of life-photonic contributions
1.5.1.	Definitions
1.5.2.	Geological periods and human development
1.5.3.	Overview of hypotheses on life's origin
1.5.3.1.	Panspermia
1.5.3.2.	Prebiotic synthesis: heterotrophic approach
1.5.3.3.	Autotrophic processes
1.5.3.4.	RNA world
1.5.4.	Selected mechanistic aspects: quantum tunnelling and prebiotic nucleotide synthesis
1.5.4.1	Panspermia-related processes
1.5.4.2	RNA world: synthesis and attachment of pyrimidine nucleotides
1.5.5.	Influence of photons on biogenesis: photochemistry and homochirality
1.5.5.1	Basic photochemistry
1.5.5.2	On the homochirality of biological molecules

of their extended use in modern experimental facilities. Bremsstrahlung also results from interactions between electrons and ions in nuclear fusion plasmas. In hot, fully ionized plasmas, the Bremsstrahlung emission coefficient is given by the Kramers formula [5]. For high-energy plasmas ($T \sim 10^7$ K), high-intensity Bremsstrahlung is observed with $\lambda \geq 1$ nm, providing an additional ultrashort wavelength source.

1.2.1 Radiation from Charges in Motion

Accelerated or decelerated charges emit electromagnetic radiation. The intensity and the spectral properties (as well as the angular distribution) are of interest for many applications. For classical, non-relativistic motion, the radiation is described by Larmor's formula [6]. The acceleration $\mathrm{d}\mathbf{v}/\mathrm{d}t$ of a charged particle, which is observed in a reference frame where $v \ll c$, leads to radiation of total power P given by [6]:

$$P = 2e^2/3c^3\,(\mathrm{d}\mathbf{v}/\mathrm{d}t)^2\,. \tag{1.1}$$

For accelerator applications, one generalizes Larmor's formula for arbitrary charge velocities. Using the relation between momentum and velocity, $\mathbf{p} = \gamma\ \mathrm{m}\mathbf{v}$ (where γ is the Lorentz factor), and the relation between radial acceleration, radius and particle velocity expressed in terms of the momentum $\mathbf{p}$, one obtains the relation $\mathrm{d}\mathbf{p}/\mathrm{d}\tau = \gamma\omega\mathbf{p}$. Here, ω is given by the circular motion by $\mathbf{v}/r_\mathrm{c}$, with r_c being the orbit radius and $\mathrm{d}t = \gamma\mathrm{d}\tau$. Obviously, the radiation emitted results from constant angular acceleration. Its power is (using $\beta = v/c$)

$$P = \frac{2e^2}{3m^2c^3}\gamma^2\omega^2 \quad \mathbf{p}^2 = \frac{2e^2c}{3r_\mathrm{c}^2}\beta^4\gamma^4. \tag{1.2}$$

The radiative energy loss per revolution is obtained by the product of the time for completing one cycle and the radiated power

$$E_\mathrm{rad} = (2\pi r_\mathrm{c}/v)P. \tag{1.3}$$

For high-energy electrons with $\beta \cong 1$, the energy loss can be given numerically:

$$E_\mathrm{rad}(\mathrm{MeV}) = 8.85 \times 10^{-2}\frac{[E\ (\mathrm{GeV})]^4}{r_\mathrm{c}\ (\mathrm{meters})}. \tag{1.4}$$

It can be shown that radiation losses from a linear acceleration are negligible. Directional momentum changes are considerably more effective for radiation emission which led to the introduction of synchrotron radiation sources, for instance.

The angular distribution of the emitted radiation narrows in the direction of electron motion nearing the velocity of light as shown in Fig. 1.1. For $\beta \to 1$, the maximum radiation intensity has an angle of $\Theta_\mathrm{max} \cong 1/2\gamma$ with respect to the electron velocity vector. At 99.9% light velocity, the angle is 1.2°.

1.2.2 Coulomb Collisions

Particles passing through matter are scattered and lose energy due to collisions. In these collisions, the particles change energy and momentum. The velocity and directional changes result in emission of electromagnetic radiation as described above. In these scattering processes, usually only the interaction of a fast particle (electron of energy E) with the Coulomb field of a fixed charge (representing the struck atom) is considered where the deflection of the incoming electron results in radiation. Due to their larger charge, nuclei deflections dominate over those of

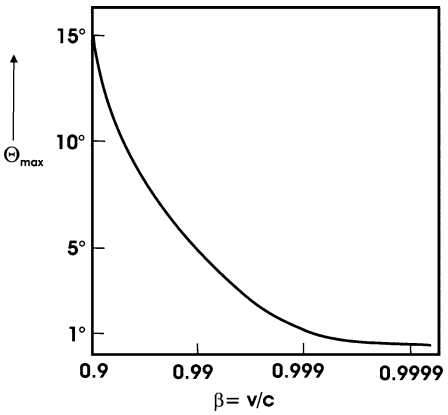


Fig. 1.1 Change of $\Theta_{\max}$ with β (see text)

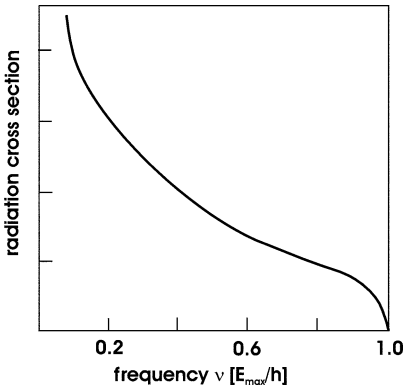


Fig. 1.2 Radiation cross section for non-relativistic Bremsstrahlung from Coulomb collisions as a function of frequency normalized to the frequency maximum $E_{\max}/h$

atomic electrons. Rutherford's formula [7] expresses that interaction, by describing momentum transfer for non-relativistic motion.

Figure 1.2 shows the radiation cross section for non-relativistic Coulomb collisions as a function of electron energy, first calculated by Bethe and Heitler [8] using the quantum mechanical Born approximation. It is observed that the cross section decreases rather linearly over a wide energy range and cuts off at the maximum frequency $\hbar/E$. For such non-relativistic electrons, however, radiative losses are negligible compared to collision losses due to electron–electron interaction [9].

1.2.3 Relativistic Bremsstrahlung: The Weizsäcker–Williams Model

Weizsäcker and Williams [10] noted that the electromagnetic fields of an electron in uniform motion at relativistic velocity are almost transverse as in the case of plane

waves. This suggested to imagine the fast electron as carrying with it a "cloud" of virtual (in the sense that they can only be observed by disturbing the electron's motion) photons that can be shed if the electron is perturbed. The model assumes that relativistic electrons carry a virtual photon spectrum with α photons per unit frequency interval. When, for whatever reason, radiation is emitted, the observed frequency spectrum will very closely follow the virtual spectrum.

An important aspect of this model is the consideration of the time or length over which the virtual photon cloud is shaken off: the more rapidly the cloud is removed, the more intense is the radiation observed in the laboratory. As concept, one uses the so-called formation time or formation length. The formation time is defined as the time an electron travels while its radiated electromagnetic wave advances one wavelength ahead of the particle. After the formation time has elapsed, the wave is a radiation which is no longer tied to its source.

If the electron would radiate all of its energy, the corresponding spectrum would be

$$\mathrm{d}E/\mathrm{d}\omega = e^2/c. \tag{1.5}$$

The number of photons of frequency ω is obtained from

$$\mathrm{d}n_\omega = \mathrm{d}E\,(\omega)\,/\hbar\omega \approx e^2/\hbar c\mathrm{d}\omega/\omega = \alpha\mathrm{d}\omega/\omega, \tag{1.6}$$

where $\alpha \cong 1/137$ is the fine structure constant. For high frequencies (for a more detailed description see [9]), a term containing $\exp(-\omega/\omega_\mathrm{c})$, where ω_c denotes the maximum or critical frequency of the spectrum, has to be included. The above expressions describe the radiation emitted over one formation length. In the case of a force which acts over several formation lengths, the photon emission spectrum is given by division by L_0, the formation length for $\omega \leq \omega_\mathrm{c}$:

$$\frac{\mathrm{d}n_\omega}{\mathrm{d}L} \approx \frac{\alpha}{L_0(\omega)}\frac{\mathrm{d}\omega}{\omega}. \tag{1.7}$$

For $\omega \geq \omega_\mathrm{c}$, the spectrum is

$$\frac{\alpha}{L_0(\omega)}\frac{\mathrm{d}\omega}{\omega}\mathrm{e}^{-\omega/\omega_\mathrm{c}}. \tag{1.8}$$

Above the critical frequency, the spectrum decreases exponentially. In practical terms, the radiative loss of high-energy electrons in matter of density ρ is usually expressed by

$$\frac{1}{\rho}\left.\frac{\mathrm{d}E}{\mathrm{d}x}\right|_\mathrm{rad} \approx -\frac{E}{X_0}, \tag{1.9}$$

where X_0 is the so-called radiation length describing the average energy loss per unit length. The resulting energy profile of the electron is given by $E \approx E_0$

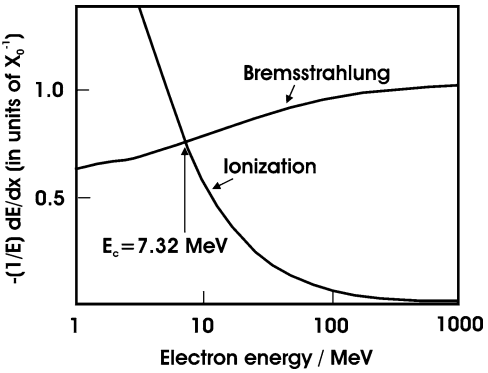


Fig. 1.3 Energy loss of electrons in Pb (Z = 82) showing the crossover at the critical energy; beyond 500 MeV, the ionization process is dominated by Möller scattering

$\exp(-\rho x / X_0)$; for these high-energy electrons, ionization effects due to inelastic collisions with atomic electrons can be ignored as long as the electron energy is above a critical energy, E_c, approximated as

$$E_c \cong \frac{610\,\text{MeV}}{Z + 1.24}, \tag{1.10}$$

where Z is the atomic number of the material. For Pb with $Z = 82$, the critical crossover energy lies at 7.32 MeV (see Fig. 1.3). Above this energy, losses occur predominantly due to inelastic collisions with nuclei producing Bremsstrahlung. It is also interesting to note that Bremsstrahlung provides the major part of the IR continuum due to e^--Hg collisions in HID (high-intensity discharge) lamps [11].

1.2.4 Coherent Bremsstrahlung

The enhancement of Bremsstrahlung occurring in a crystal when the momentum transfer from a relativistic electron to the scattering atom matches a reciprocal lattice vector is considered. The process is analogous to Bragg scattering and results – in contrast to the smooth incoherent $1/E$ spectrum (Fig. 1.3) – in a saw-tooth-like intensity distribution of the emitted photons [12]. Therefore, using crystals such as diamond, for instance, provides rather monochromatic and polarized Bremsstrahlung photons [13]. Figure 1.4 shows calculated (*dotted curve*) and measured data from coherent Bremsstrahlung (*full line*) also indicating the incoherent background (*dashed–dotted line*) for ultra-relativistic electrons impinging on a diamond crystal with 4.8 GeV.

The emitted photons, even for lower electron energies around 250 MeV, are extremely energetic and, in the latter case, emit at 60 MeV which corresponds to 0.02 pm wavelength. Such high-energy photons are presently used to investigate electromagnetic interactions of hadrons with nuclei. With regard to materials science, attempts are being made to use fast electrons and the radiation emitted by their travelling in periodic nanostructures for characterization of these

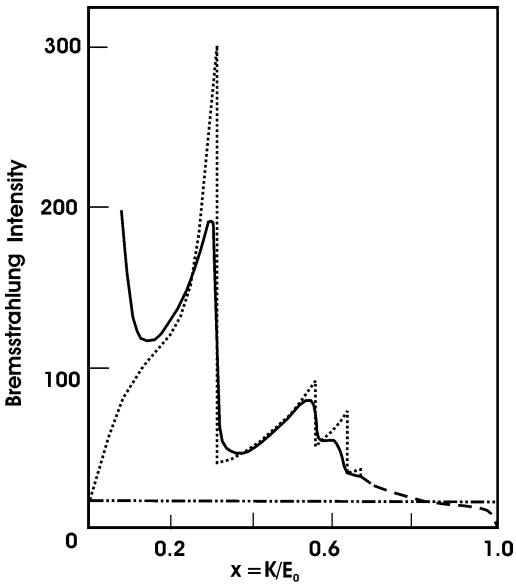


Fig. 1.4 Intensity of coherent Bremsstrahlung vs. x (the normalized photon energy, K energy of photons) for diamond (thickness 1 mm); electron beam energy $E_0 = 4.8\,\text{GeV}$; $\theta = 3.44\,\text{mrad}$; $\alpha = 0°$ (angle between incidence plane and crystal plane). *Full line*: fitted experimental data, *dotted line*: calculation, *dashed-dotted line*: incoherent Bremsstrahlung and *dashed tail*: fit to experimental data at low intensity [14]

nanostructures [15]. We will revisit this topic in the context of the characteristics of Cherenkov radiation (Sect. 2.5) in conjunction with its application to photonic crystals (Sect. 5.2.5).

The advantages of this radiation, being quasi-monochromatic, easily tunable, polarized and almost free of background, are balanced by its main drawback, i.e. the relatively low intensity compared to synchrotron radiation, for instance. On the other hand, the CXR (coherent X-radiation) characteristics depend strongly on structural features of crystalline samples, thus allowing the analysis of macromolecular crystals. Also, self-assembled systems and phase transitions in such systems can be studied. Therefore, applications for biologically relevant topics can be envisaged.

1.2.5 Polarization Bremsstrahlung

Polarization Bremsstrahlung results from the scattering of the electromagnetic field of a relativistic particle (electron) moving in a medium (solid) with atomic electrons. The interaction is also described as dynamic polarization of the target under the action of the Coulomb field of such projectiles. The effect is largest for atom types with high dynamic polarizibility [16]. Since the effective impact parameter for collisions is commensurate with atomic dimensions, the features of polarization Bremsstrahlung are affected by interatomic correlations. Therefore, the energetic and angular intensity distributions of spectra from amorphous and polycrystalline samples are supposed to differ substantially [17]. Although differences in Bremsstrahlung between amorphous and polycrystalline environment have been

attributed to polarization Bremsstrahlung [18], strong evidence was obtained only rather recently for this type of radiation [19].

1.3 Principle of Light Generation from Quantum Transitions

We consider an isolated atom that interacts with photons, described by their electromagnetic field, using the vector potential $\vec{A}$ related to the electric field by $\vec{E} = \nabla\varphi - (1/c)(\partial\vec{A}/\partial t)$. Using the gauge invariance ($\varphi = 0$) of the time-dependent Schrödinger equation

$$H\psi = \mathrm{i}\hbar\dot{\psi} \tag{1.11}$$

and the Coulomb gauge $\nabla\vec{A} = 0$ which states that the wavelength of the radiation that excites the atom is large compared to the spatial dimensions in which the excitation takes place, the proportionality $\vec{E} \propto \vec{A}_0$ results, where the latter describes the vectorial amplitude of the vector potential, related to the unit polarization vector $\vec{e}_0$ of the light by $\vec{A}_0 = A_0\vec{e}_0$. The description of photon emission from electronic transitions follows the quantum mechanical concept of perturbation theory. The Hamilton operator H that describes the electron–nucleus interaction in an atom is extended by the (small) perturbation due to electromagnetic radiation, represented by its vector potential $\vec{A}$; the resulting Hamilton interaction operator has the form [cit] $(p - (q/c)\vec{A})^2$ where p denotes the momentum operator $p = (\hbar/i)\nabla$. Multiplication yields the terms $p^2 - (q/c)\vec{A}p - (q/c)p\vec{A} + (q^2/c^2)A^2$ because momentum and position do not commutate. With the assumption of a small perturbation, the quadratic term can be neglected; the term $p\vec{A}$ vanishes because of the aforementioned Coulomb gauge. Only the second term, $\vec{A}p$, describes the interaction in the framework of first-order perturbation theory. The Schrödinger equation for this case then reads

$$\left[-\frac{\hbar^2}{2m}\Delta + V - \frac{q}{mc}\frac{\hbar}{i}\vec{A}\nabla\right]\psi = \mathrm{i}\hbar\dot{\psi}, \tag{1.12}$$

where the third term within parentheses describes the perturbation due to absorbed or emitted radiation. Elimination of the time dependence (see, for instance [20]) yields for the transition probability between states with main quantum numbers n, m:

$$W_{nm} = \frac{2\pi}{\hbar^2}\,|H_{nm}|^2\,\delta(\omega_{nm} - \omega). \tag{1.13}$$

The perturbation term H_{nm} is given in Dirac's bracket notation (see Sect. 6.2) by

$$H_{nm} = -\frac{q}{mc}\frac{\hbar}{i}\,\langle\psi_n|\,\vec{A}_0\,\mathrm{e}^{\mathrm{i}kr}\nabla\,|\psi_m\rangle \tag{1.14}$$

and the Delta function in (1.13) describes the energy conservation, i.e. that transitions do occur only if the transition energy $\hbar\omega_{nm}$ equals the wavelength of the electromagnetic radiation. It can be shown that the transition probabilities for photon absorption and emission are equal and given by

$$W_{nm} = W_{mn} = \text{const.}\, f(\omega_{nm}) q^2 |\, \langle \psi_n |\, \vec{e_0}\, \mathrm{e}^{\mathrm{i}kr} \nabla\, | \psi_m \rangle\, |^2. \tag{1.15}$$

In the case of optical transitions, with a wavelength in the visible spectral range and an atomic diameter in the 0.1 nm range, the product $\vec{k}\vec{r}$ is of the order of 10^{-3}. The series development of the exponent in (1.15) then reduces to the first term, yielding just $\vec{\mathrm{e}}_0$. Then, the bracket in (1.15) can be written in terms of the matrix elements of the momentum operator $\boldsymbol{P}$ as

$$\langle \psi_n | \vec{e_0} \nabla\, | \psi_m \rangle = \frac{\hbar}{i} \vec{e_0} P_{nm}. \tag{1.16}$$

Using the commutation relations between momentum and position operator X yields an expression for the transition probability that contains the product of charge and position, similar to a dipole that coined the name dipole transitions:

$$W_{nm} = \text{const.}\, f(\omega) q^2 \left| \vec{e_0}\, X_{nm} \right|^2. \tag{1.17}$$

The so-called dipole approximation is valid up to photon energies of about 300–400 eV, where $(2\pi/\lambda)\ 0.1\ \text{nm}$ is $\sim\ 0.1\ <\ 1$. Accordingly, many transitions observed in (soft) X-ray emission spectroscopies (XES, RIXS) can be treated within the dipole approximation (see Sect. 4.2.6). The same holds for the lower energy end of synchrotron radiation spectroscopy (SRPES) (Sect. 2.4) and ultraviolet photoelectron spectroscopy (UPS) using He I and He II lines at 21.2 and 40.6 eV.

1.4 Quantum Electrodynamical Concepts of Light Generation and Propagation

In this subchapter, the basic concepts of quantum electrodynamics (QED), the interaction of photons with electrons, will be presented, some of which have already been treated in Sect. 1.3. In Sect. 1.2, photon–matter interaction has been described using the following concept: whereas particles are described by quantum mechanics, radiation is accounted for classically (as in the Weizsäcker–Williams approach, for instance). In this way, absorption, stimulated emission, the Compton effect and Bremsstrahlung can be and have been treated. To treat the interaction of light and matter accurately, the Maxwell equations which describe electromagnetism have to be changed according to the principles of quantum mechanics. In (1.12), photon absorption is described by the classical vector potential of an electromagnetic wave

that represents the photon by incorporation of the term into the time-dependent Schrödinger equation. In that sense, we have already introduced QED in the theory of light absorption (and spontaneous emission). In the following, a short treatise of the propagation of light through matter that is not absorbing is given, based on the concepts of QED. The fascinating part here is the description of former geometric optics and interference phenomena based on the assumption of photons as particles by analyzing their trajectories associated with respective probabilities.

1.4.1 Background

QED, usually referred to as that field of physics that treats the interaction of electromagnetic radiation with charged matter within the framework of relativity and quantum mechanics, had its beginning with Dirac's article on the quantum theory of emission and absorption of radiation [21] (see also Chap. 6). The theory belongs to the class of quantum field theories where the other prominent example is quantum chromodynamics (QCD), the theory which describes interactions between quarks in high-energy physics. In QCD, the so-called strong force is mediated by gluons, similar to the photons that mediate the electromagnetic force in QED.

In 1928, the formulation of the Dirac equation [22] (see also (7.4) in Sect. 7.3) is considered the next step in the advancement of QED, followed by the path integral formulation of quantum mechanics, introduced by Feynman in 1948 [23]. This concept is related to Lagrange's approach to Newtonian mechanics based on the calculus of variations. It arose from the problem called *brachistrochrone* that occupied the Bernouilli brothers in 1696 [24]: which shape must a ramp have that a mass will slide under the force of gravity in the shortest time between two points A and B? Jakob Bernoulli recognized that any curve that has a minimum property with regard to the shortest time *globally* must also have this property *locally* as shown in Fig. 1.5.

The argument was: one assumes that the curve ACEDB is the path of least time from A to B. If C and D are two more points on that curve, then the curve part CED must have the same property as AB, i.e. being the fastest connection between C and D. If this were not the case, there must be a path, for instance, CFD, that

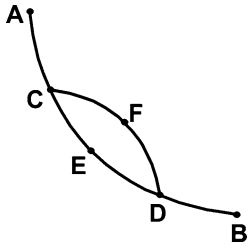


Fig. 1.5 Bernoulli's diagram for the solution of the variation problem (see text)

would provide a faster connection than the path ACEDB which is in contradiction with the initial assumption. Hence, the overall fastest path must also be fastest for all intermediate points which states that the global property holds also locally.

In geometric optics, the path of light can also be determined from a variational principle, e.g. Fermat's principle that states that light travels from any given point to another in the shortest time. This is equivalent to the statement that the phase Φ of the light wave $\xi = \xi_0 e^{i\Phi}$ varies as little as possible during the motion. In classical mechanics, the physical path followed by any object is given by the one that minimizes the integral S (called "action") where $L = T–U$ (T, kinetic; U, potential energy; v, velocity):

$$S = \int_{t_1}^{t_2} L(v = \mathrm{d}x/\mathrm{d}t, x, t)\mathrm{d}t. \tag{1.18}$$

The variation method applied to S leads to Newton's second law (often written as $F = m\mathrm{d}v/\mathrm{d}t$). To come up with a wave description of mechanics, Schrödinger guessed that the light wave phase Φ corresponds to the action S. This supposes that, according to de Broglie, frequency, energy, wavelength and momentum are related in the same way for particles with mass as for photons. One then considers a wave equation for non-zero rest mass particles. In analogy to the expression for the light wave, it was *guessed* that the wave description of mechanics might take the form

$$\psi = \psi_0 e^{iS/\hbar}, \tag{1.19}$$

where the division by Planck's constant in the exponent ensured that the argument of the exponent is dimensionless. Feynman's calculation was based on Dirac's use of the integral formulation of Huygens principle of elementary wave generation for $t_2 > t_1$:

$$\Psi(x, t_2) = \int G(x, y)\Psi(y, t_1)\mathrm{d}y. \tag{1.20}$$

$G(x, y)$ is named "kernel" or "propagator" that was determined by Feynman to be

$$G(x, y) = A_0 e^{iS/\hbar}, \tag{1.21}$$

which led him to derive the time-dependent Schrödinger equation, assuming that the action S could be approximated by (with $\Delta t = \varepsilon$)

$$S = \int L\mathrm{d}t \approx L\varepsilon. \tag{1.22}$$

Using the simplest case for a Lagrangian, i.e. the difference of kinetic and potential energy, and with some approximations, the integral in (1.20) led to the time-dependent Schrödinger equation $H\Psi = \mathrm{i}\hbar\dot{\Psi}$. Feynman realized that arbitrary times Δt such as $\Delta t = \varepsilon, 2\varepsilon, 3\varepsilon, \ldots, N\varepsilon$ could be represented by the method, repeating the integration N times. This was named a "sum over histories" and the multiple integral was called the path integral. Then, the amplitude $G(x, y)$ can be expressed for better comprehension as a sum of the amplitudes $A_0[\exp(i/\hbar) \quad S_\mu(x, y)]$, instead of the respective integrals. These amplitudes are associated with all possible paths μ that connect x and y, yielding

$$G(x, y) \propto \sum_{\mu} \exp\left(\frac{i}{\hbar} S_\mu(x, y)\right). \tag{1.23}$$

Taking fixed starting and end points a and b, respectively, the probability to travel from point a at a time t_a to b (arriving at t_b) is $P(b, a) = |G(b, a)|^2$.

From the formalism, all paths contribute equally to the probability and it has to be analyzed how, in the classical case (of geometrical optics), a particular path can dominate. In the classical limit ($S \gg \hbar$), when $S_\mu/\hbar$ becomes very large, the according phase yields a very large angle and the real/imaginary part of the phase is the cos/sin of this angle. A small change of the path δx on the classical scale results in a small change of S, but when measured in units of $\hbar$, where the angular functions will oscillate rapidly between positive and negative values, the total contribution will result in no net contribution. Therefore, paths with different neighbouring actions (S) cancel out and only those trajectories for which S is an extremum (action S_{cl}), a small change δx produces, in first order, no significant change in S. The contributions from these trajectories are nearly in phase and they do not cancel out and in this picture, the classical limit is obtained by considering those trajectories for which the action is within $\hbar$ of S_{cl}.

Hence, the least action path and their neighbouring trajectories dominate the propagation and can be regarded as the classical trajectory with an indefiniteness that shows the limitations of the classically defined path. This formalism thus provides an elegant and quite intelligible connection between classical theory and quantum mechanics in cases where the classical theory can be described by an action principle such as in geometric optics. Below, in the subsequent section, a few phenomena of geometrical optics will be represented based on the path integral formalism.

Historically, at the time of the development of path integrals, the problems of QED, related to divergent integrals in perturbation calculations, indicating that the theory fails at high energies, became obvious. Schwinger then introduced the concept of renormalization of mass, charge and wave function to avoid what was called at the time the "ultraviolet divergence". This is related to the fact that infinite terms, logarithmically divergent, contribute to the so-called "bare" electron charge and mass (m_0, e_0) which are different from that of the "dressed" high energy electrons. The renormalization replaces m_0 and e_0 by the experimentally measured

values m and e. As a consequence, the fine structure constant α is not an actual constant but instead increases with energy ($\alpha = 1/137$ for low energy and $\alpha = 1/128$ for 100 GeV, for example) [25, 26]. For the renormalization of the electron mass, the requirement that the energy of the physical state which corresponds to a free electron moving with a momentum $\vec{\mathrm{p}}$ is equal to $\sqrt{p^2 + m_{\mathrm{e}}^2}$, where m_{e} is the electrons rest mass, was introduced. In a similar manner, the *observed*charge was renormalized. QED has then been very successful in the determination of the Lamb shift [27] and of fundamental nature constants, for instance [28]. In the light of some recent experiments, however, aspects of QED have been questioned recently [29]. The development of quantum chromodynamics (QCD) as a renormalized theory was the next major step that described elementary particle physics by a quantum field theory approach [30–32]. For an impression of some further developments in modern physics, see Sect. 6.3.

1.4.2 Light Propagation Concept

In QED, the wave–particle duality of light, for instance, has been replaced by stating that light has corpuscular character. The disadvantage of this concept is that only *probabilities* for the behaviour of photons can be calculated to describe, for example, whether a photon will hit a detector. The QED concept, represented by (1.20, 1.23), can be used to *visualize* the behaviour of light in basic optics, describing seemingly simple processes such as reflection, refraction and diffraction in the framework of QED. First, one considers the probability amplitude Ψgn (1.23) and the corresponding photon paths in a reflection experiment at a mirror as shown in Fig. 1.6. To each path, an arrow can be assigned that contains the information on the amplitude (probability) and phase of the repective photon path. With the velocity of light, the arrow turns 14,400 times per cm for the red part of the spectrum. When the light reaches the detector, the turning movement of the arrow is stopped, giving a fixed direction that results from the path length. Such arrows are displayed in Fig. 1.7 for the mirror segments A–N.

The quantum physical interpretation of the experiment states that the probability for each photon paths is virtually equal, meaning that the lengths of the arrows are drawn equal. The free parameter is then the arrow's direction, given by the elapsed travelling time of the considered photon. For the reflection experiment in Fig. 1.6, the lengths of the paths S–A–D are larger than, for instance, those of S–G–D. Therefore, the time vs. mirror segment graph shows a curve which has its minimum at the intersection between segments G and H where specular reflection is known to occur in geometrical optics. The directions of the arrows for the outer mirror segments become more statistical due to the larger time difference in the outer segments compared to those in the centre of the mirror as can be seen in Fig. 1.7(*top*). The final arrow is obtained by usual vector addition as shown in Fig. 1.7(*bottom*), yielding a non-vanishing resulting vector that represents the actual path integral, visualized here as a sum of photon paths.

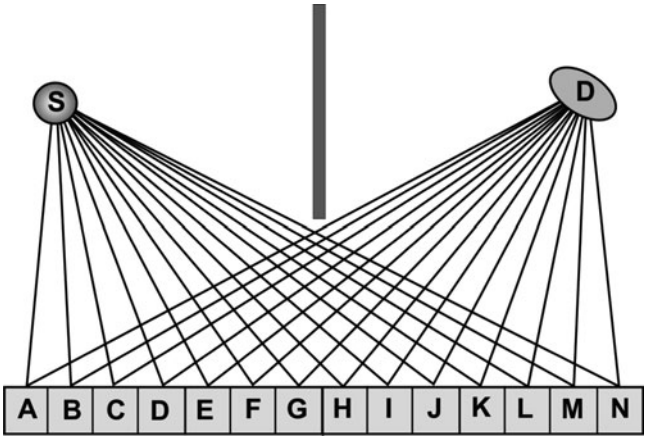


Fig. 1.6 Photon path summation in a reflection experiment with a screen between the source S and the receiver/detector D; for simplification, the reflection is assumed to occur solely at the front surface; the segments A–N indicate parts of the mirror for analysis of the temporal difference of the photon paths (see text)

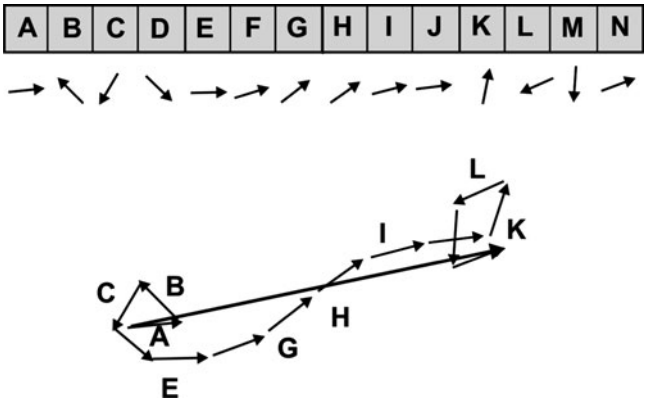


Fig. 1.7 (*Top*) Time delay for the photon paths A–N; (*bottom*) summation of the probability amplitudes of the photon paths depending on the elapsed time

The used picture can easiest be understood in analogy to the solution of the wave equation for plane waves:

$$\Psi(x,t) = A\exp[\mathrm{i}(kx - \omega t)], \tag{1.24}$$

where A is the length of the arrow and t is the temporal evolution of the trajectory up to a specific event such as the detection of a photon. The propagation process is then simulated by an arrow of a certain length, equivalent to the probability

amplitude A, and the elapsed time which defines the angular position of the arrow. It can be seen from Fig. 1.7 that the photon trajectories for larger deviations from the geometrical path cancel each other largely, and only those contributions for which the photon travelling time is very similar contribute to the final arrow from A to K. In a similar manner, the concept allows the description of the optical properties of lenses, diffraction gratings, for instance, and reflection and transmission of light.

1.4.3 Photon Generation for Quantum Information Science

With the advent of so-called *cavity* QED, it became possible to modify the spontaneous emission rate of a single atom in a resonant cavity, to alter the Lamb shift, i.e. atomic energy levels, and to induce oscillatory energy exchange between a single atom and a cavity mode, for instance. Cavity QED dates back to the work of Purcell [33] that states that spontaneous emission is not an intrinsic property of an isolated atom but rather a combined property of an atom coupled to its electromagnetic environment. Actually, Purcell showed that the enhancement of the rate f for spontaneous emission of electromagnetic radiation into the cavity is a function of the wavelength, the mode volume of a resonator V_m and the quality factor Q:

$$f = \frac{3Q\lambda^3}{4\pi^2 V_m}. \tag{1.25}$$

The enhancement in (1.27) is that over the spontaneous emission rate without a cavity. A shortened derivation of this expression will provide some understanding of this so-called Purcell effect of interaction of photons and electrons in confined spaces.

One considers an atom in a cavity V_0 which is large enough to exhibit negligible influence of the emission properties of the atom. The Purcell factor (1.25) is obtained dividing the transition probabilities for spontaneous emission by that of emission in a narrow *confined* cavity, e.g. $f = W_{\text{cav}}/W_{\text{free}}$. First, W_{free} is calculated from Fermi's golden rule in the weak perturbation limit:

$$W_{\text{free}} = \frac{2\pi}{\hbar^2} \left|M_{if}\right|^2 g(\omega), \tag{1.26}$$

where M_{if} is the transition matrix element for spontaneous emission and $g(\omega)$ is the photon density of states. With the photon modes in free space, $g(\omega) = \omega^2 V_{0/}\pi^2 c^3$. The matrix element is given by the electric dipole interaction (see Sect. 1.3). Using $E_{\text{vac}} = \sqrt{\hbar\omega/2\varepsilon_0 V_0}$ for the vacuum field (there is no internal field source within the cavity assumed), taking μ_{if} as the actual atomic dipole moment and averaging over all orientations of the dipole yield

$$\left|M_{if}\right|^2 = \frac{1}{3}{\mu_{if}}^2 E_{\text{vac}}^2 = \frac{\hbar\omega\mu_{if}^2}{6\varepsilon_0 V_0}. \tag{1.27}$$

For the emission rate W_{free}, the result is obtained by multiplication in (1.26) using $g(\omega)$ and (1.27):

$$W_{\text{free}} = \frac{\mu_{if}^2\omega^3}{3\pi\varepsilon_0\hbar c^3}. \tag{1.28}$$

One now considers an actual cavity of volume V_{cav} that has only one resonant mode which is close to the emission frequency of the assumed two-level atom. Using a normalized Lorentzian function for $g(\omega)$ and taking exact resonance, $g(\omega = \omega_0)$ can be expressed as

$$g(\omega) = \frac{2Q}{\pi\omega_0} \tag{1.29}$$

and with the matrix element for the interaction in the cavity, assuming dipole alignment,

$$\left|M_{if}\right|^2 = \frac{\hbar\omega\mu_{if}^2}{2\varepsilon_0 V_1}, \tag{1.30}$$

the emission rate is obtained from multiplication according to the golden rule above and the ratio for resonant behaviour $W_{\text{cav}}/W_{\text{free}}$ yields the (original) Purcell factor presented in (1.25). Q, the quality factor, describes the frequency quality of the resonance.

This possibility to enhance spontaneous emission from atoms in a cavity provides a tool for generation of photons with desired properties, suited, for instance, for information processing. Quantum information systems are based on superposition and entanglement of quantum mechanical states. Whereas superposition is rather easy to comprehend, entanglement needs a short description: it represents that property of a quantum mechanical state of a system of, for example, two objects for which the quantum states of the objects are closely linked in such a way that each object cannot be described any longer without taking fully into account the other object. This even holds for spatially separated objects and will be treated in more detail in Chap. 6 and, in particular, in Sect. 6.2.1. Clear evidence for entanglement between individual quantum systems has been obtained in photon down-conversion, with trapped ions (see Sects. 6.3), and in cavity QED. An example for the latter is the interaction of a two-level Rydberg atom with a cavity field mode (see above) where two qubit quantum logic gates as basic unit of quantum information processing systems can be realized.

1.5 On the Origin of Life: Photonic Contributions

In the late nineteenth century, the mysterium of the origin of life and its possible natural cause started to be intensively discussed among scientists [34] but also entered the arts. An excellent example is *Gustav Mahler's Sinphony No. 3* (1898) where the extended first movement symbolizes the transformation of inorganic matter towards life. Meanwhile, with the dramatic increase in the knowledge of biochemical processes and reactions, the topic can be discussed on a considerably advanced scientific level and examples of the presently advocated models and theories will be given below in Sects. 1.5.3.–1.5.5. The complexity of the bioorganic chemistry involved and the uncertainties regarding the actual environmental conditions that were present when life originated, however, has so far not resulted in a single widely accepted model. Rather, schools have developed that occur to maintain borders between them by advocating their respective hypotheses. These hypotheses will be reviewed below, and selected aspects, particularly those related to the influence of photons, will be treated in somewhat more detail.

Before addressing the contemporary concepts on the origin of life, a definition of the term "life" is needed. An overview of this aspect is given at the outset of this chapter. Subsequently, the predominant advocated theories on life's origin are reviewed, whereafter the possible role of photons in the development of life forms will be discussed. Among the topics considered are the homochirality of bioorganic compounds, prebiotic chemistry, metabolic concepts, the so-called RNA world and panspermia. As the latter puts the origin of life at an earlier time and possibly different conditions than those on earth several billion years ago, the influence of presumed specific conditions on earth on the creation of early life forms at a time of about –3.9 Ga is also considered.

Whether photons had a role in generating life is debated. Here, presently advocated models, theories and concepts are discussed with emphasis on possible photonic processes in the generation of biological molecules/structures.

1.5.1 Definitions

A satisfactory definition of what life is remains a challenge, and numerous approaches exist that encompass mathematical, cybernetic [35], biological and cross-disciplinary formulations. Here, we follow the latter approach by presenting a slightly modified version of the recent definition by Koshland [36] because of its generality and intelligibility although parts of it have been debated [37].

In an attempt to describe a living organism in a single sentence, one might come up with wordings like "life is defined as an organized structure that is able to carry out self sustaining metabolic reactions, responds to external stimuli, is able to protect itself and can reproduce". Actually, elegance and simplicity might not always be possible at an intermediate stage of development of a field and, possibly,

might not generally be achieved (see Sect. 7.3 where the current attempts on a unifying theory of the forces in physics are described).

The essential principles which define life might be listed as follows (1) the existence of a program that contains the (encoded) information on the ingredients of the organism and their dynamic interactions, given by the DNA. The program already contains the dynamic (fast) reactions to environment as behavioural response; (2) exchange of energy with the environment by being the so-called open metabolizing system and (3) the capability to change (improvize) in order to adapt to changes in the biosphere of the organism. This occurs by mutation and successive selection of the mutant by its environmental adaptivity; (4) protection from the environment as realized by membranes, forming an individual organism that functions in an autonomous manner; (5) the possibility to regenerate, thereby compensating thermodynamic losses of biochemical reactions and achieving stability of the overall metabolism and structure. Such regeneration must be contained in the overall program code and (6) selectiveness of chemical, biochemical (enzymatic) reactions to allow for the simultaneous occurrence of undisturbed reactions and reaction pathways.

The above preconditions appear necessary for the functioning of a living organism. These principles have been derived from inspection of life on earth but the mechanisms to arrive at a living organism might be variable. Presently, the mechanisms are far from being optimized in this intermediate stage of evolution and it might be interesting whether a different or modified set of mechanisms could result in life. This raises the question of "process tolerance" in the origin of life, i.e. the possible width of starting conditions that still result in the formation of living organisms. Hence, for a certain non-negligible width of the processing window, this could point to the possibility of life originating outside earth under quite different conditions on exoplanets that will be discussed below.

1.5.2 Geological Periods and Human Development

As introduction, the geological periods that are distinguished to describe earth's history are summarized as they are closely related to the possible origin and further developments of life on the planet. First, the *Precambrian* time is considered. Figure 1.8 gives an overview including a few comments on the milestones in the development.

Almost four billion years (Ga) passed before the first animals left their trace on earth, documented by fossils. This longest period in the history of earth (about 7/8 of the time) is named Precambrian. During that time, the most relevant biological developments took place. The Precambrian has been divided into five ages. The hadean is not actually a geological time as – except for meteorites – no rocks on earth are that old. The oldest meteorites and lunar rocks are about 4.5 billion years old and the oldest rocks known on earth are 3.8 billion years old. The hadean age therefore stands for the time when the solar system and the earth were formed.

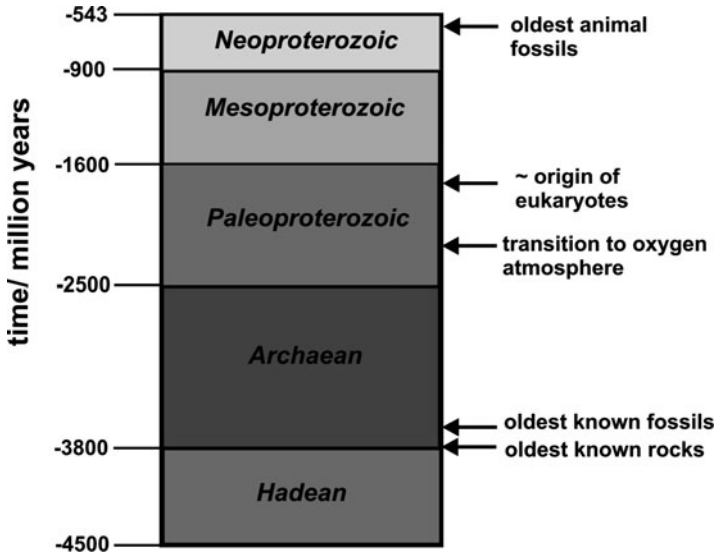


Fig. 1.8 Geological periods in the history of earth relative to the present time including some development milestones

First forms of life appeared early in the archaean. The oldest fossil findings (of bacteria microfossils) are about 3.5 billion years old. All life during this more than a billion years long period was of bacterial nature. The proterozoic period began at –2.5 billion years and it ended 543 million years ago. Available data about organisms that lived during the proterozoic period are sparse. The so-called arcritarchs were present in the late Paleoproterozoic and they are interpreted as unicelled photosynthetic protists [38]. Because almost all of their nanoscale molecular organization has been lost, only very little structural information exists. They are generally described as a group of decay-resistant organic-walled vesicular microfossils. During this time, the atmosphere became oxygen enriched by the occurrence of photosynthesis. The occurrence of eukaryotes, where a cell nucleus contains the genetic material, paved the way for multi-cellular organisms which took advantage of aerobiosis. Compared to eukaryotes, the genetic material in prokaryotes is found within a nucleoid, where a single loop chromosomal DNA is stored that lacks a nuclear envelope. Prokaryotes existed from about –3.5 Ga as evidenced by findings of cynobacteria-like fossils (see Fig. 1.8).

The oldest animal fossils date back 620–550 Mio. years. The end of the Precambrian period is characterized by a change in the carbon cycle that resulted in the ability of the new animals to synthesize hard body parts of carbonate minerals. Whether this change was due to geochemistry or genetic alterations is debated. In the Cambrian period, fossil findings show that most of the major groups of animals appear in this relatively short time between 543 and 590 Mio. years ago. Because of the relatively short time span over which this diversity of

forms developed, this development has led to the wording “Cambrian explosion”. The Cambrian period belongs to the paleozoic aera (–543 to –248 Mio. years) that is characterized by two rather little understood and debated events: the first one, as already mentioned, was the explosion of diversity of life forms near the beginning of the Cambrian period that occurred within a few million years and the second one took place at the end of the Cambrian, when, in the largest mass extinction on earth, 90% of all marine animals and 70% of all terrestrial organisms became extinct.

The more recent (bio)geological periods are condensed in Fig. 1.9. There were additional major extinction events in the Triassic and Cretaceous periods, the latter leading to the extinction of the dinosaurs about 65.5 Mio. years ago whereas mammals and bird clades survived the event with a few extinctions. The Pleistocene, characterized by advances and retreats of glaciers and also called the ice age, is also characterized by the evolution of the homo-sapiens (the first known hominids were human-like primates that evolved in eastern Africa at –5.2 Mio. years; humans and chimpanzees shared their last common ancestor about 7 Mio. years ago). Also, mammoths and sabre-toothed tigers date back to this period. The Pleistocene saw the homo-erectus (about –1.8 to –1.6 Mio. years), the homo-heidelbergensis (–0.8 Mio. years), modern man (–160 k years) and, in addition, the homo-neandethalensis. Archaic homo-sapiens evolved between 400,000 and 250,000 years ago. The search for human ancestors has been invigorated by the use of non-recombining genetic

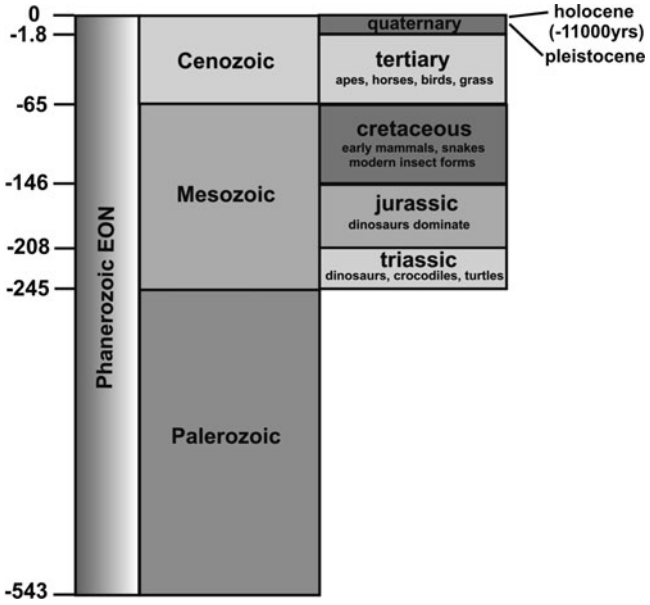


Fig. 1.9 The Phanerozoic aera and its subdivision including a few indications on the occurrence of biological forms; the palerozoic has been described above and corresponding life forms are omitted in this figure; the *left-hand side* shows the geological time span in million years until the present (0); the Holocene and Pleistocene are treated in the above text

systems such as the Y chromosome (paternally inherited part) and mitochondrial DNA (maternally inherited) [39]. The Holocene period is defined by the development of human civilization, advances in knowledge and technologies and a rather warm climate as this epoch began at the end of the last glacial period. Also, the climate variations were weaker than those in the preceding epoch but nevertheless non-negligible [40] and the collapse of the Akkadian (∼–4,600 years) as well as that of the Maya (–1,200 years) culture has been attributed to prolonged droughts [41].

1.5.3 Overview of Hypotheses on Life's Origin

1.5.3.1 Panspermia

In 1894, in a speech to the British association, Lord Kelvin proposed that impacts between planetary bodies could scatter meteoric material that contains forms of life through space eventually propagating life between the stars and throughout the universe [42]. This hypothesis, called panspermia, coined by Arrhenius in 1903, has been discussed vividly [43] and very recently, the finding of the amino acid glycine of extraterrestrial origin in the dust of the comet Wild 2 [44] has sparked new interest in this idea. In parallel research efforts, exoplanets have been identified that are believed to show surface conditions similar to those on earth, such as the recently identified planet moving around the red dwarf Gliese 581 [45].

Three major implications regarding panspermia can be drawn in the present situation (1) if the building blocks of early life did not originate from earth, the conditions for formation might have been different from those that were present on earth, but synthetic organic chemistry and biomolecule stability result in a distinct narrowing of the synthesis window; (2) if only life's building blocks but not complete life forms such as bacteria have been distributed, the organization of higher life forms must have become possible in terrestrial environments or (3) biological molecules or their performs can originate from interstellar grains as what is called *cold seeds* of life [46].

1.5.3.2 Prebiotic Synthesis: Heterotrophic Approach

As Fig. 1.8 shows, the oldest known forms of life date back 3.8 billion years. Considering the age of the earth, life, if it was not transported to earth, had a window of a few hundred million years to evolve. The questions that arise in this context concern the sources of small organic molecules that were building blocks of the first self-replicating systems and the evolution of biological organization and functioning from a supply of abiotic small organic molecules. The first question has been addressed in prebiotic synthesis experiments and in surface metabolic models (see below), whereas the second one relates to the so-called RNA world and will also be treated further below.

The heterotrophic model, which is still largely accepted, relies on the abiotic synthesis of rather complex nutrients. A heterotrophic organism uses nutrients from its environment that must have been synthesized earlier. A key experiment on the building blocks of life that initiated intense research has been performed by Miller [47]. In electric discharge experiments in reducing atmosphere (containing CH_4, NH_3, CO, H_2 and H_2O), it was found that a substantial percentage of the carbon present in the simulated early earth environment mixture (ocean–atmosphere system) has been incorporated into small organic molecules. Among them, naturally occurring amino acids were prominent reaction products. Altogether, nine different amino acids were found in this first experiment. In subsequent experiments, using more advanced analytical techniques, many more, of which some, but definitely not all, amino acids found in living organisms, could be synthesized by electric discharge experiments [48].

1.5.3.3 Autotrophic Processes

It appears widely accepted that a heterotrophic organism is simpler than an autotrophic one because the latter must be self-sustaining by producing food from inorganic compounds using light or chemical energy. The environment of the autotrophic organism, however, can then be simpler in the sense that organic compounds are not needed for survival. For an autotrophic organism, it is assumed that the system produces all its constituents from CO_2 (or HCOOH and CH_4) and H_2O. Two rather developed autotrophic models have been introduced: Morowitz suggested reduction of CO_2 to organic compounds by charge separation of electrons due to light-generated electrochemical potentials [49], whereas Wächtershäuser advocated a chemoautotrophic process by oxidative formation of pyrite according to

$$FeS + H_2S \rightarrow FeS_2 + 2H^+ + 2e^-, \tag{1.31}$$

where FeS and H_2S act as reducing agents. The reaction exhibits a strong reducing potential and the final electron acceptor is proposed to be CO_2, thus reducing carbon compounds by oxidative formation of pyrite [50]. It has been further outlined that the products of the fixation of CO_2 become bonded to the surface of pyrite where they accumulate and can undergo successive reactions, thereby establishing a surface metabolic two-dimensional reaction system. The strong adsorption proposed as well as the reduction of CO_2 to, for instance, amino acids has not yet been proven.

This points to a pronounced general shortcoming of the research on the origin of life: experimental evidences and, in particular, detailed model experiments in a variety of plausible environments are largely missing. Here, a cross-fertilization between biology, chemistry, surface science, physics and informatics could result in much more than incremental advances in our understanding.

1.5.3.4 RNA World

The phrase "RNA world" has been coined by Gilbert referring to the then recently known catalytic properties of RNA [51] that were based on the discovery of enzymic activities in RNA molecules of *Escherichia coli*. As a consequence, it was contemplated that RNA molecules could catalyze their own synthesis, leading to an RNA world where protein enzymes would not be needed at the beginning of evolution. A good example of the dominance of speculations compared to actual experimental modelling is the long-held opinion that one of the serious difficulties of the RNA world concept lies in the prebiotic synthesis of pyrimidines (cytosine, uracil) and, particularly, in the difficulty of attaching them to ribose to form pyrimidine nucleosides. This belief resulted in the postulation of a pre-RNA world, i.e. of a simpler genetic system(s) that preceded RNA. It was then supposed that the evolutionary advances by that more ancient system were "somehow" carried over into the RNA world [52]. Very recently, it has been demonstrated, however, that an altered synthesis route can circumvent the difficult condensation with pyrimidines. Here, the nucleotides have not been prepared separately but rather via reaction of comparably simple chemical precursors with subsequent simultaneous incorporation into one molecule [53]. This advance in prebiotic chemistry eliminates at least some of the increasing scepticism that developed regarding the RNA world concept due to difficult to envisage chemistry.

1.5.4 Selected Mechanistic Aspects: Quantum Tunnelling and Prebiotic Nucleotide Synthesis

In this subsection, a few subjectively selected processes are described in more detail following the above chosen sequence. They encompass aspects related to panspermia and prebiotic organic synthesis as well as the recent findings related to synthesis of RNA. Photonic processes are treated in subchapter 1.5.5.

The time span that accounts for the origin of life on earth is surprisingly short (see also Fig. 1.8): between the formation of the planet, i.e. the accretion of earth with a metal core at about –4.5 Ga, and the age of the oldest known fossilized bacterial aggregates is ~3.5 Ga, yielding a time window of 1 billion years. In addition, earth was very likely bombarded at –3.8 to 3.9 Ga (so-called late heavy bombardment, for which evidence was found on the moon) by meteorites that most likely sterilized the entire planetary surface, including submarine hydrothermal systems, due to the conversion of all oceans to steams. This results in only 400–500 Mio. years for geogenesis. As the leap from a mixture of amino acids to a bacterium is considered much larger than from bacterium to people, the question arises why the temporal development of the former has been so much faster than the development of animals and plant life which took about three billion years. This argumentation supports the concept of an extraterrestrial origin of life (1.5.3) and below some related scientific aspects are outlined.

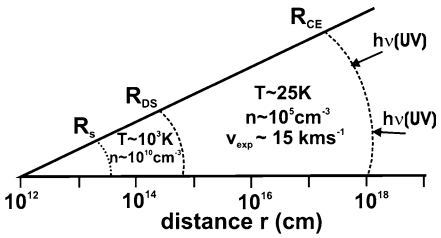


Fig. 1.10 Radial cross section of a circumstellar envelope: R_S, radius of the star; R_{DS}, radius of the dust shell near the stellar photosphere that is warm and dense ($n \sim 10^{10}$ cm^{-3}); R_{CE}, radius of the circumstellar envelope; v_{exp}, velocity of the outward flow of matter and UV irradiation is also indicated

1.5.4.1 Panspermia-Related Processes

In the later stage of stellar evolution, circumstellar envelopes are created by an extensive mass loss. These envelopes exhibit physico/chemical conditions as shown in Fig. 1.10 for a schematic cross section of an asymptotic giant branch star (AGB).

The envelope is characterized by an inner and an outer shell boundary. In the former, local thermal equilibria are possible, forming molecules such as CS, NH_3, HCN and HCCH, as has been demonstrated in IR measurements [54]. At intermediate radii, HCN and CO have been detected and at the outer radius (R_{CE}) photons and cosmic rays create radicals such as CN, CCH, C_3H and C_4H, detected in interferometer maps. The occurrence of organic molecules in dense interstellar clouds might originate from carbon-rich fragments that have been formed in circumstellar shells. Meanwhile, a plethora of carbon compounds has been identified and many of them have been detected in planetary nebulae, too. The ultimate origin of this organic material could be coupled to the carbon-rich chemistry of stellar envelopes and might have led to living systems.

The panspermia hypothesis has been corroborated by the discovery of a low-temperature quantum limit in the study of radiation-induced polymerization of formaldehyde [55]. The rate-determining step was identified as the tunnelling of the whole CH_2O molecule because proton tunnelling could be excluded by investigating the polymerization rate of deuterated formaldehyde. The effect has been described in the framework of quantum classical dynamics where electrons (mass m) are treated quantum mechanically and the dynamics of larger particles with mass M (see also Born–Oppenheimer parameter $\ell = (m/M)^{1/4}$) are approximated classically. This approach allows the correct description of photochemical and biochemical systems, for instance. The Born–Oppenheimer approximation for such systems where the nuclear motion proceeds without changes of the quantum state of the electron cloud describes the adiabatic case. Non-adiabaticity involves changes in the population of adiabatic states with a change in nuclear configuration and such distortion of the electron density in turn influences the nuclear trajectory. This so-called electron-nuclear feedback is of significant importance for many chemical systems.

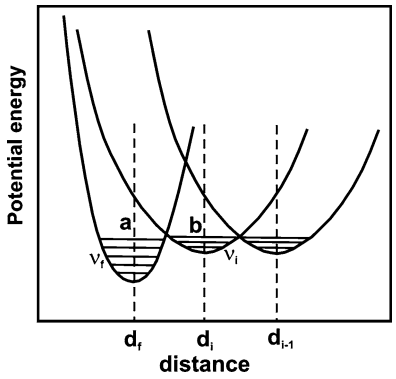


Fig. 1.11 States of formaldehyde molecules in the adiabatic approximation symbolizing the addition of a CH_2O molecule to the growing polymer chain $\ldots CH_2$–O–CH_2–O–CH_2–O– $\ldots$; also shown is the state for the subsequent molecule **b-1** that is added to the polymer chain after state **b** has made the transition to **a**; $v_{i,f}$ are vibrational indices of initial and final states

Considering the adiabatic potential wells for the case of formaldehyde in Fig. 1.11, the system would be in a stationary state where the initial state **b** is occupied and the final state **a** is empty. For small distances between initial and final states, the adiabatic approximation is corrected by a non-adiabaticity term and the transition probability **b**→ **a** is given by the expression (with i initial and f final state):

$$W_{if} = \frac{2\pi}{\hbar} L^2 F_{\upsilon i,\upsilon f} \rho_f . \tag{1.32}$$

In (1.32), $\boldsymbol{F_v}$ denotes the Franck–Condon factor that contains the vibrational states

$$F_{\upsilon i,\upsilon f} = \left| \langle \varphi_{\upsilon i}(q) \varphi_{\upsilon f}(q) \rangle \right|^2 . \tag{1.33}$$

The additional term in (1.32) is expressed via the adiabatic electron wave functions and the non-adiabaticity operator $\bar{L}$ as

$$L = \langle \psi_i \bar{L} \psi_f \rangle . \tag{1.34}$$

For the case of formaldehyde, the estimated distance d_f–d_i is $\Delta d \sim 0.4$–0.5 Å. This rather early discovered behaviour [46] regarding quantum tunnelling that can lead to finite reaction rates and thus to chemical reactivity at low and ultralow temperatures points to its possible role in chemical and biological evolution (cold seeds of life).

Finite reaction rates at temperatures close to the absolute zero, caused by quantum tunnelling, can lead to the exothermic formation of rather complex molecules

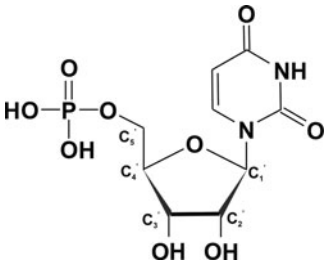


Fig. 1.12 Drawing of the ribonucleotide uridine-monophosphate; the sugar molecule ribose is connected to the uracil base at the position C_1', and the OH group at the C_5' position is forming an oxobridge with the phosphoric acid group. For polymerization, the C_3' OH group connects to the phosphate group of an adjacent nucleotide

and thus this effect suggests the possibility of the so-called cold prehistory of life. It is then plausible to consider interstellar grains as cold seeds of life. The chemical reactions in the mantles of such grains can lead to complex organic compounds, particularly under UV irradiation, such as methyl formate and glycoaldehyde. Hence, these findings support the hypothesis of a possible prebiotic evolution in interstellar clouds.

1.5.4.2 RNA World: Synthesis and Attachment of Pyrimidine Nucleotides

RNA, being considered the first informational polymer that emerged from the prebiotic environment, has been intensively investigated with respect to nucleotide synthesis and polymerization. Here, we consider the missing prebiotic link in the RNA world hypothesis, e.g. the overcoming of the reaction barrier in the formation of ribonucleotides such as uridine-monophosphate (Fig. 1.12) by synthesis of a new precursor (2-aminooxazole), thus circumventing the difficult condensations with pyrimidines.

The classical and the novel approach of nucleotide synthesis is visualized in Fig. 1.13. The classical condensation of the pyrimidines uracil and cytosine with the ribose at the $C_1' - OH$ position of the ribose ring does not virtually occur (upper part of Fig. 1.13), leading to serious doubts whether the assumption of an RNA world is realistic. The altered synthesis pathway (bottom of Fig. 1.13) does not proceed via free ribose but, instead, uses 2-aminooxazole formed by condensation between glycoaldehyde and cyanamide, thus creating parts of the sugar and the base simultaneously. Some of the intermediate steps are indicated in Fig. 1.13. The starting molecules of this reaction sequence are plausible prebiotic species and their synthesis conditions are in accordance with geochemical models on the early earth conditions. Despite the fact that glycolaldehyde and glyceraldehyde must be introduced in a temporal separate manner, this reaction scheme represents the first robust pathway from prebiotic reagents to *informational polymers* which

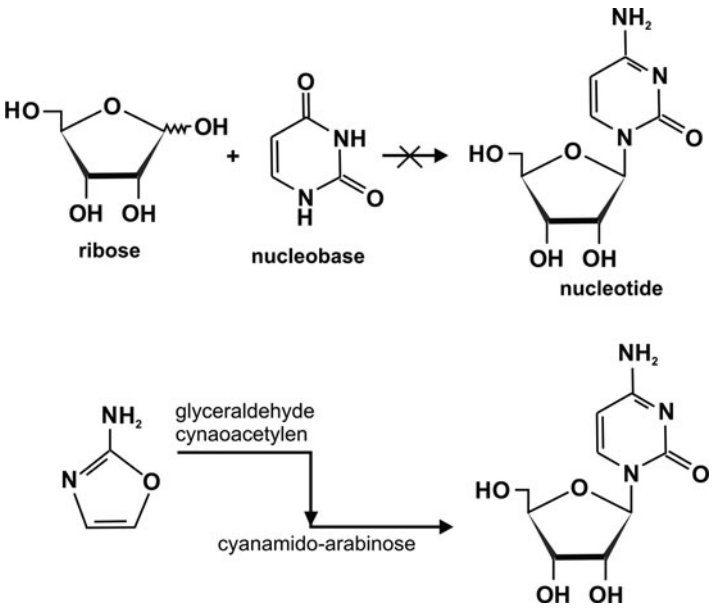


Fig. 1.13 Comparison of the synthesis chemistry of ribonucleotides upon separate condensation (*top*) of ribose and base and (*bottom*) using the joint precursor aminooxazole (see text)

also act as catalysts in complex reactions. As stated at the end of the note on autotrophic processes, the described reaction is another example of the need for model experiments that are more interdisciplinary.

1.5.5 Influence of Photons in Biogenesis: Photochemistry and Homochirality

Figure 1.10 gives an indication of the ubiquitous presence of high-energy photons in premordial times – in the circumstellar envelopes, in interstellar space and on early earth (not protected by an ozone layer). Below, two aspects of the influence of photons on the occurrence of life are considered (1) the photo-driven synthesis of prebiotic molecules and (2) the influence of photons on the *homochirality* of biological molecules. The latter topic, not yet being addressed, will be treated in more detail. The former, related to the basic photochemical reaction that results in the formation of prebiotic organic molecules (related to the heterotrophic and panspermia hypotheses, see above), will be reviewed initially.

Life, as a non-stationary, non-equilibrium state where matter, undergoing chemical as well as biological evolution, has to be kept in this state away from equilibrium. This needs a driving force which also allows living organisms to maintain their

higher order. This driving force has the function to induce higher energy states in matter compared to that of thermal equilibrium. As already argued by Oparin and Haldane [56], the UV radiation impinging on the early earth should have significantly contributed to the synthesis of prebiotic organic compounds. As high-energy radiation is also present in circumstellar envelopes and beyond, this scenario includes the synthesis in outer space and/or on exoplanets. Compared to the other energy sources available at the stage when life was formed, solar radiation had a distinct quantitative advantage as it is and has been an intense permanently available energy form.

Three types of prebiotic molecules can be discerned: small inorganic molecules in reducing early atmosphere such as CO, CO_2, H_2O, H_2 and CH_4; inorganic salts (phosphates and carbonates) and small organic molecules such as cyanides, pyridines and nucleic acid monomers. The absorption edge of these species varies from ∼320 nm (organic molecules) to ∼220 nm (inorganic salts) to ∼190 nm (small molecules). Thus, all components of the primitive reducing atmosphere were transparent to UV radiation of $\lambda > 200$ nm (corresponding to ∼6 eV) and thus allowed UV photo-induced reactions also with solid and liquid phases. Some aspects of the related chemistry will now be reviewed.

1.5.5.1 Basic Photochemistry

Photoexcitation of molecules has two primary effects (1) an increase in energy because the excited electronic states are 2–5 eV higher in energy and (2) a pronounced redistribution of the electronic charge. This change can open reaction pathways with low activation energy, thus effectively competing with the decay of the excited state. This relaxation to the ground state exhibits considerably different rates for singlet and triplet states (compare also Sects. 3.5.1. and 3.5.2), making reactions with triplet states more likely. Furthermore, a configuration-independent property of excited states is their much easier reduction and oxidation because the excitation energy must be added to the reduction/oxidation potentials. Hence, photochemical reactions are characterized by much larger rates than ground state reactions and, due to differences in electronic structure, reactions can occur that are not possible in the ground state. Figure 1.14 shows a comparison of ground state and excited state chemistry for a ketone, belonging to the carbonyl compounds as the aldehydes. The latter are known to readily polymerize under UV irradiation to form sugars or sugar-like materials and, also, react with nitrogen compounds such as HCN, NH_3 and amines.

A more detailed description of the processes related to photoexcitation of organic molecules is presented in Fig. 1.15, where the two major electronic excitations in a carbonyl group are displayed. The $\pi \rightarrow \pi^*$ transition is characterized by a much larger oscillator strength than the $n \rightarrow \pi^*$ transition.

The higher reactivity of the photoexcited molecules originates from the photonic excitation energy that leads to an electron in the former unoccupied LUMO and a hole (missing electron) in the HOMO and, therefore, the excited molecule can be

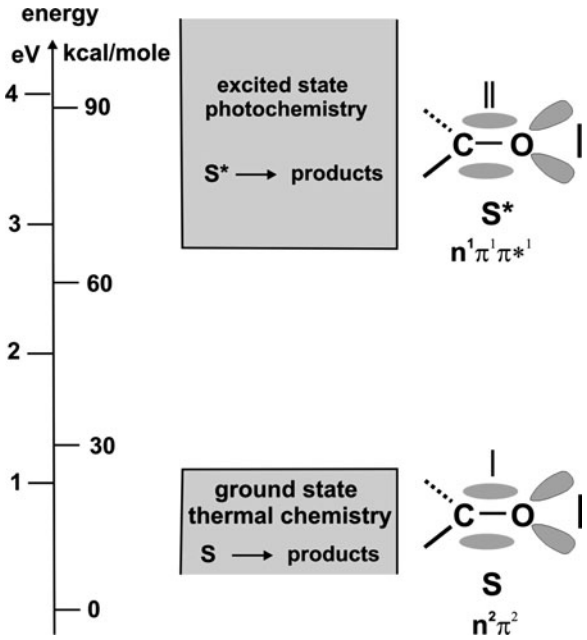


Fig. 1.14 Thermal ground- and (photo)excited state chemistry (S and S^*, respectively) taking place in two energy domains; also shown is the redistribution of electron charge upon excitation for a ketone; the molecular orbital signature of the states is indicated (n: non-bonding orbital); the ground state is a weak electrophile; the excited state ($n\pi^*$) behaves like an oxygen centred radical (see text)

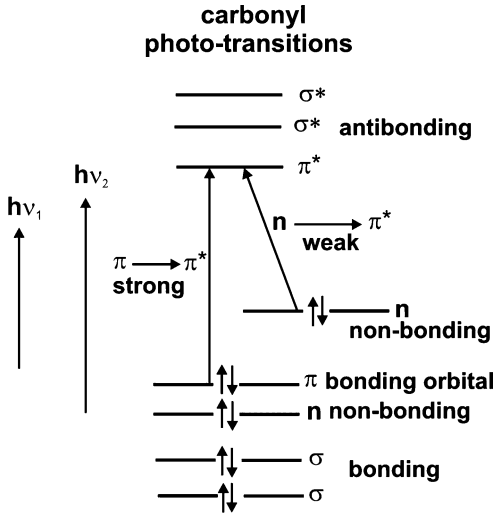


Fig. 1.15 Schematic of light-induced transitions in carbonyl groups; the two major excitations occur at 180–190 nm ($\pi \rightarrow \pi^*$) and at 280–300 nm ($n \rightarrow \pi^*$); the weak transition corresponds to an imaginary part of the (optical) dielectric constant of 10–20, whereas the strong transition is characterized by values of $\varepsilon = 10^3$–10^4; the excited state S^* in Fig. 1.13 is given by the $n \rightarrow \pi^*$ excitation; generally, the transition probability is also defined by the spin selection rule: singlet–singlet and triplet–triplet transitions have much higher probability than singlet–triplet transitions

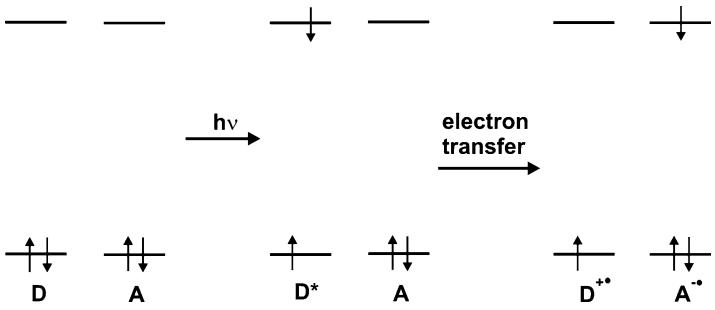


Fig. 1.16 Electron transfer reaction induced by photoexcitation of the donor D; excited donor: D^* (which is a singlet state); A: acceptor ground state; $D^{+\bullet}$: oxidized donor radical; $A^{-\bullet}$: reduced acceptor radical; the scheme does not display energetic changes due to oxidation/reduction

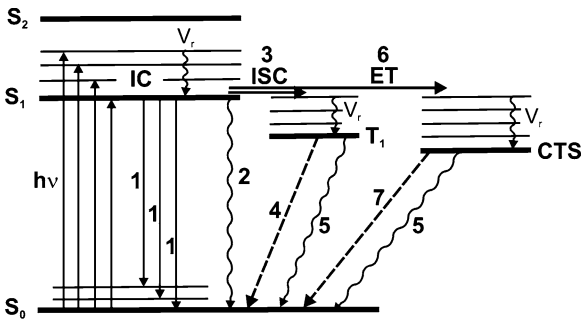


Fig. 1.17 Jablonski diagram for photo-induced electron transfer including other possible decay paths for the excited state S_1, including vibrational levels of the ground and excited state; *ET* electron transfer; *CTS* charge transfer state; V_r vibrational relaxation and *IC* internal conversion (see text)

easier oxidized or reduced by photo-induced electron transfer. A formal drawing of such reaction sequence is indicated in Fig. 1.16.

The excited molecular state of the donor and its fate with regard to the various possible decay channels is depicted in the Jablonski diagram of Fig. 1.17 (compare also Sects. 3.5.1 and 3.5.2) [57]. The locally excited singlet state (Fig. 1.16) can recombine via fluorescence (decay channel 1 in Fig. 1.17), by vibrational radiationless decay (channel 2) and by intersystem crossing (ISC), (channel 3), with subsequent phosphorescence (channel 4) or vibrational relaxation (channel 5), electron transfer (channel 6) and emission from the charge transfer state (channel 7).

Hitherto, most research on life's origin has been centred on the chemistry and the according energy sources which lead to the formation of the small building blocks of life, such as nucleobases, amino acids and amphiphiles. To a lesser extent, the actual condensation reactions that result in the linkage of monomers into biologically relevant polymers were addressed. An important research avenue is therefore the investigation of the mechanisms by which prebiotic monomers could become activated to form biopolymers. This type of research could also contribute to the discussion whether life originated on earth or in space (see Sects. 1.5.4 and 1.5.5). Here, photonic excitation is used as energy source for the activation reaction

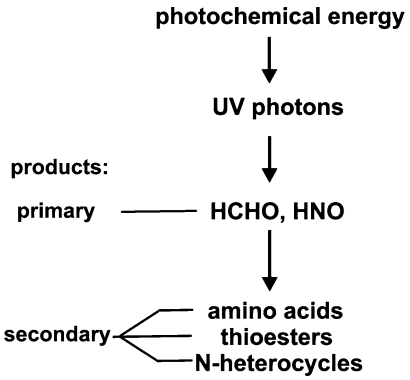


Fig. 1.18 Schematic on the prebiotic synthesis of the smaller building blocks of life by photo-induced chemistry (see text)

for biopolymer formation. Upon consideration a reaction route that results in a higher complexity than that of the starting system, the enthalpy and entropy of the system have to be assessed. Potentially destructive reactions such as the hydrolysis of polymers need to be considered as back reactions of biopolymer synthesis, and plausible mechanisms have to be formulated that allow polymer synthesis despite possible back reactions.

As an example, the photochemical synthesis of formaldehyde in an assumed prebiotic earth atmosphere is considered [58]. The presumably major atmosphere constituents were CO_2, N_2, water and less abundant H_2 and CO. The formation reactions

$$\begin{aligned} &CO_2 + h\nu \rightarrow CO + O \\ &H_2O + h\nu \rightarrow OH + H \\ &H + CO \rightarrow HCO \\ &HCO + HCO \rightarrow H_2CO + CO, \end{aligned} \tag{1.35a}$$

are counteracted by the following photolysis reactions:

$$\begin{aligned} &H_2CO + h\nu \rightarrow HCO + H \\ &H_2CO + h\nu \rightarrow CO + H_2. \end{aligned} \tag{1.35b}$$

Despite the considerably higher rate for the photodissociation reactions (1.35b), a sufficient concentration of formaldehyde can be calculated to induce UV polymerization in the oceans. For $h\nu = 2.26$ eV, a concentration of $3 \times 10^{-4} M H_2CO$ suffices for inducing UV photo-polymerization. This value is lower than that calculated for the accumulated formaldehyde in 10^7 years which results from wet removal in rain droplets, thus escaping the photolysis reaction. Of course, such models are based on a variety of assumptions about the atmospheric conditions at an early time of earth's history that can therefore only be estimated based on plausibility arguments and archaeological records (Fig. 1.18).

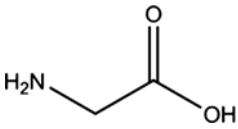


Fig. 1.19 Structure schematic of the amino acid glycine

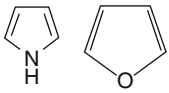


Fig. 1.20 Structure schematic of pyrrole (*left*) and furane (*right*)

Model experiments on the photochemical synthesis of particular small molecules, relevant for building blocks of large biological molecules, have not yet been intensively pursued. Experiments that model the generation of small biological molecules from admixtures, however, have been carried out to a considerable extent, following the initial work of Miller [47]. In an attempt to investigate photochemical reactions in interstellar space, analogues of the so-called interstellar ice, prepared using the knowledge of the ice composition, have been irradiated with UV light [59]. Dust particles in dense interstellar clouds accrete ice layers that contain predominantly water, CO, CO_2, NH_3 and methanol. The ice can enter circumstellar disks where planets and comets are formed. If biological molecules can be formed in a model experiment on an ice analogue whose composition resembles that of protostellar sources, the hypothesis of panspermia would be supported. UV-irradiation of an ice composed of H_2O, CH_3OH, NH_3 CO, CO_2 in a mixture of 2:1:1:1:1 resulted indeed in the formation of 16 amino acids and six of them are constituents of proteins. The most abundant amino acid found was glycine (Fig. 1.19) and the second largest contribution was from alanine. Particularly interesting is the finding of pyrroles and furanes in the experiment. Pyrrole and its precursor furane (Fig. 1.20) play a prominent role in the synthesis of biologically active compounds. Pyrrole is a building block of compounds such as porphyrines and chlorophylls and its photophysics and -chemistry is of fundamental interest in the search for the origin of life.

This experiment is an example of a model experiment that investigates possibilities of formation of prebiotic molecules in extraterrestrial environments. Its finding supports the assumption that prebiotic molecules could have been delivered to the early Earth, for instance, by comets which consist, to a large extent, of pristine interstellar material during the period of heavy extraterrestrial bombardment.

We are thus, towards the end of this chapter, led to the assumption that life could have originated in environments that at first glance appear too hostile for the formation of prebiotic molecules. The influence of photonic excitations and of low-temperature quantum tunnelling and assembly of polymers, considered together with the unusual short period of time from the cooling down of the early Earth to the occurrence of the first cellular organism, supports the panspermia hypothesis. Recently, adenine, one of the constituents of the DNA molecule, has been detected

in model experiments on the aerosol analogue of the Jupiter satellite Titan, based on measurements of voyager 1 (1980) and the Cassini/Huygens mission (2004) which lends further support to the panspermia idea [60].

1.5.5.2 On the Homochirality of Biological Molecules

Since the discovery of Pasteur of what he called *the asymmetry of life*, a plethora of ideas and concepts on its origin has been disseminated [61]. Chirality of a molecule occurs if it is characterized by non-superimposable mirror images of each other. Achiral molecules are identical to their mirror image. The two distinguishable forms are called enantiomers or optical isomers because they rotate plane-polarized light to the left or right. An equal mixture of left- and right-handed molecules is called a racemic mixture. Figure 1.21 shows the two forms (enantiomers) of alanine, an amino acid that has also been detected in the above model experiments on interstellar ice analogues.

First, models that propose an extraterrestrial origin of biomolecular homochirality presumed that UV light from an astronomical source acted on chiral molecules present in the molecular clouds from which solar systems are formed. The action is thought to produce an excess of one enantiomer. In essence, this model is based on the panspermia hypothesis. It assumes further that this material has been introduced to Earth via comets and meteorites and dust particles during the period of heavy bombardment (first few hundreds of million years of the solar system). This chiral imbalance occurred in the prebiotic material and was thus available for the development of life. The observation of an excess of a 33% of an enantiomer of alanine of the Murchison meteorite [63] appears to support the hypothesis, although the analysis of the results has been somewhat relativated [64].

The cause of the development of non-racemic mixtures upon irradiation with circularly polarized light from extraterrestrial sources is attributed to asymmetric photolysis, i.e. photochemical decomposition, of the *S* and *R* enantiomers at slightly different rates. Asymmetric photolysis occurs because of the slightly different molar absorption coefficients for left and right circularly polarized light. Because the rate of photolysis depends on the light absorption of the respective compound, this circular dichroism induces different reaction rates and thus the enantiomeric asymmetry. The enantiomer that has the higher absorption coefficient

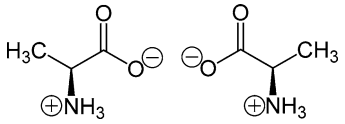


Fig. 1.21 The two mirror image forms of alanine: *left*, S-alanine; *right*, R-alanine in zwitter ionic form at neutral pH; labelling after the Cahn–Ingold–Prelog priority rules [62]

is preferentially decomposed. The efficiency to produce non-racemic mixtures can be described by the optical anisotropy factor $\gamma^* = \Delta\varepsilon/\varepsilon$, where ε denotes the frequency-dependent dielectric function: $\varepsilon = (\varepsilon_+ \varepsilon_-)/2$. ε_+ and ε_- are the dielectric functions for right (clockwise) and left rotating circularly polarized light. The small initial enantiomeric excess can be amplified to large excess by asymmetric autocatalytic processes. Generally, chiral autocatalysis is a reaction in which a chiral product that is formed from achiral precursors catalyzes its own formation. The respective enantiomer thus amplifies its own chirality and eventually, the reaction results in absolute enantioselective synthesis, i.e. the synthesis of an enantiopure product from achiral starting material without chiral additives or, for instance, asymmetric physical fields during the reaction. Indications for such reactions to occur have been recently communicated [65]. The hypothesis of extraterrestrial origin of life on Earth is additionally supported by the findings of slight enantiomeric excesses in the variety of chiral amino acids found in meteorites. The reservoir of chiral molecules, however, could be changed to achiral ones if one substitutes, for example, in an achiral amino acid such as glycine, one of the hydrogen atoms in the molecules methylene group by deuterium. This mechanism of enantioselection has been found for deuterium-enriched glycine and α-methylalanine and, in addition, it was shown that the amino acids catalyzed the formation of large excess of enantio-enriched alcohols. The chirality of the alcohols depended on that of the reaction catalyzing amino acid, pointing to the possibility that the known deuterium excess in meteorites is related to the asymmetry of otherwise chiral amino acids. Also, recently, differences in solubility of L and D amino acids and sugars, with a much higher solubility of L amino acids, for instance, have been proposed to induce enantiomeric amplification in solutions [66].

The sources that induce enantioselectivity have to possess a symmetry-breaking property which is the case, for instance, for circularly polarized light and the circularly polarized Bremsstrahlung from longitudinally polarized electrons (left-handed, because their spins are antiparallel to the propagation direction) that have been emitted with parity violation during β-decay. In addition, besides other parity-violating effects such as the parity-violating so-called z-force which results in an energy difference of enantiomers (originating from the charge–parity–time theorem on physical invariance with regard to space inversion, antiparticles and time), magnetic circular dichroism and the magnetochiral effect have to be included in the considerations.

Magnetic circular dichroism, also known as Faraday effect, of a material occurs if a magnetic field is applied parallel to the direction of a light beam. Also, materials that are achiral show magnetic circular dichroism due to the magnetic field-induced lifting of degenerate orbitals and spin states and due to the mixing of electronic states. Figure 1.22 illustrates the situation: the selection rules for left and right circularly polarized light are different with $\Delta J = +1$ (left) and $\Delta J = -1$ (right), where J denotes the total angular momentum.

Magnetochiral dichroism has been initially described in the framework of semiclassical radiation theory [67] and considers the radiation (photon) field

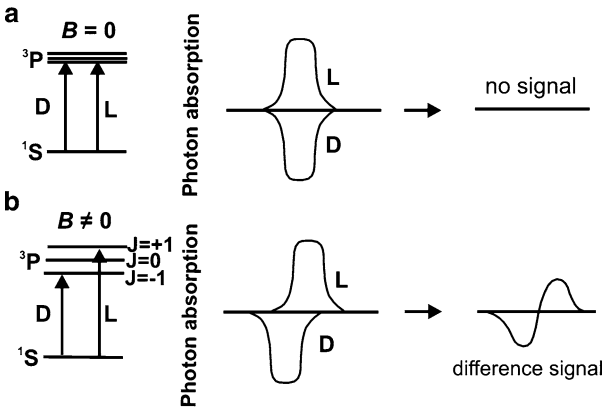


Fig. 1.22 Visualization of the change of absorption due to magnetic dichroism: (**a**) without magnetic field and (**b**) with magnetic field; D refers here to *dextro* (right circularly polarized light) and L to *levo* (left-handed rotation sense) (see text)

molecule interaction terms. Including the electric quadrupole interaction, the single photon transition probability from a ground state $|i\rangle$ to an excited (final) state $|f\rangle$ involves six terms: the electric dipole–electric dipole interaction (term 1), the electric dipole–magnetic dipole term (2), the electric dipole–electric quadrupole term (3), the magnetic dipole–magnetic dipole term (4), the magnetic dipole–electric quadrupole interaction (5) and the electric quadrupole–electric quadrupole term (6). For the magnetochiral dichroism effect, terms 2 and 3 that lead to natural dichroism are considered for the case where a magnetic field is aligned parallel or antiparallel to the propagation direction of the impinging light beam. One obtains an increment of the absorption coefficient that has the same sign for left and right circularly polarized light and thus would also occur for linearly polarized light. In a medium of chiral molecules, the according molecular tensor is odd with respect to parity; the magnetochiral effect is opposite in sign for the two enantiomers and vanishes for racemic mixtures or centrosymmetric molecules. It can be shown that this effect is independent of the polarization of the light (observable also with unpolarized light) but only depends on the handedness of the medium and on the relative orientation of the light propagation and the magnetic field (Fig. 1.23).

Thus, the magnetochiral anisotropy effect can be considered as a cross-effect between magnetic and natural optical activity. The related absorption, described by the dielectric function, is given by an additional term, proportional to $k_{\mathrm{Ph}} \cdot B$, which fulfils the symmetry requirements for a chiral influence. Similar to the enantioselectivity for circularly polarized light, an according enantioselectivity of photochemical reactions of chiral species irradiated with unpolarized light under the action of a magnetic field (Fig. 1.23) is expected. Conclusive experiments that demonstrate the effect have been carried out using paramagnetic complexes that contained transition metal or rare earth ions because they show the largest

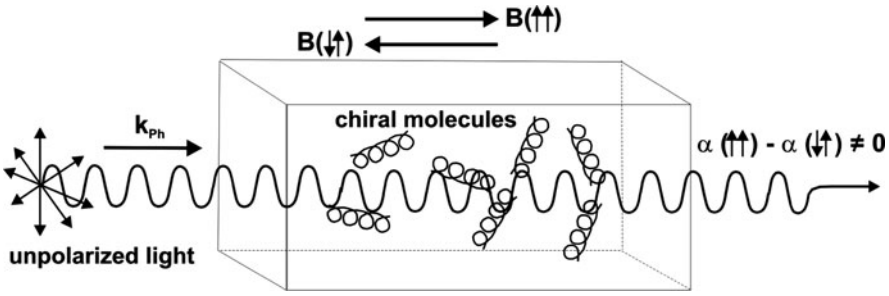


Fig. 1.23 Schematic on the origin of magnetochiral dichroism; B denotes the magnetic field oriented parallel (↑↑) or antiparallel (↑↓) to the propagation of the light beam, k_{Ph}; the absorption coefficients for both magnetic field directions, α, show a non-vanishing difference

magnetochiral dichroism [68, 69]. The first observation of the effect was made by measuring the anisotropy parameter

$$g = \frac{\partial / \partial B (I_{B\uparrow\uparrow k} - I_{B\uparrow\downarrow k})}{I_{B\uparrow\uparrow k} + I_{B\uparrow\downarrow k}} B, \tag{1.36}$$

where the subscripts are related to the nomenclature in Fig. 1.23 and I denotes the measured luminescence intensity. Figure 1.25 shows the experimental result for tris(3-trifluoroacetyl- ± -camphorato) europium (III) complexes, where the sign ± indicates the handedness. In abbreviated form, one writes $(\mathrm{Eu}(\pm)\mathrm{tfc})_3$. For this compound, the $^5\mathrm{D}_0 \rightarrow ^7\mathrm{F}_{1,2}$ strong luminescent transitions occur (see Fig. 1.24) and, also, substantial magnetic circular dichroism has been observed for the ions in aqueous solution. This material has therefore been selected for the detection of the magneto-chiral effect.

Although the enantiomeric selection processes only result in a rather small and sometimes even tiny excess of one enantiomer and amplification processes are necessary to achieve almost pure chirality, the search for the origin of homochirality must include all possibilities. In the above experiments one could also argue that the selected molecules are characterized by a particularly strong magneto-response and large magnetic fields had to be applied, which might be less plausible for situations outside the laboratory. Our knowledge about the force fields in outer space is limited despite the information accumulated by observatories and space missions. On earth, effects related to circularly polarized light (at dusk or dawn) and magnetic field are considered too weak to have an influence and also, they typically cancel out due to comparably rapid changes. Figure 1.26 summarizes some of the above considerations and includes adsorption on sites with chiral symmetry which has not been mentioned here before.

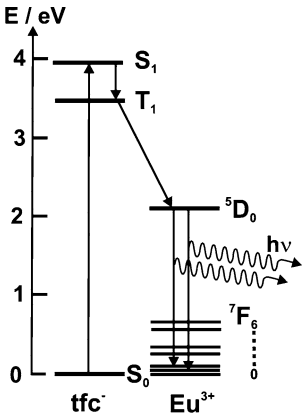


Fig. 1.24 Jablonski-type term scheme for the luminescence transition $^5D_0 \rightarrow ^7 F_{1,2}$ in the Eu-tfc system

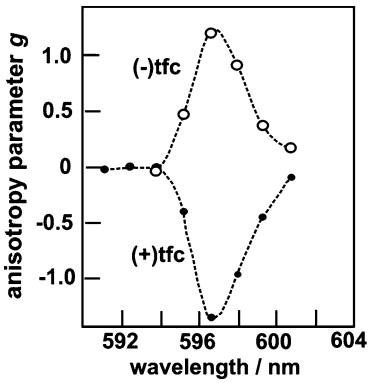


Fig. 1.25 Magneto-chiral photoluminescence anisotropy (described by g in 1.36) of europium $((+/-)\ tfc)_3$ complexes; excitation wavelength: 350 nm; magnetic field strength: 0.9 T

Consequently, as asymmetry inducing objects and force fields, only those of extraterrestrial origin are discussed. Among possible origins of circularly polarized light that might be responsible for enantiomeric selection in interstellar organic molecules, synchrotron radiation from magnetic neutron stars and circular polarization resulting from dust scattering in reflection nebulae in star formation regions have been advocated. Although the latter is an infrared phenomenon, it has nevertheless entered the discussion [70]. The circularly polarized synchrotron radiation from electrons that orbit a neutron star (pulsar) which impinges onto the organic mantles of interstellar grains can cause enantiomeric excess. It is known that the 19 L-amino acids absorb left-handed circularly polarized UV light preferentially. It is then assumed that this excess is then reaching earth via interplanetary material (dust particles, comets and meteors).

The contribution of the magnetochiral effect to the enantiomeric excess might be more significant than that of objects which produce circularly polarized light because unpolarized light and large magnetic fields are more common in the cosmos. In star-forming regions, magnetic fields range from 10^{-10}T to over 10^{-4}T.

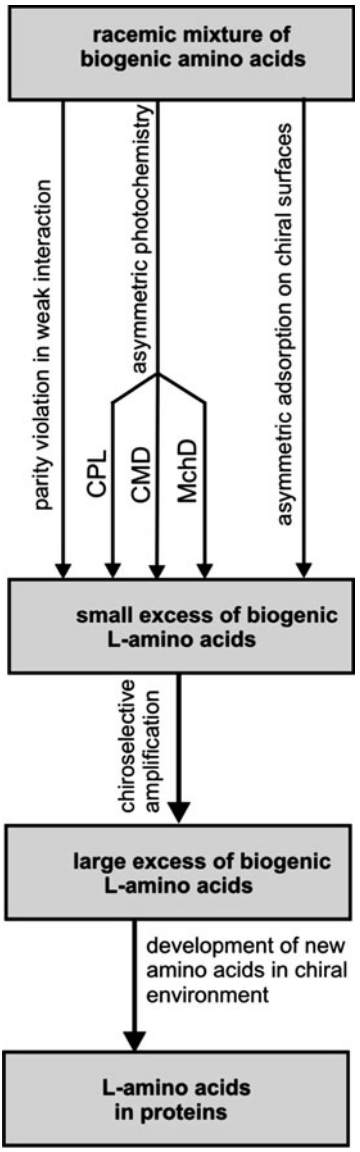


Fig. 1.26 Overview of selected alternative scenarios (*parallel vertical arrows*) for biomolecular homochirality for amino acids only; *CPL* circularly polarized light; *CMD* circular magnetic dichroism; *MchD* magnetochiral dichroism

Much stronger fields prevail at the surface of young stellar objects (0.1–1 T). The latter stars are also considered to be strong UV sources. Thus, in such areas, strong magnetic fields, UV light and organic interstellar material are present, increasing the likelihood of forming small enantiomeric excess via magnetochiral dichroism.

References

1. H. Nifenecker, J.A. Piston, High energy photon production in nuclear reactions, Ann. Rev. Nucl. Part. Sci. **40**, 113–1143 (1990)
2. M. Tardocchi et al., Gamma ray spectroscopy at high energy and high time resolution at JET, Rev. Sci. Instrum. **79**, 10E524 (2008)
3. C.F.v. Weizsäcker, Ausstrahlung bei Stößen sehr schneller Elektronen, Z. Phys. **88**, 612–625 (1934)
4. E.J. Williams, Correlation of certain collision problems with radiation theory, Kgl. Danske Videnskab. Selskab Mat.-fys. Medd. 13(4), (1935)
5. H.A. Kramers, On the theory of X-ray absorption and of the continuous X-ray spectrum, Philos. Mag. **46**, 836–871 (1923)
6. J. Larmor, On a dynamical theory of the electric and luminiferous medium, Philos. Transact. Roy. Soc. **190**, 205–300 (1897)
7. J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1980)
8. H.A. Bethe, W. Heitler, On the stopping of fast particles and on the creation of positive electrons, Proc. Roy. Soc. London A **146**, 83–112 (1934)
9. H. Bethe, Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie, Ann. d. Physik **397**, 325–400 (1930)
10. M.S. Zolotorev, K.T. McDonald, *Classical radiation processes in the Weizsäcker-Williams approximation* (Aug.25, 1999), physics/0003096, http://puhep1.princeton.edu/~mcdonald/accel/weizsacker.pdf
11. M.T. Herd, J.F. Lawler, Infrared continuum radiation from metal halide high intensity discharge lamps, J. Phys. D: Appl. Phys. **40**, 3386–3395 (2007)
12. D. Zwanziger, Energy and momentum spectral function of coherent Bremsstrahlung radiation, Phys. Rev. D **20**, 2001–2026 (1979)
13. V.B. Gavrikov, V.P. Likhachev, J.D.T. Arruda-Neto, Coherent x radiation by relativistic electrons in crystals, Phys. Rev. A **68**, 024901 (2003)
14. L. Criegee, G. Lutz, H.D. Schulz, U. Timm, W. Zimmermann, Polarization measurement of the 6-GeV coherent Bremsstrahlung from the Hamburg electron synchrotron, Phys. Rev. Lett. **16**, 1031–1033 (1966)
15. F.J. Garcia de Abajo, A.G. Pattantyus-Abraham, N. Zabala, A. Rivacoba, M.O. Wolf, P.M. Echenique, Cherenkov effect as probe of photonic nanostructures, Phys. Rev. Lett. **91**, 143902 (2003)
16. A.V. Korol, A.G. Lyalin, O.I. Obolenskii, A.V. Solovyov, The role of the polarization mechanism of radiation of atoms over a broad frequency range, Sov. Phys. JETP **87**, 251–259 (1998)
17. N.N. Nasonov, V.A. Nasonova, I.G. Popov, Polarization Bremsstrahlung from relativistic electrons moving in a small-grained medium, Phys. Atom. Nucl. **64**, 1037–1041 (2001)
18. S.V. Blazhevich, A.S. Chepurnov, V.K. Grishin, B. Ishkhanov, N. Nasonov, V. Petukhov, V. Shvedunov, Polarization Bremsstrahlung of relativistic electrons in aluminium, Phys. Lett. A **254**, 230–232 (1999)
19. Sal Portillo, C.A. Charles, Absolute doubly differential cross sections for electron Bremsstrahlung from rare gas atoms at 28 and 50 keV, Phys. Rev. Lett. **91**, 173201 (2003)
20. R. Loudon, *The Quantum Theory of Light*, 3rd edn. (Oxford University Press, 1973), pp. 241
21. P.A.M. Dirac, The quantum theory of emission and absorption of radiation, Proc. Roy. Soc. A **114**, 243–265 (1927)
22. P.A.M. Dirac, The quantum theory of the electron, Proc. Roy. Soc. A **117**, 610–624 (1928)
23. R.P. Feynman, Space time approach to non-relativistic quantum mechanics, Rev. Mod. Phys. **20**, 367–387 (1948)
24. D.J. Struik, *A Source Book in Mathematics*, *1200–1800* (Harvard University Press, Cambridge, MA, 1969), pp. 391–399

25. J. Schwinger, Quantum electrodynamics. I. A covariant formulation, Phys. Rev. **74**, 1439–1461 (1948)
26. J. Schwinger, Quantum electrodynamics. II. Vacuum polarization and self energy, Phys. Rev. 75, 651–679 (1949)
27. W.E. Lamb Jr., R.C. Retherford, Fine structure of the hydrogen atom by a microwave method, Phys. Rev. **72**, 241–243 (1947)
28. B.N. Taylor, The determination of best values of the fundamental physical constants, Philos. Transact. Roy. Soc. A **363**, 2105–2122 (2005)
29. C.T. Chantler, Discrepancies in quantum electro-dynamics, Radiat. Phys. Chem. **71**, 611–617 (2004)
30. F. Wilczek, Quantum chromodynamics: the modern theory of the strong interaction, Ann. Rev. Nucl. Part. Sci. **32**, 177–209 (1982)
31. H.D. Politzer, Reliable perturbative results for strong interactions, Phys. Rev. Lett. **30**, 1346–1349 (1973)
32. D.J. Gross, F. Wilczek, Reliable perturbative results for strong interactions, Phys. Rev. D **8**, 3633–3652 (1973)
33. E.M. Purcell, Phys. Rev. **69**, 681 (1946)
34. J.L. Bada, A. Lazcano, Prebiotic soup-revisiting the Miller experiment, Science **300**, 745–746 (2003)
35. B. Korzeniewski, Cybernetic formulation of the definition of life, J. Theor. Biol. **209**, 275–286 (2001)
36. D.E. Koshland Jr., The seven pillars of life, Science **295**, 2215–2216 (2002)
37. Y.N. Zhuravlev, V.A. Avetisov, The definition of life in the context of its origin, Biogeosciences **3**, 281–291 (2006)
38. J.W. Huntley, S.-H. Xiao, M. Kowalewski, 1.3 Billion years of acritarch history: An empirical morphospace approach, Precambrian Res. **144**, 52–68 (2006)
39. P.A. Underhill, T. Kivisild, Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations, Ann. Rev. Genet. **41**, 539–564 (2007)
40. P.A. Mayewski et al., Homocene climate variability, Quaternary Res. **62**, 243–255 (2004)
41. D.A. Hodell, J.H. Curtis, M. Brenner, Possible role of climate in the collapse of Classic Maya civilization, Nature **375**, 391–394 (1995)
42. W. Thomson (Lord Kelvin) Presidential Address to the British Association for the Advancement of Science, Nature **4**, 262 (1871)
43. F. Hoyle, N.C. Wickramasinghe, *Astronomical Origins of Life* (Kluwer, Dordrecht, 2000)
44. The Times, London, UK, August 19, 2009
45. ESO News, 22.04. 2007
46. V. Goldanskii, M.D. Frank-Kamenetskii, I.M. Barkalov, Quantum low temperature limit of a chemical reaction rate, Science **182**, 1344–1345 (1973)
47. S.L. Miller, Production of amino acids under possible primitive earth conditions, Science **117**, 528–529 (1953)
48. D. Ring, Y. Wolman, N. Friedman, S.L. Miller, Prebiotic synthesis of hydrophobic and protein amino acids, Proc. Natl. Acad. Sci. USA **69**, 765–768 (1972)
49. H.J. Morowitz, *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis* (Yale University Press, New Haven, CT, 1992)
50. G. Wächtershäuser, Before enzymes and templates: theory of surface metabolism, Microbiol. Rev. **52**, 452–484 (1988)
51. W. Gilbert, The RNA world, Nature **319**, 618–619 (1986)
52. G.F. Joyce, RNA evolution and the origins of life, Nature **338**, 217–224 (1989)
53. M.W. Powner, B. Gerland, J.D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions, Nature **459**, 239–242 (2009)
54. J.J. Keady, S.T. Ridgway, The IRC+10216 circumstellar envelope-III. Infrared molecular line profiles, Astrophys. J. **406**, 199–214 (1993)
55. V. Goldanskii, M.D. Frank-Kamenetskii, I.M. Barkalov, Quantum low temperature limit of a chemical reaction rate, Science **182**, 1344–1345 (1973)

56. A.I. Oparin, *The Origin of Life* (Macmillan Publications, New York, 1938)
57. A. Jablonski, Efficiency of anti-Stokes fluorescence in dyes, Nature **131**, 838 (1933)
58. J.P. Pinto, G.R. Gladstone, Y.L. Yung, Photochemical production of formaldehyde in earth's primitive atmosphere, Science **210**, 183–184 (1980)
59. G.M. Munoz Caro, U.J. Meierhenrich, W.A. Schutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W.H.-P. Thiemann, A. Brack, J.M. Greenberg, Amino acids from ultraviolet irradiation of interstellar ice analogues, Nature **416**, 403–406 (2002)
60. S. Pilling, D.P.P. Angrade, A.C. Neto, R. Rittner, A. Naves de Brito, DNA nucleobase synthesis at Titan atmosphere analog by soft X-rays, J. Phys. Chem. A **113**, 11161–11166 (2009)
61. L. Pasteur, *Researches on the Molecular Asymmetry of Natural Organic Products* (vol. 14, 1905), pp. 1–46 (English transl. of French original publ. by Alembic Club Reprints)
62. R.S. Cahn, C.K. Ingold, V. Prelog, Specification of molecular chirality, Angew. Chem. Internat. Ed. **5**, 385–415 (1966)
63. M.H. Engel, S.A. Macko, Isotopic evidence for extraterrestrial non-racemic amino acids in the Murchison meteorite, Nature **389**, 265–268 (1997)
64. S. Pizzarello, J.R. Cronin, Alanine enantiomers in the Murchison meteorite, Nature **394**, 235 (1998)
65. T. Kawasaki, M. Shimizu, D. Nishiyama, M. Ito, H. Ozawa, K. Soai, Asymmetric autocatalysis induced by meteoritic amino acids with hydrogen isotope chirality, Chem. Commun. **29**, 4396–4398 (2009)
66. R. Breslow, Z.-L. Cheng, On the origin of terrestrial homochirality for nucleosides and amino acids, Proc. Nat. Acad. Sci. USA **106**, 9144–9146 (2009)
67. G. Wagnière, A. Meier, The influence of a static magnetic field on the absorption coefficient of a chiral molecule, Chem. Phys. Lett. **93**, 78–81 (1982)
68. G.J.L.A. Rikken, E. Raupach, Enantioselective magnetochiral photochemistry, Nature **405**, 932–935 (2000)
69. G.J.L.A. Rikken, E. Raupach, Observation of magneochiral dichroism, Nature **390**, 493–494 (1997)
70. J. Bailey, A. Chrysostomou, J.H. Hough, T.M. Gledhill, A. McCall, S. Clark, F. Ménard, M. Tamura, Circular polarization in star-formation regions: implications for biomolecular homochirality, Science **281**, 672–674 (1998)

Chapter 2
Sources of Spectral Photon Radiation

2.1 Introduction

The chapter treats the various experimental methods to generate high-energy photons from accelerated and relativistic charges, relevant for applications in chemistry, biology, material science and physics which will be treated in Chaps. 4 and 5. It commences with photon generation in thermonuclear fusion reactions, followed by gamma quanta emission in electron orbital capture processes and Cherenkov radiation. As important and widespread light sources, synchrotron, undulator and wiggler radiation is described, including radiation from free electron lasers and their advancements using self amplified spontaneous emission (SASE) and high gain harmonic generation (HGHG). The chapter ends with an introduction to the method of laser slicing for femtosecond X-ray pulse generation. Table 2.1 summarizes the topics treated in this chapter.

2.2 Nuclear Processes

Based on the advances made in laboratory fusion experiments and with the realization of International Thermonuclear Experimental Reactor (ITER) which will most likely be followed by a prototype commercial reactor, radiation products from nuclear reactions will become available. One distinguishes the rather low-energy black body radiation emerging from the sun's photosphere, high energy photons created in the various nuclear reactions in the sun's core (surrounded by the radiation zone) and photons from thermonuclear reactions in earth-bound fusion reactors. Energy release by fusion in the sun results from fusing four H nuclei into one He nucleus. The overall reaction for the proton–proton chain is

$$4\,^{1}\mathrm{H} + 2\mathrm{e} \rightarrow {}^{4}\mathrm{He} + 2\,\text{neutrinos} + 6\,\text{photons}. \tag{2.1}$$

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_2,

Table 2.1 Content overview of Chap. 2

2.1.	Introduction
2.2.	Nuclear processes
2.3.	Beta decay
2.4.	Orbital electron capture: perturbed angular correlations (PAC)
2.5.	Cherenkov radiation
2.6.	Synchrotron, undulator and wiggler radiation
2.6.1.	Introduction to synchrotron radiation
2.6.2.	Undulator radiation
2.6.3.	The free electron laser: self-amplified stimulated emission and high gain harmonic generation
2.6.4.	Wiggler radiation and femtosecond laser slicing

The net energy of the fusion of four H atoms into one He atom is 26 MeV [1]. The initial step of this four-step reaction is highly unlikely because one of the interacting protons has to change into a neutron (thus emitting a positron and a neutrino) which requires the so-called *weak interaction.* The probability for this to happen when the protons have overcome their mutual Coulomb repulsion barrier with a distance of about 10^{-3} pm is almost zero. Where, in the reaction, are the photons created? In a subsequent reaction step, the positron collides with an electron and two photons are emitted from the annihilation reaction. Their energy is 0.511 MeV, but due to absorption within the gas, a large number of low-energy photons are eventually created. Photons are also emitted in a reaction where ^{2}H collides with a proton (^{1}H), producing ^{3}He and a photon. Finally, the recombination of two ^{3}He leads to the formation of ^{4}He. Since two ^{3}He are needed, the three reaction steps have to take place twice, resulting in the emission of six photons.

For "synthetic" nuclear reactions in laboratory experiments, the primary concern is a low fusion ignition threshold. The reaction with the lowest threshold is the D–T (deuterium–tritium) fusion:

$$\mathrm{D} + \mathrm{T} \rightarrow {}^{4}\mathrm{He}(3.5\,\mathrm{MeV}) + n(14.1\,\mathrm{MeV}). \tag{2.2}$$

It has a threshold energy of ∼50 keV [2]. The cross section of producing a photon is generally several orders of magnitude smaller than that for producing a nucleon [3]. In the D–T reaction, fusion results in the emission of γ-rays of 16.75 MeV due to deexcitation of an excited ^{5}He nucleus. As collisions take place between decelerating fusion products with energies of several MeV and the nuclei in the plasma, the kinetic energy varies in that same energy range (several MeV). The branching ratio of this reaction compared to the alternative creation of 14 MeV fusion neutrons is around 10^{-4} [4]. Because of the additionally extended energy spectrum of emitted photons, such photons are presently almost undetectable. Nevertheless, such an extremely high-energy light source with a broad spectral range might become available. Presently, one uses the so-called secondary γ-rays, emitted by a nucleus at rest. They are spectrally intrinsically narrow and are produced in a typical ITER plasma by the nuclear reaction

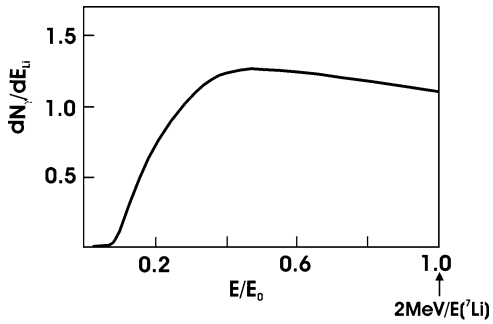


Fig. 2.1 γ-Distribution function for the ITER scenario; zero α losses have been assumed; in reality, losses will alter this spectrum

$$^4\mathrm{He}_{\mathrm{fast}} + T_{\mathrm{thermal}} \rightarrow ^7\mathrm{Li} + \gamma(2.5 + 0.478\,\mathrm{MeV}). \tag{2.3}$$

The fusion energy is 2.5 MeV and the secondary γ-ray has an energy of 0.478 MeV. If $v \ll c$ (v, velocity of the emitting ^{7}Li nucleus), the γ distribution function vs. the kinetic energy of ^{7}Li for the ITER scenario (assuming no losses due to slowing down of α particles) is shown in Fig. 2.1 [5]. The application of such radiation might presently appear far-fetched; nevertheless, considering the massive efforts in fusion research with ITER being the second largest scientific program ever established, advances in fusion reactions that emit γ-quanta at a much higher rate can be anticipated. The wavelength range accessible would cover the range from 12 to 0.6 pm and could enable preparation of extremely small dimensioned novel structures.

In stellar processes, however, high-energy photons cannot escape elastically. Therefore, one observes black body radiation for a source temperature of 5,800 K, described by Planck's formula which relates the spectral density K_v of black body radiation to the radiation frequency v by [6]

$$K_v = \frac{2hv^3}{c^2}\frac{1}{\mathrm{e}^{hv/kT} - 1}. \tag{2.4}$$

The total radiation increases according to $S = \sigma T^4$ (Stefan–Boltzmann law with temperature T in Kelvin). The constant σ is proportional to $k_\mathrm{B}^4/(c^2h^3)$. For many applications, the behaviour according to Wien's law, which describes the energetic shift of the radiation maximum with increasing temperature to shorter wavelengths, is important [7]. Accordingly, the sun and solar simulators based on Xe high-pressure lamps, for instance, have spectra with higher emission in the UV than in a tungsten iodine lamp for which the radiation temperature is about 2,500 K lower than for the sun. Figure 2.2 summarizes these observations.

For the surface temperature of the sun and atmospheric absorption, the radiation maximum is located the green spectral range. In this wavelength region, the sensitivity curve of the human eye has its maximum.

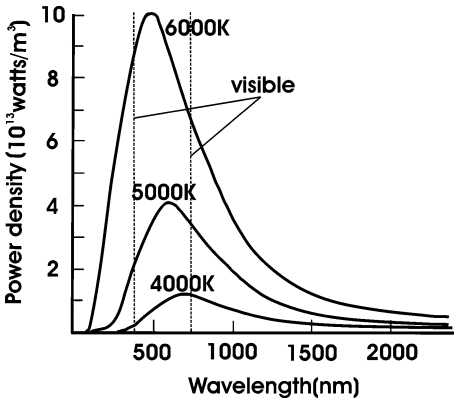


Fig. 2.2 Influence of temperature on black body radiation characteristics; the visible spectral regime is indicated by the *dotted lines*

2.3 Beta Decay

In β-decay, an unstable nucleus with atomic number Z transforms into a nucleus with atomic number $Z \pm 1$ either by electron or positron (and neutrino) emission or by electron capture. We will shortly describe the former two processes:

$$Z \rightarrow (Z \pm 1) + \mathrm{e}^{\mp} + \nu. \tag{2.5}$$

Here, e^+ denotes the positron and ν the neutrino. Subsequently, the electron-capture process will be treated in more detail since it has found numerous applications in the materials science of intermetallic alloys, metals, semiconductors and surfaces.

Photon emission during β-decay was first observed by Aston [8] and is often termed "internal Bremsstrahlung" since it can be viewed as the sudden creation (acceleration) of a fast charged particle, accompanied by emission of electromagnetic radiation. Since the energy released is shared between the electron and the neutrino, the electron energy is limited to values below those given by the mass defect. The ratio of radiation to particle energy is rather low as seen in Fig. 2.3.

The principal similarity of the curve, shown above, with the behaviour shown in Fig. 2.2 results from the Ansatz that the creation of an electron in a time interval τ can be treated analogously to an acceleration of an electron undergoing a Coulomb collision with regard to the emission of radiation.

2.4 Orbital Electron Capture: Perturbed Angular Correlations (PAC)

The method presented here is an example of a high-energy resolution, atomically local nanosecond analysis of surfaces and material phases that are non-cubic. It is based on radioactive probe atoms that, upon transmutation, emit γ-quanta that carry

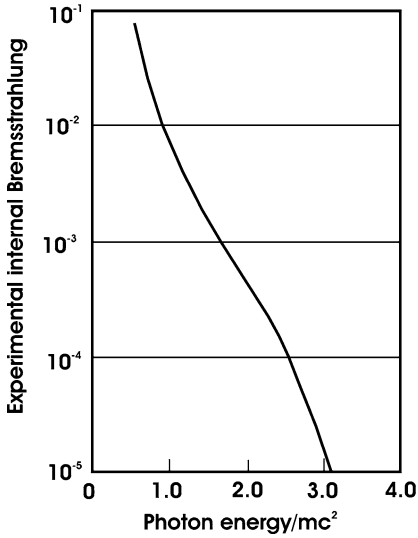


Fig. 2.3 Experimental probability ϕ of internal Bremsstrahlung per β disintegration for the β-decay ^{90}Sr–^{90}Y vs. photon energy [9]; in the determination of ϕ corrections due to the experimental arrangement have been taken into account

information on the atomistic environment of the probe nuclei. A dilution of $<10^{-12}$ suffices for bulk analysis allowing also surface investigations and suppressing transmutation doping. The frequently used radioactive ^{111}In atom is described here in detail, since examples will be given in Sect. 4.3.

Instable nuclei, such as ^{111}In, characterized by proton excess, which cannot convert the energy by positron emission ($p \rightarrow n + e^+$), capture an inner orbital electron, usually from the K shell. In this conversion of a proton into a neutron, the mass difference between the progeny and the parent atom is released as γ-radiation and neutrino emission [10]. A small number of radionuclides have, amongst their decay characteristics, the so-called γ–γ cascade in the nuclear state, formed between the emission of the two γ quanta characterized by a lifetime in the order of nanoseconds and a significant nuclear quadrupole moment [10]. Figure 2.4 shows a decay scheme for ^{111}In.

Nuclei in states with an angular momentum quantum number $I > 1/2$ have a non-spherical charge distribution. The origin of an electric quadrupole moment relates to this non-spherical distribution of the nuclear electric charge, which might be imagined as an elongated or flattened ellipsoid [11]. If such a nucleus resides or is put into a non-uniform electric field, it experiences torque and precesses. The result is a splitting of the magnetic resonance line. Such hyperfine interactions give detailed information about extremely local electrical field distributions as will be shown below.

Figure 2.4 shows that ^{111}In decays with a half-time of 2.8 days into ^{111}Cd. The indium nucleus is in an excited state and rids itself of the excess energy by emission of the γ-quanta and of X-ray photons with ∼23 keV energy. The

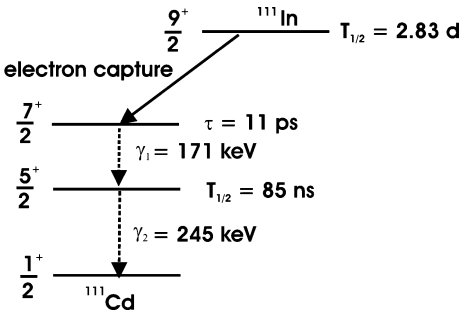


Fig. 2.4 γ–γ Emission in orbital electron capture β-decay of ^{111}In to ^{111}Cd; the electron capture process, the successive decay cascade with lifetime τ, half-lives $T_{1/2}$ and γ-quanta energies are given; states are characterized by their angular momentum and parity

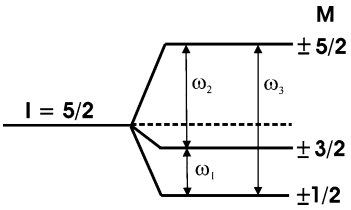


Fig. 2.5 Hyperfine splitting of the $I = 5/2$ core level due to an axially symmetric electric field gradient (see text)

photons are not of interest for the following considerations. Instead, we consider lifetime and quadrupole moment of the experimentally relevant $5/2^+$ isomer level (metastable atomic state). With $T_{1/2} = 85$ ns and a quadrupole moment of $Q = 0.83 \times 10^{-24}$ cm^2, the ^{111}In atom belongs to the most favourable probes. It can be shown that an angular correlation between the emitted γ-quanta exists, i.e. the emission of the second γ-quantum with respect to the direction of the first is not isotropic but has a defined angular distribution in a coincidence measurement.

For the In radionuclide, the electrical hyperfine interaction leads to a threefold splitting of the $I = 5/2$ isomeric level. For an axially symmetric electric field gradient, the energetic splitting E_Q of these hyperfine levels is given by the product of the largest component of the electric field tensor, V_{zz}, and the nuclear quadrupole moment Q (Wigner–Eckart-theorem) [12, 13]:

$$E_Q = eQV_{zz}\frac{3M^2 - I(I+1)}{4I(2I-1)}, \tag{2.6}$$

where I is the nuclear spin (5/2) and M the projection of I onto the z-axis with $M = \pm 1/2, 3/2$ and 5/2. The expression $eQV_{zz}/\hbar$ is known as quadrupole coupling constant. The hyperfine splitting is shown in Fig. 2.5.

The figure shows that the perturbed system is characterized by three transition frequencies. The non-isotropic angular correlation of the unperturbed system now experiences a precession of the angle-dependent radiation characteristics. The transition frequencies ω_n which allow an assessment of the strength of the interaction are given by

$$\omega_n = \frac{3}{20\hbar} eQV_{zz} n \quad \text{with} \quad n = 1, 2, 3 \tag{2.7}$$

For polycrystalline samples, the oscillation frequency depends only on the absolute value of the electric field gradient due to the averaging over many orientations. The measured spectrum has then the form

$$R(t) = \text{const} \sum_{n=0}^{3} a_n \cos n\omega_0 t, \tag{2.8}$$

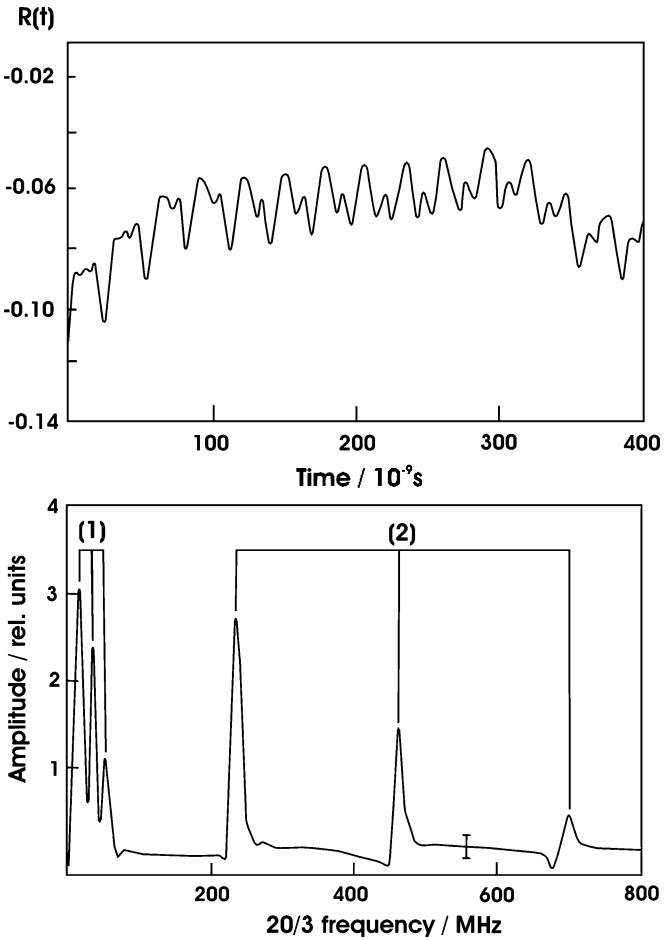


Fig. 2.6 PAC spectrumobtained from 12 coincidence spectra of In-rich copper indium disulfide with In-excess of 20–30% (*top*) and Fourier transform of $R(t)$ showing the precession frequencies (*bottom*); note that each of the two phases is characterized each by a set of three frequencies according to the energy scheme in Fig. 2.5

where $\omega_0 = \omega_1$, i.e. the lowest transition or precession frequency. A measured $R(t)$ spectrum is evaluated by Fourier analysis and fitting the measured data with a root mean square routine to the theoretical curve [14]. For details of the coincidence measurements, the reader is referred to [15]. Because of the large quadrupole moment of the indium atom, the number of probe atoms needed for a reliable analysis can be kept in the concentration range of 10^{10} atoms cm^{-3}. Hence, the effect of transmutation doping due to Cd formation in the decay of ^{111}In is negligible in semiconductors. The method allows highly sensitive surface investigations since the probe atoms contribute only with a factor of 10^{-5} to the surface of an atomic layer.

In Fig. 2.6, a typical $R(t)$ spectrum and its Fourier transform are shown for an In-enriched ternary chalcopyrite semiconductor, $CuInS_2$, an important material for photovoltaic solar cells [16, 17]. The obtained hyperfine parameters allow the identification of metallic ($\omega_1/2\pi = 18$ MHz) and the compound InS ($\omega_2/2\pi = 233$ MHz). The method is powerful and sensitive in analyzing point defects and foreign phases in materials, including buried interfaces and surfaces. Its application in the optimization of a ternary chalcopyrite absorber material for photovoltaics is given in Sect. 4.3, where the room temperature phase diagram for the Cu–In–S system, including intermetallic phases and the In-rich compounds, is presented.

2.5 Cherenkov Radiation

Already being observed by Pierre and Marie Curie in the early years of the last century, this radiation appeared as bluish-white continuum when transparent bodies were placed close to a radioactive source. It was the experimental work of Cherenkov [18] and the theoretical interpretation by Tamm and Frank[19, 20] that led to the description of this effect. Cherenkov radiation is unrelated to Bremsstrahlung which is emitted from the moving electron itself and results from collisions with atoms in the respective medium. Cherenkov radiation, in contrast, is actually emitted by the medium through which a relativistic particle moves and is caused by the action of the field of the fast particle on the medium.

The coherent response of a medium to the passage of a relativistic electron can be viewed as closely linked to the so-called density effect which describes the reduction of the expected energy loss of ultra-relativistic particles in dense media. This effect was theoretically first treated by Fermi [21]. In dense media, the dielectric polarization of the atomic electrons influences and alters the electromagnetic fields of the traversing particle to characteristic fields in the dielectric. This, however, is only true when the so-called impact parameter b (the distance between the considered atomic electron and the passing particle) is very large with respect to atomic dimensions, so that many bound electrons are located between the trajectory of the particle and the considered atomic electron. The energy loss dE/dx for such "distant collisions" can be calculated using the medium's dielectric constant $\varepsilon(\omega)$

in a continuum approximation. The integrand of the expression for the energy loss contains the part

$$\left[1 - \frac{1}{\beta^2 \varepsilon(\omega)}\right] \mathrm{e}^{-(\lambda+\lambda^*)b}, \tag{2.9}$$

where $\lambda^2 = (\omega^2/c^2)(1 - \beta^2\varepsilon(\omega))$. Only for an imaginary λ, the above expression becomes independent of b, meaning that not all the energy is deposited near the path of the projectile but that a part escapes to infinity as radiation. This is only true if $\beta^2\varepsilon(\omega) > 1$ or, in more explicit form, if the inequality $v > (c/\sqrt{\varepsilon(\omega)})$ holds. The inequality demands that the speed of the particle has to be larger than the phase velocity of electromagnetic radiation in the medium in order to observe Cherenkov radiation of frequency ω. The radiative energy loss is given by

$$\left.\frac{\mathrm{d}E}{\mathrm{d}x}\right|_{\mathrm{rad}} = \mathrm{const} \int\limits_{\beta > \sqrt{1/\varepsilon}} \omega \left(1 - \frac{1}{\beta^2 \varepsilon(\omega)}\right) \mathrm{d}\omega. \tag{2.10}$$

Equation (2.10) gives the result obtained initially by Frank and Tamm. The spectral emission features depend on the value of $\varepsilon(\omega)$ with respect to c^2/v^2 (β^{-2}). Figure 2.7 shows the real part of $\varepsilon(\omega)$ in the neighbourhood of a resonance at ω_0 where the imaginary part of the dielectric function becomes large. It is observed that the radiation is emitted in bands close to the resonances (*dashed region*). For $\beta \cong 1$, however, these bands can be quite extended.

The emission of Cherenkov radiation is restricted to cases in which $n > 1$. Since this condition is normally not fulfilled in the soft X-ray region, that spectral region has been neglected. It was, however, later realized that due to the resonant optical behaviour at absorption edges, Cherenkov radiation can be generated in rather narrow spectral segments. Particularly interesting is the spectral region of the water window where water is transparent but absorption from the carbon K edge (4.37 and 2.33 nm for oxygen) allows in situ investigation of biological samples. Figure 2.8 shows a result for low Z elements and 10 MeV electrons [22].

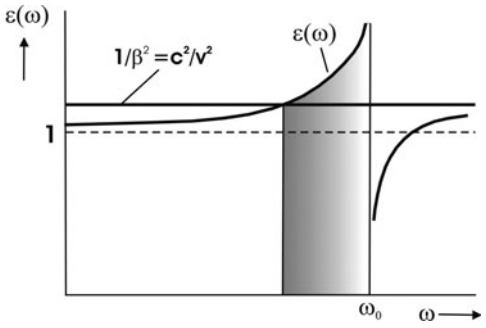


Fig. 2.7 Schematic of spectral characteristics for Cherenkov radiation occurring in bands with $\varepsilon(\omega) > 1/\beta^2$

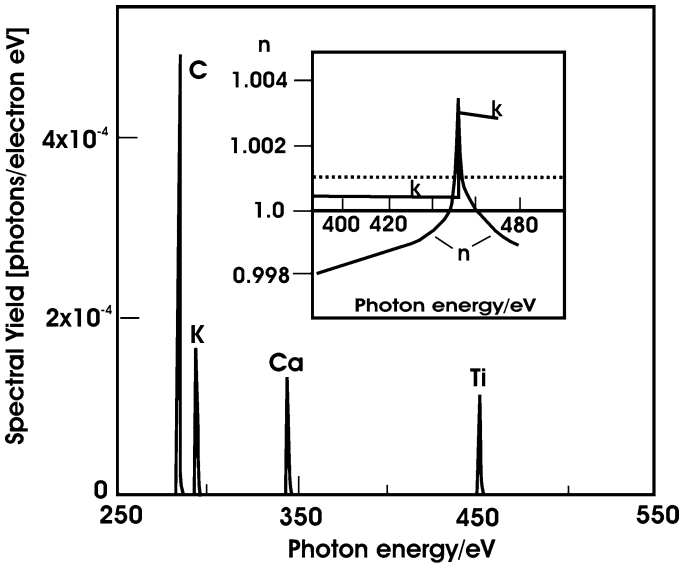


Fig. 2.8 Predicted Cherenkov spectral yield in the C–O water window for 10 MeV electrons for a series of low Z materials; the *inset* shows $n-1$ (real part) and k (imaginary part) of the refractive index for Ti near the L edge at 454 eV; the *dotted line* denotes the emission threshold for 10 MeV electrons

The angular characteristic of the radiation in the far field approximation, where the fields become transverse, is given by the condition

$$\cos\Theta_c = \frac{1}{\beta\sqrt{\varepsilon(\omega)}} = \frac{1}{\beta n}; \quad \left(n = \sqrt{\varepsilon}, \text{refractory index}\right). \tag{2.11}$$

For $\beta^2\varepsilon > 1$, the inequality for the emission angle $3\pi/2 < \Theta_c < \pi/2$ holds. The resulting angular distribution narrows with increasing particle velocity. The measurement of the Cherenkov angle Θ_c allows the determination of the particle velocity if suited band-pass filters are used. Such counting devices are employed extensively in high-energy physics (see below).

The overlapping wavefronts for the relativistic particle interfere constructively. Emission occurs in a cone of light from each point of the particles trajectory. These fields become identifiable after the particle has advanced by a formation length $L_0 = v\,t_0$, i.e. the distance over which the electron moves one wavelength ahead of the projection of the field motion onto the direction of the particle.

Recently, electron energy losses due to Cherenkov radiation have been used to probe photonic nanostructures [23, 24]. These observations will be treated in Sect. 5.2.5 in which the characteristics of photonic crystals are described.

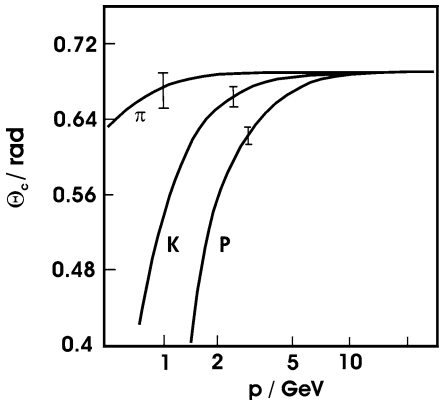


Fig. 2.9 Cherenkov cone size as a function of particle momentum showing distinct bands for pions (π), kaons (K) and protons (P); with increasing momentum, the common maximum Cherenkov angle is reached ($\beta \rightarrow 1$) (see text)

The energy–momentum selectivity of Cherenkov radiation has resulted rather early in its use in high-energy cosmic and γ-ray detectors [25]. From the correlation between particle energy and momentum (making use of the cone-shaped coherent photon emission), identification of high-energy particles becomes possible. Figure 2.9 shows three distinct bands that correlate the Cherenkov angle Θ_c and momentum **p** in GeV (/c) for a ring imaging Cherenkov counter [26]. For the case of *transition radiation*, the reader is referred to the review of Artru and coworkers [27].

2.6 Synchrotron, Undulator and Wiggler Radiation

2.6.1 Introduction to Synchrotron Radiation

With the advent of third-generation synchrotron sources, the so-called insertion devices such as wigglers and undulators have been used to produce energetic photons at high light intensities. They produce incoherent and coherent radiation, respectively. In Sect. 1.2, radiation from circular motion was introduced and the basic formula for relativistic motion, based on Lienard's extension of Larmor's work [28], was given (1.6). The spectral features of synchrotron radiation are characterized by the critical frequency ω_c given by $\gamma^2 eB/4\pi m_e$. Detailed analysis shows that the emission peaks at $\omega = 0.29\omega_c$. To derive the spectral dependence of the radiation, one uses (1.6) and derives expressions for the formation length for frequencies below and at $\omega \geq \omega_c$. The formation length concept only applies to electron motion on a straight line; therefore the concept of formation time, resulting in a formation angle for circular motion, is applied. With $L_0 = (2\lambda/\Theta^2 + 1/\gamma^2)$, the formation lengths for each case can be derived from the respective formation

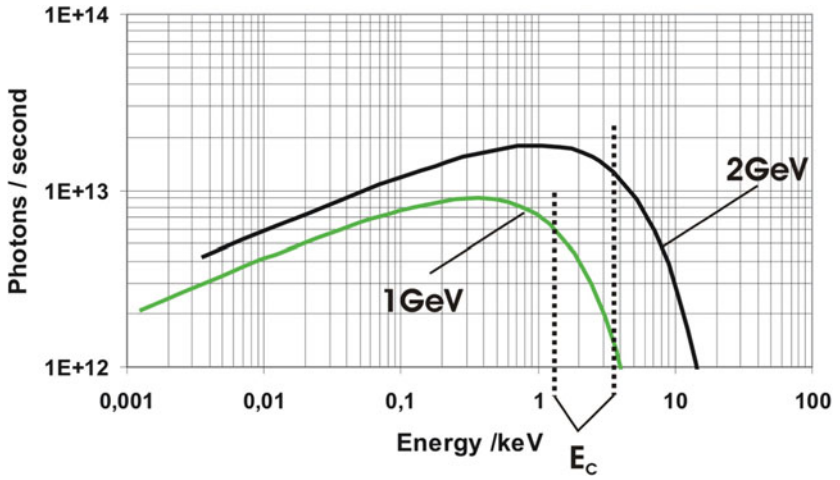


Fig. 2.10 Spectral distribution function for synchrotron radiation including the critical energies given by $E_c(\text{keV}) = 0.665B$ (T), $E^2(\text{GeV}^2)$ for 1.87 T (1 GeV) and 1.3 T (2 GeV) for two electron energies

angle expressions. For $\lambda \gg \lambda_c$, $L_0 \approx \gamma^2\lambda_c(\lambda/\lambda_c)^{1/3}$ for the low-energy part of the spectrum, and $L_0 \approx \gamma^2\sqrt{\lambda\lambda_c}$ for $\lambda \leq \lambda_c$. The resulting energy spectrum is given by a term proportional to $\omega^{1/3}$ for the low-energy branch and $\mathrm{d}E(\omega)/\mathrm{d}l \approx \sqrt{\omega}\mathrm{e}^{-\omega/\omega_c}$ for frequencies at and above the critical frequency. The differential is formed with respect to a unit path length l for which photons are radiated. A typical spectrum is shown in Fig. 2.10 in double logarithmic representation for the photon flux. At the critical energy, the photon flux (photons/s mrad and 0.1% bandpass) is equal above and below that value.

Relativistic effects result in a change of the angular distribution of emitted radiation from a dipole characteristic (in the frame of the moving electron) to one resembling a sweeping search light with dominant emission tangentially in the direction of the moving electron (in the laboratory frame). Photons are radiated within an angle of the order $1/\gamma(\gamma = 1/\sqrt{1-\beta^2})$ given by $\Theta \approx m_0c^2/E \rightarrow \Theta = 0.5\,\text{mrad}/E(\text{GeV})$. Synchrotron radiation is linearly polarized in the plane of electron acceleration and its time structure is given by the size of the micro-bunches (typically 20–100 ps) that are separated by a few ns. For ultrashort time resolution, either the current pulse duration of the linear injecting accelerator can be reduced or the principle of *laser slicing* (see Sects. 3.6.3 on the *free electron laser* and 3.6.4 on *wigglers*) can be applied.

2.6.2 Undulator Radiation

In an undulator, magnets are arranged such that the electron encounters on its path in a region of transverse magnetic field whose magnitude oscillates with the spatial

period λ_0. An electron traversing the undulator experiences transverse oscillations with that period. This oscillation results in the emission of undulator radiation (Fig. 2.11).

Compared to a wiggler, the modulation of the electron path is considerably reduced, resulting in a longer light pulse rather than in a series of short bursts. The spectrum is bunched up into harmonics and narrows with the number of undulator periods: $\Delta E/E = 1/nN$ (n, harmonic; N, undulator periodicity). In technical realization for planar undulators, the magnets are moved vertically and a wide range of photon energies can be swept in an experiment. Figure 2.12 shows the spectral characteristic of the undulator U49/2 at the storage ring Bessy II, with a spectral width between 80 and 2,000 eV [29].

The physical principle is based on special relativity: the electron entering an undulator with almost the speed of light "sees" the undulator length L_U contracted by the Lorentz factor γ. The wavelength of the emitted radiation would be $\lambda \approx L_U/\gamma N$; because of the electron motion when seen from the laboratory frame, however, the wavelength further shrinks due to the Doppler effect by $\approx 2\gamma$. Hence,

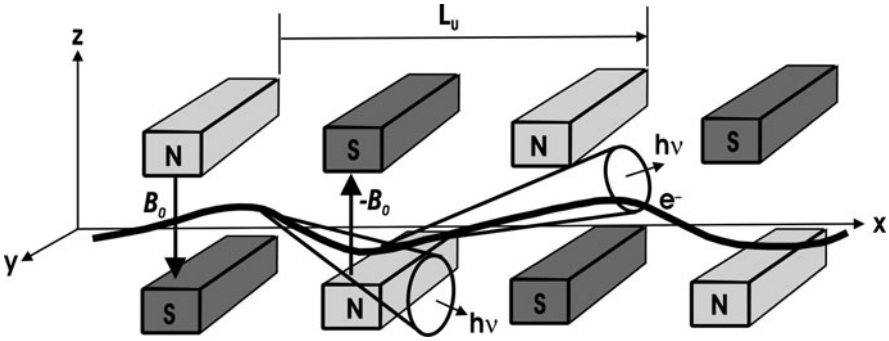


Fig. 2.11 Schematic of an electron trajectory and photon emission in an undulator with period L_U

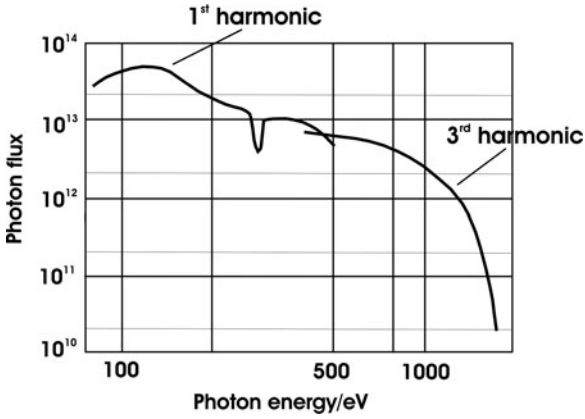


Fig. 2.12 Spectral photon flux density of U49/2 at Bessy II

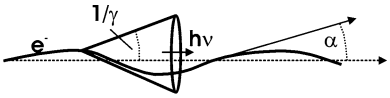


Fig. 2.13 Schematic of the undulator condition (see text); $1/\gamma$, aperture angle of the light conus; α, electron wiggling angle

the actual wavelength of undulator radiation is $\lambda_{\mathrm{rad}} \approx L_{\mathrm{U}}/2\gamma^2 N$. This results in a wavelength transformation factor of $>10^6$ from undulator periodicity to the frequency of the emitted light.

For an undulator, the wiggling angle α smaller than the photon emission angle $1/\gamma$ as can be seen in Fig. 2.13. In that case, interference effects result from the fact that the angular electron excursion lies within the $1/\gamma$ cone of the emitted light. The deflection parameter K describes this behaviour: undulator radiation is obtained for $K < 1[K = 0.934 L_{\mathrm{U}}\,(\mathrm{cm})\,\mathbf{B}_0(T)]$. Observation of the radiation at an angle θ off axis yields constructive interference at the wavelengths

$$\lambda_{\mathrm{ph}} = \frac{L_{\mathrm{U}}}{2\gamma^2}\left(1 + \frac{K^2}{2} + \gamma^2\Theta^2\right). \tag{2.12}$$

The undulator spectra exhibit pronounced peaks in photon emission that include higher harmonics. In spectroscopic experiments, the undulator gap and thus the peak in the spectrum are varied synchronously with the energy settings of the monochromator to achieve optimum irradiance conditions, resulting, for instance, in a spectrum as shown in Fig. 2.12.

2.6.3 The Free Electron Laser: Self-Amplified Stimulated Emission and High Gain Harmonic Generation

Advanced FELs are based on the principle of self-amplified spontaneous emission (SASE) [30] and HGHG. We first review the operational principles under which SASE occurs: when an electron beam with short and intensive bunches passes through an undulator, along the beam, spontaneous synchrotron radiation (Bremsstrahlung) is emitted. Under well-defined conditions, the electron beam can amplify the electromagnetic field of the synchrotron radiation. The necessary resonance condition relates photon wavelength λ_{ph}, beam energy E_{B}, the spatial undulator period L_{U} and the parameter $K(K = \mathrm{e}\mathbf{B}_0 L_{\mathrm{U}}/2\pi m_{\mathrm{e}}^0 c)$ which characterizes the undulator's field strength ($\mathbf{B}_0$) and periodicity. The resonance condition

$$\lambda_{\mathrm{ph}} = \frac{L_{\mathrm{U}}}{2\gamma^2}\left(1 + \frac{K^2}{2}\right), \tag{2.13}$$

shows that the photon energy can be tuned via the Lorentz factor γ and thus by the beam energy E_B. For the high electron velocities involved, (2.13) can be approximated by $\lambda_{ph} \sim L_U/2\gamma^2$. This condition is fulfilled for an electron velocity for which the time to pass a photon wavelength equals the time a photon passes an undulator period: $t(\lambda_{ph}(\text{electron})) = t(L_U(\text{photon}))$. For SASE to occur, the undulator must have a minimum length. Then, the electric field of the radiation, emitted in its initial segment, can act on the electrons along a sufficiently long path (see Fig. 2.14). Hence, the spontaneous emission of synchrotron radiation in the first and subsequent undulator segments results in intensity amplification in the following segments. Due to the mutual interaction of both fields, the electrons concentrated in a region of extension $\sim\lambda_{ph}$, sometimes denoted as discs, will produce coherent radiation due to simultaneous emission of light from electrons in a given layer. As a result, very short and intense X-ray flashes are produced that have the properties of laser light.

The radiation intensity is proportional to n_e^2 (n_e, the number of electrons). The radiation power increases exponentially with path length z as the wave is moving along the undulator: $P(z) = A\, P_0 \exp(2z/L_g)$; P_0 is the effective input power, A an input coupling constant and L_g denotes the so-called field gain length. As example, the SASE intensity at the TESLA test facility (terra electron volt energy superconducting linear accelerator) at DESY was – in the year 2000 – about 300 times larger than the spontaneous emission [31]. Figure 2.15 compares spontaneous and SASE emission at the TESLA TTF around the optimum beam position. The observed FEL gain was $G \sim 3,000$. The radiation in the UV region near 100 nm observed at the TTF was the smallest wavelength hitherto achieved with an FEL. It should be noted, however, that the shown data are values that were averaged from considerable noise, due to poor temporal coherence because the coherence time is much smaller than the pulse duration. This led to the development of HGHG FELs, described further below. For SASE FELs, the next development stage includes a length increase of the Linac to, for instance, 300 m and electron energies at >1 GeV resulting in FEL radiation down to 6 nm (soft X-ray region).

The time structure of the emitted light originates from the interaction of the electron beam and the radiation, both propagating through the undulator introducing a correlation on the scale length of $2\pi L_c \cdot L_c = (\lambda_{ph}/L_U)L_G$ is called the

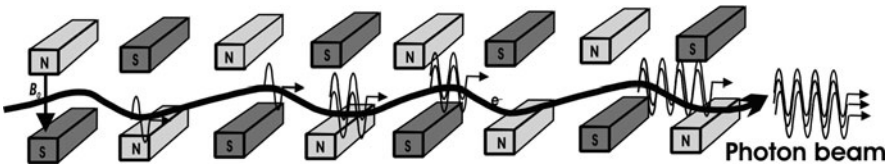


Fig. 2.14 Schematic of the SASE processin an FEL; electron discs emit synchronized electromagnetic radiation generating an intense laser beam

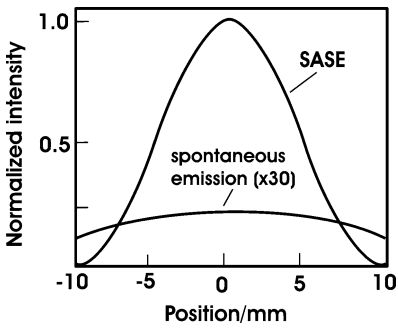


Fig. 2.15 Comparison of spontaneous undulator radiation intensity and FEL-SASE emission intensity; the position was measured with a photodiode (0.5-mm aperture) at a distance of 12 m from the undulator; note that the SASE data show considerable noise around the drawn averaged curve

cooperation length [32]. L_G is referred to as instability growth rate or gain length [30]. Nanometer long intense bunches are accomplished by so-called self-bunching of a longer injected bunch. Electrons that move in the transverse magnetic field of alternating polarity in the undulator can amplify an existing electromagnetic radiation field $\boldsymbol{E}$. In this case, $\mathbf{v}_e \boldsymbol{E} \neq 0$; hence, energy transfer between the electron beam and the radiation field occurs. The resulting longitudinal density modulation of the electron beam leads to micro-bunching. This FEL interaction produces spikes in the radiation pulse as well as a random intensity distribution. The number of the spikes is $N_{sp} = L_{Bunch}/2\pi L_c$ and is given by the ratio of the rms bunch length and the cooperation length. Hence, the time structure can be thought to result from the "slippage" of the electrons with respect to the radiation they produce and from the fact that the FEL starts from a fluctuation of the initial longitudinal particle distribution.

Because the laser amplification in the SASE mode is driven to saturation in a single pass of the electron beam through the undulator due to the lack of high-reflectance mirrors for X-rays, the temporal profile of an SASE-FEL shows a statistical, spiky structure as shown in Fig. 2.16 where coherence time is less than the pulse duration.

HGHG is an approach where the high intensity and spatial coherence of SASE are provided but, in addition, high temporal coherence is achieved [34]. The principal arrangement is schematically given in the top of Fig. 2.17. The electron beam that passes through the modulator is modulated in its energy by interaction with a seed laser pulse. In the shown initial realization, a CO_2 laser with a wavelength of 10.6 μm has been used. Using an optical chopper, laser pulse durations can be reduced to 10–100 ps. Initially, the pulse duration was 100 ns (CO_2 oscillator pulse length). The electron bunch enters the undulator (modulator) with a magnetic period λ_U and is overlapped with the pulse seeding laser of wavelength λ_L. The laser modulates the electron energy with the periodicity of the laser wavelength when the resonance condition holds (compare also (2.13))):

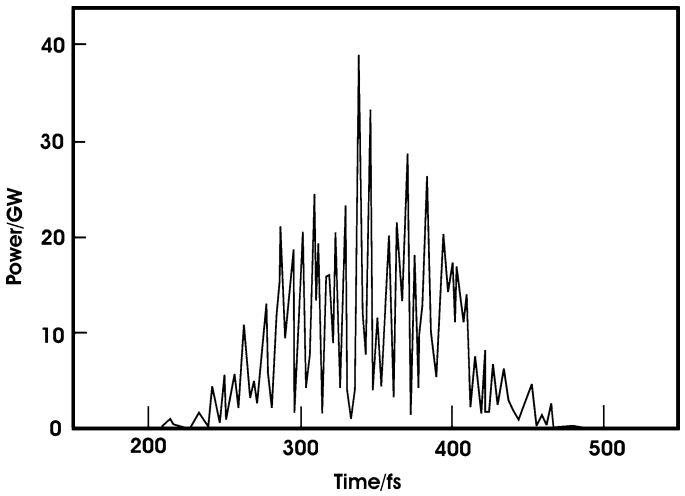


Fig. 2.16 Temporal output characteristic of an SASE-FEL, calculated using a full 3D time-dependent simulation code [33]

$$\lambda_{\mathrm{L}} = \lambda_{\mathrm{U}} \frac{1 + K^2}{2\gamma^2}. \tag{2.14}$$

Here, γ and K are, as above, the Lorentz factor and the undulator parameter, respectively. This resonance condition, where the laser wavelength λ_{L} equals the wavelength of spontaneous undulator radiation, is usually tuned by changing K between values of 1 and 15, accomplished by varying the vertical separation of the undulator magnets. The oscillatory energy modulation has the period length λ_{L} and is given by

$$(\Delta E)^2 = 4\pi\alpha A_{\mathrm{L}} E_{\mathrm{L}} \frac{\frac{K^2}{2}}{1 + \frac{K^2}{2}} \frac{\Delta\omega_{\mathrm{L}}}{\Delta\omega_{\mathrm{U}}}, \tag{2.15}$$

where α is the fine structure constant, A_{L} the beam energy and E_{L} the photon energy. The bandwidth ratio $\Delta\omega_{\mathrm{L}}/\Delta\omega_{\mathrm{U}}$ is usually expressed by the ratio of the undulator periods and the optical cycles in the laser pulse, $N_{\mathrm{U}}/N_{\mathrm{L}}$. In the dispersion section, the optical energy modulation is transformed into an optical density modulation of the electron bunch where the particle path length becomes energy dependent according to $\Delta z = R_{\mathrm{Di}}(\Delta E_{\mathrm{el}}/E_{\mathrm{el}})$. The deviation Δz from the path length z of an electron with nominal energy E by the energy modulation is additionally proportional to the so-called momentum compaction factor of the dispersion element, R_{Di}. Without considering collective effects (self-interaction, plasma oscillations), the amplitude of the density modulation can be approximated for the European XFEL, for instance, by the expression

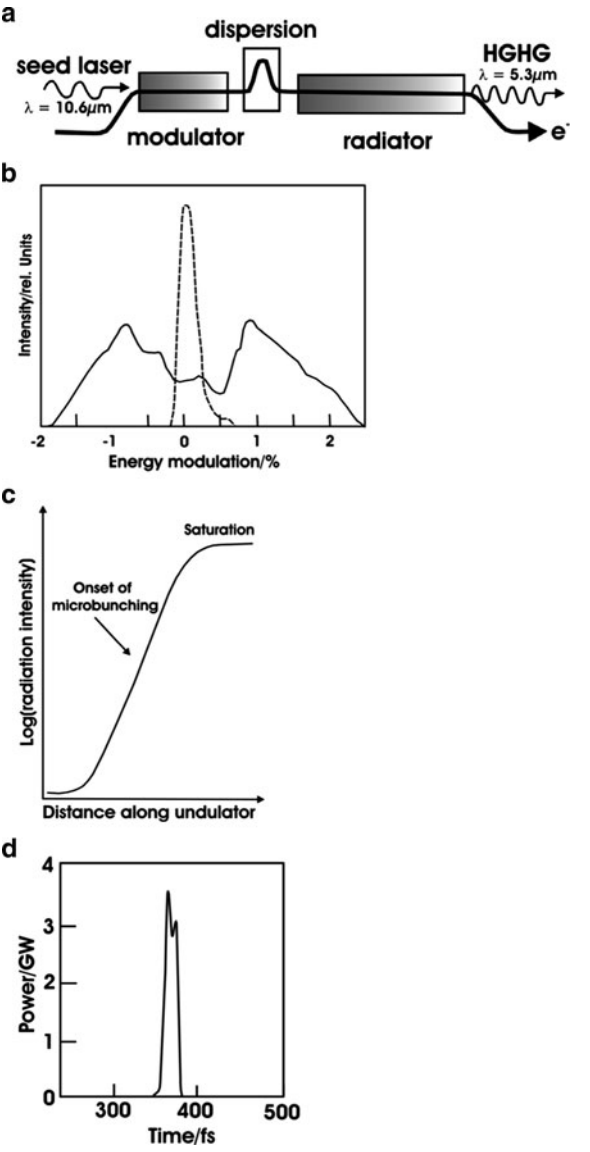


Fig. 2.17 Operation principle and output characteristic of a high gain harmonic generation free electron laser: (**a**) Arrangement with a seed laser pulse form a CO_2 laser with 10.6 μm wavelength that modulates the energy of electron beam (*full black line*) in the first undulator, called modulator. In the dispersion magnet, the energy modulation is converted into a spatial density modulation that is coherent. In the second undulator, radiation of the nth harmonic of the seed laser wavelength is generated and amplified to saturation; here, the second harmonic with half the wavelength of the seed laser has been realized. (**b**) Energy modulation for a seed CO_2 laser interaction in the modulator (*solid line*) and without seed laser (*dashed line*) (horizontal axis in % energy spread). (**c**) Schematic *curve* showing the increase of photon beam intensity until saturation in the radiator is reached; also shown is the onset of micro-bunching. (**d**) High gain harmonic generation output (see text)

$$A_i = \frac{2\pi R_{\mathrm{Di}}}{\lambda}\frac{(\Delta E)_i}{E}, \tag{2.16}$$

where an exponential term being close to unity has been neglected. With the data for the XFEL with electron energy of 2 GeV, $E = 4 \times 10^3$ (in units of the rest mass), a 0.5-MeV modulation at $\lambda = 400$ nm (second harmonic of a Ti:Sapphire laser) and $R_{\mathrm{Di}} \sim 30$ μm, a modulation amplitude of $A_i \sim 0.1$ is obtained. This density

modulation, occurring at the seed laser wavelength, is changed in the subsequent radiator, an undulator that is tuned to a higher harmonic n of the density modulation and it emits radiation at the preinstalled fraction of the seed wavelength. In the example shown in Fig. 2.17, frequency doubling has been achieved in the initial device. This output of the radiator element can be used as a seed pulse for the next stage in HGHG. By cascading several of such stages, short target wavelengths can be realized.

In order to discuss the phenomenon of saturation in the radiator, necessary for maximum intensity output, the interaction of the electron beam in an undulator in SASE configuration is shortly reviewed: in this situation, where an interaction between the electron beam and the emitted radiation is sustained, the net energy transfer from the electron beam to the photon beam leads to a back-interaction of the photon beam on the electron beam which induces a periodical modulation of the longitudinal beam density called micro-bunching. The density modulation produces a coherent fraction of light, generated within the micro-bunch and accordingly, the amplitude of the electric field of the photon beam is proportional to the number of electrons in the micro-bunch, as described above in the explanation of (2.13). The density modulation leads to an increase in radiation intensity, and the increased intensity in turn causes a higher beam density modulation. The gain length, mentioned above, can be expressed by $L_G = \lambda_U/4\pi\sqrt{\rho}$, where r denotes the so-called FEL parameter [30] which is a function of the normalized peak current of the electron bunch, the undulator parameter K, the wavelength of the radiation and the transverse rms size of the electron beam. The scaling parameter r is typically of the order of 10^{-3} for visible wavelengths and shorter ones; therefore, the gain length is typically about 100 undulator periods [35]. For the undulator saturation length, the relation $L_{sat} \sim 10\, L_G$ holds. For the TESLA-TTF SASE-FEL, saturation was reached after $\sim$14 m at a gain length of 1.35 m and for a radiation wavelength of 98.1 nm [36].

In the example given in Fig. 2.17, the intensity amplification of the seed laser pulse was about 24, with an input power of 0.7 MW and an output at half wavelength (5.3 μm) of 17 MW. In X-ray FEL development, cascading, where the radiator output serves as a seed pulse for the next stage, allows frequency multiplication. Assuming, for example, that third harmonic bunching is optimized in the dispersion element, the electron beam, seeded with an ω_0 pulse, would radiate at ω_3 after the long radiator undulator. In a second cascade, the original seed wavelength would be reduced to $\lambda_{HG} = 1/9\,\lambda_L$ and to 1/27 in a third stage. Thus, a wavelength reduction by a factor of 10–30 is feasible without additional acceleration between the stages. With seeding from a Ti:sapphire laser, wavelengths of about 30 nm can be realized. With meanwhile commercially available Ti:sapphire lasers that emit at 200 nm, the wavelength region around 1 nm becomes accessible. Very recently, the first XFEL operation has been reported [37].

In summary, the implementation of free electron lasers (FELs) as radiation source allows the investigation of processes that take place in the femtosecond range and towards the attosecond range with a large range of wavelengths down to the X-ray

Table 2.2 Overview of possible research areas that can advance by using free-electron laser light sources

Scientific field	Processes
Atoms/Ions/Molecules/Clusters	Multiphoton events; multiple ionization; creation and spectroscopy of excited states
Plasma physics	Dynamics, electronic and geometrical structure of matter; solid-density plasmas
Condensed matter physics	Plasma diagnostics; ultrafast dynamics; electronic structure
Materials science	Disordered materials and soft matter; dynamics of hard materials
Chemistry	Structure and dynamics of nanomaterials; reaction dynamics in solid/liquid systems; catalysis
Biology	Single particle/molecule imaging; dynamics of biomolecules; proteomics
Optics and non-linear phenomena	Non-linear effects in atoms, solids, gases and liquids; high-field science

region. We give here a brief overview of the topics that can be addressed with these light sources in Table 2.2.

Already in the year 2004, experimental realizations included that at $\sim$100 nm wavelength, a power of about 1 GW and sub-100 fs pulses were obtained [38]. The technique which has been employed to produce these short pulses has been termed *Laser Slicing* [39]. It is based on electron–photon interaction in a wiggler and will be described below. With femtosecond high-energy FELs, structural changes on the timescale of a molecular vibration, atomic motion associated with phase transitions, ultrafast chemical reactions and coherent vibrational motion become accessible.

2.6.4 Wiggler Radiation and Femtosecond Laser Slicing

Increasing the magnetic field strength beyond that in an undulator leads to distortion of the sinusoidal transverse electron motion due to relativistic effects. The radiation from one oscillation period does not overlap with the successive one and hence there is no constructive interference. The radiation can then be considered as the summing up of the synchrotron radiation from N separate magnets. If the dimensionless parameter $\eta = (eB_0/m\omega_0 c)$ for transverse electron motion in an oscillatory field is larger than 1, the motion of the electron in its average rest frame is relativistic and higher multipole radiation becomes relevant. This results in higher harmonics and a corresponding line spectrum. For very strong fields, merging into a continuous spectrum from IR to hard X-rays with rather constant spectral intensity occurs. A typical wiggler spectrum is shown in Fig. 2.18. Above the critical photon energy $\omega_c \propto B\gamma^2/\hbar$, the spectral emission drops exponentially.

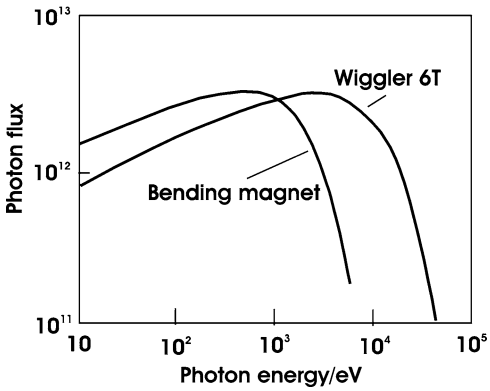


Fig. 2.18 Spectral characteristics of wiggler and bending magnet radiation; $E = 1.5\,\text{GeV}$; $I = 100\,\text{mA}$

A rather recent development using wiggler radiation is the generation of high-energy femtosecond pulses. It is based on the violation of the Lawson–Woodward theorem of electron interaction with a laser photon field [40, 41] which states that "the net energy gain of a relativistic electron interacting with an electromagnetic field in vacuum is zero". Since ponderomotive effects have been neglected in the formulation of the theorem, energy transfer to highly relativistic electrons ($v \approx c$) in a bunch can occur for the high electric field of a femtosecond laser. A Ti:sapphire laser system ($\lambda = 800\,\text{nm}$, $\Delta t \approx 50\,\text{fs}$) induces an energy modulation of the electrons as they traverse the wiggler [42]. The modulation occurs by higher-order laser field inhomogeneities, present in the pulse, which result in a net force acting on the centre of the oscillation of the particle. This force, known as ponderomotive force, is described by the relation

$$m\mathrm{d}v/\mathrm{d}t = \mathrm{q}\nabla\Phi_{\mathrm{POND}}, \tag{2.17}$$

$$\Phi_{\mathrm{POND}} = \frac{e}{4m\omega^2}|E_0|^2, \tag{2.18}$$

where $E_0(\mathbf{r})$ is the spatially varying amplitude of the wave field. The ponderomotive force is thus proportional to the gradient of the intensity of the wave field. As a result, electrons in the bunch which happen to be within the interaction volume with the laser pulse (see Fig. 2.19) move away from the centre position, thus creating a cavity that is known to emit terahertz radiation ([43], and references therein). This spatial separation from the rest of the bunch allows selection when passing through a bending magnet. In this dispersive element, the deflection of the main bunch and of the slower and faster electrons allows filtering of those electrons which have interacted in the short pulse time with the laser field.

Filtering is done by shading off the main bunch (i.e. the beam core region) with an aperture, thus achieving spatial separation by a dispersive bend. A schematic of the method is shown in Fig. 2.19. The resulting X-ray emission is of the timescale

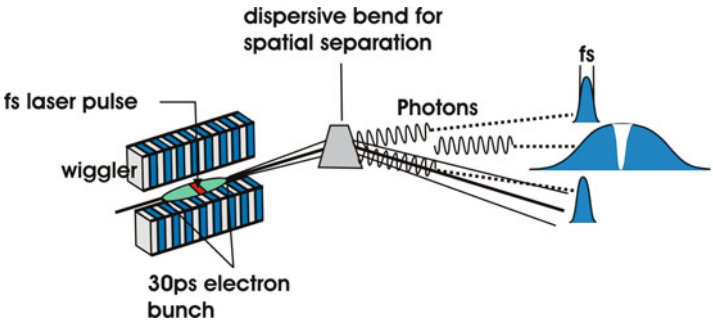


Fig. 2.19 Schematic of laser slicing technique for femtosecond X-ray pulse generation in storage rings

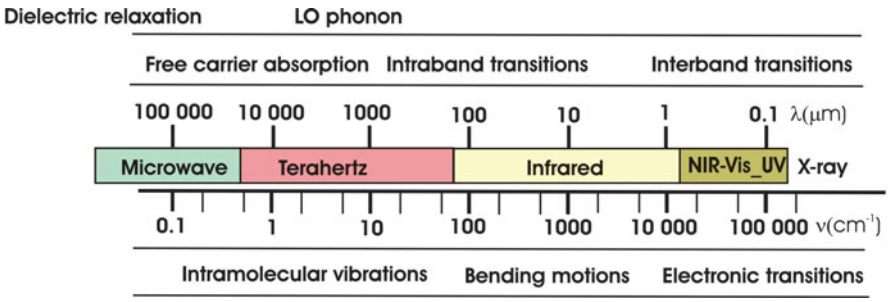


Fig. 2.20 Overview of the electromagnetic spectrum including the terahertz regime

of 100–200 fs. A first system is in operation at ALS (Advanced Light Source, Berkeley); others are being constructed (Bessy II, Berlin Adlershof, SLS, Swiss Light Source). More generally, ponderomotive forces are of widespread interest, for instance, in plasma physics, and are discussed in astrophysics where the influence of the neutrino ponderomotive force on electrons and other particles in a dense stellar fusion plasma is thought to contribute to supernovae dynamics [44].

Future developments of powerful light sources aim at the terahertz spectral regime, energetically located between electronics and photonics (see Fig. 2.20). FELs operating as long wavelength source are considered to complement high-energy short wavelength sources as have been described above. THz FELs will allow one to probe collective processes and their *function* in complex systems in a fashion complementary to analyzing *structure* via X-rays.

Typical envisaged applications include functional dynamics experiments by tuning to specific intramolecular vibrational modes for studying protein folding, the investigation of photosynthesis and metabolic pathways, and also superconductor properties, spintronics, medical imaging and detection of explosives and bio-agents [45].

References

1. S.B. Patel, *Nuclear Physics: An Introduction* (New Age International (P) Ltd. Publishers, 1991), p. 322 (reprint, 2006)
2. R.J. Bickerton, The purpose, status and future of fusion research, Plasma Phys. Control. Fusion **35**, B3–B21 (1993)
3. F.E. Cecil, D.M. Cole, F.J. Wilkinson III, S.S. Medley, Measurement and application of DDγ, DTγ and D^3Heγ reactions at low energy, Nucl. Instrum. Meth. Phys. Res. B **10/11**, 411–414 (1985)
4. J.M. Mack et al., Remarks on detecting high-energy deuterium–tritium fusion gamma rays using a gas Cherenkov detector, Radiat. Phys. Chem. **75**, 551–556 (2006)
5. M. Haegi, E. Bittoni, A. Fubini, S. Rollet, Gamma diagnostics on charged fusion products in a thermonuclear plasma, Nucl. Fusion **35**, 1625–1630 (1995)
6. M. Planck, Über das Gesetz der Energieverteilung im Normalspektrum, Ann. d. Phys. **309**, 553–563 (1901)
7. W. Wien, Ueber die Energievertheilung im Emissionsspectrum eines schwarzen Körpers, Ann. d. Phys. **294**, 662–669 (1896)
8. G.H. Aston, The amount of energy emitted in the γ-ray form by radium E, Proc. Cambridge Philos. Soc. **23**, 935–941 (1927)
9. B. Singh, S.S. Al-Dagazelli, Production of internal Bremsstrahlung accompanying β-decay from Sr^{90} and Y^{90}, Phys. Rev. C **4**, 2144 (1971)
10. G. Schatz, A. Weidinger, *Nuclear Condensed Matter Physics* (Wiley, New York, 1995)
11. E.M. Purcell, The lifetime of the $2^2S_{1/2}$ state of hydrogen in an ionized atmosphere, Astrophys. J. **116**, 457–462 (1952)
12. C. Eckart, The application of group theory to the quantum dynamics of monatomic systems, Rev. Mod. Phys. **2**, 305–380 (1930)
13. E.P. Wigner, Einige Folgerungen aus der Schrödingerschen Theorie für die Termstrukturen, Z. Physik **43**, 624–652 (1927)
14. P. Bloomfield, *Fourier Analysis of Time Series: An Introduction* (Wiley, New York, 2000)
15. C. Dey, A perturbed angular correlation spectrometer for materials science studies, Pramana J. Phys. **70**, 835–846 (2008)
16. M. Brüssler, H. Metzner, K.-D. Husemannn, H.J. Lewerenz, Phase identification in the Cu–In–S system by γ–γ perturbed angular correlations, Phys. Rev. B **38**, 9268–9271 (1988)
17. H. Metzner, M. Brüssler, K.-D. Husemann, H.J. Lewerenz, Characterization of phases and determination of phase relations in the Cu–In–S system by γ–γ perturbed angular correlations, Phys. Rev. B **44**, 11614–11623 (1991)
18. P.A. Cherenkov, Visible emission of clean liquids by action of γ radiation, Doklady Akad. Nauk SSSR **2**, 451 (1934)
19. I.E. Tamm, I.M. Frank, Coherent radiation of a fast electron in a medium, Doklady Akad. Nauk. SSSR **14**, 107–112 (1937)
20. A.P. Kobzev, The mechanism of Vavilov–Cherenkov radiation, Phys. Part. Nucl. **41**, 452–470 (2010)
21. E. Fermi, The ionization loss of energy in gases and condensed materials, Phys. Rev. **57**, 485–493 (1940)
22. W. Knulst, M.J. van der Wiel, O.J. Luiten, J. Verhoeven, High-brightness, narrowband, and compact soft X-ray Cherenkov sources in the water window, Appl. Phys. Lett. **83**, 4050–4052 (2003)
23. C. Luo, M. Ibanescu, S.G. Johnson, J.D. Joannopoulos, Cherenkov radiation in photonic crystals, Science **299**, 368–371 (2003)
24. F.J. Garcia de Abajo, Optical excitations in electron microscopy, Rev. Mod. Phys. **82**, 209–275 (2010)
25. W. Galbraight, J.V. Jelly, Light pulses from the night sky associated with cosmic rays Nature **171**, 349–350 (1953)

26. I. Arino et al., The HERA-B ring imaging Cherenkov counter, Nucl. Instrum. Meth. Phys. Res. A **516**, 445–461 (2004)
27. X. Artru, G.B. Yodh, G. Mennessier, Practical theory of the multilayered transition radiation detector, Phys. Rev. D **12**, 1289–1306 (1975)
28. J.P. Blewett, Synchrotron radiation 1873–1947, Nucl. Instrum. Meth. Phys. Res. A **266**, 1–9 (1988)
29. R. Follath et al., Comissioning of the U49/2 PGM1 beamline, Synch. Rad. Instrum. CP **705**, 348–351 (2004)
30. R. Bonifacio, N. Narducci, C. Pellegrini, Collective instabilities and high-gain regime in a free electron laser, Opt. Commun. **50**, 373–378 (1984)
31. J. Andruszkow et al., First observation of self-amplified spontaneous emission in a free electron laser at 109 nm wavelength, Phys. Rev. Lett. **85**, 3825–3829 (2000)
32. R. Bonifacio, L. De Salvo, P. Picrini, N. Piovella, C. Pellegrini, Spectrum, temporal structure, and fluctuations in a high-gain free-electron laser starting from noise, Phys. Rev. Lett. **73**, 70–73 (1994)
33. S. Reiche, GENESIS 1.3: a fully 3D time dependent FEL simulation code, Nucl. Instrum. Meth. A **429**, 243–249 (1999)
34. L.-H. Yu et al., High-gain harmonic-generation free-electron laser, Science **289**, 932–934 (2000)
35. S.V. Milton et al., Exponential gain and saturation of a self-amplified spontaneous emission free-electron laser, Science **292**, 2037–2041 (2001)
36. V. Ayvazyan et al., Generation of GW radiation pulses from a VUV free-electron laser operating in the femtosecond regime, Phys. Rev. Lett. **88**, 104802 (2002)
37. P. Emma et al., First lasing and operation of an angström-wavelenght free-electron laser, Nature Photon. **4**, 641–647 (2010)
38. J. Feldhaus, The FEL program at DESY: from first results at 100 nm wavelength to a True 1À X-ray Laser, Phys. Scripta T **110**, 413–419 (2004)
39. R.W. Schoenlein et al., Generation of femtosecond pulses of synchrotron radiation, Science **287**, 2237–2240 (2000)
40. P.M. Woodward, A method for calculating the field over a plane aperture required to produce a given polar diagram, J. Instrum. Electr. Eng. **93**, 1554 (1946)
41. J.D. Lawson, Lasers and accelerators, IEEE Trans. Nucl. Sci. **NS-26**, 4217 (1979)
42. A.A. Zholents, M.S. Zolotorev, Femtosecond X-ray pulses of synchrotron radiation, Phys. Rev. Lett. **76**, 912–915 (1996)
43. N. Yamamoto et al., Study of the coherent terahertz radiation by laser bunch slicing at UVSOR-II electron storage ring, Proc. IPAC' **10**, 2570–2572 (2010)
44. R. Bingham, L.O. Silva, J.T. Mendonca, P.K. Shukla, W.B. Mori, A. Serbeto, Neutrino plasma coupling in dense astrophysical plasmas, Plasma Phys. Control. Fusion **46**, B327–B334 (2004)
45. G.R. Neil et al., Production of high power femtosecond terahertz radiation, Nucl. Instrum. Meth. Phys. Res. A **507**, 537–540 (2003)

Chapter 3
Photon Generation in Electronic Transitions: Lasers and Nanoscopic Sources

3.1 Introduction

As an extension of the methods to generate photons, this chapter elaborates on electromagnetic radiation that originates from energy dissipation in electronic transitions between quantum states. These processes encompass fluorescence, luminescence and laser action. First, short wavelength radiation emitted from lasers will be treated and, subsequently, processes such as X-ray fluorescence, chemo- and bioluminescence will be presented in Sect. 4.5. Light emission in scanning tunnelling experiments and a short review on the manipulation of electromagnetic radiation using so-called metamaterials follow. Subsequent to the description of the somewhat mystical properties of metamaterials, the chapter closes with an overview of so-called mysterious light sources that stimulated the fantasy of people for many centuries. The section sequence has been chosen to cover photonic processes from high energy to high spatial resolution. A summary of the section headings is given in Table 3.1.

3.2 X-Ray Plasma Lasers

The basic operation principle of a laser involves (1) population inversion, resulting in an excited metastable state, (2) subsequent stimulated emission and (3) optical confinement in a Fabry–Perot cavity [1]. At the core of laser action is the stimulated emission process, first introduced by Einstein [2] and later theoretically described by Göppert-Mayer [3]. Stimulated emission occurs when an incident photon strongly couples with an excited electron and causes its decay to a lower level, thereby emitting two photons of the same energy, direction and phase as the incoming photon (Fig. 3.1). The coherence properties of the laser result from these photon correlations. Actually two-photon emitting systems have been employed to test Bell's inequality, regarding a discussion on causality in quantum

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_3,

Table 3.1 Overview of the content of Chap. 4

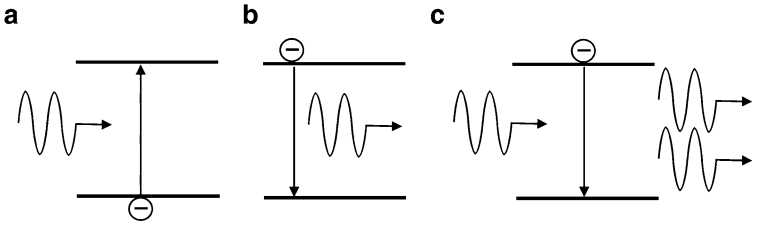


Fig. 3.1 Schematic of stimulated emission (**c**) compared to spontaneous emission (**b**) and absorption (**a**) of photons by an electronic system (see text)

mechanics (see Chap. 6) [4, 5]. The inverse process, e.g. two-photon absorption, has found widespread application and is now used in biology as multi-photon microscopy [6, 7].

The laser systems treated below have been ordered with respect to their energetic and temporal properties. Being closer in emission energy to FELs, we begin with the soft X-ray lasers, followed by VUV excimer lasers and the Nova system.

Earlier examples of envisaged applications of X-ray plasma lasers were soft X-ray projection lithography for fabrication of integrated circuitry with dimensions between 30 and 70 nm that were already reached in 2002 [8], low photon energy radiobiology and soft X-ray contact microscopy of living biological species [9], for instance. Typical energies allow to tune into the "water window", i.e. the spectral region between 2.33 and 4.37 nm given by the carbon K-edge located at 4.37 nm and the oxygen K edge at 2.33 nm (Fig. 3.2). As seen in the figure, photons in the

window range are absorbed by C but not by oxygen in water, thus maximizing the contrast between water and biological molecules. The accessibility of this spectral range allows the study of living biological organisms.

The plasma laser concept goes back to the work of Matthews, Rosen, Suckewer, and their coworkers [10–12]. Laser action is based here on achieving population inversion in plasmas. Two operational modes have been realized: collisional excitation and collisional recombination. In the former, the laser upper levels are predominantly excited by direct electron impact collision from the ground state of the ion stage of interest. Usually, plasma ionization leads to highly charged ions that are stripped from their outer electrons until a rather stable electron configuration of noble gas character is reached. In many cases, this is the Ne-like ion shell: Ne-like Se, Ge as well as Cu and Ag ($Z = 29, 47$) show laser action. The populated upper laser levels are metastable with respect to radiative decay to the ground state and the lower-lying laser levels are depopulated by strong dipole-allowed transitions (labelled E_1 in Fig. 3.3). In the given example, collisional excitation (not shown in

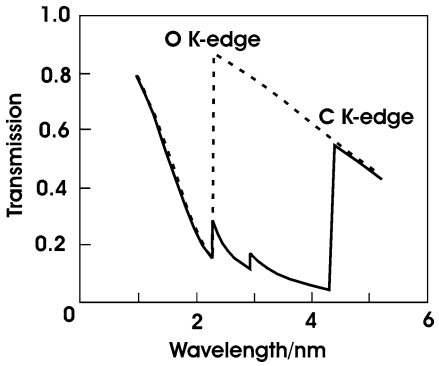


Fig. 3.2 Transmission in the water window of oxygen (in water) and carbon for 1-μm thickness

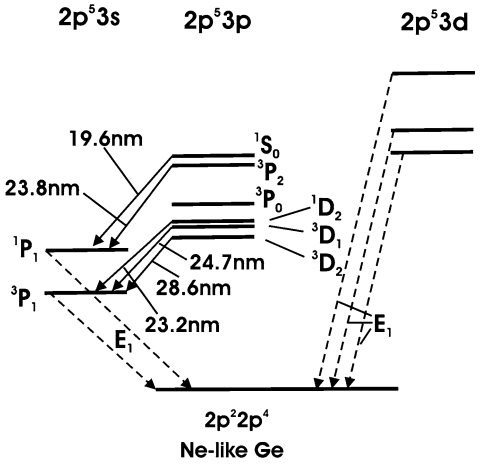


Fig. 3.3 Simplified Grotrian diagram of Ne-like Ge showing de-excitation processes only; nomenclature: terms are given in Russell–Saunders notation with $L = 0, 1, 2, 3$ represented by S, P, D, F and the total spin $2S + 1$ and total angular momentum $J = L + S$ written as indices in front and behind the angular momentum letter, respectively: $^{2S+1}P_J$, for instance; main terms are also represented in j–j coupling

Fig. 3.3) populates the upper 3p levels from the Ne-like ion ground state and also by cascades from higher states, such as the 3d levels.

Population inversion is achieved by the very fast decay of the laser 3s levels ($^1P_1$, $^1P_3$) to the ionic ground state.

In collisional recombination lasers, the upper laser level is populated following the recombination of ions according to a three-body interaction

$$X^{Z+1} + 2e \rightarrow X^{Z^*} + e \tag{3.1}$$

that populates highly excited bound levels of the ion with charge Z. Realization of ion lasers for the XUV to X-ray spectral region requires a high gain medium in order to compensate cavity efficiency and gain coefficient (G) losses at decreasing wavelength:

$$G = N_U\sigma_{em} - N_L\sigma_{abs} = \text{const}\, S\lambda\sqrt{\frac{M}{T}}IN_U \tag{3.2}$$

N_U and N_L denote the densities in the upper and lower state, respectively, connected with the cross sections (σ) for induced emission and absorption; S is the transition oscillator strength, M the ion mass, T the plasma temperature and I is the so-called inversion factor. Attaining emission of X-ray photons necessitates judicious adjustment to meet several crucial criteria: the gain condition that determines the pumping threshold for lasing requires a larger gain than loss coefficient. Translated into a condition for the population difference $N_0 > N_{th}(N_0$, stationary population density difference), where N_0 increases with the pumping rate R. The threshold population difference N_{th} is defined as the ratio of loss coefficient α_r and transition cross section $\sigma(v)$:

$$N_{th} = \frac{1}{ct_{ph}\sigma(v)}, \tag{3.3}$$

where t_{ph} denotes the photon lifetime. Using $\sigma(v) = (\lambda^2/8\pi t_{sl})g(v)$, with t_{sl} spontaneous lifetime and $g(v)$ the transition lineshape, it follows that

$$N_{th} \propto \frac{1}{\lambda^2 t_{ph}}. \tag{3.4}$$

Therefore, as one wants to decrease λ, the threshold population density difference must be increased quadratically with the laser wavelength, making laser action increasingly difficult.

Early X-ray lasers were realized by line – focussing a pulsed pump laser beam on a target, typically a thin metal foil, creating a plasma and within 2–3 pulses, the collisional population inversion was achieved in the formed elongated plasma column. Spontaneous emission is amplified through the plasma column, resulting in a propagating monochromatic X-ray beam with small divergence. An early X-ray system uses the plasma of C^{6+} ions generated by an infrared CO_2 laser pulse [12].

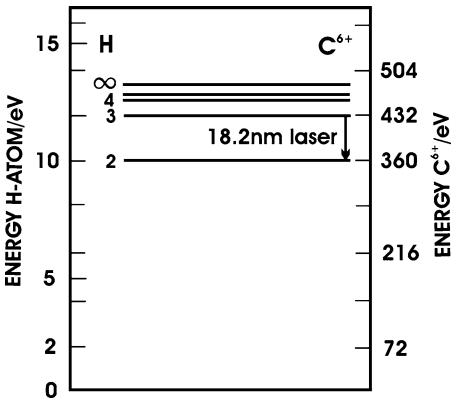


Fig. 3.4 Energy levels of H-like C^{6+} in comparison to H

This C VI H_α laser is a collisional recombination laser. Figure 3.4 shows the energy scheme for C^{6+} in comparison to the Rydberg series for the H atom. Population inversion occurs during cooling of the plasma towards the end of the 50-ns laser pulse by capture of electrons in the $C^{6+} q = 3$ level and a simultaneous dearth of electrons in the $q = 2$ level. The energy of the emitted photons is given by

$$E_{\mathrm{ph}} = \frac{m_\chi Z^2 e^4}{2\lambda^2 \left(\frac{1}{2^2} - \frac{1}{3^2}\right)}. \tag{3.5}$$

Here, m_r is the reduced mass of the atom and $Z = 6$. The spontaneously emitted photons with $E_{\mathrm{ph}} = 68\,\mathrm{eV} (\lambda = 18.2\,\mathrm{nm})$ cause stimulated emission of X-ray photons from other ions, leading to an amplified spontaneous emission. The overall laser gain, defined as $G(v) = \mathrm{e}^{\gamma(v)d}$ (γ amplifier gain coefficient, d total length of interaction region), yields $G \sim e^6$ in this case. As a result, a 20-ns pulse of soft X-ray amplified spontaneous emission with an energy of 2 mJ, a power of 100 KW and a divergence of 5 mrad was obtained.

More recent developments involve transient pumping schemes which use an initial lower intensity pulse of longer duration to create the plasma and a subsequent high-intensity femtosecond pulse which creates the population inversion through electron-ion collisional excitation. The first realization using the so-called chirped pulse amplification to expand and compress the high intensity pulse was achieved with Ne-like Ti [13].

Another rather recently developed area is the channelling of intense optical fields in plasmas where ultrashort laser pulses, injected into relativistic plasmas, generate a wakefield which can be used to accelerate particles [14]. For intense laser pulses, the oscillating motion of the plasma electrons becomes relativistic, leading to self-channelling due to an increase of the electron effective mass for $v \approx c$. The resulting change of the effective refractory index in the plasma and electron cavitation due to ponderomotive forces lead, similar to optical waveguides, to self-guiding that results in relativistic self-focussing of terawatt laser pulses [15].

3.3 Vacuum Ultraviolet Lasers

The energy range considered here is shown as a part of Fig. 3.5. As a rough estimate, the demands in sub-quarter-micron lithography define the wavelength range of the systems, e.g. the molecular fluorine and compound excimer lasers (noble gas halides.). Applications extend those of nanolithography; they encompass bio-particle detection based on their UV absorption characteristics, molecular uncaging through UV photolysis and glass and plastic patterning, for instance. The wavelength region considered for the lasers with single emission lines transforms to an energy range of ∼4.0–7.9 eV, thus spanning energies of many atomic and molecular excitations as also the fundamental absorption edge of many bio-molecules [16]. It also is the wavelength region presently used in the downscaling of structures in electronic devices. These laser systems belong into the category of excimer lasers, a word composed from *excited dimer*. Excited diatomic molecules are stable only in their excited states and quickly dissociate when a transition to the ground state occurs.

Consequently, large population inversions can be reached since the ground state is depleted by dissociation. The excited states, however, have usually a much shorter lifetime compared to the metastable states of more conventional lasers and require high pumping rates. The laser action is characterized by high power pulses in the blue or UV spectral region after excitation by fast electron beam discharges.

Because of its high-energy radiation, the physics and operational principle of the fluorine laser (emission at $\lambda = 157$ nm) is presented first, followed by a description of noble gas-halide excimer laser action. Limitations in pattern resolution of the KrF, ArF and similar lasers (emission lines at $\lambda = 248$ nm and 193 nm, respectively) will eventually lead to their replacement. In particular, free electron lasers have the potential to reach shorter wavelength regions, as indicated in Fig. 3.5. A 20-nm radiation is available since 2005 at the TTF2 FEL at DESY and generation of 6nm

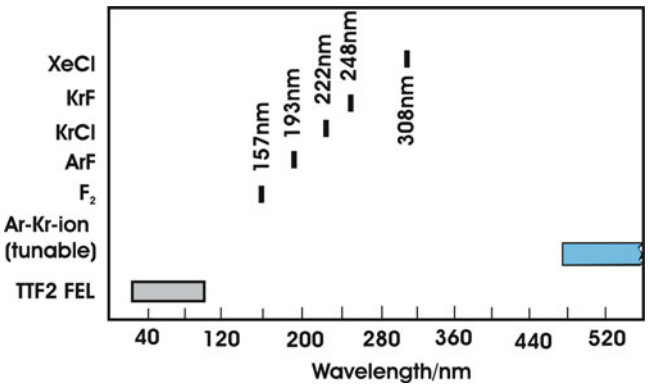


Fig. 3.5 Overview of emission lines of selected VUV lasers including the Tesla test facility free electron laser (TTF2 FEL) and a tunable ion laser with 676 nm $\leq \lambda \leq$ 476 nm

(corresponding to 200 eV) radiation is planned. Ultimately, undulator synchrotron radiation that allows access to the water window (see above) is anticipated.

3.4 The Molecular Fluorine Laser

3.4.1 Short Note on Terminology

The determination of the rotational, vibrational and electronic energy levels of diatomic molecules follows from solving the time-independent Schrödinger equation $H\psi = E\psi$, where the Hamilton operator can be written in the Born–Oppenheimer approximation (separation of electronic, vibrational and rotational energies) as $H = H_e + H_{vib} + H_{rot}$ and the according eigenfunctions are written as a product: $\psi = \psi_e \cdot \psi_{vib} \cdot \psi_{rot}$. For molecules with closed shells (no unpaired electrons), the eigenvalues are obtained by $E(v, J) = T_e + G(v) + F_v(J)$ where the terms describe the electronic (T_e), vibrational and rotational energies, respectively. The term nomenclature for linear molecules, such as F_2, is adapted from that of atoms with the difference that for the molecules, the total angular momentum component with regard to the dinuclear axis has to be considered, too. That component is denoted by $\hbar\Lambda$. For $|\Lambda| = 0, 1, 2, \ldots$, one uses the symbols Σ, Π, Δ, ..., where the value of Λ results from the individual values of the momentum component, λ_{el}, of the electrons of the molecule. The component of the total spin angular momentum with regard to the internuclear axis yields singlets, doublets and triplets, for instance, denoted by the superscripts 1, 2, 3, respectively. The additional $+/-$ superscript and the g/u subscript relate to the symmetry behaviour of the orbital function (symmetric, anti-symmetric, gerade, ungerade). For the H_2^+ molecule, with one electron in a σ orbital that exhibits rotational symmetry around the H–H axis, λ_{el} is zero. Therefore, $\Lambda = 0$ and the term symbol for this molecule is Σ. With the spin multiplicity of 2 (one electron with spin $\pm 1/2$), the state is written as $^2\Sigma$. The parity of the single occupied orbital is even (g) and the molecular term is, therefore, also labelled with the subscript g. The singly ionized hydrogen molecule orbital is, hence, written as $^2\Sigma_g$. In addition, one distinguishes electronic ground and excited states. They are labelled by the capital letters X, A, B, C, ..., for the case that the excited states possess the same term multiplicity (spin quantum number). If the multiplicity is different, then lower-case letters (a, b, c, ...) are used.

3.4.2 Operational Principles

The F_2 laser oscillates on a transition between two bound molecular electronic states with a spectral emission at 157.52 nm and 157.63 nm for discharge excitation (Fig. 3.6). Excited fluorine molecules F_{2*} form the upper $^3\Pi_g$ level of the F_2 laser. Under the usual high pressure conditions, collisional relaxation occurs and only the lowest vibrational level of the upper electronic level is highly enough populated.

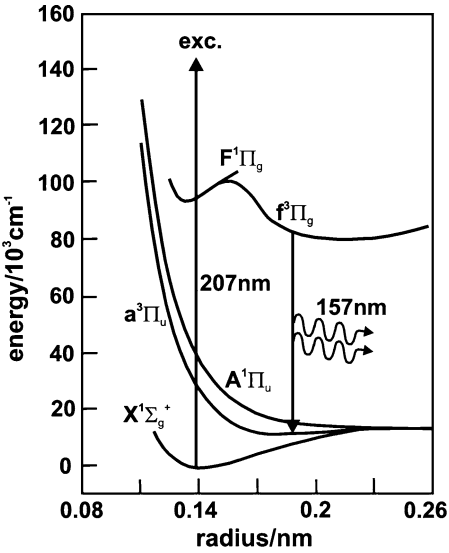


Fig. 3.6 Energy term scheme for the molecular fluorine laser

In discharge pumped lasers, the formation of F_{2*} results from either a neutral reaction $F_2 + F^* \rightarrow F_{2*} + F$ or an ionic three-body recombination reaction in the He–F_2 gas mixture $F^+ + F + He \rightarrow F_{2*} + He$. At least four possibilities to form excited F* atoms have been postulated, among them are excited rare gas halide species (HeF*) which rapidly decay to yield F*. Another route is the formation of F^- by dissociated capture of low-energy electrons by F_2 resulting in F* after further electron capture or by reaction with F_{2+}. Faris et al. [17] have first attributed the sharp fluorescence spectrum at 7.9 eV to the laser transition $f^3\Pi_g \rightarrow a^3\Pi_u$ (for nomenclature, see Sect. 3.5). The quenching mechanisms for the upper and lower states are either due to reactions with F_2 electrons (upper and lower level) or due to collisional de-excitation by He atoms (lower level only). Therefore, an increase of the buffer gas pressure results in an accelerated decay of the lower laser level. Simultaneously, metastable F* atoms can be increasingly formed from HeF* complexes, thus increasing the formation rate for the upper laser level. Also the rate for the ionic reaction increases with the buffer gas pressure. The relative concentration of F_2 in the laser gas mixture has to be optimized due to the adverse effect of increasing the formation rates of the upper level and its quenching by F_2. It turned out that efficient operation is obtained at higher buffer gas pressure. In an early experiment [18], the VUV pulse energy was shown to increase as a function of He pressure from about 30 mJ (depending on the charging voltage V_c at 3 bar up to 180 mJ at 11 bar for $V_c = 80$ kV).

Commercially available F_2 lasers emit more than 10 W at a 1,000-Hz repetition rate with dual wavelength output (at $\lambda = 157.523$ and 157.629 nm). Beam delivery systems that connect to the laser comprise CaF_2 beam shaping optics and N_2 purging to prevent absorption by molecular oxygen and water vapour. Lithography patterns in 100-nm Shipley resist on silicon for the 70-nm technology mode have been

fabricated rather early [[19], and references therein] showing regular resist pattern formation down to 70 nm and, using a fluoro-polymer-based resist, an ultimate resolution of 55 nm has been obtained.

The development of broadly tunable VUV lasers uses frequency shifting of tunable laser sources. Frequency mixing with optically nonlinear solids is difficult due to their absorption in the VUV energy range; therefore, gaseous media are used, which results in a reduction of the laser efficiency. In nonlinear optics, the higher terms of the electrical polarization P, given by the series

$$P(E) = \chi_0 E + \chi^{(2)} E^2 + \chi^{(3)} E^3 + \cdots \tag{3.6}$$

are used. Here, $4\pi\chi_0 = \varepsilon_0 - 1$ is the dielectric susceptibility for the linear case. In crystals with a symmetry centre or for isotropic materials, $\chi^{(2)} = 0$; hence, in gaseous media, the so-called three wave mixing term vanishes and the third term on the right-hand side of (3.6) contains the four wave mixing process. The according susceptibility, written in Dirac notation [20], is given by

$$\chi^{(3)} = \frac{\langle \psi_g \left| \vec{p} \right| \psi_a \rangle \langle \psi_a \left| \vec{p} \right| \psi_b \rangle \langle \psi_b \left| \vec{p} \right| \psi_c \rangle \langle \psi_c \left| \vec{p} \right| \psi_g \rangle}{(\Omega_{cg} - \omega_1 - \omega_2 - \omega_3)(\Omega_{bg} - \omega_1 - \omega_2)(\Omega_{ag} - \omega_1)}. \tag{3.7}$$

The Dirac brackets denote dipole transition matrix elements between the involved states ψ; $\vec{p}$ is the momentum operator $((\hbar/i)\nabla)$. The denominator describes the energy relations, including lifetime (energy) broadening Γ, for example, $\Omega_{ag} = \omega_{ag} - i\Gamma/2$. Among the three possibilities to set one of the real parts of the denominator of (3.7) to zero, only for $\omega_1 + \omega_2 = \omega_{bg}$, an enhanced susceptibility without linear absorption of incident or generated radiation is obtained. Hence, the incident radiation has to be tuned to a two-photon resonance [21]. Tunable radiation is obtained at the sum or difference frequency $\omega_4 = 2\omega_1 \pm \omega_2$ using tunable radiation at a second frequency ω_2 after tuning one of the incident frequencies such that $2\omega_1 = \omega_{bg}$ (Fig. 3.7).

In compound excimer lasers, such as KrF-, ArF- and XeF-based lasers, the noble gas and the halide in its molecular form X_2 (X = F, Cl) form compound molecules in hot plasma environment. Lifetimes of excited states are in the 1–10-ns range, long enough to achieve pulsed laser action. Rather high-power pulses can be realized because the cross section for stimulated emission is comparably low, allowing the built-up of a large population inversion.

3.5 Photon Emission in Chemical, Biological and Solid-state quantum systems

This section addresses the principles and some recent applications in using light in the visible and near IR region to study dynamics of chemical reactions, to extract information on metabolic processes in biological organisms [23], to analyse surface reactions [24] and to monitor quantum confinement [25].

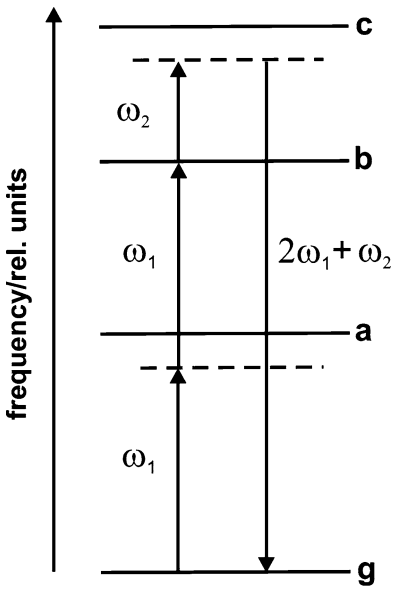


Fig. 3.7 Term scheme for achieving tunable laser radiation from frequency mixing [22]; the frequency ω_1 is tuned such that $2\omega_1$ corresponds to the frequency of an allowed two-photon excitation; ω_2 is variable; susceptibility enhancement is achieved by the resonance process where $2\omega_1 = \omega_{bg}$ (see text)

Chemiluminescence, as described here, refers to light emission due to energy release in a defined chemical reaction. We distinguish between the purely chemical phenomenon in oxidation processes involving radicals or reactive oxygen species and electroluminescence, e.g. electrical current-induced light emission. Compared to the incandescent emission by hot objects, chemiluminescence is often referred to as cold light, indicating that the energy dissipation is dominated by adiabatic electron recombination.

Bioluminescence is among the most important tools in life sciences and the associated technologies. It is applied in microbiology, food and environmental testing, immuno- and enzyme assays and in gene expression assays. The technique of fluorescence labelling of biological molecules in conjunction with nano-technological supports (solid or biological) is a rapidly developing field with numerous discipline-bridging applications.

3.5.1 Chemiluminescence

Chemilumiscence in molecular systems is commonly described by a Jablonski diagram (Fig. 3.8) [26]. The excitation energy is supplied either by a chemical reaction or by photons, distinguishing between chemiluminescence and fluorescence/phosphorescence. For better intelligibility, we shortly review the nomenclature: Singlet and triplet states refer to antiparallel or parallel spin configurations, respectively. The triplet states are energetically lower-lying states. They are labelled S and T in Fig. 3.26. Only the electronic states (S, T) and the vibrational levels (v_s),

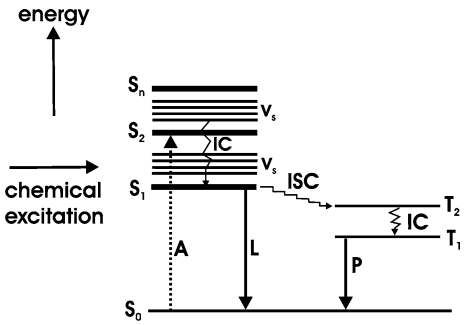


Fig. 3.8 Jablonski diagram for chemiluminescence, phosphorescence (P) and fluorescence (L): A absorption of photons, T triplet states, S singlet states (see text)

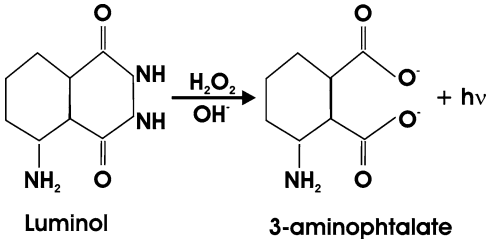


Fig. 3.9 Chemiluminescence reaction of luminol

but not the rotational levels with even smaller splitting, are shown. Energy loss by internal conversion (IC) as well as luminescence (L) and phosphorescence (P) are shown.

The transitions $S_n \rightarrow S_0$ are spin-allowed and fast, resulting in luminescence. Upon intersystem crossing (ISC), spin-forbidden transitions $T_1 \rightarrow S_0$ to the ground state have low probability and are slow, leading to phosphorescence with time windows of about 10^{-4}s to minutes. A well-known example is the reaction of luminol with hydrogen peroxide in alkaline medium to 3-aminophtalate which leads to emission of blue light (see Fig. 3.9).

The chemiluminescence quantum yield Φ_{CL} is the product of the probability of generating an excited state, W_{ex}, the fluorescence quantum yield Φ_{ph} and the reaction probability W_R, describing the fraction of starting molecules which undergo the luminescent decay rather than a side reaction

$$\Phi_{CL} = W_{ex} \times \Phi_{ph} \times W_R. \tag{3.8}$$

In diatomic molecular reactions, such as in collisional excitations to form stable excited states for excimer laser action, the description of the energy levels uses total angular momentum J, total spin S, the molecular symmetry under inversion ("gerade" and "ungerade" meaning even or uneven transformation) and wave function symmetric or antisymmetric character. For $J = 0, 1, 2, 3\ldots$, the symbols $\Sigma, \Pi, \Delta, \Phi$ are used. The total spin of 0, 1/2, 1, 3/2, 2, … is indicated by small numbers 1, 2, 3, 4, 5, … such as $^1\Sigma$, for instance. The wave function character and the state symmetry are noted by a plus or minus sign and a "g" or "u" behind the

angular momentum symbol, reading $^3\Sigma^+{}_g$ for a molecular level of the H_2 molecule as an example. Often, one also finds X, A, B, C, etc. in front of the Σ, Π, Δ, etc. symbols used to describe the ground state (X) and subsequent excited states (A, B, C, ...).

Chemiluminescence in oscillatory reactions has been reported for the autocatalytic Belousov–Zhabotinsky reaction [27]. Autocatalysis refers to the catalysis of a reaction by its products. It occurs if the reaction rate for the reaction $A \rightarrow B$ is given, for instance, by $k_{AB} = a[A][B]$. Hence, the reaction rate increases when the product is formed. As an example, consider the model autocatalytic Lotka–Volterra mechanism (where the first two steps are autocatalytic):

$$\begin{array}{lll} \text{(i)} & A + I_1 \rightarrow 2I_1 & d[A]/dt \propto -[A][I_1] \\ \text{(ii)} & I_1 + I_2 \rightarrow 2I_2 & d[I_1]/dt \propto -[I_1][I_2] \\ \text{(iii)} & I_2 \rightarrow B & d[B]/dt \propto [I_2] \end{array} \tag{3.9}$$

Here, $[I_1]$ and $[I_2]$ are the concentrations of reaction intermediates. If the concentration of A is kept constant by steady supply of the consumed reactant, the variables are the concentrations of the reaction intermediates, since B does not contribute to the reaction once it has been formed. In chemical oscillations, autocatalysis provides the positive feedback. The periodic variation of I_1 and I_2 has some similarity with the fox-and-rabbit scheme: at some stage, there is little I_1 (rabbits) present, but the reaction (i) generates more and the production of I_1 autocatalyses the generation of even more I_1. Then reaction (ii), e.g. the formation of I_2, begins slowly because $[I_2]$ is small (which could be envisaged as a population increase of foxes due to more rabbits as food). Autocatalysis, however, leads to a surge of I_2 which removes I_1 (in this case by eating). Since now less rabbits are "available", the concentration of I_2 decreases, slowing reaction (ii). As less I_2 is present to remove I_1, the latter can surge forward (multiply) again, resulting in an oscillatory cycle. The reaction forms $HBrO_2$ which is a reactant in the first step of the bromate oxidation in acidic solution [27]. It is noteworthy that Belousov's manuscript was initially rejected and that he subsequently published it in the journal of the institute for theoretical and experimental biophysics at Moscow university, where it was not reviewed [28]. Figure 3.10 shows the periodic variation of the solution redox potential and the chemiluminescence in opposite phase to it. The origin of the luminescence has so far only been speculated about.

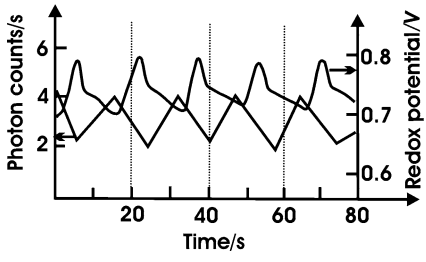


Fig. 3.10 Periodic photon emission from the oscillating Belousov–Zhabotinsky reaction

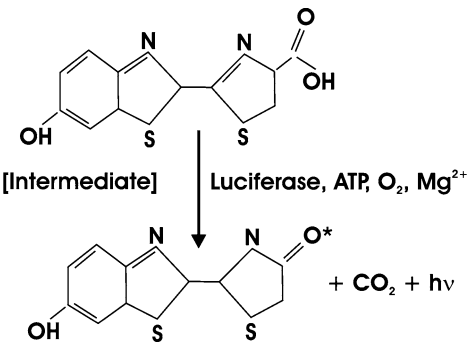


Fig. 3.11 Scheme of light emitting reaction by oxidation of luciferin

3.5.2 Bio- and Electroluminescence

In bioluminescence, numerous observations of light emission from marine bacteria, pony fish and the firefly (lampyridae), for instance, are known. Luciferin is the generic term for a wide variety of compounds that emit light when oxidized in the presence of the enzyme luciferase (Fig. 3.11):

ATP provides the energy to reset the system after oxidation. Dioxetanone is the energy-rich intermediate. Its breakdown releases the energy needed to enter the excited state of oxyluciferin (lower part of Fig. 3.11) and CO_2. The efficiency of photon emission per oxidized luciferin molecule is close to 1. The emission colour varies and differences between 560 and 630 nm are discussed to be related to the ternary structure of the catalytic site [29]. The photons generated by these biochemical reactions are used in various ways: in photodynamic therapy (PDT), one makes use of the fact that photosensitizing agents are collected more readily in cancer cells than in normal cells.

The uptake and retention of a photosensitizer in a tumour is followed by irradiation with light of wavelength 650–800 nm, which results in the formation of singlet oxygen which is highly toxic to cells, initiating the necrosis of the tumour. Figure 3.12 shows a simplified Jablonski diagram for the generation of singlet oxygen. The porphyrin sensitizer is excited to $^1S^*$. If the lifetime of the excited state is long enough, the energy can be dissipated not only by fluorescence (L) but by intersystem crossing to the triplet state of the sensitizer $T_1(^3S^*)$. If the energy released is larger than the excitation energy needed to convert the available triplet oxygen $^3O_2$ into singlet oxygen $^1O_2$ as shown here, the latter will be formed, leading to tumour death.

In biological resonant energy transfer, the photons from bioluminescence are used to attain information on protein–protein interaction and conformational change. Even ultra-weak photon emission in the visible and near IR accompanying metabolic processes can be used to obtain physiological information or to extract information on regulation processes in biochemical reactions [23].

Electroluminescence can be observed in electrochemical and photoelectrochemical systems, for instance. It is, of course, also used in solid-state devices and

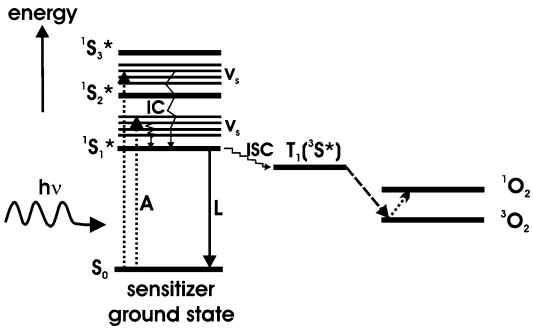


Fig. 3.12 Jablonski diagram for a generation of cell-toxic singlet oxygen in PDT

a wealth of literature covers this subject (see for example [30] and references therein). We restrict the considerations to the reactive interface encountered in electrochemistry. In silicon electrochemistry, porous Si is formed and sustained current oscillations are observed at higher anodic potentials in fluoride containing solution [31]. It has been found that the integral thickness of the anodic oxide, the surface roughness at the oxide/electrolyte contact and the integral over the light-induced minority carrier concentration vary periodically with fixed phase with regard to the oscillating current [32]. The oscillations themselves have been described by a combination of system theory, Markov processes [33] and probability considerations. Macroscopic oscillatory behaviour is obtained from the synchronization state of the system which is the Markov chain generated by probability distributions for the oscillator periods [34, 35]. A periodical variation of the photoluminescence from the $Si/SiO_x^-SiO_2$/electrolyte system has been observed for which the photoluminescence is off phase to the current by 180° [36]. The luminescence behaviour has been attributed to surface electronic properties at the Si/oxide interface under slowly varying surface electrostatic potential.

3.5.3 Photon Emission from Solid-State Quantum Systems

The optical analysis of *quantum effects* due to reduced dimensionality has almost exclusively been done on semiconductors. Two-dimensional quantum well structures and quantum dots have been investigated. The origin of luminescence of InGaN insertions in GaN-based light emitting structures, for instance, has been attributed to InGaN nano-domains with quantum dot properties [37]. In InGaAs quantum well structures, the appearance of luminescence rings with diameter of several hundred microns has recently been reported [38, 39]. Also, magneto-luminescence showing a well-resolved Zeeman effect from InP quantum dots formed by self-organization on InGaAs/GaAs quantum wells [40] has been reported. Electron energy quantization in metallic nano-particles by optical methods in the form of discrete electronic states has been observed only rather recently: in two-photon excited luminescence, discrete lines from colloidal silver aggregates in

broadband anti-Stokes luminescence (in general occurring by thermal, multiphoton or Auger recombination-induced excitation of higher energy photons [41]) have been found [42]. Such solid-state quantum systems will be described in Chap. 5 in more detail.

3.6 Light Emission in Scanning Tunnelling Microscopy

3.6.1 Introduction

Light emission in electron tunnelling events had been reported about three decades ago for metal/oxide/metal (MOM) junctions that contained noble metals such as Au or Ag [43]. Appreciable light emission intensity was only observed for rough electrodes, which led to the argument that the emitted photons resulted from the radiative decay of surface plasmons. Plasmonic excitations result from the collective long-range Coulomb interaction between electrons and the corresponding resonance energy for a nearly free electron ensemble is given by

$$\omega_\mathrm{p} = \sqrt{\frac{e^2 n_\mathrm{el}}{\varepsilon_0 \varepsilon_\mathrm{b} m^*}} \tag{3.10}$$

for the volume plasmon mode. ω_p denotes the volume plasmon resonance frequency, n_el is the concentration of conduction band electrons, m^* their effective mass and ε_b is the dielectric function that describes the contribution of bound electrons. For high-enough excitation frequencies, the energy-dependent dielectric function is given by

$$\varepsilon(\omega) = \varepsilon_\mathrm{b}\left(1 - \frac{\omega_\mathrm{p}^2}{\omega^2}\right). \tag{3.11}$$

For the case that $\omega < \omega_\mathrm{p}$, $\varepsilon(\omega)$ is negative and the impinging electromagnetic radiation undergoes total reflection. If $\omega > \omega_\mathrm{p}$, the collective excitation of the electrons cannot follow the frequency of the radiation and the solid (metal) becomes transparent if interband transitions are not possible in the considered energy range. In the case of Ag, the volume plasmon resonance is located energetically at ∼3.8 eV and the first interband transition at the L-point of the Brillouin zone occurs at 4.1 eV, which makes it possible to observe the plasmon in near UV optical spectroscopy or by photoemission-into-electrolytes [44] directly and rather undisturbed. A typical result is shown in Fig. 3.13 for ultrathin Ag films prepared in situ by underpotential deposition [45] where films grow in a layer-by-layer mode (Franck–van der Merwe growth). Besides the occurrence of the Ag volume plasmon peak, an energetic shift of the resonance with film thickness is observed [46]. It can be seen that obviously, size quantization contributes to the shift of the volume plasmon resonance from the

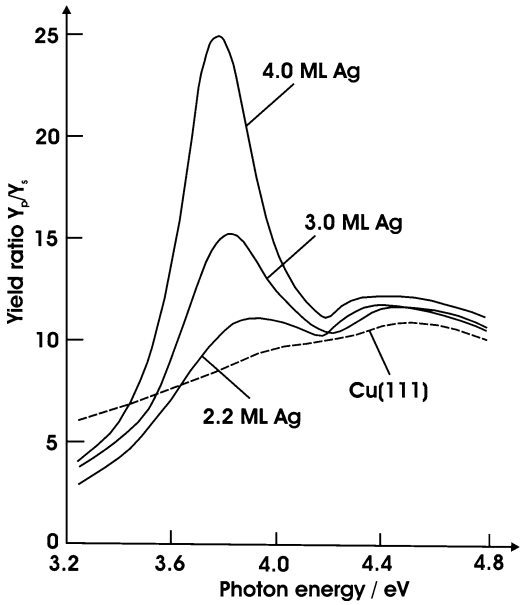


Fig. 3.13 Volume plasmon resonance of ultrathin Ag films grown on Cu(111) by underpotential deposition, measured by photoemission-into-electrolyte yield spectroscopy; Y_p photoemission yield for p-polarized light, Y_s yield for s-polarized light, electrolyte 1M H_2SO_4, blackened Pt reference electrode; work function at the electrolyte contact 3 eV, angle of incidence: 67.5°

very small coverage of 2.2 to 4 ML where the bulk value is practically reached. Because the transverse volume plasmons cannot be excited with s-polarized light, the yield ratio represents a very sensitive measure for the development of the plasmon resonance with thickness.

An overview of the various plasmonic excitation modes is given in Fig. 3.14 where the transverse and longitudinal plasmon modes for a free electron system, including thin films, are contrasted with the surface plasmon excitation and the so-called pair excitation regime. For a review, the reader is referred to the pioneering work of Raether [47]. Here, in the context of light emission in STM experiments, the main interest is on surface plasmon excitation and other highly local light emitting processes.

The volume plasmon dispersion which can be approximated for small values of the wave vector k by the expression

$$\omega = \omega_p \left(1 + \frac{3v_F^2}{10\omega_p^2} k^2\right) \tag{3.12}$$

allows excitation by photons. v_F in (3.12) is the electron Fermi velocity. For the case of Ag, an energy–wave vector diagram is shown in the above figure where the photon dispersion intersects the plasmon dispersion. If one uses the parameters of Ag and the approximation of a free electron gas, the volume plasmon resonance is energetically located at $\hbar\omega = 9$ eV, whereas the experimentally observed resonance is located at 3.78 eV. This pronounced difference can be attributed to the screening

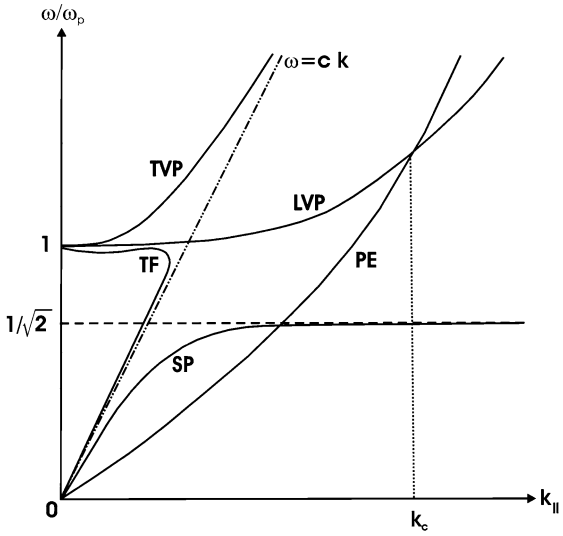


Fig. 3.14 Dispersion relations for photons and volume as well as surface plasmon excitations in a free electron gas; the photon dispersion is shown for grazing incidence. *TVP* transverse volume plasmons, *LVP* longitudinal volume plasmons, *SP* surface plasmon, *TF* volume plasmon for thin films where the coupling between front and back side drastically changes the dispersion relation of the transverse volume plasmons, *PE* denotes so-called pair excitations in the region on the *right-hand side* of the corresponding *curve* which defines a critical cut-off wave vector $\mathbf{k}_c$. Above this value, collective plasmon excitation should not be possible. Electron energy loss spectroscopy indicates that longitudinal oscillations are also possible in the region of electron–hole excitation. Therefore, the *curve* has been extended into that region

behaviour of the d-electrons of silver. In a simple mechanical analogy, one can assume that the screening results in a reduction of the Coulombic interaction forces, which results in a lower resonance frequency of the collective vibration.

For surface plasmon excitation, the dielectric constant of the interfacing medium, ε_W, has to be taken into account:

$$\omega_S = \frac{\omega_p}{\sqrt{1+\varepsilon_W}} \tag{3.13a}$$

which yields at the metal–vacuum interface the relation:

$$\omega_S = \frac{\omega_p}{\sqrt{2}}. \tag{3.13b}$$

At the Ag/aqueous electrolyte contact, the surface plasmon occurs at $\hbar\omega = 3.5\,\text{eV}$ for rough silver films. The dispersion relation of surface plasmons, presented in Fig. 3.14, shows that the photon dispersion and the surface plasmon do not cross, meaning that the momentum transfer from the photon is too small to excite surface

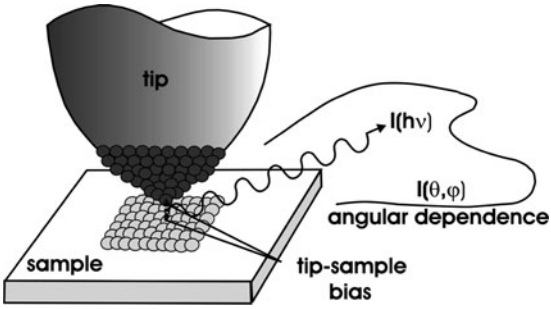


Fig. 3.15 Schematic of the experimental principle for photon emission in a scanning tunnelling experiment; $I(\theta, \varphi)$ denotes the angular intensity distribution; $I(h\nu)$ is the spectral characteristics of the light intensity (see text)

plasmons. The situation is different for the volume plasmon of silver, for instance, where the cross-over of the photon dispersion $E = hc\mathbf{k}$ (or $\omega = c\mathbf{k}$) with the volume plasmon allows excitation of the plasmon with p-polarized light [48]. For surface plasmons, an additional momentum source is needed which is usually provided by surface roughness that acts as source for the momentum component parallel to the surface, $\mathbf{k}_{//}$. Such coupling would show up in Fig. 3.14 as additional $\mathbf{k}_{//}$ vector distributions, depending on the roughness autocorrelation function [49]. Surface plasmon excitation at rough Ag films is in widespread use in sensor technology [50].

The inventor of the so-called topographiner, a precursor of the scanning tunnelling microscope, R. Young, proposed very early that light emission from his device should be possible [51]. A schematic of a typical experiment using a scanning tunnelling microscope is shown in Fig. 3.15, where the intensity distribution, the angular radiation characteristics and the influence of the applied voltage in the tunnelling experiment are sketched.

The first report of light emission in an STM experiment performed on rough polycystalline Ag films at the electrolyte contact showed unusually intense light emission of about 10^{-3} photons per tunnelling electron [52], whereas the photon yield was of the order of 10^{-5}–10^{-6} in MOM junctions. Therefore, the possibility to combine the analytical properties of optical spectroscopy with sub-nanometer spatial resolution was envisaged and rather soon after the inception of the method, single molecule analysis has been achieved [53]. In the section below, the initial work will be reviewed, followed by an overview on the existing theoretical concepts and subsequently, applications to nanoscale photonics on single organic molecules will be given.

3.6.2 Initial Work on Photon Emission in STM Experiments at Ag Films

A few years after the inception of STM as a method with atomic resolution, the first photon emission experiments have been performed using an STM apparatus. This work pioneered the area of emission of spectral radiation from a highly local

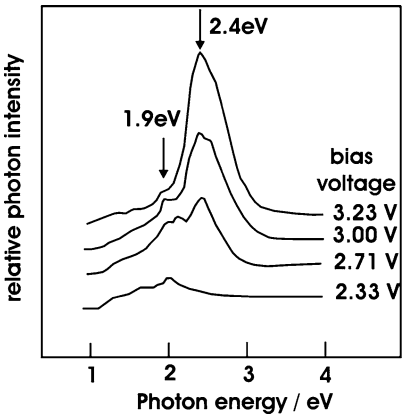


Fig. 3.16 Photon emission between 1 eV and 4 eV with two peaks located at 2.4 eV and 1.9 eV. Note that peak positions are virtually independent of the applied voltage, i.e. the energy of the tunnelling electrons

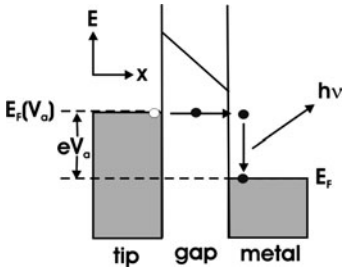


Fig. 3.17 Energy scheme for inelastic tunnelling between an STM tip and a metal sample with light emission in the course of electron relaxation to the metal Fermi level

probe and has been already cited above [52]. Because of the already at that time achieved near-atomic resolution of the spectroscopic information, the authors coined the acronym STOM (scanning tunnelling optical microscopy), which, however, was used rather seldom. Two modes of light emission have been discerned: (1) the isochromat mode where photons of constant energy are detected upon varying the electron energy via the applied bias to the tip (see Figs. 3.17 and 3.19) or (2) the fluorescence mode for which the spectral characteristic of the emitted light is determined for a single electron injection energy. In the following, only the latter will be treated as the quantum efficiency QE of the process in the range between 1 and 4 eV is much larger, with values for QE of the order of 10^{-3} and higher when an Ir tip was used over a condensed Ag film in UHV. Fluorescence spectra are shown in Fig. 3.16.

Below the quantum cut-off for $h\nu = eV_a$ (see Sect. 3.6.3 below), two distinct features at $h\nu \sim 1.9$ eV and at 2.4 eV are seen which, in their energetic position, do not depend on the tunnelling voltage, although the signal intensity varies considerably with the voltage. This invariance rules out that the features are related to the density of final states in the film. It could be shown that these peaks represent a resonance that involves more than one channel. It was tentatively proposed that these features result from the surface plasmon modes at these energies that mediate the enhanced photon emission around these two energies. This supposition was deduced from the existing knowledge at that time on local plasmon modes that have been confined by microstructures. Their resonance energy was located below that of free

surface plasmons and actually, for small Ag particles above an Al substrate, plasmon modes at 1.9 and 2.4 eV had been predicted [54]. The authors stated in addition that "local surface plasmon modes are efficient radiation emitters as opposed to surface plasmons", a statement that has now widely been recognized.

3.6.3 Theoretical Concepts

In STM light emission under the appropriate bias between tip and sample, electrons can be injected into the sample and the emitted photons are analysed with regard to their intensity and their spectral as well as angular characteristics, as schematically shown in Fig. 3.15. The method is characterized by some analogy with inverse photoemission and cathodoluminescence. The differences are the virtually atomic resolution of the STM method and the comparably low energy of the injected electrons. Whereas the energy of inverse photoemission is in the range of several tens of eV and is considerably larger in cathodoluminescence (keV), in STM light emission, this energy is restricted to the low eV range. This favours nondestructive measurements and provides the examination of vibrational excitations of molecules, as will be discussed below in Sect. 3.6.4.

The main photon emission channels are: (a) inelastic electron tunnelling and (b) luminescence (hot electron injection) and one distinguishes the light emitting processes at metal (STM-tip)—insulator (tunnel gap)–metal junctions and at metal–insulator–semiconductor or molecule junctions. In any light emitting STM system, the applied voltage V_a between tip and sample defines the maximum attainable photon energy due to energy conservation according to

$$h\nu \leq eV_a. \tag{3.14}$$

In limited cases, this inequality is violated. They will be discussed at the end of this section. The tunnelling process into a metal has been schematically drawn in Fig. 3.17.

Most of the electron excess energy due to the applied voltage is released by phonon scattering and heat and typically, only one in about 10^6–10^3 electrons releases its energy as photons. Models of light emission in the inelastic tunnelling mode of (planar) metal–insulator–metal structures have been based on the current fluctuations during tunnelling that occur on the time scale of optical periods. Light emission by direct radiation and by excitation of surface plasmons that decay into photons is possible [55, 56]. In the case of STM experiments, the macroscopic dielectric functions of tip and sample were used in the earlier theoretical descriptions, similar as in the case of planar structures. They also predicted radiation resulting exclusively from current fluctuations in the STM tunnelling gap [57]. In contrast to the planar junction, the translational symmetry is broken in an STM experiment and the normal modes of electromagnetic radiation that is excited in the

tip-sample gap are localized surface plasmons where the dominant part of the light emission stems from the subsequent decay of these surface plasmons.

3.6.4 STM-Induced Photon Emission on a Molecular Scale

Single molecule optical spectroscopy is important for ultrasensitive chemical identification, probing of local molecular environments and for biomolecule identification and dynamics, for instance. In addition, the conformational structure–property relationship is of particular importance in the field of molecular spectroscopy. In general, photon emission in STM from molecular adsorbates deposited on metals has shown that the emission is strongly quenched and that the broadening of the molecule's electronic levels by the presence of the metal makes detection and identification difficult [58]. Sub-molecular resolution could be observed by introduction of an ultrathin "spacer" between an organic molecule and the metallic substrate. This has been initially realized by the growth of a 0.5-nm-thick alumina film on a crystalline NiAl substrate. The oxide interphase results in a decoupling of the metal–molecule interaction, allowing for enhanced photon emission in a constant current STM experiment. It also turned out that the use of a Ag tip increased the photon emission intensity by more than an order of magnitude yielding quantum efficiencies of $\sim 10^{-4}$. Figure 3.18 shows the molecular structure scheme of the used ZnEtiol molecules. Because of the relatively flexible structure of the porphyrine macrocycle and the comparably inhomogeneous surface of alumina, the molecules experience non-uniform local conditions and deformations of the porphyrin macrocycle. They are known to result in pronounced changes in optical and vibrational properties, for instance. Constant current STM images of a series of molecules show differences in the apparent relative height of the four molecular lobes which are indicated and numbered by the toned ellipses in Fig. 3.18. It has been shown that these differences in the lobes reveal themselves in the emitted photon spectra [59].

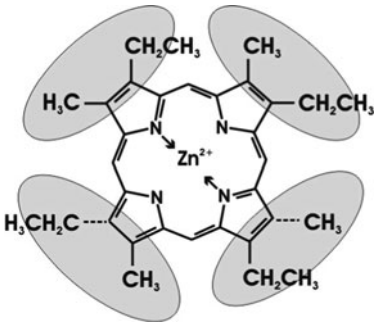


Fig. 3.18 Structure schematic of ZnEtiol with four lobes indicating the symmetry of the molecule

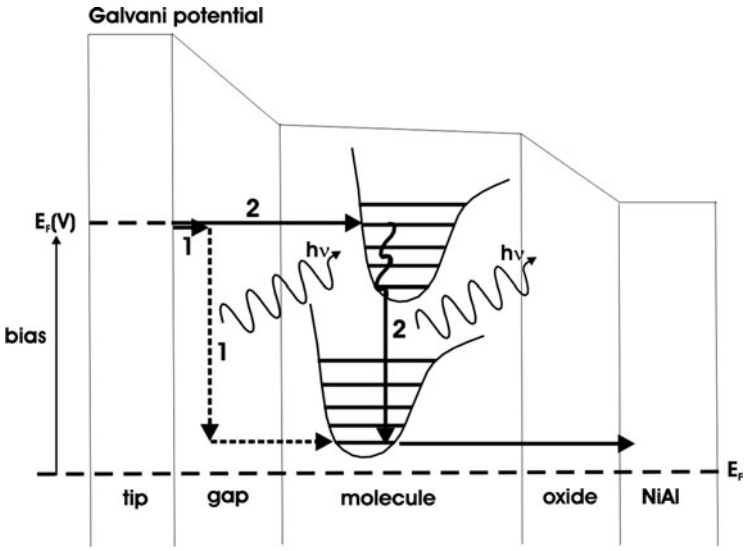


Fig. 3.19 Experimental condition and pathways for photon emission in STM tunnelling into a macromolecule; process 1: inelastic tunnelling mode; process 2: fluorescence mode; the Galvani potential is the electrostatic potential (vacuum level) and a certain potential drop across the molecule has also been indicated; layer thicknesses: gap: 0.6 nm; molecule: 0.2 nm; oxide: 0.5 nm (approximate values)

The principal possibilities for STM-induced photon emission from an adsorbed macromolecule are shown in Fig. 3.19 where two unoccupied electronic states with their vibronic splitting are displayed. Process 1 refers to inelastic tunnelling from the tip into the lower unoccupied state with simultaneous plasmon excitation. In process 2, elastic electron tunnelling into the higher molecular orbital occurs, resulting in an electronically and vibronically excited anionic molecular state. The radiative fluorescence transition to the lower state excites a plasmon. The plasmons subsequently decay into photons that are detected in the far field with their spectral characteristics (see Fig. 3.15).

Also shown in Fig. 3.19 is the process that the molecule can undergo vibrational relaxation (waved line in the upper electronic level) before the radiative transition (2) to the lower electronic state. In the final step, the excess electron tunnels through the aluminium oxide film into the NiAl substrate. The Galvani potential, indicated in the figure, represents the variation of the electrostatic potential across the junction. Besides the potential drops across the tunnelling gap and the oxide, a drop across the molecule is also assumed, whose actual value is uncertain as it relates to the local dielectric constant.

In Fig. 3.20, the spectral response collected with a Ag tip over one of the molecule's lobes is shown. First, one observes that submolecular resolution of spectral features is achieved. Second, the recorded spectrum is obtained from the overall structure, i.e. the light emission from the NiAl is contained in the spectrum

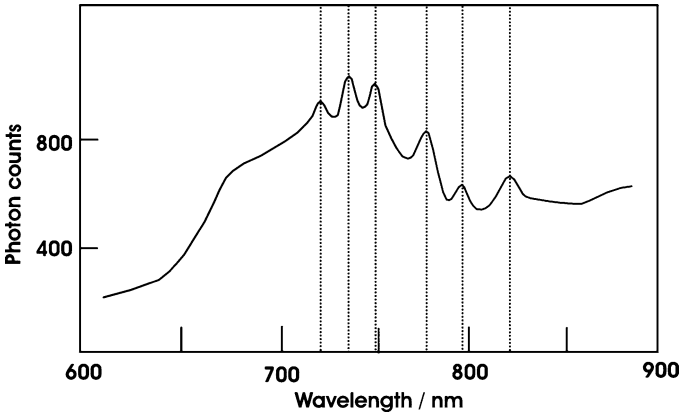


Fig. 3.20 Light emission spectrum from one of the lobes of the molecule shown in Fig. 3.18 taken with a Ag tip at a bias of 2.35 V and at a current of 500 pA for 300 s; the NiAl light emission spectral characteristic is an underlying feature but does not show the sequence of the sharp lines as indicated by the *vertical bars*; signals using a W tip are lower by a factor of ∼30

displayed. Since the NiAl substrate emission does not show the observed fine structure, indicated by the vertical position markers, these features are attributed to vibrational progression patterns and the process has been explained as a fluorescence event.

There exists, meanwhile, a considerable variety of structures that have shown photon emission in STM experiments where the tunnelling electrons provide a source for excess energy (hot) carriers. Besides the already-discussed single molecule vibrational spectroscopy, luminescence has been observed from metallic and semiconducting samples as well as from clusters and quantum well structures [60]. In the latter, a mechanism has been invoked that involves hot electron injection and interband transitions between two quantum wells, whereas typically, at bulk metallic samples, an inelastic tunnelling process excites a collective excitation, i.e. the coupled tip-sample plasmon mode that decays radiatively [24].

3.7 Metamaterials

3.7.1 Introduction

The traditional description of the interaction of electromagnetic radiation with matter, particularly in the optical frequency range, is based on averaging over the atomic details. This concept describes the medium, inhomogeneous on the atomic scale, by a homogeneous one where the electric and magnetic permittivities ε and μ, respectively, define the optical properties. Such interactions have been introduced

in Sect. 1.4, albeit with a different theoretical background. The actual nanoscopic structural organization of the sample does not necessarily have to originate from atoms or molecules; instead, any collection of nano-dimensioned objects that is inhomogeneous would result in a macroscopic response as described by ε and μ. With the advent of micro- and nanoscience [61, 62], possibilities exist to construct and fabricate samples (optical "materials") with desired properties. Based on a rather early theoretical treatise [63], the increased possibilities for realization of novel optical materials have resulted in a surge of experimental and theoretical work and various applications are envisaged or have already been realized [64]. Such applications encompass, for instance, resolution in optical imaging below the diffraction limit (superlens) [65], cloaking [66], optical nanolithography (near-field) [67], integrated communication systems and novel materials for telecommunication [68]. Below, the basic aspects and selected examples of metamaterials will be introduced.

Consider the classical dispersion relation between the frequency ω (energy) and the wave vector k of an electromagnetic wave for an isotropic medium:

$$k^2 = \frac{\omega^2}{c^2} n^2. \tag{3.15}$$

The square of the index of refraction, n^2, is given by $n^2 = \varepsilon\mu$. Without losses (absorption), n, ε and μ are real numbers. Although a simultaneous sign change of ε and μ would leave (3.15) unaffected, it turns out that substances with negative ε and μ exhibit properties that are strikingly different from those substances for which both permittivities are positive (classical optical media). In order to analyse these differences, the Maxwell equations for monochromatic *plane waves* are considered. They read

$$\vec{k} x \vec{E} = \frac{\omega}{c} \mu \vec{H}, \tag{3.16a}$$

$$\vec{k} x \vec{H} = -\frac{\omega}{c} \varepsilon \vec{E} \tag{3.16b}$$

for the electric and magnetic field $\vec{E}$ and $\vec{H}$, respectively. If $\varepsilon > 0$ and $\mu > 0$, then $\vec{E}$, $\vec{H}$ and $\vec{k}$ form a right-handed set of vectors. If, however, $\varepsilon < 0$ and $\mu < 0$, this set of vectors is left-handed. This has consequences for the propagation of electromagnetic waves in such medium: the energy flux is given by the Poynting vector $\vec{S}$ with

$$\vec{S} = \frac{c}{4\pi} \vec{E} x \vec{H}. \tag{3.17}$$

Because the Poynting vector forms a right-handed set with the vectors for the electric and the magnetic field, the wave propagation vector $\vec{k}$ and the Poynting vector $\vec{S}$ are collinear. The opposite is true for the left-handed materials which have a resulting negative group velocity moving towards the light source when light is

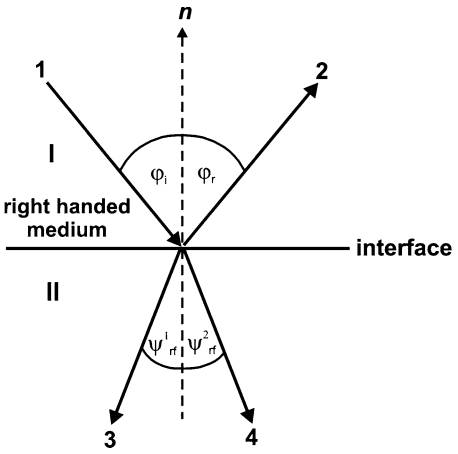


Fig. 3.21 Schematic for the passage of light through two adjacent media for the case that both media are right-handed (rays 1, 2 and 4) and medium II has a larger index of refraction and for refraction upon entering a left-handed medium II; φ_1, φ_r, impinging and reflected beam, respectively; Ψ^1_{rf}, Ψ^2_{rf}, refracted rays for a left- and right-handed medium, respectively; the interface is located in the *xy*-plane, the surface normal *n* defines the *z*-direction

impinging on such a substance, which leads to what is called a "reversed Doppler effect" [69] (e.g. an increase in frequency when the source moves away from the observer) and an reverse Cherenkov effect [70].

The refractory properties of left-handed materials can be described by a modification of Snell's law (for geometry, see Fig. 3.21):

$$\frac{\sin\varphi_1}{\sin\varphi_2} = \frac{p_2}{p_1}\left|\sqrt{\frac{\varepsilon_2\mu_2}{\varepsilon_1\mu_1}}\right|. \tag{3.18}$$

In (3.18), p_1 and p_2 describe the symmetry of the two media with regard to their rightnesses. If the sign of p_1 and p_2 is different, right- and left-handed media are adjacent and the index of refraction of the left-handed medium is negative relative to vacuum. Whereas the boundary conditions for the *xy*-direction remain unaltered independent of the handedness of the medium, the *z*-component changes sign. In a typical refraction experiment, the reflected beam is identical for, both, the right- and left-handed medium. The *refracted* beam, however, is located on the opposite side of the *z*-axis (perpendicular to the geometrical surface) than for the "classical" case of refraction in right-handed materials (Fig. 3.21).

3.7.2 Superlenses

In a conventional lens, the imaging is achieved by increased refraction for the outer parts of the lens compared to the central part on the axis. This classical action of the lens is often approximated by the passage of light rays through a prism, for example, for the outer part which is further away from the optical axis of a convex lens. In physical terms, the lens acts by providing a phase correction for each Fourier component of the Fourier expansion of the object to be imaged. Then, on the other side of the lens, the image is generated by reassembling the field components. Thus,

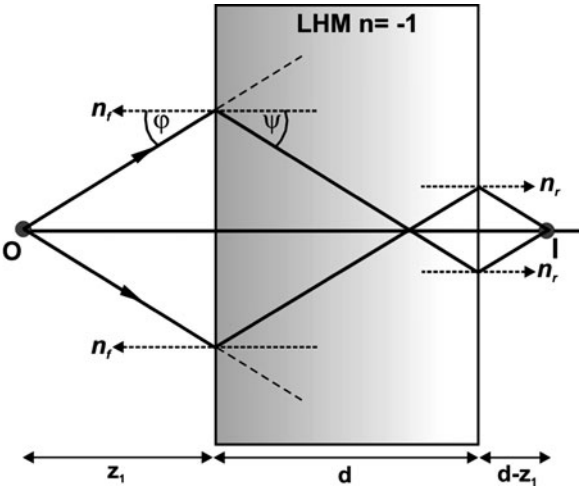


Fig. 3.22 Imaging properties and light focussing of a left-handed negative index material; LHM: left-handed material; $\boldsymbol{n}_{\mathrm{f}}$, vector normal to the front surface; $\boldsymbol{n}_{\mathrm{r}}$, vector normal to the rear surface; *dashed lines* indicate conventional refraction (which does not occur for $n = 1$); z_1, distance between object O and the front surface of the slab; d, thickness of the slab; I: second image, obtained outside the slab; the angles φ and ψ have the same meaning as in Fig. 3.21 (see also text)

the focussing of electromagnetic waves using media with positive refractive index requires curved surfaces. The resolution, achievable with such lenses, is limited by the diffraction limit because the finer details of the image result from high-k (wave vector) components of the Fourier series that reconstructs the image. For smaller wave vectors with $\omega^2 c^{-2} > k_x^2 + k_y^2$, the relation $k_z = +\sqrt{\omega^2 c^{-2} - k_x^2 - k_y^2}$ holds (the z-axis is defined as collinear with the axis of the lens). High k vectors determine the details of the image and, because of $\omega^2 c^{-2} < k_x^2 + k_y^2$, exponential damping occurs, resulting in evanescent waves where the wave vector is now given by $k_z = +\mathrm{i}\sqrt{k_x^2 + k_y^2 - \omega^2 c^{-2}}$. These waves decay effectively before reaching the image plane. In conventional lenses, contributions of these evanescent waves are absent and only objects that have sizes larger than about half the wavelength of the radiation can be focussed.

In negative index materials, a different focussing behaviour prevails. Starting from Fig. 3.21, one can construct the ray propagation through, for instance, a slab with parallel sides. As Fig. 3.22 shows, the left-handed material with $n = -1$ and $\mu = -1$ bends light towards the opposite direction than in conventional materials. This results in a focussing effect using a slab without any curvature. Actually, a double focussing effect is noted. The second image occurs at a distance given by $d - z_1$, where z_1 is the distance between object and slab and d denotes the thickness of the slab (see Fig. 3.22). It has been shown that the medium in Fig. 3.22 exerts the property on the electromagnetic waves to amplify the evanescent waves (instead of cancelling in conventional media) and accordingly, both, the propagating and the evanescent waves contribute to the resolution of the image. This allows near perfect reconstruction of the object into an image limited only by apertures and perfection of the slab (lens) surface. Figure 3.23 shows measured data for a 2D version of a negative index material where sub-wavelength focussing has been achieved [71]. It is seen that the image resolution is well belowthe diffraction limit.

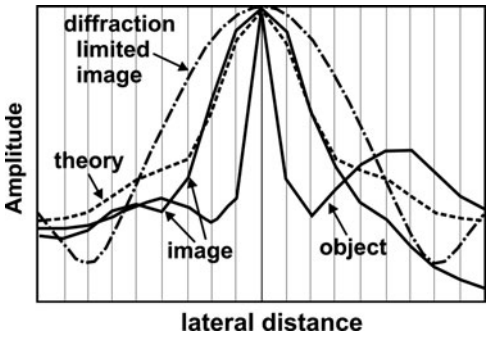


Fig. 3.23 Experimental result on imaging with left-handed negative index materials [71]. The signal amplitude is plotted vs. the lateral distance across the image plane; the object is reproduced with an accuracy below the diffraction limit

Negative refraction has been demonstrated at microwave frequencies [72] and, more recently, attempts have been made to move towards optical frequencies because of the possibility to attain imaging of visible objects below the diffraction limit. A promising approach uses the well-known porous alumina templates where the pore sizes can be controlled between 10 and 200 nm and the aspect ratio is high. Subsequent synthesis of Au nanowires in these pores leads to a regular array of well-defined Au nanowires that are perpendicular to the surface plane of alumina and parallel to each other [73]. Such a medium is optically strongly anisotropic and is also characterized by two surface plasmon resonances, a longitudinal and a transverse mode. It has been shown that for wavelengths larger than that of the longitudinal mode, the Au nanowire/porous alumina system represents a negative refractory index metamaterial in the frequency range that covers the visible spectrum as well as spectral regions on both sides beyond the visible energy range. It can, thus, be used for superlens imaging.

The correlation of the superlens imaging with surface plasmon modes originates from the condition for surface plasmon excitation (see also (3.13b)) $\varepsilon = -1$. This concept has been used to demonstrate focussing action of a so-called quasi-electrostatic field (in an approach without placing conditions on the magnetic susceptibility μ) via surface plasmon excitation [65]. The imaging properties of the device which consists of a slab of Ag and that focuses light which is in resonance with the surface plasmon frequency of Ag are shown in Fig. 3.24.

The resolution of the image in the lower part is limited by the imaginary part of the dielectric function of Ag, i.e. the absorption properties of the material. The dielectric function that has been used was $\varepsilon \sim 5.7 - 81/\omega^2 + 0.4\mathrm{i}$ for focussing at $h\nu = 3.48\,\mathrm{eV}$, the surface plasmon resonance, where the imaginary part prevents ideal reconstruction.

3.7.3 Invisibility Cloaks

Towards the end of this chapter, issues are addressed that occupied the fantasy of humans throughout history. Among them is the dream of invisibility which, although potentially dangerous, becomes increasingly realistic, as will be shown

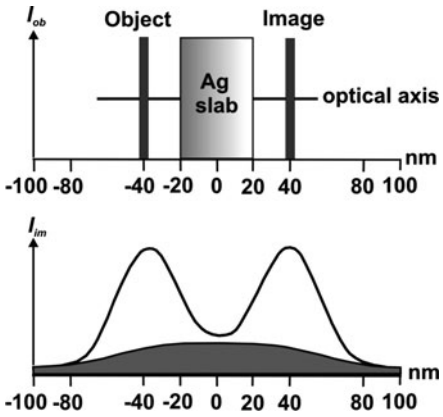


Fig. 3.24 Superlens operation based on resonant excitation of Ag surface plasmons using a Ag slab: *top*: object intensity distribution I_{ob} along the optical axis; *bottom*: image intensity distribution, where the *grey area* denotes imaging without the Ag slab (amplification of evanescent waves); the *full line* shows imaging well below the diffraction limit

in this section, and, for example, the observation of so-called mysterious lights (see Sect. 3.8). A straightforward connection between metamaterials and cloaking becomes apparent when inspecting the behaviour of a cylindrical lens with a shell and the optical properties $\varepsilon_{\text{core}} = 1$ and $\varepsilon_{\text{shell}} = -1$ (plus a small imaginary part). In the so-called quasistatic approximation [74], if a collection of point dipoles is positioned close to the lens, it, as well as the lens, becomes essentially invisible to an impinging uniform electric field [75]. Re-inspection of Fig. 3.24 shows that with negative index materials, imaging as well as cloaking should be possible. Cloaking occurs if the distance between object and lens (slab) is larger than half of the slab width; accordingly, the figure based on the original work of Pendry [65] depicts a situation at the border between cloaking and imaging. More generally, metamaterials with sub-wavelength details of structure can be constructed such that the refractive index is controlled at every point in the material. With specific internal designs, such material would allow impinging electromagnetic radiation to meander and bend a path around an object.

The general principle of cloakingis based on the form invariance of the Maxwell equations. A coordinate transformation which would, for example, squeeze a three-dimensional volume like a sphere into a shell that surrounds the volume that should be concealed, will only affect the components of the permittivity tensor $\vec{\varepsilon}$ and the permeability tensor $\vec{\mu}$ which become both, anisotropic and spatially varying. If one can realize such complex material properties in a sub-wavelength regime, the concealed volume and the surrounding shell (cloak) have the properties of free space when viewed from the outside. The cloakthen does not scatter incoming waves and also does not produce a shadow in the transmitted beam, which would result in some visibility. A typical design consists of compression of a cylindrical region of extension $0 < r < b$ into an annular region with $a < r' < b$ (see Fig. 3.25).

A transformation that accomplishes the desired change is given by (see Fig. 3.25)

$$r' = \frac{b-a}{b} r + a; \quad \vartheta' = \vartheta; \quad z' = z \tag{3.19a}$$

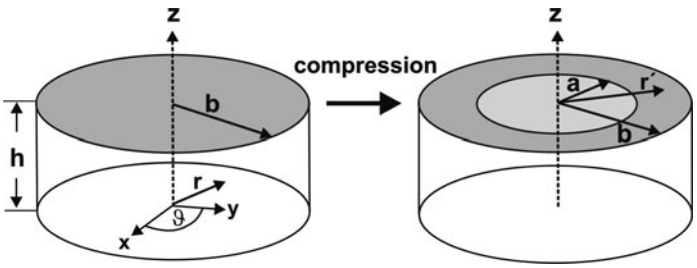


Fig. 3.25 Geometrical considerations for coordinate transformation with regard to cloaking; note the compression of the volume from $2\pi b^2 h$ to $2\pi(b^2–a^2)h$; note that there is no variation along the z-direction

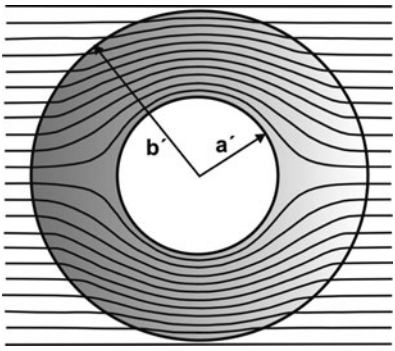


Fig. 3.26 Schematic of ray trajectories in a cloak (*grey*) surrounding a sphere (*white*) to be concealed in a 2D cross section; the rays are collinear with the Poynting vector (3.17)

The transformation results in spatially varying ε and μ and the new values are given by

$$\varepsilon_{r'} = \mu_{r'} = \frac{r'-a}{r'} \quad \varepsilon_{\vartheta} = \mu_{\vartheta} = \frac{r'}{r'-a} \quad \varepsilon_z = \mu_z = \left(\frac{b}{b-a}\right)^2 \frac{r'-a}{r'}. \tag{3.19b}$$

All the components of the permittivity and permeability tensors change as a function of radius posing considerable demands on the design of a suited metamaterial. In a simplified version, with the electric field in the direction of the z-axis, one can show that only $\mu_{r'}$ varies radially across the structure. The above cylindrical system has been described because it has facilitated experimental implementation. In Fig. 3.26, resulting ray trajectories are shown for a spherical object with its surrounding cloak. For the calculation, $b' \gg$ l, the probing wavelength, has been assumed [66].

It is observed that the rays are diverted within the cloak (grey area between two spheres of radius a' and b') and emerge on the far side of the system without deflection from their original parallel arrangement. The electromagnetic radiation passes around the cloak in a manner as if it would have travelled through empty space and, because no radiation can enter the concealed space, an observer is led

to the conclusion that the concealed volume is empty. Based on the above model considerations, a plethora of metamaterials can be developed and designed for desired optical/electromagnetic properties. Naturally, it is of considerable interest to develop cloaking systems which are active in a broad spectral range, covering the spectral region from the microwave region up to the visible/UV [67, 68]

3.8 "Mysterious Lights"

This section treats predominantly light emission processes in nature that have been observed over centuries and that were the source of many speculations on their origin. Due to the variation of observation conditions, most observations are treated separately [76, 77]. The sometimes mysterious character of these photon emission processes is reflected by the names given to the respective "glow", lightning or "fire". As mankind has lived with these mysteries, some of them are shortly reviewed here, including explanations of the nature of the respective light emission mechanism. We will see that still nowadays, interpretations and explanations vary considerably even in the scientific community. Therefore, only a few examples with their presumed origins are presented here and their hypothetical character is emphasized when necessary. In addition, related to its still rather obscure origin and generation mechanism, a short note is included in this section on what has been called "biophotons" (ultraweak photon emission attributed to biochemical reactions). The considerably more intense bioluminescence has already been described above in Sect. 3.5.

3.8.1 Fox Fire

This light is attributed to the behaviour of rotten vegetation that sometimes burst into flames and is also called *swamp gas*, *spook-* or *ghostlights* they refer to lights that are sometimes seen at night or twilight over bogs. The lights have a flickering appearance and tales of their receding upon approaching have further increased human fantasy over their origin and meaning [78]. Among possible explanations, chemiluminescence via oxidation of phosphine and methane produced by the decay of organic material, bioluminescence of fungi (honey fungus) and atmospheric discharges induced by tectonic strain have been proposed. The latter is also known as earthquake lights which are believed to result from piezoelectricity induced by the strain acting on piezoelectric components in the soil (such as quartz) that results from tectonic movement. The electricity, generated by shifting and grinding of rocks within the Earth's crust, is believed to be channelled upwards in columns of vaporized water to the surface where it leads to atmospheric charges that produce photons. The formation of methane is exemplarically described by the decomposition reaction of organic molecules:

$$C_6H_{12}O_6 \rightarrow 3CH_4 + 3CO_2. \tag{3.20}$$

Depending on the available amount of oxygen, various reactions are possible. For sufficiently large oxygen supply, the reaction

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \quad (\Delta H^0 = -891\,\mathrm{kJmol}^{-1}) \tag{3.21}$$

can take place. In the overall reaction that includes reaction intermediates, the chemiluminescent species is an excited CO_2 species. The spectral features are rather consistent with the description of the appearance of the light, noted often as a bluish cold flame. Another possibility involves the burning of phosphanes (PH_3, P_2H_2) that have been formed in anaerobic sediments. The confirmation that phosphane is produced in the biosphere could explain the ignition of CH_4 by autoignition of diphosphane (P_2H_2) offering another plausible mechanism for the lights occurring over swamp-like grounds or, for instance, cemeteries [79].

3.8.2 Earthquake Lights

Prior to large earthquakes, the earth emits transient signals, sometimes strong, but more often, subtle and fleeting. The signals may consist of local magnetic field variations, electromagnetic emissions over a wide range of frequencies and of a variety of atmospheric and ionospheric phenomena. Often, electromagnetic emission in the form of visible light that accompanies or precedes earthquakes has (until rather recently) been considered folklore. For centuries, reports have been collected about flashes, glows and various fireworks that where temporarily connected with large earthquakes. Only after such lights have been recorded via photography (originally in Japan during a month-long earthquake in the mid-1960s) and after other careful observations, scientific examination of its origin became possible. A characteristic feature of these lights is their occurrence before, during and after earthquakes.

Observed colours are white, blue or red and the light can be as bright as moonlight. A variety of other visible light-related occurrences has been reported, among them fire balls and fire columns, luminous funnels, clouds and vapours, flashes, electrical discharges and sparks [80]. Their form shows a large variety, including bands, rays and sheets, for example. It was also found that the lights were accompanied by radio noise in the frequency range from 10 to 20 kHz. The lights have been observed sometimes weeks before the actual earthquake at a distance up to hundreds of kilometres away from the epicenter. Lights were observed in connection to the Kobe earthquake (1995) [81] and, recently with that in Aquila and on many other occasions [82, 83].

Besides photon emission, nonseismic events include diverse observations, such as ground-hugging fog, low-frequency electromagnetic radiation, local anomalies of the earth magnetic field of up to 0.5%, temperature anomalies in the range of

several degrees and changes in the plasma density of the ionosphere. In addition, strange animal behaviour and human premonition of impending earthquakes have been noted on several occasions [84]. Many of the phenomena listed above require the flow of electric currents in the ground. When, for instance, local magnetic field anomalies occur in the time period before a seismic event, electric currents of considerable magnitude are required which did not flow earlier. The origin of the electric currents has been vaguely attributed to the "earthquake preparation process" but the actual mechanism has remained elusive. If electromagnetic radiation occurs in or around the area where an earthquake is to occur, currents that vary with time are required. Therefore, models have been developed that address the generation of electrical currents via mechanical processes in the earth's crust. The more prominent models are based on piezoelectricity, streaming potentials and on so-called p-hole (surface) conductivity of rocks [85] and will be treated below.

The origin of piezoelectricity is based on the mechanic behaviour of rocks in the upper part of the earth's crust where, at moderate temperature, rocks are brittle. Pushing a rock against another one, deformation will set in. According to the mechanics of stress-induced strain, the regions of deformation will cover the elastic regime followed by plastic deformation (generation and movement of dislocations) and, finally, under the action of tectonic forces that maintain pushing, failure occurs by massive cracking of blocks of rocks (compare Fig. 3.27).

The onset of plastic deformation defines generation of dislocations which begin to move under the exerted force. This region is characterized by plastic flow (Fig. 3.27) and it moves over to a regime where coalescence of dislocations, resulting in microcracks, occurs. These microcracks grow into larger cracks that coalesce into fissures and, as a consequence, the volume of the rock increases. This is what is called "buckling" of the earth's surface, observed by geodesic methods [86], indicating that rocks deep under the surface of the earth enter the phase of volume expansion. With continuing tectonic pressure, the brittle system

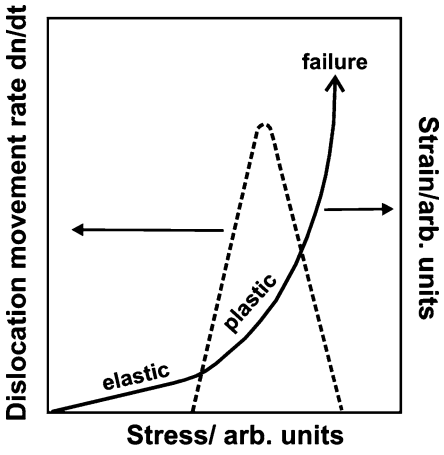


Fig. 3.27 Schematic stress–strain relation for a brittle material such as rocks in seismic events indicating the material changes (*full line-ordinate on the right*) and schematic rate of dislocation movement (*dotted line-left*)

becomes increasingly unstable, indicated by the increasing slope of the strain curve in Fig. 3.27, evolving into criticality and the possibility o sudden failure.

Two possible mechanisms for the origin of electric charge have been discussed in the related literature (1) piezoelectricity and (2) so-called streaming potentials (see below). The former is a fundamental property of minerals that have a non-centrosymmetric symmetry, such as, quartz crystals which belong to the naturally occurring rock-forming minerals. Modern examples are the III–V nitrides, used for LEDs (GaN, used for the blue diode [87, 88]). Exerting forces, in particular, crystal directions on quartz, leads to small atom displacements along polar axes that generate voltages on opposite surfaces. Granite, for instance, contains a high amount of quartz. Pressurizing by seismic movement will lead to piezoelectric voltages but, because of the random orientation of the crystals and crystallites, the resulting piezoelectric voltage will be very small and typically be cancelled out. Streaming potentials [89] arise from the movement of fluids through narrow openings under considerable external force. Some electric charges will absorb on the capillary walls and the fluid will retain the charges of opposite sign. During fluid motion, charge separation can generate high voltages that cause sparks in insulating liquids. In the natural system of porous and/or stressed rocks, the voltages that can be built up are limited by the ionic conductivity of the water. In the presence of the salts that can be dissolved in water, the maximum achievable voltages are estimated to reach only the range of several millivolts and the corresponding electrical fields are too small for generating electrical currents of significant magnitude.

A possible explanation can be given by the more recent p-hole model, mentioned above, which uses solid-state physical concepts to explain the generation of larger fluctuating electrical currents in the ground that evoke electromagnetic radiation. This model has been inferred from earlier observations on MgO, prepared by thermal decomposition of $Mg(OH)_2$, where it was found that H_2 was evolved in the process. It was expected, however, that hydroxyl ions OH^- would react to split off water according to $2OH^- \rightarrow O_2^- + H_2O$. The reaction towards hydrogen evolution meant that electrons were needed to reduce water or protons ($2H^+ + 2e^- \rightarrow H_2$). Therefore, the more likely reaction scheme was based on the formation of peroxy bonds: $2OH^- \rightarrow O_2^{2-} + H_2$. This reaction, where OH^- bonds are converted to peroxy bonds at temperatures of about 400–500°C, is not limited to MgO. It takes place, for instance, in silicate minerals that have crystallized in a moist or aqueous environment. In quartz, the $O^- - O^-$ bond that occurs in MgO would be replaced by a so-called peroxy link: $O_3Si - O - O - SiO_3$ [90]. It has then been assumed that the O^- sites in an O^{2-} matrix represent a defect electron or hole (in the terminology of semiconductor physics) and that the corresponding energy state is located within the valence band. This assumption needs more substantiation. It is also possible that the peroxy bridge-related states are located in the energy gap acting as deep acceptors that can be thermally or vibrationally activated. The p-hole model is based on the consideration of the intact $O^- - O^-$ bond as the source of a "positive hole pair", which is self-trapped as long as the bonds are not broken.

A further argument made concerns the redistribution of the positive charge of activated bonds (by a seismic event, for example) to form a thin charged surface

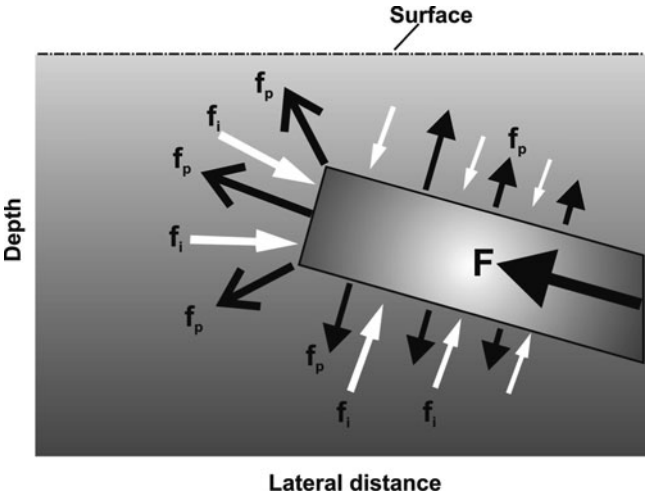


Fig. 3.28 Schematic of the reaction of a rock (quartz crystallite, symbolized by the inclined slab) to an exerted force F; f_p: outflow of positive charge (p-holes); f_i: influx of compensating charges; surface denotes the earth's surface

layer due to differences in the static dielectric constants of the material and the ambient. Apparently, the positive hole pairs can be activated and then released by acoustic waves, indicating that this possibility also exists in earthquake and pre-earthquake phases. The following three claims have been made regarding the p-hole model (1) in stressed rocks and minerals (silicates), positive hole pairs can be released by activation, resulting in positive charge in the valence band of otherwise insulating material; (2) the charge can spread to the surface and surrounding rocks; (3) charge "clouds" can arrive at the earth's surface. It is further argued that process (2) can lead to current pulses in the earth's crust, due to the flow of positive carriers and the built up electrical field that counteracts the flow leading to polarization (surrounding rocks more positively charged, source more negatively charged). As a result, a reverse flow of negative charges towards the p-hole source will take place that reduces the polarization field. This is indicated in the schematic of Fig. 3.28.

When charge clouds associated with p-holes reach the surface, high electric fields $\boldsymbol{E}$ can be built up at the surface–air interface, which are particularly high at convex topographical points of the earth's surface because $\boldsymbol{E} \propto V/r_c$ (V, potential difference, r_c, curvature radius). Candidates for field enhancements are hill tops, ridges and mountain peaks. As a result, an atmospheric layer near the surface will become ionized, leading to a simultaneous conductivity increase. As the electric fields near concave locations will be lower, substantial lateral field gradients exist that are supposed to lead to corona discharges and flashes, for instance. The p-hole model, thus, predicts that ionized, electronically excited air molecules emit light which is regarded as a possible cause of various luminous phenomena which have been labelled earthquake lights. Another phenomenon, the ball lighting that

is observed to precede strong thunderstorms and earthquakes, is attributed within the framework of the p-hole model to what is called “free-floating volumes of ionized air” that have detached from the ground when clouds of p-holes reach the surface, resulting in high electric fields. Several other phenomena can obviously be explained based on this model, among them the high-frequency noise that has been noted prior to earthquakes by radio operators; this static is attributed to high-frequency electromagnetic emissions that accompany the discharges mentioned above.

It should be noted, however, that, although the p-hole model represents the first general concept that explains a variety of hitherto apparently unrelated phenomena, the model has not yet been fully accepted by the geo-science community. Basically, this appears to be due to the relatively little solid-state physics information with regard to the effect of peroxy bonds on the conductivity of minerals. The claim that the according energy levels are located within the valence band of, for instance, quartz crystallites needs further corroboration, possibly by solid state theory. Another aspect is the propagation of the positive charge to the surface of a rock. One might invoke other concepts for the generation of surface charge. As an example, an ultrahigh conductivity of ultrathin water films has been observed in STM experiments [91]. The introduction of the p-hole model has, on the other hand, led to a microscopic consideration of the origin of earthquake and pre-earthquake related phenomena and, thus, paved the way for further microscopic theories.

3.8.3 St. Elmo’s Fire

The name for this phenomenon stems from St. Erasmus who was an early Christian martyr and the patron saint of sailors [92]. In rough weather, seamen attributed the blue glow around the mast tops as a sign of protection by the saint. St. Elmo’s Fire is known as a type of continuous electric spark due to a glow discharge. It is also similar to the glows found inside fluorescent tubes, mercury vapour streetlights or old orange-display calculators, for instance.

Physical descriptions of St. Elmo’s fire have ranged from a ghostly dancing flame to natural fireworks. It is usually being described as a blue or bluish-white colour attached to fixed, grounded conductors and has a lifetime of minutes. The flame is heatless and nonconsuming, occasionally accompanied by a hissing sound. These latter properties promote the myths of spiritual presence.

St. Elmo’s Fire and normal sparks can appear when a high electrical voltage affects a gas. It therefore can be seen during thunderstorms when the ground below the storm is electrically charged and when a high voltage between clouds and the earth surface exists. The ignition of the process needs about $30\,\mathrm{kV\,cm^{-1}}$, but at sharp points with a small curvature radius, triggering appears at lower voltages (see below). The electric field at tips is given by $\vec{E} = V/r_\mathrm{c}$, where r_c denotes the curvature radius and V the existing voltage. The glow of the discharge results from

plasma formation. The plasma, a partially ionized gas, consists of spatially separated positive and negative charges and non-ionized gas particles. Globally viewed, it is neutral. Inelastic electron collisions with the atmospheric molecules result in excitation and ionization (complete removal of an electron) of gas molecules. The glow is a result of the de-excitation of collision-excited gas molecules A. The initial excitation step is described by $e^-(E_{\mathrm{Kin}}) + \mathrm{A} \rightarrow e^-(E_{\mathrm{Kin}} - E_{\mathrm{inel}}) + \mathrm{A}^*$ where E_{Kin} and E_{inel} are the kinetic energy and the inelastic energy loss of the interacting electron e^-. With $\mathrm{A}^* = \mathrm{A} + E_{\mathrm{inel}}$, the excited molecule A^* decays according to $\mathrm{A}^* \rightarrow \mathrm{A} + h\nu$, and the emitted photons produce the observed glow. The colour of the glow depends on the type of gas involved. Our atmosphere consists predominantly of nitrogen and oxygen, and this mixture shows a blue/violet glow when exposed to high voltage fields. Air is a good electric insulator. However, if a notable electric field exists at an end of a conducting rod (see above on the increase of the electrostatic field), air molecules in its vicinity can be ionized and discharges are intensified at the end of pointed metal rods.

St. Elmo's fire is also known as corona discharge – i.e. a discharge of electricity that ionizes the air surrounding a sharp object and, in the course of plasma formation as described above, causes the air to glow. Because static electricity in the atmosphere can induce charge in the considered object, current can flow between the object and the atmosphere, thereby ionizing the molecules of gas in the air, forming a plasma, excited by electron collisions, at voltages as low as $200\,\mathrm{V\,cm^{-1}}$. The voltage developed is much lower than that necessary for lightnings, and St. Elmo's fire usually causes no damage.

With the age of flight, the fire has appeared along the wing tips, propellers, and antennae of aircraft, often disrupting radio communications. There is even a theory that the Hindenburg Zeppelin disaster may have been sparked by St. Elmo's fire igniting leaking hydrogen. Figure 3.29 shows a spectrum of St. Elmo's fire observed in a NASA mission on board of a Westwind 2 jet aircraft upon flying through a cloud at 13.83-km altitude. The spectrum was recorded using an at that time highly sensitive monochromatic and television system detectors aboard the aircraft, observed on the night of February 26, 1995 over Peru [93].

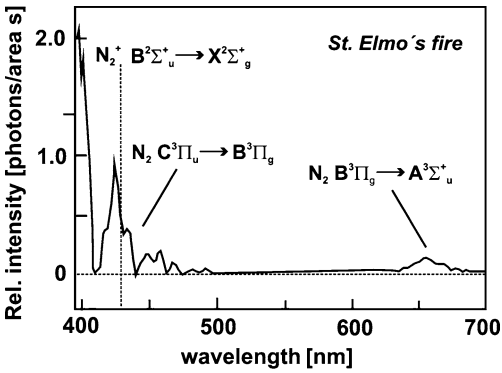


Fig. 3.29 Spectrum of St. Elmo's fire recorded from the front of paraboloidal wingtip fuel pods and from the top-front and rear corners of the vertical fins on the pods in the upper troposphere (after ref. [93])

The spectrum shows some analogy with the electronic/molecular transitions in molecular nitrogen but, probably related to the considerably different experimental conditions in the upper troposphere, frequency shifts are seen, particularly for the ionized N_2 molecule.

3.8.4 Novaya–Zemlya Effect

The effect has been first observed by G. de Veer in the sixteenth century (in the year 1596) on the Russian arctic island between the Barents and the Kara Sea, after which the effect has been named. The effect is observed in arctic regions and related to the motion stability of the extremely cold air. The stillness of the air results in very smooth and even boundaries between different atmospheric layers that vary due to temperature, for example. The main observation is that the sun appears to rise considerably earlier than it actually should be seen and, in addition, the shape of the sun can appear as a rectangle instead of a circle. This is explained by the projection of the sun's image onto the atmospheric layers at a time when the sun is still several degrees below the horizon.

The phenomenon has been described using so-called *Ray-Tracing* calculations [94]. The Novaya–Zemlya effect represents an example of what one calls a mirage and is caused by a strong temperature inversion in the atmosphere. The optical calculation of the effect necessitates knowledge of the change of the refractory index of the atmosphere with temperature. Assuming that the refractive index is known for the relevant heights in the atmosphere, one attempts to trace the light backwards from the observer's eye towards the imaged object. Thus, the curvature of the path of the light from object to observer is determined. The curvature of a considered ray is defined as the inverse of its radius and, taking into account the surface curvature of the Earth, one writes $c_{\mathrm{Te}} = 1/r_{\mathrm{A}} + 1/R_{\mathrm{E}}$, where R_{E} denotes the radius of the earth. Under usual atmospheric conditions, the temperature gradient is such that T decreases with height in the range of $-6\,\mathrm{m°C\,m^{-1}}$. Then, the light curvature $1/r_{\mathrm{A}}$ is about 6 times larger than that of the Earth. Backward tracing of such a ray, shown in Fig. 3.30, (assumed to be horizontal at the position of the observer) yields a downward bending of the light ray that is considerably smaller than the curvature of the Earth's surface and the ray "escapes". For such a behaviour, c_{Te} is positive and the ray appears to be bent away from the Earth.

For a strong temperature inversion, with vertical temperature gradients in the range of $+0.11°\mathrm{C}$, c_{Te} becomes negative within such layer and near horizontal rays are bent (back) towards the Earth's surface. Ray tracing is possible if the dependence of the refractive index, n, on height h and position x is known; $n(h, x)$ is obtained from the temperature profile, which itself depends on the temperature of the central isotherm within the profile (see Fig. 3.31a), T_{CI}, the height where the central isotherm is positioned above ground, $h_{\mathrm{CI}}(x)$, the total temperature difference across the inversion region, $\Delta T(x)$ and a parameter that describes the steepness of the inversion, $a(x)$. The resulting multiple total reflection at the temperature

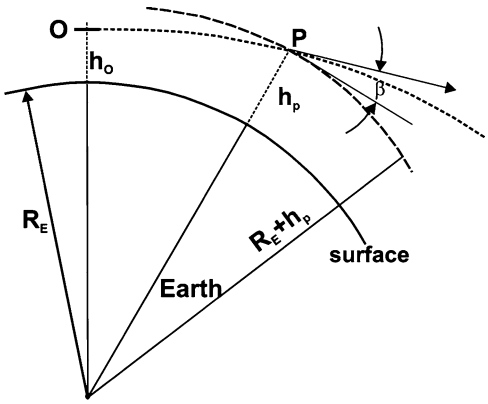


Fig. 3.30 Geometry for description of ray tracing in atmospheric phenomena for positive c_{Te}; O observer, h_O height of observer above the earth's surface, h_p actual height of ray (*dotted line*) at point P, β angle of the light ray with the local horizon at point P (see text)

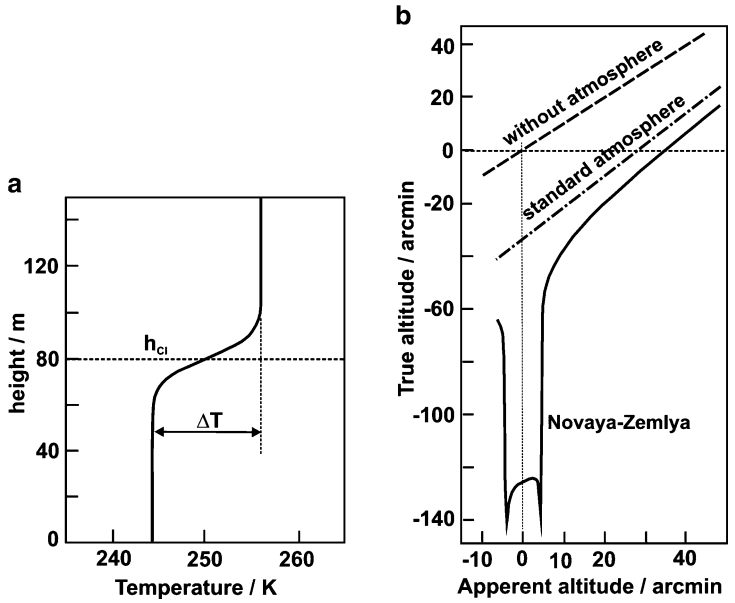


Fig. 3.31 (**a**) A typical schematic temperature profile with an atmospheric inversion layer; h_{CI} denotes the height for which the temperature profile has its inflection point; (**b**) image transformation curves for three situations (1) without atmosphere, showing the 1:1 relation between object appearance and actual position; (2) the influence of the standard atmosphere on the image generation and (3) the situation for the Novaya–Zemlya inversion layer (see also text)

inversion layer can then be calculated and one obtains the so-called transformation curves that relate apparent and true altitude of an object, shown in Fig. 3.31b, for the case of a *superior* image (the strong temperature inversion is located *above* the observer). The parameters are here the temperature difference, taken to be 5°C and h_{CI} which was 45 m. The cold layer temperature was −23°C. The image shows the

appearance of the sun at sunset at about −6 arcmin when the actual position is −130 arcmin below the horizon. In a normal atmosphere, the influence yields an apparent altitude of +27 arcmin at the true altitude 0° (compare lower part of Fig. 3.31).

3.8.5 Bio-Photons

The emission of photons in the course of biological processes remains, mostly related to the low light levels observed, a rather obscure phenomenon for which a straightforward explanation has yet to be provided. Therefore, this effect is treated here under the heading "mysterious light sources", because of the ambiguity related to its origin and due to the fact that it is difficult to find reliable spectral analyses on bio-photonic emission.

The observation of ultraweak radiation from biological specimen dates back far into the last century, when a photon emission process named "mitogenic radiation" has been first reported (see Gurwitsch [95]). Meanwhile, it is well established that all living systems emit spontaneously ultraweak light. The early assumption was that dividing cells emit a very weak UV radiation which was thought to be involved in the cell division process, possibly initiating it. This hypothesis has not yet been substantiated or revoked despite many efforts [96]. Meanwhile, the so-called bio-photon emission has been found in organisms ranging from plants, microorganisms, to animals and humans. It has been reported that all kinds of stress (chemical, mechanical, thermal), the cell cycle and cancer growth result in bio-photonic activity [97]. There are indications that, in humans, bio-photons could contribute to the functioning and operation of the nervous system. In plants, bacteria and animal cells, cell-to-cell communication based on bio-photons has been observed. In the early days of research on this topic, it has been shown, for instance, that the induction of mitosis (cell division) at the tip of a root propagated to a second root, triggered by UV light. More recently, the question arose whether neural cells interact at least partly by the weak biological radiation.

The main issue in research on bio-photons is the unambiguous connection of the ultra-low light intensities observed with processes that occur in biology. The light levels to be analysed in their spectral behaviour are at and below 10^{-10} mW cm^{-2}. In the spectral range between about 350 and 800 nm, the corresponding photon counts range between 1 and 1,000 photons per second and cm^2. Signal noise and artefacts can generally not be excluded. It is difficult to find a typical spectral behaviour in the literature which mostly presents time dependences or spatially resolved images from obviously spectrally integrated signals (for example in the range between 350 and 800 nm). An example is presented in Fig. 3.32 where the photon emission from Sinorhizobium fredii (SF 2048) bacteroids is shown, whose spectral features (see arrows in Fig. 3.32) have been attributed to singlet oxygen that is formed during oxidative metabolism in the cells [98].

After having treated photon generation from discrete transitions in lasers, in chemical and biological systems but also in highly localized processes using STM,

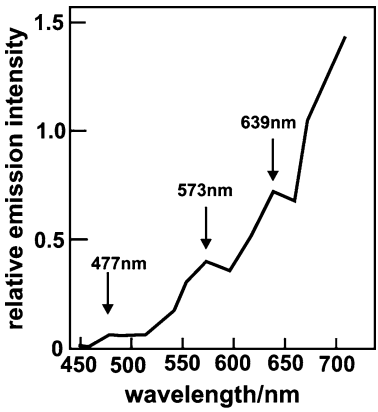


Fig. 3.32 Ultraweak photon emission spectrum, corrected for photomultiplier and spectral filter response, of Sinorhizobium fredii bacteroids (after ref. [98])

the next chapter treats the generation of high energy photons for surface science investigations and applications in materials science, including the analysis of chemical and electrochemical processes with synchrotron radiation. It, thus, completes the part of this book that predominantly deals with the generation and application of photons in various fields of science.

References

1. T. Maiman, Stimulated optical radiation in ruby, Nature **187**, 493–494 (1960)
2. A. Einstein, Zur Quantentheorie der Strahlung (On the Quantum Theory of Radiation), Physika Zeitschrift **18**, 121–128 (1917)
3. M. Göppert-Mayer, Über Elementarakte mit zwei Quantensprüngen, Ann. Phys. (Leipzig) **9**, 273–294 (1931)
4. J.D. Franson, Bell inequality for position and time, Phys. Rev. Lett. **62**, 2205–2208 (1989)
5. W. Perrie, A.J. Duncan, H.J. Beyer, H. Kleinpoppen, Polarization correlation of the two photons emitted by metastable atomic deuterium: a test of Bell's inequality, Phys. Rev. Lett. **54**, 1790 (1985)
6. F. Helmchen, W. Denk, Deep tissue two photon microscopy, Nat. Methods **2**, 932–940 (2005)
7. W.R. Zipfel, R.M. Williams, W.W. Webb, Nonlinear magic: multiphoton microscopy in the biosciences, Nat. Biotechnol. **21**, 1369–1377 (2003)
8. H. Daido, Review of soft x-ray Laser researches and developments, Rep. Prog. Phys. **65**, 1513–1576 (2002)
9. P. Albertano et al., Atmospheric pressure soft X-ray source, for contact microscopy and radiobiology applications, SPIE Int. Conf. Opt. Sci. **3157**, 164–175 (1997)
10. D.L. Matthews et al., Demonstration of a soft X-ray amplifier, Phys. Rev. Lett. **54**, 110–113 (1985)
11. M.D. Rosen et al., Exploding-foil technique for achieving a soft X-ray laser, Phys. Rev. Lett. **54**, 106–109 (1985)
12. S. Suckewer, C.H. Skinner, H. Milchberg, C. Keane, D. Voorhees, Amplification of stimulated soft X-ray emission in a confined plasma column, Phys. Rev. Lett. **55**, 1753–1756 (1985)
13. P.V. Nickles, V.N. Shlyaptsev, M. Kalachnikov, M. Schnürer, I. Will, W. Sander, Short pulse X-ray laser at 32.6nm based on transient gain in Ne-like titanium, Phys. Rev. Lett. **78**, 2748–2751 (1997)

14. R. Wagner, S.-Y. Chen, A. Maksimchuk, D. Umstadter, Electron acceleration by a laser wakefield in a relativistically self-guided channel, Phys. Rev. Lett. **78**, 3125–3128 (1997)
15. T.-Y. Chien et al., Prepulse controlled splitting of relativistically self-guided channel and suppression of Raman forward scattering instability, Phys. Plasm. **11**, 1173–1177 (2004)
16. P. Facci, D. Alliata, L. Andolfi, B. Schnyder, R. Kötz, Formation and characterization of protein monolayers on oxygen-exposing surfaces by multiple step self-chemisorption, Surf. Sci. **504**, 282–292 (2002)
17. G.W. Faris, M.J. Dyer, D.L. Huestis, W.K. Bischel, Two-photon spectroscopy of the $F^1\Pi_g$ and $f^3\Pi_g$ states of molecular fluorine, J. Chem. Phys. **97**, 5964–5969 (1992)
18. S.M. Hooker, A.M. Haxell, C.E. Webb, Influence of cavity configuration on the pulse energy of a high-pressure molecular fluorine laser, Appl. Phys. B **55**, 54–59 (1992)
19. T. Itani, W. Wakamiya, Progress in 157 nm lithography development for 70-nm mode, Microelectr. Eng. **61–62**, 49–55 (2002)
20. P.A.M. Dirac, Lectures on quantum mechanics, Belfer Graduate School of Science Monographs Series **2** (1964)
21. R.T. Hodgson, P.P. Sorokin, J.J. Wynne, Tunable coherent vacuum-ultraviolet generation in atomic vapors, Phys. Rev. Lett. **32** 343–346 (1974)
22. M. Maeda, Short-lifetime measurement of molecules in the vacuum ultraviolet using high-resolution laser spectroscopy, Jpn. J. Appl. Phys. **24**(1985) 717–722
23. M. Kobayashi, H. Inaba, Photon statistics and correlation analysis of ultraweak light originating from living organisms for extraction of biological information, Appl. Optics **39**, 183–192 (2000)
24. J.K. Norskov, D.M. Newns, B.I. Lundqvist, Molecular orbital description of surface chemiluminescence, Surf. Sci. **80**, 179–188 (1979)
25. J.D. Holmes, K.P. Johnston, R.C. Doty, B.A. Kogel, Control of the thickness and orientation of solution-grown silicon nanowires, Science **287**, 1471–1473 (2000)
26. J. Zimmermann, A. Zeug, B. Röder, A generalization of the Jablonski diagram to account for polarization and anisotropy effects in time-resolved experiments, Phys. Chem. Chem. Phys. **5**, 2964–2969 (2003)
27. O. Inomoto, T. Ohya, S. Kai, Intensity oscillation of chemiluminescence in ferroin-catalyzed Belousov–Zhabotinsky reaction, Forma **15**,213–218 (2000)
28. B.P. Belousov, A periodic reaction and its mechanism, Sb. Ref. Radiats. Med. 145–147 (1959)
29. T. Wilson, J.W. Hastings, Bioluminescence, Annu. Rev. Cell Devel. Biol. **14** 197–230 (1998)
30. Y.K. Su et al., InGaN/GaN blue light-emitting diodes with self-assembled quantum dots, Semicond. Sci. Technol. **19**, 389–392 (2004)
31. H.J. Lewerenz, G. Schlichthörl, Light-induced oscillating reactions of silicon in ammonium fluoride solutions: Part 1. Simultaneous photocurrent and excess microwave reflectivity measurements, J. Electroanal. Chem. **327**, 85–92 (1992)
32. H.J. Lewerenz, Spatial and temporal oscillation at Si(111) electrodes in aqueous fluoride-containing solution, J. Phys. Chem. B **101**, 2421–2425 (1997)
33. A.T. Bharucha-Reid, *Elements of the Theory of Markov Processes and Their Applications*, (McGraw-Hill, New York, 1960)
34. J. Grzanna, H.J. Lewerenz, H. Jungblut, A model for electrochemical oscillations at the Si/electrolyte contact Part I. Theoretical development, J. Electroanal. Chem. **486**, 181–189 (2000)
35. J. Grzanna, H.J. Lewerenz, H. Jungblut, A model for electrochemical oscillations at the Si/electrolyte contact Part II. Simulations and experimental results, J. Electroanal. Chem. **486**, 190–203 (2000)
36. J. Rappich, V.Y. Timoshenko, Th. Dittrich, Correlation between surface non-radiative recombination and current oscillation at p-Si(100) during electropolishing in fluoride solution, Ber. Bunsenges. Phys. Chem. **101**, 139–142 (2000)
37. I.L. Krestnikov, N.N. Ledentsov, A. Hoffmann, D. Bimberg, Quantum dot origin of luminescence in InGaN-GaN structures, Phys. Rev. B **66**, 155310 (2002)

38. D. Snoke, S. Denev, Y. Liu, L. Pfeiffer, K. West, Long-range transport in excitonic dark states in coupled quantum wells, Nature **418**, 754–757 (2002)
39. L.V. Butov, A.C. Gossard, D.S. Chemla, Macroscopically ordered state in an exciton system, Nature **418**, 751–754 (2002)
40. R. Rinaldi et al., Zeeman effect in parabolic quantum dots, Phys. Rev. Lett. **77**, 342–345 (1996)
41. Y. Xu, H. Naramoto, K. Narumi, K. Miyashita, T. Kamiya, T. Sakai, Strong anti-Stokes luminescence from H^+-irradiated diamond, Appl. Phys. Lett. **83**, 1968–1970 (2003)
42. V.P. Drachev et al., Quantum size effect in two-photon excited luminescence from silver nanoparticles, Phys. Rev. B **69**, 035318 (2004)
43. J. Lambe, S.L. McCarthy, Light emission from inelastic electron tunneling, Phys. Rev. Lett. **37**, 923–925 (1976)
44. J.K. Sass, H.J. Lewerenz, Photoemission yield spectroscopy of metal electrodes, J. Phys. Colloques **38**, C5-277–C5-284 (1977)
45. D.M. Kolb, M. Przasnyski, H. Gerischer, Underpotential deposition of metals and work function differences, J. Electroanal. Chem. Interf. Electrochem. **54**, 25–38 (1974)
46. J.K. Sass, S. Stucki, H.J. Lewerenz, Plasma resonance absorption in interfacial photoemission from very thin silver films on Cu(111), Surf. Sci. **68**, 429–435 (1977)
47. H. Raether, *Plasmons on Smooth and on Rough Surfaces and on Gratings* (Springer, Berlin, 1988)
48. R. Kötz, D.M. Kolb, The appearance of the bulk Plasmon in thin silver overlayers monitored by electrochemical modulation spectroscopy, Surf. Sci. **97**, 575–585 (1980)
49. M. Rasigni, G. Rasigni, G. Palmari, Surface plasmon and autocorrelation functions for rough surfaces of silver deposits, Phys. Rev. B **23**, 527–531 (1981)
50. J. Homola, S.S. Yee, G. Gauglitz, Surface plasmon resonance sensors: a review, Sensor. Actuat B: Chem. **54**, 3–15 (1999)
51. R.D. Young, J. Ward, F. Scire, The topografiner: an instrument for measuring surface microtopography, Rev. Sci. Instrum. **43**, 999–1011 (1972)
52. J.H. Coombs, J.K. Gimzewski, B. Reihl, J.K. Sass, R.R. Schittler, Photon emission experiments with the scanning tunnelling microscope, J. Microscop. **152**, 325–336 (1988)
53. R. Berndt et al., Photon emission at molecular resolution induced by a scanning tunnelling microscope, Science **262**, 1425–1427 (1993)
54. R.W. Rendell, D.J. Scalapino, Surface plasmons confined by microstructures on tunnel junctions, Phys. Rev. B **24**, 3276–3294 (1981)
55. B. Laks, D.L. Mills, Photon emission from slightly roughened tunnel junctions, Phys. Rev. B **20**, 4962–4980 (1979)
56. A. Takeuchi, J. Watanabe, Y. Uehara, S. Ushioda, Prism coupled light emission from tunnel junctions containing interface roughness: theory, Phys. Rev. B **38**, 12948–12958 (1988)
57. P. Johansson, R. Monreal, P. Apell, Theory for light emission from a scanning tunnelling microscope, Phys. Rev. B **42**, 9210–9213 (1990)
58. W.L. Barnes, Fluorescence near interfaces: the role of photonic mode density, J. Mod. Optics **45**, 661–699 (1998)
59. X.H. Qiu, G.V. Nazin, W. Ho, Vibrationally resolved fluorescence with submolecular precision, Science **299**, 542–546 (2003)
60. G. Hoffmann, J. Kliewer, R. Berndt, Luminescence from metallic quantum wells in a scanning tunneling microscope, Phys. Rev. Lett. **87**, 176803 (2002)
61. S.A. Campbell, H.J. Lewerenz (eds.), *Semiconductor Micromachining*, Fundamental Electrochemistry and Physics, vol. 1; Techniques and Industrial Applications, vol 2 (Wiley, Chicester, 1998)
62. G.L. Hornyak, H.F. Tibbals, J. Dutta, J.J. Moore, *Introduction to Nanoscience and Nanotechnology*, (CRC, Boca Raton, FL, 2008)
63. V.G. Veselago, The electrodynamics of substances with simultaneously negative values of ε and μ, Sov. Phys. USPEKHI **10**, 509–514 (1968)
64. F. Cappolino, Applications of Metamaterials (Routledge, London, 2009)
65. J.B. Pendry, Negative refraction makes a perfect lens, Phys. Rev. Lett. **85**, 3966–3969 (2000)

66. J.B. Pendry, D. Schurig, D.R. Smith, Controlling electromagnetic fields, Science **312**, 1780–1782 (2006)
67. V.M. Shalaev, Optical negative-index metamaterials, Nat. Photonics **1**, 41–48 (2007)
68. M. Duran-Sindreu, F. Aznar, A. Vélez, J. Bonache, F. Martin, Analysis and applications of OSSR- and OCSSR-loaded transmission lines: A new path for the design of compact transmission line metamaterials, Metamaterials **4**, 139–148 (2010)
69. S.H. Lee, C.M. Park, Y.M. Seo, Z.G. Wang, C.K. Kim, Reverse Doppler effect of sound, Phys. Rev. Lett. **104**, 045 301 (2010)
70. H. Chen, M. Chen, Flipping photons backward: reversed Cherenkov radiation, Mater. Today **14**, 34–41 (2011)
71. D.R. Smith, J.B. Pendry, M.C.K. Wiltshire, Metamaterials and negative refractive index, Science **305**, 788–792 (2004)
72. R.A. Shelby, D.R. Smith, S. Schultz, Experimental verification of a negative index of refraction, Science **292**, 77–79 (2001)
73. L. Menon, W.T. Lu, A.L. Friedman, S.P. Bennett, D. Heiman, S. Sridhar, Negative index metamaterials based on metal-dielectric nanocomposites for imaging applications, Appl. Phys. Lett. **93**, 123 117 (2008)
74. A.A. Zharov, N.A. Zharova, On the electromagnetic cloaking of (nano)particles, Bull. Russ. Acad. Sci.: Phys. **74**, 89–92 (2010)
75. A. Alú, N. Engheta, Plasmonic and metamaterial cloaking: physical mechanisms and potentials, J. Optics A: Pure Appl. Optics, **10**, 093 002 (2008)
76. J. Lee, Bioluminescence: The first 3000 years, J. Sib. Fed. Univ. Biol. **1**, 194–205 (2008)
77. T. Wessel-Berg, Ball lighting and atmospheric light phenomena: a common origin? J. Sci. Explor. **18**, 439–481 (2004)
78. J.W. Hastings, Biological diversity, chemical mechanisms, and the evolutionary origins of bioluminescent systems, J. Mol. Evol. **19**, 309–321 (1983)
79. G. Gassmann, D. Glindemann, Phosphane (PH_3) in the biosphere, Angew. Chem. Int. Ed. **32**, 761–762 (1993)
80. J.S. Derr, Earthquake lights: A review of observations and present theories, Bull. Seismol. Soc. Am. **63**, 3177–2187 (1973)
81. M. Kamogawa, H. Ofuruton, Y.-H. Ohtsuki, Earthquake light: 1995 Kobe earthquake in Japan, Atmosph. Res. **76**, 438–444 (2005)
82. C. Fidani, The earthquake lights (EQL) of the 6 April 2009 Aquila earthquake, in Central Italy, Nat. Haz. Earth Syst. Sci. **10**, 967–978 (2010) Aquila earthquake
83. F. St-Laurent, The Saguenay, Québec, earthquake lights of November 1988–January 1989, Seismol. Res. Lett. **71**, 160–174 (2000)
84. H. Tributsch, *When the Snakes Awake* (MIT, Cambridge, MA 1983)
85. F.T. Freund, Rocks that crackle and sparkle and glow: Strange pre-earthquake phenomena, J. Sci. Explor. **17**, 37–71 (2003)
86. W.F. Brace, W. Pauling, C. Scholz, Dilatancy in the fracture of crystalline rocks, J. Geophys. Res. **71**, 3939–3953 (1966)
87. H. Schulz, K.H. Thiemann, Crystal structure refinement of AlN and GaN, Sol. St. Comm. **23**, 815–819 (1977)
88. H. Masui, S. Nakamura, S.P. DenBaars, U.K. Mishra, Nonpolar and semipolar III-nitride light-emitting diodes: achievements and challenges, IEEE Transact. Electron Dev. **57**, 88–100 (2010)
89. Y. Bernabé, Streaming potentials in heterogeneous networks, J. Geophys. Res. **103**, 20827–20841 (1998)
90. F. Freund, Towards a unified solid state theory for pre-earthquake signals, Acta Geophysica **58**, 719–766 (2010)
91. R. Guckenberger, M. Heim, G. Cevc, H.F. Knapp, W. Wiegräbe, A. Hillebrand, Scanning tunneling microscopy of insulators and biological specimen based on lateral conductivity of ultrathin water films, Science **266**, 1538–1540 (1994)
92. J. Eyers, *Don't Shoot the Albatross!: Nautical Myths and Superstitions* (A&C Black, London, 2011)

93. E.M. Wescott, D.D. Sentman, M.J. Heavner, T.J. Halliman, D.L. Osborne, The optical spectrum of aircraft St. Elmo's fire, Geophys. Res. Lett. **23**, 3687–3690 (1995)
94. S.Y. van der Werf, G.P. Können, W.H. Lehn, Novaya Zemlya effect and sunsets, Appl. Optics **42**, 367–378 (2003)
95. A.A. Gurwitsch, A historical review of the problem of mitogenic radiation, Experentia **44**, 545–550 (1988)
96. B. Devaraj, M. Usa, H. Inaba, Biophotons: ultraweak light emission from living systems, Curr. Opin. Sol. State Mat. Sci. **2**, 188–193 (1997)
97. Y. Sun, C. Wang, J. Dai, Biophotons as neural communication signals demonstrated by in-situ biophoton autography, Photochem. Photobiol. Sci. **9**, 315–322 (2010)
98. X. Shen, X. Han, J. Tian, F. Zhao, L. Xu, X. Li, Spontaneous luminescence from soybean Rhizobium bacteroids, FEMS Microbiol. Lett. **81**, 335–340 (1991)

Chapter 4
High-Energy Photons for Surface/Interface Analysis and Materials Science

4.1 Introduction

In this chapter, a selection of examples is presented in which some of the light sources that were introduced in Chap. 2 are employed for a specific application. The chapter begins with the introduction to synchrotron radiation and related experimentation. Particularly with the introduction of third-generation synchrotrons, characterized by a substantially increased photon flux, experiments such as X-ray emission spectroscopy (XES) have become possible that allow, for example, the investigation of the energetic structure of the metal centres in metalloproteins. Again, the bridging character of the content is emphasized and examples are included from the areas of chemistry, physics, materials science and biology. Another important aspect is the investigation of the solid–liquid interface and of electrochemical processes that will be treated in Sects. 4.2.2 and 4.2.3. Such studies gain increasing importance due to the renewed interest in the generation of fuels from sunlight in integrated structures. The chapter closes with the application of a hyperfine interaction method, perturbed angular correlations (PAC) that has been successfully applied to the development of efficient thin film solar cells based on the ternary chalcopyrite $CuInS_2$ (Table 4.1).

4.2 Synchrotron Radiation Experiments

From the numerous experiments that use synchrotron radiation in surface physics and chemistry [1], investigations with strong interdisciplinary character are selected here. A particular interdisciplinary field is photoelectrochemistry [2, 3]. Its appeal results from the possibility of generating solar energy conversion structures that operate in the photovoltaic or in the photoelectrocatalytic mode. Experiments that probe the solid–liquid interface are presently of particular interest because of the absence of sound theoretical background regarding photoelectrocatalytic reactions.

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_4,

Table 4.1 Overview of the content of Chap. 4

More generally, synchrotron radiation experiments have the potential to be applied beyond the areas of surface chemistry and photoelectrochemistry, allowing, for instance, detailed structural and electronic investigations of biological molecules.

4.2.1 Introduction: Experimental Considerations

The experiments presented below focus on the elucidation of underlying reaction mechanisms in soft solution processing of semiconductors. A detailed understanding of dissolution and deposition processes provides means for preparation of advanced nanostructures, which, if prepared by self-organized electrochemical processes, are scalable in electrochemical low (room) temperature processes. Nanostructures with feature sizes far below the 100-nm range can be realized rather easily [4]. The use of synchrotron radiation for the study of the solid/liquid interface also extends to (electro)catalysis research [5, 6], epitaxial film growth in solution by so-called underpotential deposition [7] and to the physico-chemical analysis of processing steps in the preparation of nanoemitter solar cells [8].

Generally, experiments that investigate the solid–liquid interface by a UHV analysis method, such as photoelectron spectroscopy (PES), are characterized by the particular difficulty to obtain meaningful results after sample exposure to solutions. The two hitherto realized successful systems use (1) a multistep differentially pumped high-pressure system where the solid surface reactions can be studied in moist ambient, for example [9], or (2) a closed *in-system* photoelectrochemical approach that allows photoelectrochemical experimentation in inert gas atmosphere and that uses a sophisticated transfer technique for sample transfer into the UHV [10]. An experimental facility for investigation of processes in semiconductor interface engineering has been established at the undulator beamline U49/2 at

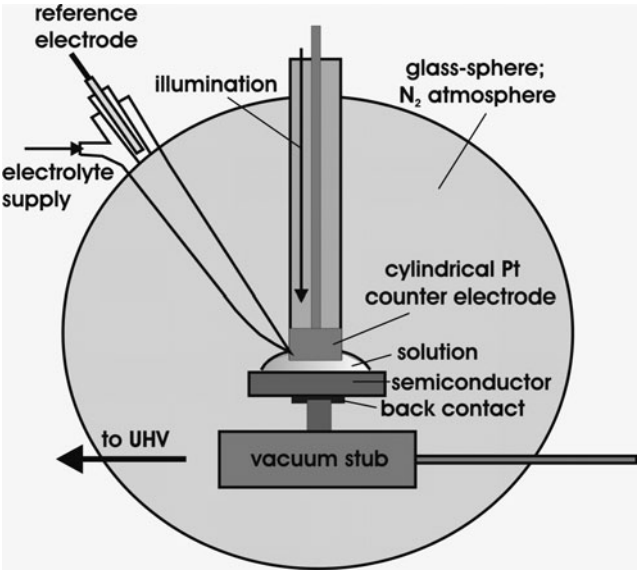


Fig. 4.1 Electrochemical cell for UHV surface analysis experiments: electrolyte droplets are supplied through the Luggin capillary that simultaneously serves as reference electrode contact; a cylindrical Pt counter electrode is used where the mounting rod serves as light guide for photoelectrochemical experiments; N_2 jet blowing of the electrolyte drop interrupts the experiment, followed by rinsing–drying and transfer to UHV

the electron synchrotron Bessy II. The system has been specifically designed for wet processing and subsequent UHV surface analysis. When equipped with a specifically designed *in-system* electrochemical cell (Fig. 4.1), ultra-low surface contamination levels (∼0.1 monolayer after electrochemical current flow) allow investigation of wet chemical treatments and (photo)electrochemical processes in detail. High surface sensitivity methods such as UPS, LEED, HREELS and synchrotron radiation photoelectron spectroscopy (SRPES) can then be employed (Fig. 4.2) and details of the surface chemistry and electronics can be assessed.

High surface sensitivity chemical analyses by SRPES and the possibility to probe the surface region below the topmost atomic layer by appropriate tuning of the photon energy allow virtually complete assessment of the surface condition. In industrial processing of semiconductors, wet processing steps are abundant and many procedures result from empirical developments. For future advanced devices, detailed knowledge of the treatment induced electronic and chemical changes will become increasingly important.

The selected examples include (1) evolution of nanostructures on single crystalline Si, where a metastable mesa-type topography and step bunching is observed [4, 11, 12] and (2) nanopore structures obtained by oscillating photocurrents. These experiments are mostly based on SRPES in combination with CM-AFM and in selected cases with Brewster angle analysis [13]. For an introduction, Fig. 4.3 shows

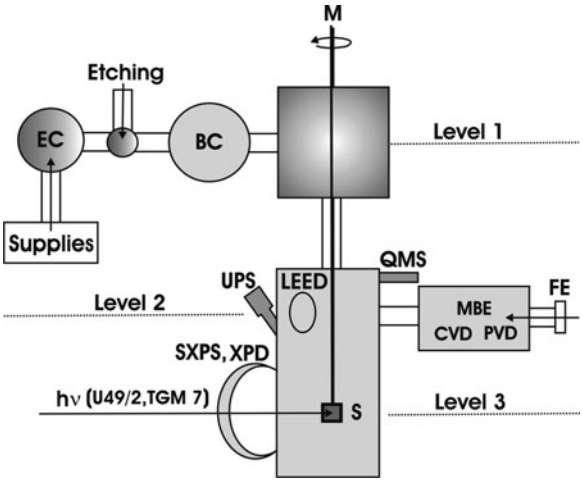


Fig. 4.2 *In-system* surface analysis apparatus for photoelectron spectroscopy At BESSY II after (photo)electrochemical surface conditioning; this solid–liquid analysis system is usually attached to the undulator beamline U49/2; the electrochemistry vessel is shown on the *top left-hand side* (*EC*); *S* sample position; *SXPS*, *XPD* soft X-ray photoelectron spectroscopy, X-ray diffraction, respectively; *FE* fast entry lock; *MBE*, *CVD*, *PVD* molecular beam epitaxy and chemical– physical vapour deposition, respectively; *LEED*, *UPS*, *QMS* low-energy electron spectroscopy, ultraviolet photoelectron spectroscopy, quadrupole mass spectroscopy, respectively; *M* manipulator; *BC* buffer chamber

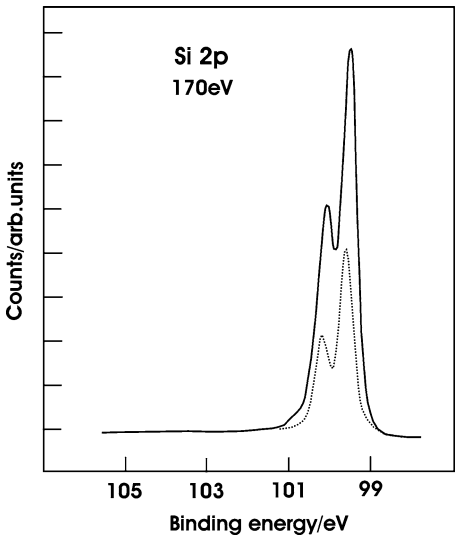


Fig. 4.3 SRPES of electrochemically prepared (1 × 1)-Si(1 1 1):H with deconvolution of the envelope curve (for assignments see text)

an SRPE spectrum after electrochemical hydrogen termination of Si(1 1 1) at a photon energy of 170 eV where the mean inelastic electron scattering length is at its minimum, resulting in a surface sensitivity in the monolayer range.

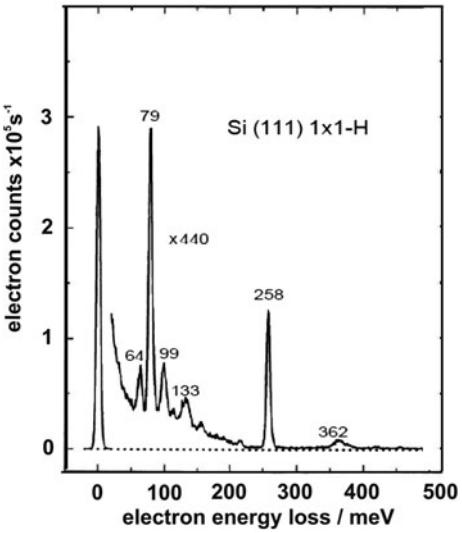


Fig. 4.4 High-resolution electron energy loss spectrum of an electrochemically hydrogen terminated Si(1 1 1) surface; the electrochemical conditioning procedure consisted of (1) current oscillations at +6 V anodic potential and (2) subsequent two-step surface transformation using a transient anodic dark current at lower potential (+0.5 V) at pH 4.0 and subsequently at pH 4.9; the absolute value of the dark current at the end of the treatment was a measure of the surface quality. Values in the nA range ensured high structural and electronic quality

Deconvolution shows the bulk Si 2p core level line at 99.6 eV and two signals shifted by ∼0.25 eV (1) and 0.5 eV (2); the latter signal is not shown in the figure. Signal (1), indicated by the dotted line in Fig. 4.3, is attributed to the surface core level shift due to partial charge redistribution from Si surface atoms to the more electronegative H atoms [14]. Signal (2) can be related to kink site atoms (with two backbonds to the Si lattice) that bind to H and OH, and signals indicative of a larger binding energy shift are presented below. Assignments of lines in the deconvoluted spectra become possible by predictions on the surface chemistry, partial charge calculations from electronegativity and structural data using, for instance, the density functional theory (DFT) approach [15] and by vibrational analysis of the surface. Details are presented below in Sect. 4.2.4.

An HREEL spectrum of an electrochemically H-terminated surface is shown in Fig. 4.4. The prominent features are the Si–H bending mode at 78 meV and the Si–H stretching mode at 258 meV. Phonon modes are seen at an energy loss of 64 meV. The Si–H_2 scissor mode (113 meV) shows the presence of Si atoms on the (1 1 1) surface with only two backbonds, being twofold terminated by hydrogen (=Si–H_2). The signals at 99 meV, 133 meV correspond to Si–OH or Si–F species and bridging oxygen in a backbond, respectively. At 362 meV, the presence of a methyl group is indicated where the dangling bond is terminated by a CH_3 group instead of H [16, 17]. It should be noted that the cross section for impact excitation of the

Si–H bending mode is smaller by a factor of ~25 compared to the excitation cross section of the oxidic lines. Therefore, the surface shown can be considered to be H-terminated in a (1 × 1) manner for about >90%. Contrary to frequent statements in the literature, an ideal H-terminated surface cannot be prepared, even for the close packed (1 1 1) surface. This is due to the presence of atomic terraces with step edges and kink sites that could partly be terminated by OH or hydrocarbon species. Recently, the preparation of an OH-free H-terminated Si (1 1 1) surface has been achieved by a consecutive chemical etch in concentrated NH_4F and HF, evidenced by SRPES at the highest possible surface sensitivity, adjusted by the photon energy [18].

In the following, fundamental aspects of PES, in particular related to semiconductor surface analysis and of HREELS, are outlined before electrodeposition is treated as an example for soft solution processing. Subsequently, nanostructure evolution on silicon and device preparation using synchrotron radiation is presented, followed by XES experiments for analysis of spurious films after wet treatment of the solar cell absorber material copperindium-disulfide ($CuInS_2$).

4.2.2 Background on Photoelectron Spectroscopy and High-Resolution Electron Energy Loss Spectroscopy

4.2.2.1 Photoelectron Spectroscopy with Semiconductors

In PES, the kinetic energy of the emitted electrons is measured. The Einstein relation

$$E_{\mathrm{Kin}} = h\nu - E_{\mathrm{B}}, \tag{4.1}$$

defines the relation between photon energy hν and the electron (binding) energy, E_{B}, of the excited state. For investigations on semiconductors, one usually relates E_{B} to either the Fermi energy, E_{F}, or to the valence band maximum, E_{VB}. An energy scheme of the situation with semiconductors is presented in Fig. 4.5. Because the Fermi level of the sample adjusts to that of the analyser of the UHV apparatus [20], the binding energy is often related to the Fermi level:

$$E_{\mathrm{Kin}} = h\nu - \Phi - E_{\mathrm{B}}, \tag{4.2}$$

where Φ is the work function of the sample. If one uses the valence band maximum as reference level, the electron affinity χ and the semiconductor band gap E_{g} have to be included (see Fig. 4.5):

$$E_{\mathrm{Kin}} = h\nu - \chi - E_{\mathrm{g}} - E_{\mathrm{B}}. \tag{4.3}$$

In Fig. 4.5, the different definitions for the binding energy are contrasted.

The Fermi level of semiconductors can be changed by doping and surface charges [21]. Such a situation is displayed in Fig. 4.6 where a depletion region has

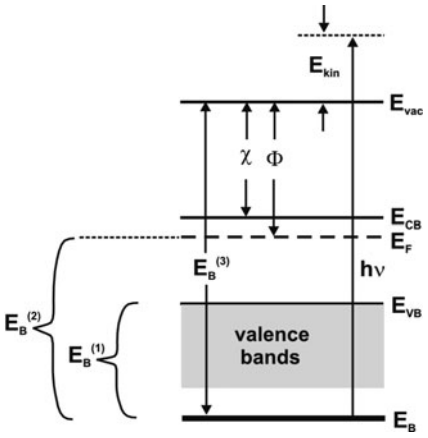


Fig. 4.5 Schematic of the important energetic parameters in photoelectron spectroscopy of semiconductors; E_{Vac}, vacuum level, E_{VB}, E_{CB}, valence band maximum and conduction band minimum, respectively; Φ, χ are the work function and electron affinity, respectively; the three $E_{\mathrm{B}}^{(1,2,3)}$ denote the possibilities to assign the binding energy; here, $E_{\mathrm{B}}^{(2)}$ will be used (see also text); the energy level E_{B}, from which electron(s) are excited, has been chosen to be energetically located outside (i.e. below) the valence band region and thus represents a core level such as, the Pt $4f_{7/2\text{-}5/2}$ level at 71 eV [19]. Therefore, the symbol E_{CL} is used for energy levels below the valence band minimum in Fig. 4.6

been formed for an n-type semiconductor, resulting in the linear–parabolic variation of the energy bands due to the charged depletion layer. The energetic difference between the situation at flatband, where the Fermi level position E_{F}(fb) is given only by the doping level (and the material properties), and that for a depletion layer, $E_{\mathrm{F}}^{\mathrm{bb}}$ is given by the change of the Fermi level position at the surface with respect to the valence band maximum. It is, therefore, possible to assess the electronic situation at the surface of a semiconductor by determination of the onset of the valence band emission relative to the Fermi level. This is the reason why one usually plots the count rate with regard to the Fermi level of the analyser to which the sample adjusts. For a given doping level, E_{F}(fb) can be calculated, provided that the necessary parameters such as the effective density of states at the conduction and valence band edges are known (requiring the knowledge of the effective masses). If, for an n-type semiconductor, the measured energetic distance between the Fermi level at flatband is larger than the calculated value, an accumulation layer has been formed. If this distance is smaller, a depletion layer exists as has been chosen for the illustration in Fig. 4.6. It should be noted that, upon illumination under UHV conditions, the probe light can induce a photovoltage [22] which leads to a reduction of the band bending. Therefore, the measurement of the onset of the valence band emission with respect to the Fermi level does not necessarily allow direct determination of the value of the band bending.

Typical depletion layers have a spatial extension between 100 nm and 1 μm. The depletion layer width in Fig. 4.6, in which the energy bands are bent, can be assumed

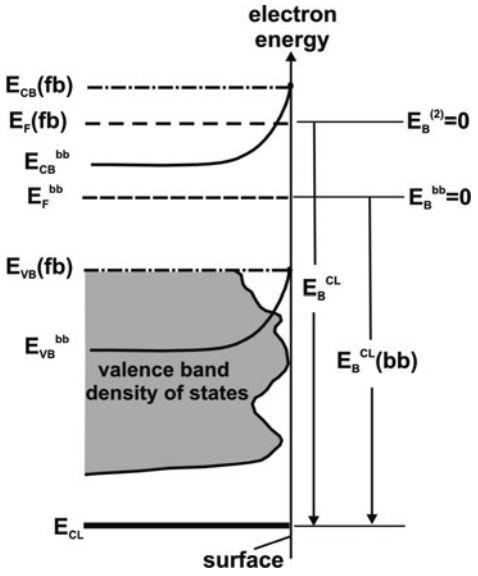


Fig. 4.6 Energy scheme for a semiconductor at flatband (fb) condition and with a depletion layer band bending (bb); the Fermi level without charging, E_F(fb), corresponds to $E_B^{(2)}$ in Fig. 4.5; the measured binding energy, E_B^{CL}, to the core level E_{CL} is larger than that for a given band bending, E_B^{CL}(bb); (*dashed dotted lines*) conduction and valence band at the flatband situation; (*full parabolic lines*) conduction and valence band for band bending (see also text)

to extend, for instance, 200 nm into the semiconductor. PES measures the properties of the solid with high surface sensitivity. This property of PES results from the scattering behaviour of the excited photoelectrons and is a function of their kinetic energy. The information on the sample electronic states, obtained from elastic, i.e. unscattered photoelectrons, is only obtained for electrons that have been generated at or near the upmost surface layers of the sample. The related inelastic mean free scattering length, λ_{esc}, is plotted in Fig. 4.7. It is a measure of the surface sensitivity of a PES experiment. The values vary between 0.3 nm $< \lambda_{esc} <$ 3–4 nm, depending on the kinetic energy of the photoelectrons. It should be noted that in experiments that use single photon energy light sources such as Mg K_α in commercially available surface analysis systems, the information on sample constituents is typically not from the same depth in the material. For a metal oxide, for instance, the O 1s core level, is located at binding energies of ~533 eV and, if the metal core level is located at a considerably lower energy such as 100 eV, the information on the compound elements is obtained for different probing depths of the material.

In comparison, SRPES provides the possibility to tune the photon energy and adjust it to yield ultrahigh surface sensitivity or to obtain information from the same probing depth (see also Sect. 2.6.2).

Because of the surface sensitivity of PES, where the escape depths range between 0.3 and 3 nm, the space charge region is not probed by this method; instead, only the topmost 1–10 atomic layers of the sample surface are analyzed. Therefore, the method samples the region that has been indicated by the small spherical dots in Fig. 4.6 that give the energetic position of the semiconductor bands at the surface.

The excitation of electrons in a photoemission experiment that translates the density of states in the initial state into a energy distribution curve (EDC) is shown

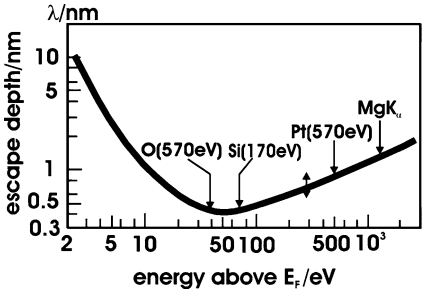


Fig. 4.7 Dependence of the inelastic mean free scattering length of photoelectrons on energy above the Fermi level; for the O1s, Si 2p and Pt 4f level, at binding energies of 533, 100 and 71 eV, respectively, the photon energies for highest surface sensitivity are indicated; the photon energy for Pt has been increased to larger values beyond the dip in the cross section due to a Cooper minimum in the excitation cross section [23] allowing measurements of small amounts of Pt; also shown is the escape depth of photoelectrons from the Si 2p core level, excited by Mg K_α radiation that is larger by approximately a factor of 4 compared to the highest attainable surface sensitivity around 150–170 eV

in Fig. 4.8. The EDC is obtained by measuring the photoelectron yield (counts per energy interval) vs. the kinetic energy of the electrons that reach the analyzer [24]. For generality, also excitation processes that are preferably observed at metallic samples are included. The EDC is characterized by the original band structure, revealing the K, L levels and the valence band. In addition, typical features, such as the inelastic background, shake up (not shown: shake off) satellites and plasmon excitations, are seen. Also, the KLL Auger decay, located energetically between the K and the L core levels, is shown. The principle of Auger emission is shown together with a different representation of the photoemission process in Fig. 4.9. The core hole in the K shell, generated by the photoelectron excitation, recombines with an electron from the L shell (level L_1). The corresponding energy is transferred to a third electron in the L shell (level $L_{2,3}$), the Auger electron, that leaves the solid with the remaining kinetic energy. Another possibility is X-ray emission, resulting from the recombination of an electron from the L_1 level to the K shell. The emission yield (Auger electrons vs. X-ray emission) is a function of the atomic number [25] and for lighter elements, the Auger process clearly dominates. The cross-over of Auger vs. XE yield is around $Z = 35$. For Zr with $Z = 40$, the X-ray yield is about twice as large as that for the Auger process (see also Sect. 4.2.6). Plasmonic effects have been discussed in Sect. 3.6. The plasmon excitation is a resonant process that leads to enhanced photoelectron emission due to the decay of the plasmon into single particle excitations [26, 27].

The investigation of the surface chemistry becomes possible due to the chemical shift of core level signals, depending on the oxidation state of the respective element, as introduced below using the so-called microcapacitor model (see also (4.15) in Sect. 4.2.4). The correlation of the observed shift in XPS signals with partial oxidation of sample constituent elements is more generally described by the following parameters:

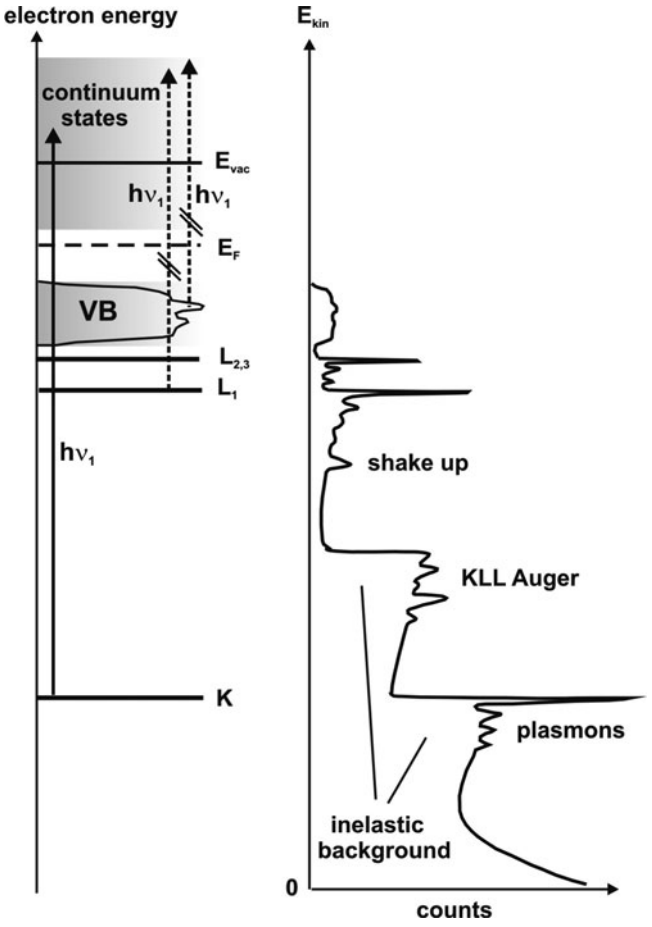


Fig. 4.8 Photoelectron excitation processes from core level states (K, L shell) and from the valence band of a semiconductor (*left hand side*); excitation with the photon energy $h\nu_1$ results in an energy distribution curve (*right hand side*) that represents a superposition of the core level excitations, secondary electrons, Auger processes, Plasmon satellites and shake-up processes (see text)

$$E_B^{exp} = E_B + \Delta E_{CH} + \Delta E_{BB} + \Delta E_M + \Delta E_R. \tag{4.4}$$

The binding energy, measured in an X-ray photoemission experiment; E_B^{exp}, is the sum of the binding energy of the element; E_B, the shift of the core level position due to ionization and chemical interaction with the environment; ΔE_{CH}, a (possible) band bending in the case of semiconducting samples, ΔE_{BB}, the Madelung term ΔE_M, and of the relaxation energy ΔE_R. The second term on the right-hand side of (4.4), the chemical shift ΔE_{CH}, is determined by the effective charge of the respective atom and by the influence of the adjacent atoms. Using a simplification

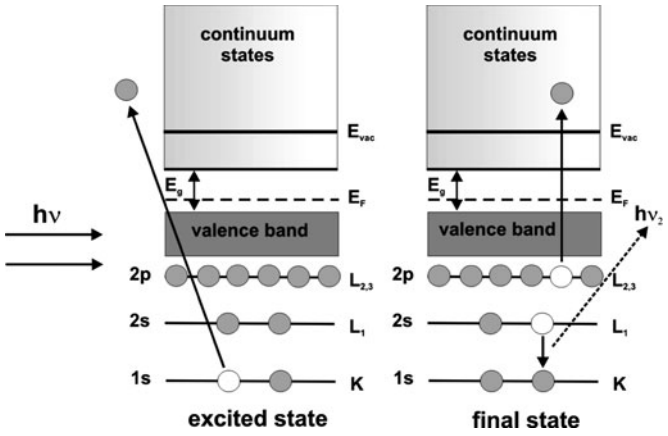


Fig. 4.9 Energy schematic for the photoemission process (*left hand side*) and a KLL Auger process (*right-hand side*); (*grey circles*) electrons; (*white circles*) holes; the transition energy $L \rightarrow K$ is not to scale, compared to the excitation $L_{2,3} \rightarrow$ continuum; also shown is X-ray emission (XES) with photon energy $h\nu_2$, a process with considerably lower cross section than the Auger process

in a spherical charge model, the energy change can be qualitatively calculated: the energy of a spherical charge is $E = z_V e^2/r_V$, (z_V, number of valence electrons; r_V, average radius of the valence electrons). A change of the valency due to an altered valence electron distribution (chemical interaction) yields $\Delta E = \Delta z_V e^2/r_V$. The term Δz_V contains partial charges and, for simplicity, r_V has been assumed to remain unchanged. Including an interatomic potential change, ΔV_{ij} (V_{ij} potential of atom i on atom j), the binding energy shift reads

$$\Delta E_B = \frac{\Delta z_V e^2}{r_V} - \Delta V_{ij}. \tag{4.5}$$

The energetic shift due to semiconductor band bending has already been discussed (see Fig. 4.6 and the corresponding text). The Madelung term defines an extra-atomic contribution that results from the additional electrical potential of the atoms of the sample beyond the direct the nearest neighbour atoms [28] which act on the excited atom. This term is given by $\Delta E_M = \sum_i \rho_i / R_i$ (ρ_i, charge on neighbour atoms (ions in ionic crystals) at positions R_i). For the example of SiO_2, the ionicity is 51%. It should be noted that the Madelung term also depends on the partial charge of the excited atom because the charge correction is also a function of the charge of the excited (central) atom and the strict separation of the terms in (4.4) is a simplification. Since the interatomic distances are also influenced by the charge of the excited atom, it is difficult to accurately determine the binding energy change with atomic charge.

The relaxation effects, ΔE_R, contain the dynamic reaction of the system's electrons, resulting from an electron density shift of the neighbouring atoms towards the oxidized (photo-excited) atom, induced by the core hole formation in the photoexcitation process. This excitation results in an unbalanced nuclear charge. The increase in positive charge pulls the electron levels of the probe atom to higher binding energies and the valence electronic charge can react, depending on its polarizability, to the formed local potential well. The extra-atomic relaxation energy can be approximated by $\Delta E_R = q^2/2r_{SCR}(1 - 1/\varepsilon_M)$, where r_{SCR} denotes the minimum electronic screening distance and ε_M is the dielectric constant of the surrounding medium. As Auger parameter shifts, $\Delta\alpha$, are directly correlated to the polarizability of a chemical state, it has been shown that $\Delta\alpha = 2\Delta E_R$ [29]. A slow relaxation of the electrons in outer shells can result in an insufficient screening of the core hole and, thus, yields a higher measured binding energy. The shake-up and shake-off processes describe the transfer of energy of the final state configuration (due to insufficient screening of the core level electron that underwent photoemission) to excite electrons from valence levels either to bound states (shake-up) or to free states (shake-off). The effect is seen as a seemingly increased binding energy.

4.2.2.2 High-Resolution Electron Energy Loss Spectroscopy (HREELS)

HREELS uses monochromatized low-energy electrons (primary energy in the range of a few electron volts) that probe the surface species of a sample under UHV conditions. The interaction of low-energy electrons with species on the surface of metals or semiconductors can be divided into three basic mechanistically different processes (1) impact scattering, (2) negative ion resonance and (3) dipole scattering. Impact scattering can be observed in off-specular directions and negative ion resonances occur when the impinging low-energy electrons are trapped in unoccupied orbitals of surface species where they form a short-lived intermediate negatively charged ion state [30]. The dipole interaction (3) is long range and the dipole losses are observed in the specular direction (see Fig. 4.10). Each interaction results in a loss of the primary energy of the probing electrons. This loss is detected with a resolution in the meV range using elaborated time-of-flight detectors [31], allowing a highly accurate assessment of the species on the surface. The method has found application in the analysis of initial steps in the thermal oxidation of Si, for instance [32] and in the analysis of H-terminated Si surfaces as shown above in Fig. 4.4. In the experiments described here, the energy loss has been measured in the specular direction. The applicability of the method to semiconductors depends on the material conductivity. If the conductivity $\sigma = en\mu$ is high enough, the image charges can follow the motion of the low-energy probe electrons. Since σ depends on the carrier concentration (which can be adjusted by doping) and mobility, crystalline substrates with sufficient doping levels can be used. These conditions were fulfilled for the analysis of H-terminated surfaces (Fig. 4.4).

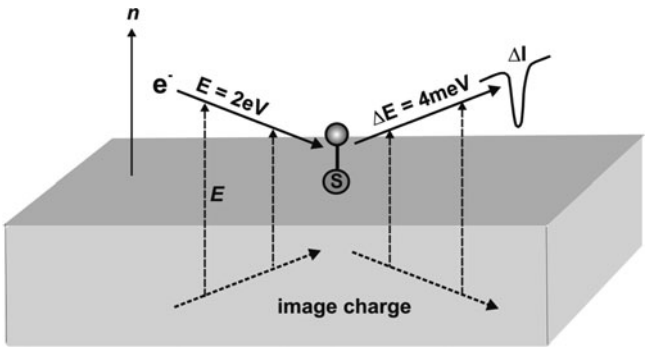


Fig. 4.10 Schematic of an HREELS experiment in the dipole mode (specular reflection); the primary energy, E, and the energy loss, ΔE, upon interaction of the impinging electron with the adsorbate (sphere) on the surface atom S that leads to a loss ΔI in the spectrum of the reflected electrons; E denotes the electric field between electron and its image charge, n is the surface normal

4.2.3 Soft Solution Processing: Basics

Modern history of wet processing of semiconductors dates back to the early 1950s, when shaping of silicon became an issue in the preparation of integrated circuits [33]. Today, micromachining techniques are ubiquitously used in the semiconductor industry [34]. The advantages of solution processing result from its nondestructiveness, low-temperature experimentation and the high achievable pattern resolution and reproducibility. In electrochemical conditioning, accurate control of parameters such as charge, potential and current is possible, allowing defined changes into the submonolayer range at surfaces. Compared to shaping by vacuum machining techniques, solution processing generally produces superior electronic devices and allows preparation of conformal coatings even of ultrathin films. The striving for extreme miniaturization makes solution methods (particularly electrochemical treatments) promising for the fabrication of nanowires, nanoparticles, quantum dots and nanocomposite structures. Upscaling of nanostructures by use of self-organized processes is a principal possibility [35, 36]. Soft solution processing knows three major routes: etching, passivation and plating. For etching, one distinguishes between electroless methods, favoured by industry due to their facilitated integration into a process line and electrochemical dissolution. Photoetching in the currentless and the electrochemical mode has been used abundantly for the improvement of surface topography and electronic behaviour [37]. Passivation refers to passive film formation, typically by oxidative processes. Such films can be corrosion products of the substrate material or produced by deposition. A specific example is electropolishing, where a dynamic equilibrium between dissolution and oxidation maintains an ultrathin film on a surface [38] and the etchstop technique for p-type Si to produce device structures (see below) [39]. Anodic films are often disordered due to substantial lattice mismatch, whereas certain plating systems can be described

by a two-dimensional electrocrystallization process [40]. Such 2D processes are of considerable renewed interest in the formation of nano-dimensioned islands, vertically quantum confined systems and for highly ordered crystalline ultrathin layers on semiconductors and ferroelectrics [41].

The mechanisms contributing to electrodeposition are (1) diffusion of ions from solution to the electrode surface, (2) electron transfer reduction of ions, (3) partial or complete stripping of the ion's solvation shells, forming ad-atoms (adsorbed atoms), (4) diffusion of ad-atoms along the surface, (5) cluster formation of ad-atoms with formation of so-called critical nuclei (see below), (6) incorporation of ad-atoms at lattice sites and (7) growth of the nuclei to mesosocopic size and beyond. In vapour-phase deposition, homogeneous nucleation is described by the *Kelvin equation* which relates equilibrium vapour pressure p^*, the radius r of the formed droplet, the surface free energy γ and the molar volume of the condensed phase:

$$\ln\left(\frac{p^*(r)}{p^*(\infty)}\right) = \left(\frac{2\gamma V_{\mathrm{mol}}}{RT}\right)\frac{1}{r}. \tag{4.6}$$

The equilibrium vapour pressure ($p^*(r)$) depends on the curvature radius of the formed nuclei. Spontaneous growth occurs if the equilibrium vapour pressure is exceeded, i.e. when the vapour is supersaturated. The supersaturation ratio p/p^* is related in this strictly thermodynamic picture to the Gibbs free energy change:

$$\Delta G = -RT\ln\frac{p}{p^*}. \tag{4.7}$$

The overall free energy change is obtained by summation of the bulk and surface contribution. It can be shown that the bulk contribution is negative for $p > p^*$, whereas the free energy change of the surface, $\Delta G_{\mathrm{surf}} = 4\pi r^2\gamma$, is positive. For a given p^*, surface and volume contribution combine to produce a maximum in ΔG dependent on the radius of the nucleus formed as shown in Fig. 4.11. One defines the critical radius and the critical free energy by $r_c = r(\partial\Delta G/\partial r = 0)$ and $\Delta G(r = r_c) = \Delta G_c$. Stability demands that the radius of the nucleus, $r > r_c$. From

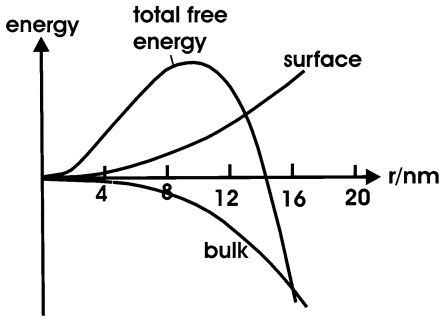


Fig. 4.11 Schematic of the contributions of surface and volume parts to the total free energy F in homogeneous nucleation from the gas phase for a spherical Hg droplet at $p/p^* = 3.2$, $T = 300\,\mathrm{K}$; F (surface) and F (bulk) are shown reduced by a factor of 10

Fig. 4.11, it is obvious that this necessitates overcoming of the activation barrier for homogeneous nucleation, ΔG_c, by thermal fluctuations.

In electrochemical phase formation, the growth occurs at the boundary of the substrate and the solution; hence, the process is heterogeneous and the equilibrium is established here by the electron transfer reaction leading to deposition. In the simple case of a spherical nucleus, the total free energy is then

$$\Delta G_{\text{tot}} = \frac{4\pi r^3 z\eta F\rho}{3M_w} + 4\pi r^2\gamma. \tag{4.8}$$

In (4.8), η denotes the potential applied beyond the equilibrium value for the reduction reaction $M^{z+} + ze^- \leftrightarrow M^0$, called the overpotential. ρ, M_w and F are density, molecular weight and Faraday constant, respectively. Differentiation of (4.8) with respect to the radius r results in an expression for the free energy of activation with the proportionality $\Delta G_c \propto \gamma^3/\eta^2$. The classical theory of nucleation, thus, states that the increase of the overpotential quadratically decreases the activation barrier $(1/\sqrt{\Delta G_c})$.

The reaction kinetics are described by considering monolayer discs of given radii that can be formed using the time dependence of the current during growth. If one defines the number of active sites for electrocrystallization on a particular surface by N_0 and the rate of incorporation by k_i, two modes can be distinguished (1) in *instantaneous* growth where the number of nuclei formed is time independent and clusters grow radially; (2) *progressive* growth where clusters grow and, in addition, their number increases linearly with time. For instantaneous growth, the current dependence is $I_I = \text{const.} N_0 k_i^2 a_0 t$, where a_0 is the vertical lattice constant of the deposit. For progressive growth, the current is given by $I_P = I_I k_{\text{nuc}}\, t$, where k_{nuc} is the rate constant for nucleation of critical nuclei. Hence, the progressive growth shows a quadratic time dependence. If one is interested in the growth of monolayers instead of nano-sized discs, the overlap of growing nuclei is described by the Avrami theorem which connects the normalized developing surface coverage (area) A_S by overlapping discs to the total surface areas of the discs A_T (as if they were not connected): $A_S(t) = 1 - \exp(-A_T(t))$. For monolayer formation, the temporal profile of the normalized current for instantaneous and progressive growth is shown in Fig. 4.12. Plotted are currents and time related to their maximum values, I_{max}, t_{max}, during growth obtained from the relation $I(t) \propto \mathrm{d}A(t)/\mathrm{d}t$ and using Avrami's relation:

$$I_S/I_{\text{max}} = t/t_{\text{max}} e^{-\frac{\left(t^2 - t_{\text{max}}{}^2\right)}{2t_{\text{max}}{}^2}}, \tag{4.9}$$

(for instantaneous nucleation) and

$$I_S/I_{\text{max}} = (t/t_{\text{max}})^2 e^{-\frac{2\left(t^3 - t_{\text{max}}{}^3\right)}{3t_{\text{max}}{}^3}}, \tag{4.10}$$

for progressive growth. Of course, the progressive growth mode results in an earlier completion of the monolayer growth as is also seen in the figure below.

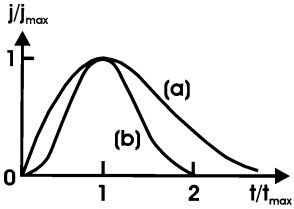


Fig. 4.12 Transient currents for nano-disc monolayer growth in two-dimensional electrocrystallization for instantaneous (*a*) and progressive (*b*) nucleation

4.2.4 Nanostructure Formation on Silicon

Preparation of advanced Si micro- and nanostructures requires detailed knowledge of dissolution mechanisms that includes reaction intermediates and rate-determining steps. Historically, the often used "complementary selectivity" in etching between the material itself and its oxide has been used to fabricate numerous devices (see below) [42]: whereas Si readily dissolves in alkaline solution, silicon oxide etches only very slowly. Fluoride-containing solutions, on the contrary, etch the oxide but barely attack silicon. Compared to this chemical behaviour, electrochemical treatment can lead to dissolution of Si, for instance, in fluoride containing solutions and it can drastically reduce the Si etch rate in alkaline electrolytes [43].

The basic (photo)electrochemistry of Si in acidic HF containing and in alkaline electrolytes is displayed in Fig. 4.13. Only the current branches, needed for the following analyses, are presented: in the cathodic range (negative current), the behaviour in alkaline solution is displayed and the anodic branch shows the behaviour in dilute ammonium fluoride solutions at moderate acidic pH (∼4). Four parts can be discerned in Fig. 4.13 (1) the cathodic region, where H_2 evolution occurs and so-called step bunching is observed [11, 12]; (2) the potential region before the first current maximum of the anodic branch where porous Si is formed during so-called divalent dissolution; (3) the electropolishing regime and (4) oscillatory behaviour at increased anodic potentials.

Below, the surface topographies for (1), (2) and (4) will be closer analysed: applying an anodic potential below the first (photo)current maximum and illuminating n-type Si leads to porous silicon formation in fluoridic electrolytes [44, 45].

For oxide etching, the overall reactions are

$$SiO_2 + 6\,HF \rightarrow SiF_6^{2-} + 2\,H_2O + 2\,H^+, \tag{4.11}$$

$$SiO_2 + 3\,HF_2^- \rightarrow SiF_6^{2-} + 2\,OH^- + H^+. \tag{4.12}$$

The contribution of HF to the etch rate k_e is about four times smaller than that of HF_2^-:

$$k_e = A[HF] + B[HF_2^-] + C. \tag{4.13}$$

The concentrations of HF and the hydrogen bridge compound HF_2^-, indicated by the parentheses in (4.13) are influenced by the solution pH:

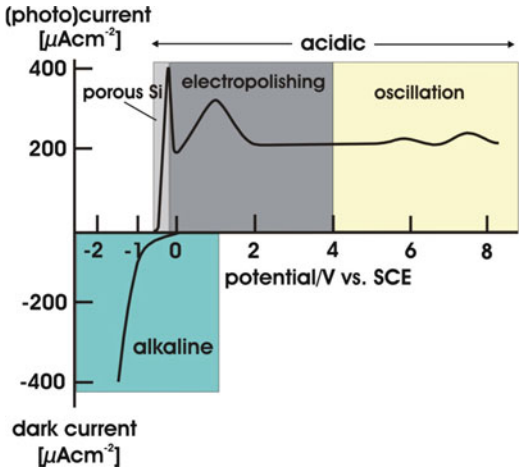


Fig. 4.13 (Photo)electrochemical behaviour of Si (1 1 1) in 2 M NaOH (cathodic branch, labelled alkaline) and in 0.1 M NH_4F (pH 4) (anodic branch; the respective potential regions associated with the electrode behaviour are indicated; potentials are referred to the saturated calomel reference electrode (SCE); the shift in current onset is due to differences in pH

$HF \Leftrightarrow H^+ + F^-$ with the dissociation constant $K_1 = [H^+][F^-]/[HF] = 1.3 \times 10^{-3}\,\mathrm{mol\,l^{-1}}$. For the HF_2^- species, one obtains for $HF + F^- \Leftrightarrow HF_2^-$ with $K_2 = [HF][F^-]/[HF_2^-] = 0.104\,\mathrm{mol\,l^{-1}}$. The constants A, B and C have been originally determined to have the values 0.25 $(\mathrm{nm\,s^{-1}})$ $(\mathrm{l\,mol^{-1}})$, 0.966 $(\mathrm{nm\,s^{-1}})(\mathrm{l\,mol^{-1}})$ and -0.014 $(\mathrm{nm\,s^{-1}})(\mathrm{l\,mol^{-1}})$, respectively [46].

Noticeable concentrations of the hydrogen bridge compound HF_2^- exist for solutions with pH between ~3 and 5. In alkaline electrolytes, silicon dissolves by solvolytic splitting of backbonds of surface atoms by water molecules, a process that occurs initially, preferably at kink sites:

$$Si + 3H_2 \rightarrow HSi(OH)_3 + H_2. \tag{4.14}$$

The dissolution reaction is a complex mixed chemical–electrochemical multistep reaction [47, 48]. For analysis of the proposed dissolution models, SRPES can be used to disentangle the dissolution reaction sequence if the observed surface core level shifts can be attributed to Si surface species whose partial charge has been determined [49].

Understanding nanostructure formation on Si by (photo)electrochemical conditioning allows one to assess the contributions from important parameters of this model system. In particular, the early stages of porous Si formation are of interest. For a comprehensive analysis, the chemical as well as the structural surface changes with increasing dissolution are relevant here. Therefore, AFM and SRPES are employed here, the latter using the photoelectrochemistry – SRPES experiments with the modified surface analysis system shown in Figs. 4.1 and 4.2. For a detailed analysis, the use of dilute fluoride containing solutions has been shown to allow the investigation of the very initial stages of the formation of porous Si where the first observable nanopits, for example, are formed. Figure 4.14 shows contact-mode AFM data that have been recorded for increasing the anodic polarization of

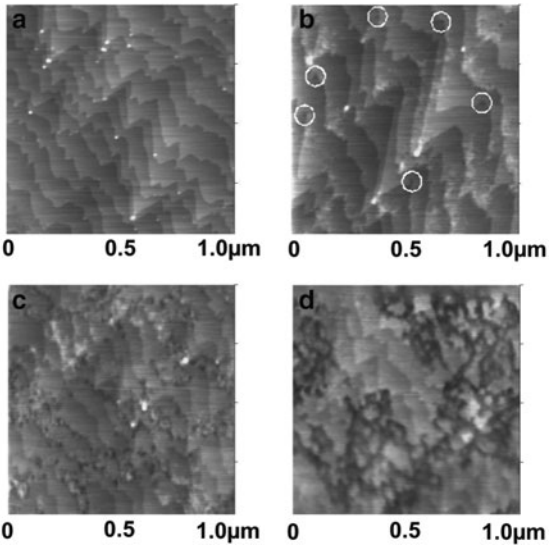


Fig. 4.14 AFM images of nanotopography development on *n*-Si(1 1 1) by photoelectrochemical conditioning in dilute NH_4F at pH4 (**a**) H-terminated surface; (**b**) 0.1 V anodic of open circuit potential (compare Fig. 4.13); (**c**) after 15 s at 0.1 V; (**d**) after 60 s at 0.1 V

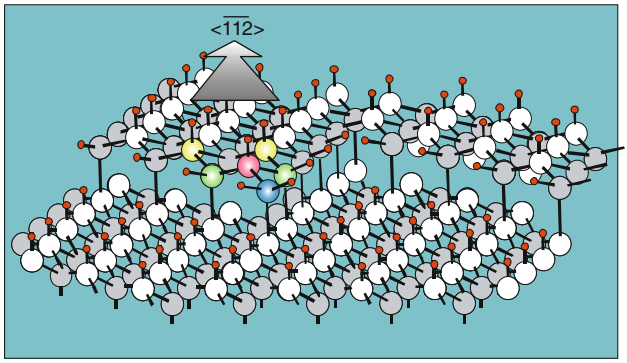


Fig. 4.15 Atomistic model of Si dissolution in acidic fluoride containing electrolytes (see text)

illuminated n-Si (1 1 1) from the open circuit potential to the region before the first photocurrent maximum. The associated nanotopography, from the H-terminated surface (Fig. 4.14a) to the stronger corroded one (Fig. 4.14d), shows the successive development of pits and corrosion. It is observed that the (1 × 1) H-terminated surface with a characteristic zigzag structure (where single Si–H bonds are exposed at the step edges) changes the alignment of the step edges as predicted in a model [50] and initial pitting occurs at the re-entrant sites of the structure, indicated by the white circles in Fig. 4.14b, which is then followed by pitting along the more aligned step edges for slightly increased charge passed. In a later stage, remnants of terraces are visible which lead to a metastable stage forming nano-sized flat Si mesas [51]. Figure 4.15 shows a schematic visualization of the successive terrace dissolution process that commences at the kink site atoms at the tips of the terraces.

The investigation of the initial *chemical* stages of porous Si formation becomes possible by combined (photo)electrochemistry-UHV surface analysis using the *in-system* set up shown in Figs. 4.1 and 4.2. The photoelectrochemical dissolution process was interrupted at three anodic potentials below and near the first current maximum according to the AFM experiments and the samples were transferred directly without any further contact to ambients into the buffer chamber of the UHV surface analysis system and subsequently into the analysis chamber. The results are presented in Fig. 4.16. In these SRPES experiments, the high spectral resolution and the tunability of the photon energy allow analysis of surface species predominantly based on their chemical shift [52].

Figure 4.16 shows the Si $2p_{1/2}$–$2p_{3/2}$ core levels, recorded for a photon energy of $hv = 170$ eV, for the three surface conditions. These conditions correspond closely to those of the AFM images of Fig. 4.14. As seen in Fig. 4.7, at this photon energy the mean inelastic scattering length for photoelectrons is very small ($\approx$0.4 nm). Therefore, elastically emitted electrons originate from the topmost surface region, providing a very high surface sensitivity. In these experiments on initial corrosion of Si, float zone material (FZ) has been used as it has an oxygen content that is lower by two orders of magnitude compared to the standard material of microelectronics, Czochralski (CZ) Si. In addition, FZ-Si exhibits much less growth defects. This is important if one wants to investigate topographic details of the dissolution process which should not be affected by corrosion at structural or chemical defects (oxygen in Si backbonds). An electrochemically H-terminated sample (Fig. 4.16a), a slightly anodized sample according to situation (c) in Fig. 4.14, with dissolution charge corresponding to 1 atomic bilayer (Fig. 4.16b), and the surface after emersion near the first photocurrent maximum, where pit formation and terrace dissolution is extended (Fig. 4.16c), are compared with regard to the deconvoluted signals under the respective envelope photoelectron spectrum.

It can be seen that even for the H-terminated sample, the ratio of the spin orbit split 3/2 and 1/2 peaks deviates from the 2:1 value expected by spin–orbit coupling $J = L \pm S = 3/2$ or 1/2 for a p orbital ($L = 1$), as can be seen from the envelope curve. For pure material, the difference in area is given by the multiplet splitting, $2J + 1$, i.e. 4:2. The deviation seen in the figure indicates that Si surface atoms have bonded to (foreign) surface species and hence only a part of the overall Si 2p line is due to fully bulk coordinated Si. For part (a), the spectral deconvolution reveals two additional features which are shifted vs. the bulk line ($E_B = 99.2$ eV) by 0.23 and 0.8 eV. Inspection of (b) and (c) for successively roughened surfaces (compare Fig. 4.14) shows additional lines that are shifted towards higher binding energies by 1.0, 2.3 and $\sim$3.5 eV. The straightforward assignments of the observed deconvoluted spectra are the surface core level shift, already mentioned above (compare Fig. 4.3) and the occurrence of a signal shifted by 3–4 eV with respect to the bulk Si line. The latter can be attributed to the formation of an anodic silicon oxide, SiO_2. Inspection of the respective line shows that oxide formation takes place at potentials that are close to the first current maximum and indeed, a transition from the so-called divalent Si dissolution that leads to roughening and porous Si formation to a tetravalent mode where four holes are consumed to form SiO_2 has

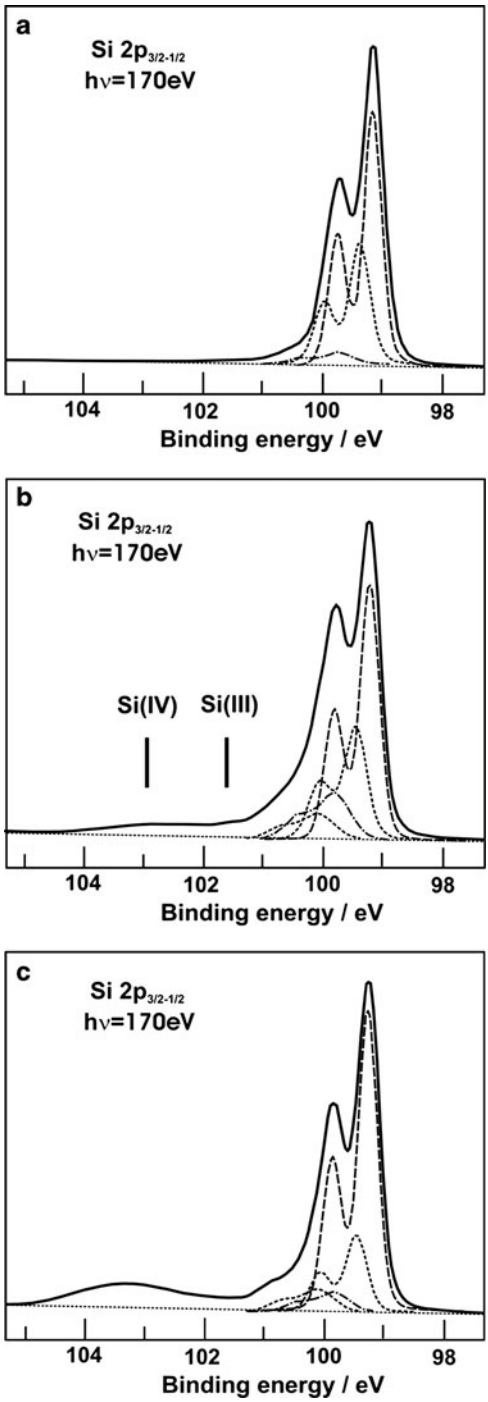


Fig. 4.16 Comparison of synchrotron radiation photoelectron spectra on initial stages of porous silicon formation for the Si 2p core level; photon energy $h\nu = 170\,\text{eV}$; (**a**) electrochemically H-terminated FZ *n*-Si (1 1 1) surface; (**b**) after 1 bilayer dissolution in the divalent dissolution regime; (**c**) emersion near the first photocurrent maximum in Fig. 4.13; the line position of the Si $2p_{3/2}$ peak has been adjusted to 99.2 eV to exclude band bending effects that vary with the surface condition (compare Fig. 4.6) thus allowing comparison of the chemical changes induced by the photoelectrochemical treatment; *full line*: envelope curve; *dashed line*: signal attributed to bulk Si; *dotted line*: with maximum closer to the $2p_{3/2}$peak: surface core level shift; *dashed–dotted line*: signal attributed to =Si–H–OH surface species; *dotted line*: shifted to higher binding energies: signal attributed to =Si–H–F surface species

been observed [53]. The chemical identification of the additionally observed lines, indicated by a lower binding energy dotted line, a dashed and a further dotted one (see figure caption) in Fig. 4.16, becomes possible by determining the partial charge transfer from Si to the various possible reaction intermediates according to proposed dissolution mechanisms [48–50]. For the Si–H bonds on the (1 1 1) surface, a simple model that allows calculation of the potential drop in a microcapacitor has been rather successful [14]. Using Sanderson electronegativities [54], the potential drop across an interface that consists of single H-terminated former dangling Si bonds (n_B) has been qualitatively calculated using the expression

$$V = \frac{q n_B \Delta\rho d_{Cov}}{\varepsilon\varepsilon_0 A}. \tag{4.15}$$

In (4.15), $\Delta\rho$ denotes the partial charge transfer between Si and H due to electronegativity differences that have been determined to be $\Delta\rho \sim 0.1$; d_{Cov} denotes the sum of the covalent radii of Si and H; hence, $d_{Cov} = 0.11\,\text{nm}$ (Si) + $0.042\,\text{nm}$ (H) $= 0.153\,\text{nm}$ [55]. For n_B, the number of Si–H bonds on a (1 1 1) surface is taken, yielding $8 \times 10^{14}\,\text{cm}^{-2}$. The largest difficulty is the appropriate assessment of the static dielectric constant at the interface, ε. Often, a mixture of that of the adjacent media is taken. In the case of H-terminated Si, a value of 6 has been assumed [14, 54]. The energetic shift is obtained by multiplication with the electronic charge and yields values of $\sim$250 mV. Such shifts are indeed observed (see Figs. 4.3 and 4.16). For the interaction of Si dangling bonds at kink sites and step edges with electronegative species such as OH and F, DFT calculations have been used to determine the partial charge of the reacted Si surface atoms [56]. Such calculations have been carried out for a number of possible Si surface species that have been taken from a model which, originally, has been formulated based mostly on chemical intuition. In Fig. 4.17, the proposed reaction mechanism [50] in conjunction with DFT calculation of the partial charge on the Si kink site atoms on the surface (DFT) is shown in detail.

The reaction commences at kink sites on the (1 1 1) surface where only two backbonds to the Si exist, as shown in Fig. 4.11. A light-generated excess hole which reaches this site produces a Si radical by splitting off a proton which enters into the solution. The consecutive step assumes water splitting by the radical and electron injection from a solution species into the conduction band (e^-_{CB}). Electron injection processes at semiconductor/electrolyte interfaces are well documented [57]. These two initial steps are the electrochemical steps involving the transfer of two charges; because of this, the corresponding potential regime is called the divalent dissolution range of silicon. In this potential range, the electrode is not covered by an oxide as has been found in XPS and SRPES measurements [58, 59]. Steps (3) and (4) are of chemical nature and involve ligand exchange and solvolytic splitting of backbonds. The subsequent solution reactions are condensed in step (5) and lead to the soluble silicon hexafluoride anion. For the Si surface atoms with one hydrogen and one hydroxide terminating the dangling bonds, the partial charge is 0.45 and after ligand exchange OH $\leftrightarrow$ F, $\rho = 0.54$.

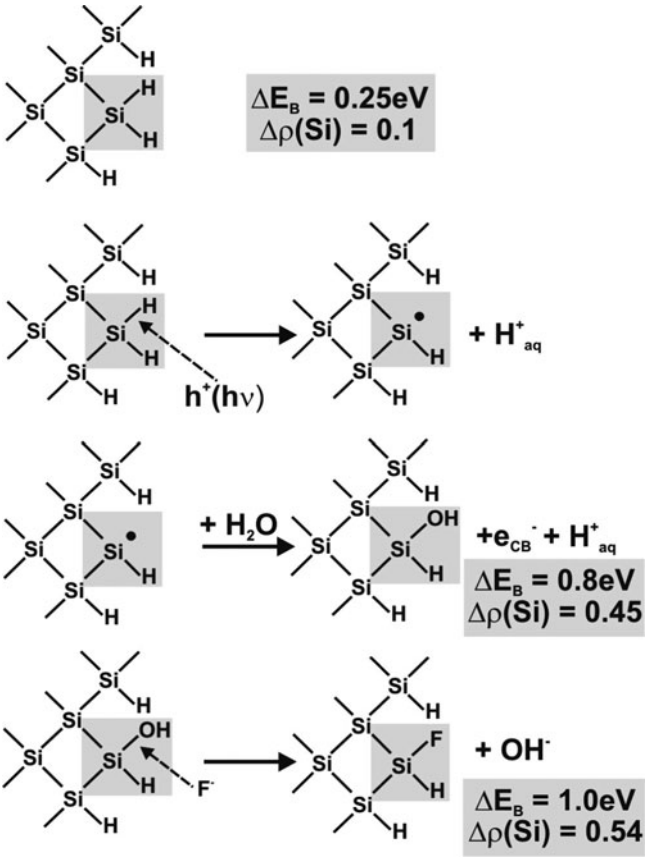


Fig. 4.17 Schematic on the first three steps of the divalent dissolution of illuminated *n*-Si in fluoride containing acidic solution; the dissolution commences at a kink site (*shaded grey*); also given are the observed binding energy shifts ΔE_B, measured by SRPES and the corresponding partial charge change of the Si kink site atom, $\Delta\rho$; the top image shows the unaffected kink site with a partial charge change due to Si–H bonds (the number of kink sites is considered too small to yield an measurable signal on a smooth H-terminated surface); the further reaction steps are discussed in the text

Subsequently, further chemical processes take place; they comprise (1) solvolytic splitting of the first backbond, leading to a probably short lived intermediate that is characterized by one remaining bond to the underlying Si lattice, (2) dissolution of the Si kink site species by an additional reaction with water with a simultaneous release of hydrogen. H_2 evolution has been indeed observed at anodic potentials under illumination in the divalent dissolution regime, (3) reaction with HF, (4) water, (5) HF with the final result of the conversion of the tetrahedral Si compound (sp^3 hybrid) into an octahedral one (sp^3d^2 hybrid) in solution. The partial charge changes could partly be followed by SRPES. Of course, only those intermediates that exist

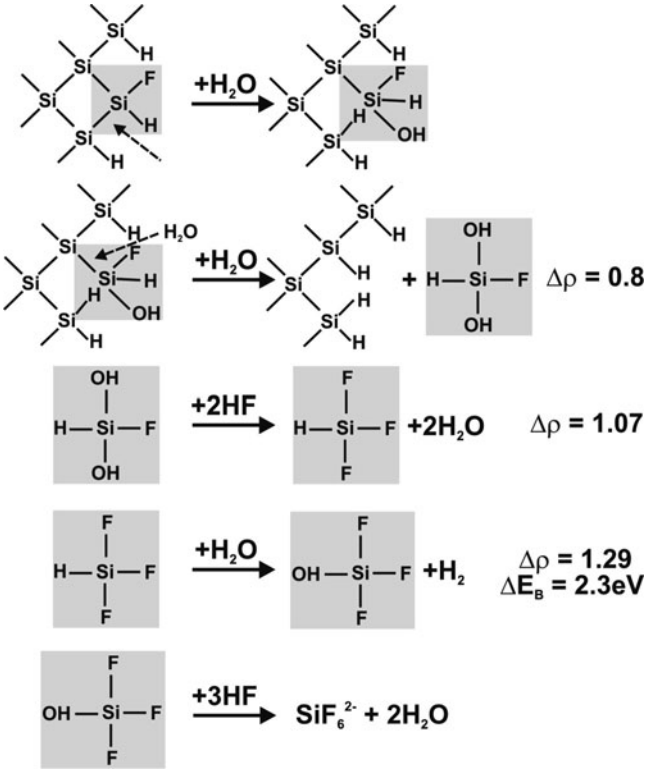


Fig. 4.18 Schematic on the final (purely chemical) reaction steps of the divalent dissolution of illuminated *n*-Si; also given are partial charge changes and measured binding energy shifts (see text)

long enough in detectable amounts and that have a sufficiently strong binding energy to survive the rinsing, drying and outgassing procedure before introduction of the sample into the UHV analysis chamber can be observed. For the species with $E_B = 2,3\,eV$, this was possible. It is assumed that the reactions with water are fast, resulting in short lived intermediates that could also further react in the rinsing procedures. Therefore, the products $His(OH)_2F$ and $HSiF_3$ could not be determined. Future experiments on the detailed study of the semiconductor–liquid interface using high pressure XPS [60] might allow a more direct measurement of reaction intermediates as such experiments can be made in moist environments (Fig. 4.18)

The formation of pits and the electronic state energies of the various reaction intermediates have not been detailed in the model. For pitting to occur, fully coordinated Si atoms have to be dissolved in a three-dimensional process. The finding of initial nanoscopic pits at re-entrant sites in the used dilute ammonium fluoride solutions at pH 4.5 suggests a specific double layer arrangement with counterions adsorbed at specific sites. The force balance that exists between the

electrostatic attraction of negative ions to the slightly positively charged surface (at small anodic potentials), hydrophobic–hydrophilic repulsion and possible van der Waals binding has been postulated to result in an anisotropic electrostatic field perpendicular to the surface [61]. Upon illumination, excess minority carriers become deflected to these sites where they are able to oxidize the topmost Si atoms. If two positive charges reach a surface atom, this will suffice to weaken the backbonds such that successive solvolytic splitting with dissolution can occurs. The dissolved species has not yet been identified. It appears likely that it is the species $H(OH)_2F$, formed by two anodic charges, too, although in a different process. The subsequent reactions then lead to the formation of tetra- of hexa-fluoridic compounds (SiF_4, SiF_6^{2-}).

In Fig. 4.19, the relation between observed binding energy shift and the DFT calculations yields a linear relationship. The slope of the binding energy shift with the surface Si atom partial charge change is approximately 2 and the linear dependence of all measured and calculated species lends strong support to the dissolution model of [50]. Since only strongly bound surface species or precipitates withstand the rinsing and drying procedure of the sample transfer from electrochemical environment into UHV (see Figs. 4.1 and 4.2), the finding of the species =Si–H–OH and =Si–H–F indicates either that the subsequent steps, ligand exchange and two-step solvolytic backbond splitting are possibly rate determining or that the bonding is strong enough to withstand the sample transfer procedures. It can also be seen from Fig. 4.16 that the relative peak area for the signal with $\Delta E_B = 0.23\,\text{eV}$ decreases with anodization; this shows a ceasing of the (1×1) H-terminated areas upon roughening. Energetically, the kink site at which terrace dissolution according to Fig. 4.17 originates should be located within the Si energy gap, allowing photogenerated holes to be trapped at such sites and consecutively setting off the reaction. The subsequent formation of the surface radical results in a pronounced shift to lower electron energies due to the high oxidation power. Energetically, the radical should be able to split water as shown in Step (1) of the reaction scheme. The postulated electron injection into the conduction band will then need a substantial activation energy. This can, for instance, be provided by a nucleophilic substitution reaction (SN2) in which a penta-coordinated Si transition complex is formed, as has been suggested for Si dissolution in alkaline media [48]. The judicious choice of the electrochemical parameters necessary to analyse the details of the dissolution reaction has, in consequence, led to the preparation of nanopores and nano-sized mesa-type self-organized features [51].

It has been demonstrated that the step-bunched Si surfaces possesses rather unique properties with regard to the immobilization of proteins and even complete photosystems [62–64]. On these surfaces, single atomic terraces transform into large smooth terraces with multi-layer high step edges as seen in Fig. 4.20. Step heights typically reach 15 bilayers, i.e. $\sim$5 nm. Interestingly, the step bunched surface of n-Si(1 1 1) is found to be in accumulation as inferred from the XP valence band spectrum in Fig. 4.21 [65]. From the nominal doping, the energetic difference between Fermi level and the conduction band edge is 0.26 eV. The spectrum onset shows an energetic distance of 1.06 eV between valence band maximum (E_{VB})

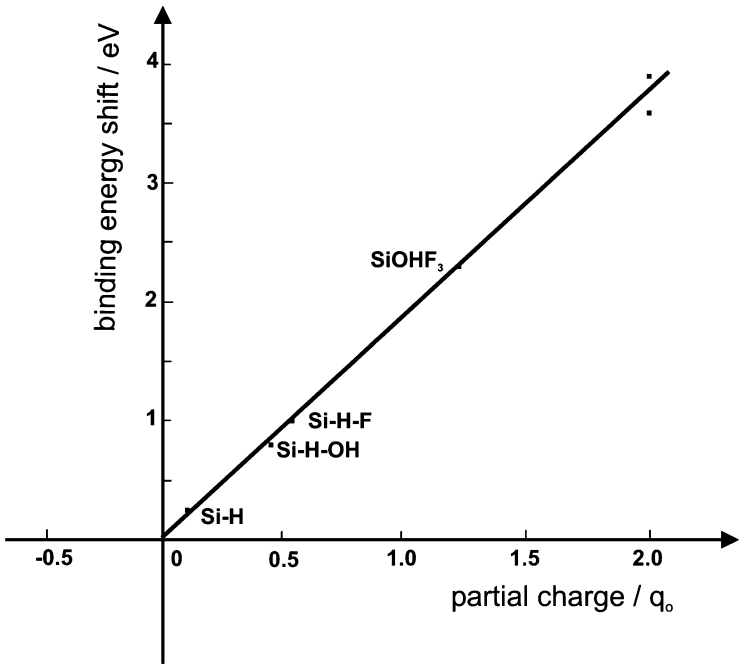


Fig. 4.19 Chemical shift determined in SRPES measurements vs. Si partial charge (in units of the elementary charge q_0) for a series of dissolution reaction intermediates (see text); the two points for $q_0 \sim 2$ correspond to Si in silicon dioxide

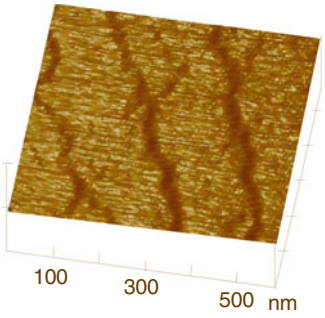


Fig. 4.20 Contact mode AFM image of an electrochemically prepared step bunched Si surface

and the Fermi level at the surface, measured by XPS. From the doping and the energy gap (1.12 eV), a position of E_F of 0.86 eV above E_{VB} is expected for the flatband situation. The data show, however, that the Fermi level is closer by 0.2 eV to E_{CB}, evidencing either the presence of an accumulation layer that remains stable in UHV environment or doping by H-diffusion due to H_2 evolution during the preparation of these surfaces. SRPES allows us to analyse the depth dependence of the change of the Galvani potential upon variation of the photon energy and

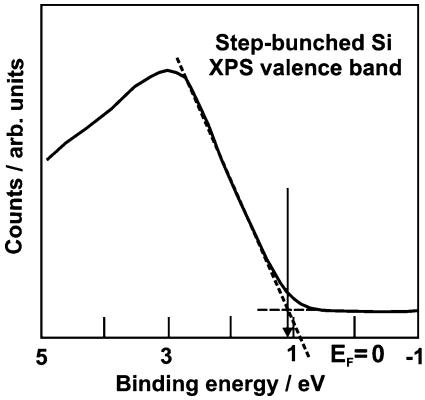


Fig. 4.21 XP-valence band spectrum of step-bunched Si, recorded for a photon energy of 150 eV; binding energy is given as usual, with respect to the Fermi level; the extrapolation of the emission onset shows the energetic distance between valence band maximum and the Fermi level.

analysis of the peak position of the Si 2p line. As result, a depth of about 2 nm below the topmost surface has been extrapolated [65], supporting the concept of an accumulation layer that is expected to be localized at the surface in a depth of about 1–3 nm, whereas H diffusion profiles are expected to extend considerably deeper into the material. Conclusive statements can be expected from ERDA (eleastic recoil detection analysis) measurements that allow profiling of H in materials [66].

An accumulation layer or a highly n-doped layer at the surface (by H^- incorporation [67]) would generate a two-dimensional electron gas at/near the surface. Besides for fundamental studies of semiconductor–biomolecule junctions, these structures are also of interest for the further investigation of the quantum Hall effect [68, 69] as well as for semiconductor spintronics [70].

The energy band diagram for this situation is presented in Fig. 4.22, assuming an accumulation layer with a thickness of about 2 nm at the surface. In this case, as can be seen, the bands of the n-type semiconductor are bent such that the Fermi level is closer to the conduction band at the surface than in the bulk, due to the doping concentration.

4.2.5 Device Preparation by Wet Processing

Although industrial applications tend towards chemical etching treatments, fabrication of advanced devices often uses mixed chemical/electrochemical processing, similar to the hybrid dissolution mechanism described above. In electroless etching of Si, the semiconductor is first oxidized and the oxide is removed in the same solutions. Etchants such as CP4 which contain a mixture of H_2O_2, HNO_3, HF and CH_3COOH are able to remove, for instance, damaged surface layers from cutting and polishing of wafer material.

In micro- and nanomachining, the central etching property exploited is the crystallographic anisotropy showing that the etch rate decreases in the series (1 1 0),

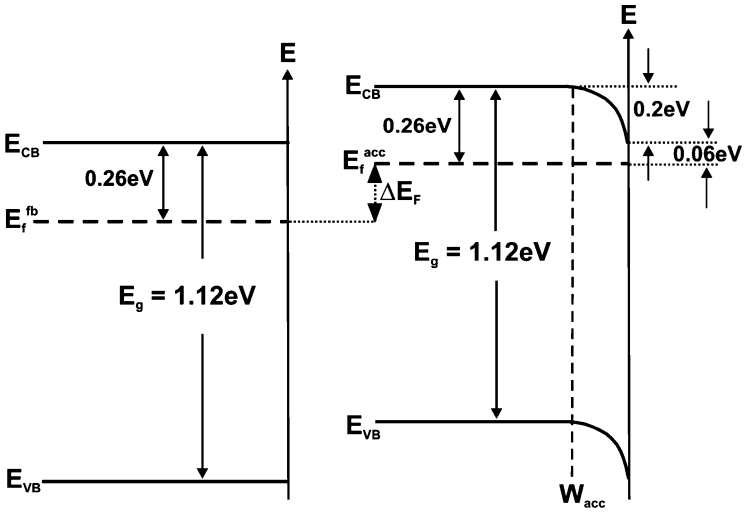


Fig. 4.22 Energy band diagram perpendicular to the Si (1 1 1) surface for flatband situation (*left*) and after step bunching, constructed using the photoelectron spectroscopy data of Fig. 4.21 and [65]; E_{CB}, E_{VB}, conduction and valence band, respectively; E_f^{fb}, E_f^{acc}, Fermi level for the flatband and the accumulation condition, respectively; ΔE_F, difference in energetic position of the Fermi levels due to doping and after step bunching treatment; W_{acc}, extension of the accumulation layer (see also text)

(1 0 0) and (1 1 1) orientation. The latter surface is most close and densely packed which is often taken as an argument to explain its higher stability. A simple example is the preparation of V-shaped grooves on Si(1 0 0). A SiO_2 mask was aligned in stripes along ⟨1 1 0⟩ direction and etching was applied at the openings in this mask where Si was exposed to alkaline solution. The anisotropy of the etching process resulted in the formation of inclined (55°) (1 1 1) facets as side walls of the V shape with dimensions into the submicrometer range. Such Si V-groove chips are used for alignment purposes in coupling fibres to semiconductor lasers by sandwiching a fibre cable between two V-groove shaped chips. The actual alignment is done using a negative V groove pattern, clamped to the positive pattern, as alignment guide [71]. Surface patterning pertains also to the fabrication of anti-reflection structures for solar cells. An early example is the early Comsat non-reflecting (NR) cell. Alkaline etching of the Si(1 0 0) base surface resulted in the preparation of (1 1 1) facets in a similar manner as for the V-shaped grooves. Adjusting the dimensions of the surface structure to the main part of the solar spectrum resulted in a strongly reduced reflectivity and in the refraction of the impinging light rays due to deflection of the rays away from the direction of the geometrical surface normal. As a result, the red sensitivity of the solar cell increased since the light-induced minority carriers were generated closer to the surface. Figure 4.23 shows the geometry of the surface texturing [72].

More recently, the formation of high-quality inverted pyramids has contributed to high efficiencies in silicon solar cells. The structuring has been applied to the

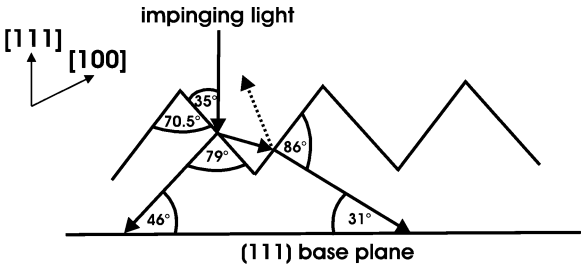


Fig. 4.23 Microtopographic surface texture of the Comsat NR cell for enhanced light coupling into crystalline Si solar cells

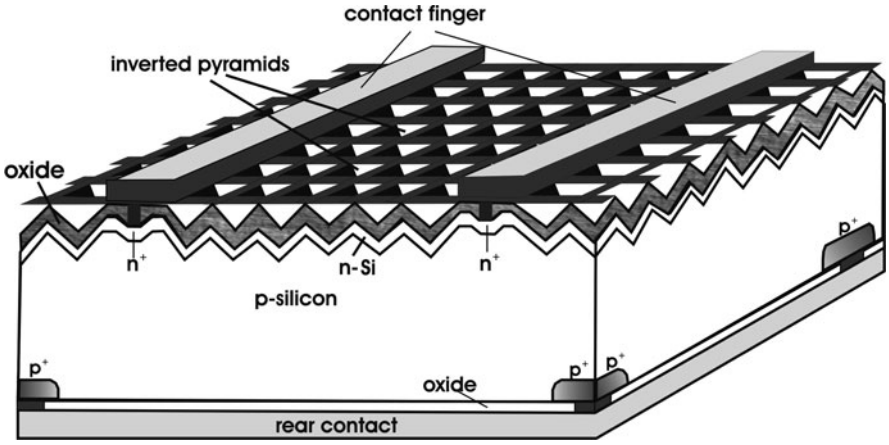


Fig. 4.24 Schematic of a silicon based PERL solar cell

so-called PERL solar cell (PERL: passivated emitter rear locally diffused), with multicrystalline silicon as base material. Quite different from traditional solar cell configurations using planar homo- and heterojunctions, the operational principle of this cell is based on the collection of the light generated carriers (electrons) at rather small highly doped n^+ emitter contacts in stripe form at the top surface. The ohmic back contact is made by highly doped locally diffused p^+ regions into the rear of the absorber material. A solar-to-electrical conversion efficiency of 19.6% was obtained with this structure [73]. Figure 4.24 shows an example.

Shaping of semiconductors by passivation is usually done by so-called etch stop techniques. The term passivation stands here for the selective inhibition of further materials dissolution (see also above). Etch stop techniques can be realized by a considerable variety of combinations of semiconductors and solutions and can also be implemented using dry methods such as plasma etching [74]. Etch stops, used in micro- and sub-microfabrication, derive their name from the implementation of material boundaries at which an etching reaction stops. Such etch stop boundaries must be based on a physico-chemical difference between materials or parts of them resulting in selective etch rates. In general, two requirements have to be fulfilled for

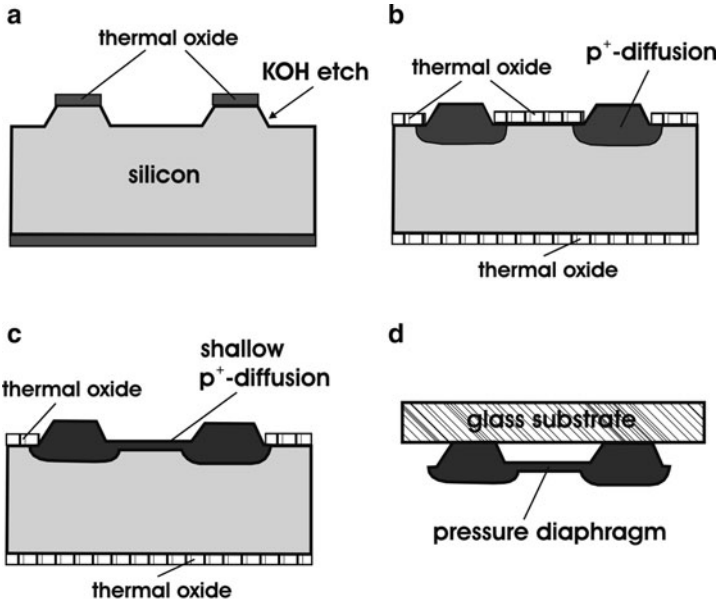


Fig. 4.25 Preparation steps for a pressure sensor based on p^+-Si (see text)

a successful microstructuring by etch stop processing: (1) control of the placement of areas with different etch rates, (2) a differential, i.e. selective dissolution method, the latter being of crucial importance for the process. In nanofabrication, the patterning process (1) also becomes inherently increasingly difficult. We will describe here two examples, one each for a device prepared by a chemical and an electrochemical etch stop.

Fabrication of a silicon pressure sensor uses the observation that the etch rates k_e in KOH and even more so in EPW (ethylendiamine–pyrochatecol–water) are depending on the boron dopant concentration in the semiconductor. For highly boron-doped Si (resulting in p^+-Si) with concentrations $>6 \times 10^{19}\,\mathrm{cm}^{-3}$, $k_e \sim 0$. Despite several explanatory attempts, a clear understanding of this phenomenon is still missing. Nevertheless, the effect has been used for the preparation of a Si-based pressure sensor, for instance. In a four-step procedure shown in Fig. 4.25, a pressure diaphragm was fabricated. In the first step, the selectivity of KOH to etch Si, but much less SiO_2, was used producing a shallow textured surface. Then, the oxide is etched (typically in HF containing solutions) and new oxide is formed by patterning (Fig. 4.25b). Subsequently, the diffusion of B into a depth of 10–20 μm creates a p^+ region. With appropriate dimensioning of the distance between the p^+ regions, an annealing step leads to a redistribution of the boron as shown in part (c) of the figure. After stripping (etching) of the oxide and subsequently the rest of the Si wafer, the structure shown in (d) is obtained. The thin connecting bridge between the initially B-diffused regions works as the pressure diaphragm which is then bonded to a glass substrate.

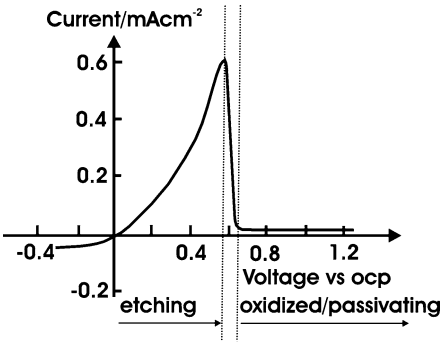


Fig. 4.26 Current–voltage curve of p-Si in 44% KOH at 66°C

Other applications of the p^+ etch stop include the fabrication of SOI (silicon-on-insulator) structures which as alternative substrate material for chip integration [75]. Because of the high boron concentration, induced stress and stress-induced defects allow only the preparation of micromechanically active structures since the p^+ etch stop layer is practically useless for electronic applications.

Among the electrochemical etch stops, we select one based on passivating so-called anodic oxides, i.e. oxides which are formed by applying positive (anodic) potentials to a silicon electrode which acts as an anode in an electrochemical cell:

$$\text{p-Si} + 4h^+_{\text{VB}} + 2\text{H}_2\text{O} \leftrightarrow \text{SiO}_2 + 4\text{H}^+. \tag{4.16}$$

This type of anodic oxidation is also applicable, though to a lesser extent, to Ge, III–V and II–VI semiconductors. The principle is demonstrated in Fig. 4.26. Upon reaching a defined potential positive from the (zero-current) rest potential (anodic overpotential), passivation due to massive oxide formation sets in. The formed oxide insulates the Si substrate from the electrolyte, the charge transfer stops and the current drops to almost zero. Since KOH etches the oxide very slowly, the passivation is not as complete as for EDW.

Etch stops based on this type of passivation behaviour have also been applied to Si pn junctions. This is done by applying an anodic potential to the n-type Si side. As shown in Fig. 4.27, as a result, the Fermi level of the n-Si, $E_{\text{F}}(V_{\text{a}})$, is located energetically below the passivation energy whereas the voltage drop across the reversely biased junction leaves the p-type side at etching potentials. The result is an etch stop once the etch front reaches the n-type side of the junction. A number of microstructures have been fabricated by this and similar methods using either KOH or EPW. Improvements have been made using four-electrode arrangements to control the potential floating of the p-type side. In another application, an n-type layer was generated by biasing a p-type substrate into inversion at an MOS (metal–oxide–semiconductor) [76].

Using small local contacts such as can be provided by scanning electrochemical microscopy (SECM) by contact metal plating or by in situ scanning probe microscopy (SPM) too, new structures with very small lateral dimension can be realized.

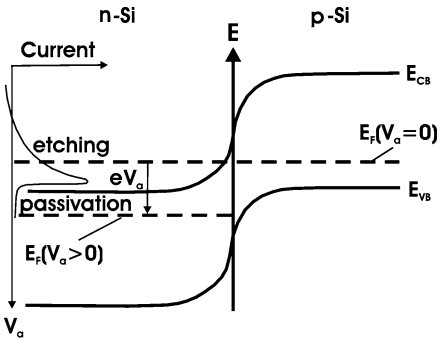


Fig. 4.27 Energy band diagram of a pn-junction biased anodically at the n-type side; V_a applied potential, E_{CB}, E_{VB} conduction band and valence band edge, respectively; $E_F(V_a)$ Fermi level on n-type side after applying the voltage V_a; E_F Fermi level at p-type side; etching occurs from the right-hand side

A high photon energy application, based on plating, is the fabrication of micromotor gears and plastic devices by the LIGA process. The acronym stands for lithography and galvanic abformung, the latter being the German word for shaping. The process uses photolithography techniques and illumination by X-rays [77, 78]. Figure 4.28 shows a typical processing sequence. It makes use of a thick (10–1000 μm) X-ray resist and an according mask. The resist used here was PMMA (polymethylmethacrylate). For the thickness range indicated, the optimum wavelength choice lies between 0.1 and 1 nm; hence, photons with energies of about $1.2\,\text{keV} \leq h\nu \leq 12\,\text{keV}$ from synchrotron storage rings are used. Because of the short wavelength of the radiation and the thickness of the photosensitive polymers, aspect ratios of 300:1 can be obtained. In the second step in Fig. 4.28, the illuminated polymer is dissolved, leaving a mould for electroplating of metallic structures (step 3, galvanoforming). Subsequently, the polymer filling is removed and a gate plate is attached. In the original process, low-viscosity materials such as polymers, plolyimides or other plastics were used for mould filling (step 5). After removal of the metal, a plastic structure, held by the gate plate, has been fabricated.

In microelectronics, higher storage capacity of memory devices (DRAMS) and increasingly faster so-called clock speeds of microprocessors demand correspondingly smaller features. The advancement of this technology depends on: (1) better resolution leading to decreased feature sizes in the lithography step and (2) the possibility to manufacture ICs with dimensions in the far submicrometer range. The latter necessitates the development of dry or wet etching techniques with extremely high shape accuracy and simultaneous electronic quality.

In wet processing, it has been demonstrated that the use of dilute electrolytes results in very local dissolution processes. In this way, pores and moulds of dimensions below 100 nm have been prepared [79]. It appears that preparation of stable nanometre-sized features demands an ideal chemical termination of the semiconductor dangling bonds. Figure 4.29 shows AFM images of self-organized arrays of mesa-type metastable features on Si(1 1 1) and the assignment of the terminating crystal facets stabilizing the side walls of the mesas. It can be observed that the nanosized structures can be constructed from the stable (1×1) H-terminated surface configuration of Si {1 1 1}, {1 1 0} and {1 1 3} faces [17, 80, 81].

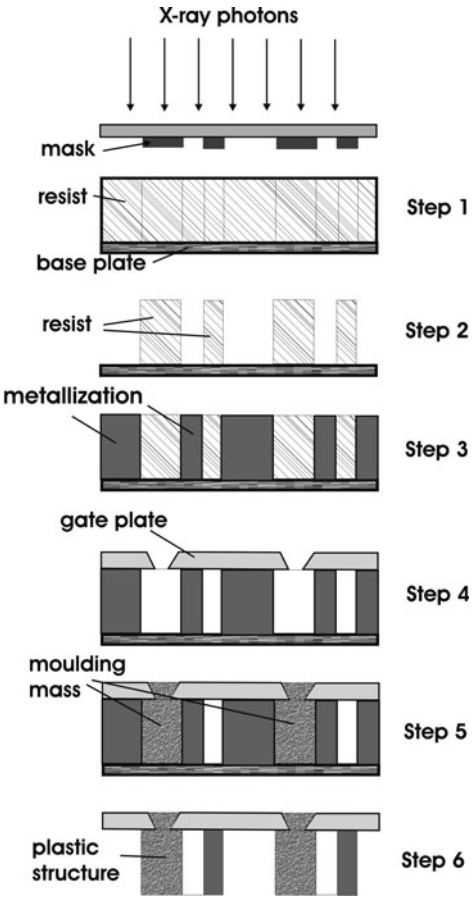


Fig. 4.28 Schematic of a simplified sacrificial LIGA process using synchrotron X-ray lithography for the original metallic mould: *Step 1*: irradiation; *step 2*: development; *step 3*: galvanoforming; *step 4*: mould fabrication; *step 5*: mould filling; *step 6*: unmoulding

4.2.6 X-Ray Emission Spectroscopy: Interface Analysis in Solar Cells and Characterization of Biological Specimen

With the development of third-generation synchrotron radiation sources, XES [82] has experienced a renaissance. In the basic process, the excitation of a core level electron and the subsequent decay of an energetically higher located electron to that core hole, the lifetime of the core hole (1–100 fs) is largely governed by Auger decay; accordingly, the cross section for X-ray fluorescence can be as low as 0.02%. With high photon flux excitation in modern synchrotrons and high sensitivity detection with regard to intensity and photon energy, various interesting new research avenues can be opened. Studies of magnetic circular dichroism, molecular femtosecond dynamics, in situ electrochemical processes and investigation of water/salt/alcohol mixtures have been performed. Due to the average X-ray attenuation length of ∼100 nm, the analysis of complex structures,

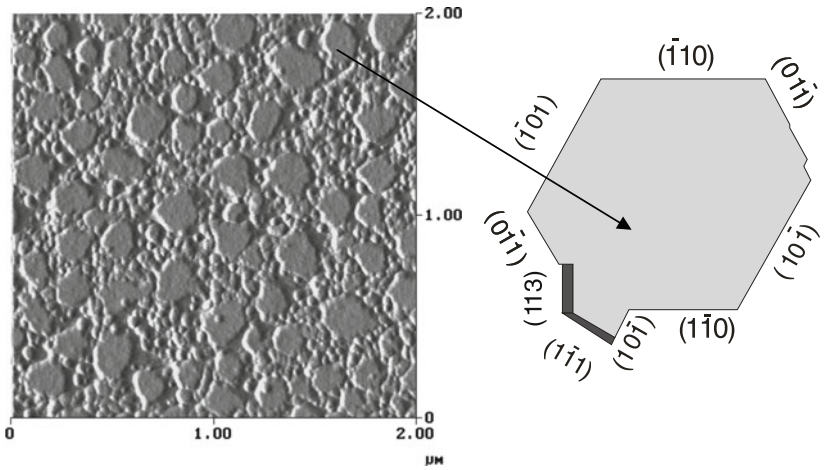


Fig. 4.29 Mesa-type self-organized nanostructures on illuminated n-Si(1 1 1) formed electrochemically under potential control in dilute NH_4F at pH 4

buried junctions and strongly textured materials becomes possible. We select here the investigation of (1) interface conditioning in a polycrystalline solar cell where the signal is expected to originate from surface layers and from "hidden" deeper lying regions between several micrometer large grains and (2) functional transition metals biological molecules.

The basic process is shown in Fig. 4.30: upon X-ray excitation, a core electron is excited leaving an excited core hole (⊕) behind. Decay by an Auger process results in a second excited electron and the filling of the core hole. In XES, the decay from occupied states, here denoted as "bonding" which could also be the valence bands of metals semiconductors, etc. to the core hole, results in a fluorescence by emission of an X-ray photon. This spectroscopy is based on the second-order optical process in which a core electron is excited by the incident photon and the excited core state decays by emission of the fluorescence photon. Several modifications of the method have been implemented: the process described in Fig. 4.30 is often termed normal XES (NXES), where the core electron is excited into the continuum of states. The according final state in this spectator decay is a valence band hole, for instance. The resonant techniques use tuneable photon energies to excite into binding but unoccupied electronic states. Two processes are discerned: resonant elastic and resonant inelastic X-ray scattering (REXS, RIXS, respectively). Figure 4.31 shows a comparison of NXES, REXS and RIXS. In REXS, the final state is the ground state due to participator decay (recombination of the excited electron to the core hole), whereas in RIXS the spectator decay leaves one electron and one hole as final state.

For the soft X-ray regime considered here, the dipole approximation, where the divergence term of the vector potential of the photon field can be neglected (Coulomb gauge), holds as the wavelength of the light is large compared to the dimensions in which electronic excitation and decay occur. The according

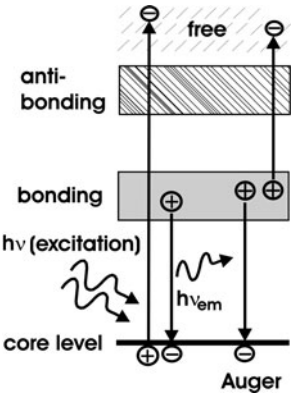


Fig. 4.30 Schematic of the competitive Auger and X-ray emission processes; $h\nu_{em}$ denotes the emitted photons due to electron recombination

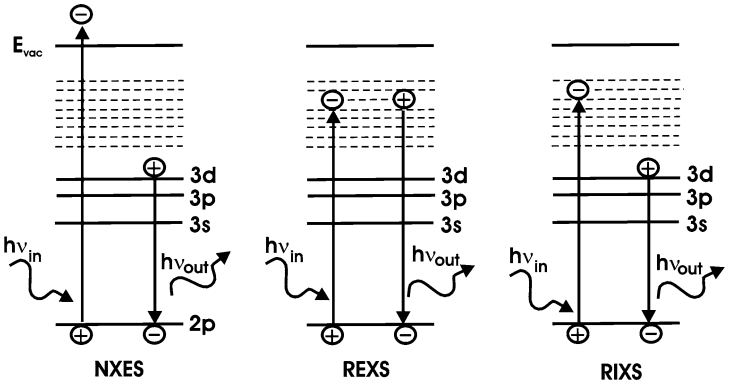


Fig. 4.31 Schematic of "normal" X-ray emission, REXS and RIXS; the *dashed line* field indicates unoccupied but binding states of the same (3d) angular character

selection rules for the angular momentum and the magnetic quantum number are $\Delta l = \pm 1$, $\Delta m = 0, \pm 1$, respectively. Hence, transitions between s and p states and p and d states are allowed, whereas s to d and s–s, p–p transitions, etc., are forbidden. In solids, such selection rules are relaxed for transitions to the valence and lower conduction bands due to hybridization of the wave functions. In LCAO terminology (linear combination of atomic orbitals), this corresponds to summing up the contributions of $|s\rangle$, $|p\rangle$ and $|d\rangle$ states, for instance, to describe a particular wave function. A d-type core level electron can then be excited to such energy level by interacting with the p-type part of the wave function since the dipole transition matrix elements are obtained by summation. The validity of the dipole approximation implies also momentum conservation which allows, for instance in the case of RIXS, one to study electronic bulk band structures [83].

4.2.6.1 Examples I: Interface Engineering for Solar Cells

Ternary chalcopyrite semiconductors such as $CuInX_2$, (X = S, Se) and their alloys with Ga are promising candidates for terrestrial photovoltaic application in thin film devices [84, 85]. Whereas $CuInSe_2$ has been rather rapidly evolved as an efficient thin film solar cell, efficient thin film cells based on the sulphide analogue have been developed over a decade later [86]. The preparation of $CuInS_2$ solar cells with efficiency larger than 10% is made from Cu-rich material with deleterious Cu–S phases, highly p-type, on the surface. The presence of covellite, CuS, leaves these cells photoinactive and a cynide etch has been developed to remove the deleterious phases. Since present $CuInS_2$ solar cells are characterized by toxic components such as a CdS buffer layer and the toxic KCN processing step, wider application is problematic. Therefore, attempts are made to replace the toxic factors by nontoxic or at least less toxic ones. A promising route is the replacement of the KCN etchant by an electrochemical treatment since electrochemistry allows the well-defined steering of processes as shown already above. Several electrochemical treatments have been developed; most of them, however, did yield very low or only moderate (3%) solar cell efficiencies. In an attempt to understand the origin of these restrictions, surface analysis using X-ray photoelectron spectroscopy (XPS) was difficult to evaluate because of the rather small changes of the Cu $2p_{3/2}$ core level (932.7 eV) and the LMM Auger line (337 eV) for the notorious Cu_2S compound and the host, $CuInS_2$. Cu_2S (djurleite) is a remnant after the removal of covellite (CuS) and resisted various corrosion treatments. Its removal was considered pivotal for attaining higher solar cell efficiencies. A clearer picture of the processing with regard to Cu_2S could be obtained from NXES experiments [87] where the spectator decay from the S 3s bands to the S 2p core levels was analysed. Figure 4.32 shows the principle energy

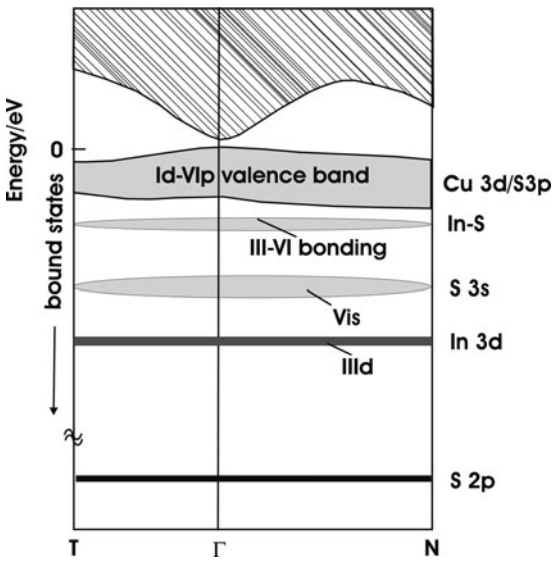


Fig. 4.32 Generic energy band diagram for I–III–VI_2 (ABX_2) chalcopyrite semiconductors including symmetry points of the I–III–VI_2 Brillouin zone; on the *right hand side*, the occupied states for $CuInS_2$ are indicated

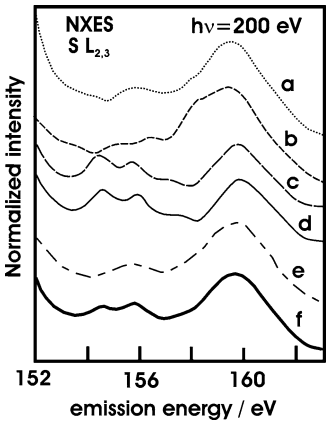


Fig. 4.33 Comparison of NXES data for standards and after chemical–electrochemical treatments; the energy region of transitions from the upper valence band of the S $L_{2,3}$ XE spectra for as-deposited $CuInS_2$ (*a*), a CuS reference (*b*), A $CuInS_2$ crystal (*c*), after KCN etching of the thin film (*d*) and for two (photo)electrochemical etching procedures (*e*) and (*f*) where the former was an alkaline treatment that did not completely remove the Cu–S surface phases and the latter was an advanced treatment in a vanadium redox electrolyte that resulted in a Cu–S free surface (see also text) as seen by comparison with the crystal signature in (*c*)

band structure for ABX_2 chalcopyrite semiconductors (X = S, Se), where the X s-band (3s for the sulphide compounds, 4s for the selenides) is about 2 eV wide and about 6 eV energetically separated from the In–X bonding orbitals (B^{III}–X^{VI} band).

Using an excitation energy of 200 eV, the spectral range of interest is given by the valence band maximum and the core level energy for S 2p (∼162 eV for the 3/2 peak). The lowest energy of the $L_{2,3}$ transition is about 146 eV expected from band structure analysis [88]. The valence band maximum is at 162 eV and the 3s → 2p decay peaks at the allowed $\Delta m = 1$ transition at 148 eV. Figure 4.25 shows a data set of NXES for various treatments also comparing the effect of the cyanide etch with the electrochemical procedures. The most marked differences are found in the peak position around 148 eV and in the upper valence band features displayed in the right hand side of Fig. 4.33. One sees that the unetched $CuInS_2$ (CIS) sample, the Cu_2S reference and the CIS crystal show pronounced differences in the region between 158 and 153 eV. The KCN etched sample shows striking similarity with the signal from the crystal and the electrochemical treatment labelled V1-etch [89] gives a similar result but the feature at 157.5 eV is missing. The K1-etch shows similarity with the Cu_2S reference indicating that, although CuS has been removed, a Cu–S phase remnant remained. Indeed, cells after the K1-etch showed negligible efficiencies whereas after the V1-etch (photoanodic treatment in vanadium redox electrolyte) 8% solar conversion efficiency were obtained.

The V1-etch was shown to exhibit less Cu depletion than the KCN etch and it is assumed that the In-rich surface after the cyanide treatment plays a role in the higher efficiency (11.3% compared to 8%).

4.2.6.2 II: Transition Metals in Biological Molecules

Metalloproteins contain transition metals as the centre for biological functions. Such d-block elements actually determine biological functions by change of valency, spin and by hybridization with ligands. Obtaining direct information about metal d-states, their energy and binding environment is of considerable interest in biochemistry. Optical analysis is often hampered by dipole-forbidden d–d transitions. In XES, however, the possibility exists to excite a p-level ($l = 1$) and to observe the decay d $\rightarrow$ p which is dipole allowed. A problem with application of this technique to biological specimen is the radiation damage. Using a rather radiation resistant molecule, such as myoglobin and freezing them at temperatures well below 77 K, allows identification of the valence and spin state including ligand changes as shown in Fig. 4.34 for ferris and ferric ionic states with high- and low-spin configuration. The dominant transition was from the metal d-states to the Fe $2p_{3/2}$ core level. In this RIXS experiment, the fluorescence intensity is plotted vs. the energy loss with regard to the excitation energy. The shifts are attributed to differences in the hybridization of Fe 3d states with the ligand valence states which reveal themselves in the excitation into unoccupied d-states and in the decay from occupied states towards the excited core hole.

The origin of the shift of the peak in Fig. 4.34 is schematically indicated in Fig. 4.35. The different ligands of the centre ion of myoglobin induce energetic changes of the respective metal d-bands. In an RIXS experiment, these changes become visible because the recombinative transition from the Fe 3d states to the 2p core level is dipole allowed. The X-ray emission lines show a change in their frequency that is related to the metal-bonding situation. In Fig. 4.34, the energy difference between the incoming excitation signal and the respective emitted soft X-rays are plotted. This allows energetic assignment of the metal ion d-band

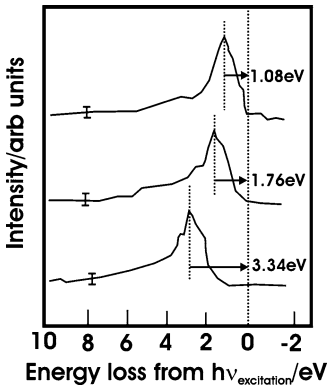


Fig. 4.34 Energy loss for different configurations of the centre iron at the heme site of myoglobin: (*top curve*) low-spin configuration Fe^{3+}–CN; (*middle*) high-spin configuration Fe^{3+}–OH_2; (*bottom*) low-spin configuration Fe^{2+}–CO

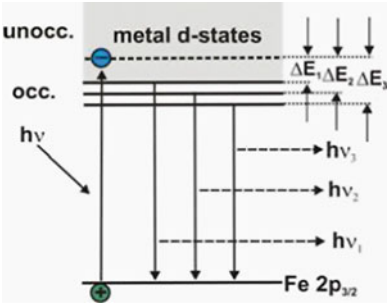


Fig. 4.35 Energy scheme for the X-ray emission lines observed in Fig. 4.34; the decay following excitation into unoccupied metal d-states (*unocc.*) from the lower occupied levels (*occ.*) to the Fe2p core level is dipole allowed and ligand-induced shifts are observed in the emission $h\nu_1, \ldots, h\nu_3$ due to the energetic changes of the d-band position with the ligands

positions for the different ligand configurations. The high-spin configuration is located 0.64 eV below the low-spin configuration and the low-spin system Fe^{2+}–CO is energetically shifted downwards by another 1.58 eV.

Photosynthetic water oxidation sustains most life on earth, providing flora and fauna with oxygen. The according reaction

$$2\,H_2O \rightarrow O_2 + 4\,H^+ + 4\,e^-, \tag{4.17}$$

is catalyzed in the oxygen evolving complex (OEC) of photosystem II (PS II) which contains four Mn atoms as active sites for catalytic water oxidation. The multivalency of Mn (II, III, IV) provides a high degree of redox and chemical reaction variability. RIXS on the OEC of photosystem II allows analysis of the electronic structure of Mn within the complex and allows investigation of the question whether oxidation can be described in a localized model using Mn oxidation states [90]. Recent results indicate that the observed shifts can be attributed to the number of Mn d-electrons and that the changes in the spectra between the S_1 and S_2 states in the OEC of PS II are related to transfer of a delocalized electron of the OEC to the PS II. The respective electron can obviously not be assigned to just one atom contradicting the hitherto favoured picture of localized oxidation. The problem of radiation damage can be alleviated by moving the samples after short (in the second time scale) measurements, but it remains a basic topic to be addressed in future work.

4.3 Perturbed Angular Correlations for Phase Analysis: The Cu–In–S System

In this section, we present three examples of the use of high-energy electromagnetic radiation in materials science, biology and surface science. They use the subsequent and correlated γ-quanta emission in electron orbital capture (Sect. 2.4.) that is the basis of the PAC method.

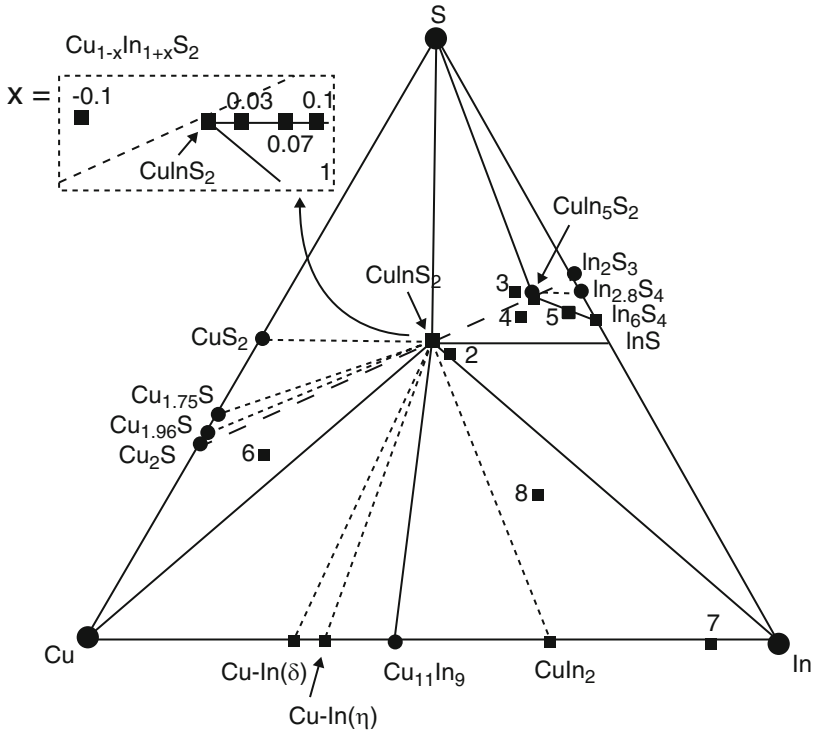


Fig. 4.36 Gibbs phase triangle for the Cu–In–S system including binary and intermetallic compounds; the extension of the homogeneity range is schematically indicated based on existing experimental data

By crystal growth using radioactive ^{111}In atoms, incorporated prior to preparation of polycrystalline samples, non-cubic phases can be identified by the resulting electric field gradient acting on the radioactive probe atoms. The method has been introduced in Sect. 2.4, and a result for In-rich $CuInS_2$ has been shown in Fig. 2.6. Here, results and their evaluation of series of measurements on the various binary and ternary In-containing phases that constitute the room temperature phase diagram of the Cu–In–S system will be presented. Whereas the tetragonal distortion of the chalcopyrite structure, a zincblende structure with an ordered cationic Cu–In sublattice could not be identified by PAC indicating a cubic symmetry for $CuInS_2$, little was known about the extension of the homogeneity range (where the material crystallizes as a single phase). Figure 2.6 shows signatures of foreign phases for large In excess. By comparison with measurements on standards, these phases could be identified as metallic In and InS.

The room temperature phase diagram of a ternary compound is derived from Gibbs' phase rule:

$$P + F = C + 2. \tag{4.18}$$

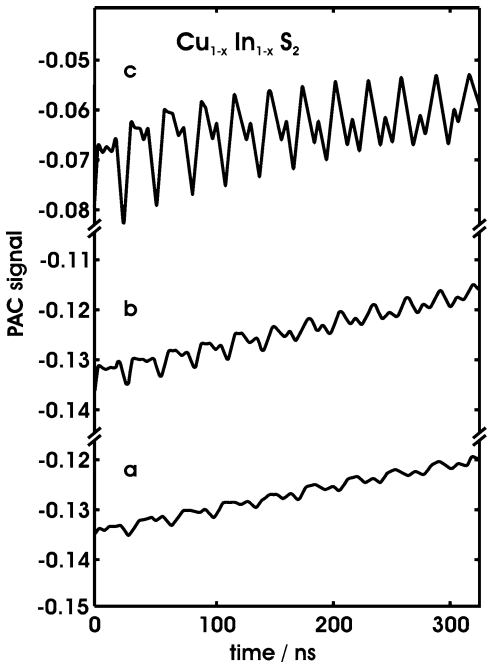


Fig. 4.37 Perturbed angular correlations for samples with In excess: (*a*) $x = 0.03$; (*b*) $x = 0.07$; (*c*) $x = 0.10$ (see also [91,92])

It states that the number of phases (P) and the number of degrees of freedom (F) is equal to the number of independent components (C) added by two. The number of components is obtained from the relation

$$C = K - (a + b), \tag{4.19}$$

where K denotes the total number of molecule types; a is the number of independent chemical equilibria, b is the number of additional restricting concentration conditions within one phase. Upon preparation from the binary constituents Cu_2S and In_2S_3, $C = 3$ and since one chemical equilibrium $Cu_2S + In_2S_3 \Leftrightarrow 2CuInS_2$ exists, $P + F = 4$. The vapour pressure of solids can be neglected; therefore, the temperature is the independent parameter that can be chosen freely. Hence three phases exist in equilibrium at room temperature. The situation is drawn in Fig. 4.36.

The Gibbs phase triangle is characterized by (pseudo) binary cuts, also called conodes and upon preparation outside the homogeneity range of $CuInS_2$; three phases coexist in equilibrium at room temperature. Also shown are the binary and ternary constituents of the system and the tie-lines (conodes), identified by a combination of PAC and X-ray diffraction.

The phase diagram has been obtained from data, such as shown in Fig. 4.37, where the extension of the homogeneity range in the direction between $CuInS_2$ and InS is investigated. It can be seen that for an In-excess of 0.03, already the signature of three phases is found: $CuInS_2$ (as an increasing background), In and InS.

This states that the homogeneity range in this direction is confined to regions smaller than $x = 0.03$. On the other side of the stoichiometrie point of $CuInS_2$, experiments have shown that up to $x = -0.1$, only single phase $CuInS_2$ exists. These results demonstrate that the homogeneity range of the compound is anisotropic within the phase triangle. This is indicated in the insert of Fig. 4.36. The existence of a conode between $CuInS_2$ and InS follows from the investigation of samples with nominal composition $Cu_4In_5S_{8+\delta}$ (point 2) and the already mentioned series of measurements on $Cu_{1-x}In_{1+x}S_2$ (see Fig. 4.37). Analysis of point 4 ($Cu_{11}In_{35}S_{54}$) yields conodes between $CuInS_2$ and In_6S_7. The latter is additionally confirmed by the evaluation of point 5 for $Cu_5In_{84}S_{111}$. Lines that have not been directly identified but which are very likely to exist, are drawn as dashed lines.

In the subsequent chapter, materials science aspects are extended to photonic crystals, their basic physics including analyses of their energetic structure. Also, the properties and characterization of biological matter, already introduced here in Chap. 4 shortly, will be extended to photosynthesis and the properties of semiconductor–enzyme junctions.

References

1. S.S. Hasnain, J.R. Helliwell, H. Kamitsubo, Fifty years of synchrotron radiation, J. Synchrotron Radiat. **4**, 315 (1997) (editorial)
2. H. Gerischer, Charge transfer processes at the semiconductor–electrolyte interface in connection with problems of catalysis, Surf. Sci. **18**, 97–122 (1969)
3. S.R. Morrison, *Electrochemistry at Semiconductor and Oxidized Metal Electrodes* (Plenum, New York, 1980)
4. K. Skorupska, Ch. Pettenkofer, S. Sadewasser, F. Streicher, W. Haiss, H.J. Lewerenz, Electronic and morphological properties of the electrochemically prepared step bunched Silicon (1 1 1) surface, Phys. Stat. Sol. (b) **248**, 361–369 (2011)
5. W.T. Grubb, Catalysis, electrocatalysis and hydrocarbon fuel cells, Nature **198**, 1183 (1963)
6. D. Teschner, R. Schlögl et al., High-pressure X-ray photoelectron spectroscopy of palladium model hydrogenation catalysts. Part 1: Effect of gas ambient and temperature, J. Catal. **230**, 186–194 (2005)
7. G. Scherb, J. Zegenhagen, K. Uosaki et al., In-situ X-ray standing-wave analysis of electrodeposited Cu monolayers on GaAs(0 0 1), Phys. Rev. B **58**, 10800–10805 (1998)
8. T. Stempel, A. Munoz, K. Skorupska, M. Lublow, M. Kanis, H.J. Lewerenz, Surface chemistry and nanotopography of step-bunched silicon surfaces: in-system SRPES and SPM investigations, Electrochem. Soc. Trans. **19**, 403–407 (2009)
9. D. Teschner, A. Knop-Gericke, R. Schlögl et al., The roles of subsurface carbon and hydrogen in palladium-catalyzed alkyne hydrogenation, Science **320**, 86–89 (2008)
10. H.J. Lewerenz, Surface scientific aspects in semiconductor electrochemistry, Chem. Soc. Rev. **26**, 239–246 (1997)
11. K. Skorupska, M. Lublow, M. Kanis, H. Jungblut, H.J. Lewerenz, On the surface chemistry of silicon under reducing conditions: an SRPES investigation, Electrochem. Commun. **7**, 1077–1081 (2005)
12. S.P. Garcia, H. Bao, M.A. Hines, Etchant anisotropy controls the step bunching anisotropy in KOH etching of silicon, Phys. Rev. Lett. **93**, 166102 (2004)

13. M. Lublow, H.J. Lewerenz, Analysis of variable scale surface roughness on Si(1 1 1): a comparative Brewster angle, ellipsometry and atomic force microscopy investigation, Trans. Inst. Metal Finish. **83**, 238–247 (2005)
14. P. Perfetti, C. Quaresima, C. Coluzza, C. Fortunato, G. Margaritondo, Dipole-induced changes of the band discontinuities at the SiO_2–Si interface, Phys. Rev. Lett. **57**, 2065–2068 (1986)
15. W. Kohn, Density functional theory: fundamentals and applications, in *Highlights of Condensed Matter Theory*, ed. by F. Bassani, F. Fumi, M.P. Tosi (North Holland, Amsterdam, 1985), pp. 1–15
16. H.J. Lewerenz, T. Bitzer, M. Gruyters, K. Jacobi, Electrolytic hydrogenation of silicon: a high resolution electron loss spectroscopy investigation, J. Electrochem. Soc. **140** , L44–L46 (1993)
17. K. Jacobi, M. Gruyters, P. Geng, T. Bitzer, M. Aggour, S. Rauscher, H.J. Lewerenz, Hydrogenation of Si(1 1 3) surfaces by (photo)electrochemical treatment, Phys. Rev. B **51**, 5437–5440 (1995)
18. M. Lublow, T. Stempel, K. Skorupska, A.G. Munoz, M. Kanis, H.J. Lewerenz, Morphological and chemical optimization of ex-situ NH4F conditioned Si(1 1 1)-(1 × 1):H, Appl. Phys. Lett. **93**, 062112 (2008)
19. S. Garbarino, A. Pereira, C. Hamel, É. Irissou, M. Chaker, D. Guay, Effect of size on the electrochemical stability of Pt nanoparticles deposited on gold substrate, J. Phys. Chem. C **114**, 2980–2988 (2010)
20. J. Azoulay, Photoelectron spectroscopy-principles and applications, Vacuum **33**, 211–213 (1983)
21. H.J. Lewerenz, K. Schulte, Combined photoelectrochemical conditioning and surface analysis of InP photocathodes: II. Photoelectron spectroscopy, Electrochim. Acta **47**, 2639–2651 (2002)
22. M. Alonso, R. Cimino, K. Horn, Surface photovoltage effects in photoemission from metal-GaP(1 1 0) interfaces: importance for band bending evaluation, Phys. Rev. Lett. **64**, 1947–1950 (1990)
23. S.L. Molodtsov, S.V. Halilov, V.D.V. Servedio, M. Richter, C. Laubschat, Solid state effects in photoionization cross sections: Cooper minima in photoemission, Nucl. Instrum. Meth. Phys. Res. A **470**, 274–277 (2001)
24. J. Ristein, W. Stein, L. Ley, Photoelectron yield spectroscopy on negative electron affinity diamond surfaces: a contactless unipolar transport experiment, Diamond Rel. Mat. **7**, 626–631 (1998)
25. L.C. Feldman, J.W. Mayer, *Fundamentals of Surface and Thin Film Analysis* (Prentice Hall, Upper Saddle River, 1986)
26. K.L. Kliewer, Nonlocal effects in photoemission studies with nonnormally incident light, Phys. Rev. Lett. **33**, 900–903 (1974)
27. M. Perner et al., Optically induced damping of the surface plasmon resonance in gold colloids, Phys. Rev. Lett. **78**, 2192–2195 (1997)
28. K. Hübner, Chemical bond and related properties of SiO2. III. Core-level shifts in SiOx, Phys. Stat. Sol. (a) **42**, (1977) 501–509
29. T.D. Thomas, Extra-atomic relaxation energies and the auger parameter, J. Electron. Spectrosc. Rel. Phenom. **20**, 117–125 (1980)
30. P.Y. Timbrell, A.J. Gellman, R.M. Lambert, R.F. Willis, Negative ion resonance selective mode enhancement in the HREEL spectrum of C2H2 on Pd(1 1 1), Surf. Sci. **206**, 339–347 (1988)
31. H. Ibach, *Electron Energy Loss Spectrometers* (Springer, Berlin, Heidelberg, 1991)
32. H. Ikeda, Y. Nakagawa, M. Toshima, S. Furuta, S. Zaima, Y. Yasuda, Initial oxidation of H-terminated Si(1 1 1) surfaces studied by HREELS, Appl. Surf. Sci. **117**,109–113 (1997)
33. A. Uhlir Jr., Electrolytic shaping of germanium and silicon, Bell System Tech. J. **35**, 333–347 (1956)
34. S.A. Campbell, H.J. Lewerenz (eds.), in *Semiconductor Micromachining*, vols. 1, 2 (Wiley, Chichester, New York, 1998)
35. J. Grzanna, T. Notz, H.J. Lewerenz, Model for current oscillations at the Si/electrolyte contact: extension to spatial resolution, ECS Trans. **16**, 173–180 (2008)

36. J. Grzanna, H. Jungblut, H.J. Lewerenz, A model for electrochemical oscillations at the Si/electrolyte contact Part I. Theoretical development, J. Electroanal. Chem. **486**, 181–189 (2000)
37. R. Tenne, V. Marcu, Y. Prior, Photoelectrochemical etching of compound semiconductors: wavelength dependence, Appl. Phys. A **37**, 205–209 (1985)
38. H.J. Lewerenz, H. Jungblut, S. Rauscher, Surface analysis of the electropolishing layer on Si(1 1 1) in ammoniumfluoride solution, Electrochim. Acta **45**, 4615–4627 (2000)
39. S.D. Collins, Etch stops, in *Semiconductor Micromachining*, ed. by S.A. Campbell, H.J. Lewerenz. Techniques and Industrial Applications, vol. 2, (Wiley, Chichester, New York, 1998)
40. E. Budevski, G. Staikov, W.J. Lorenz, *Electrochemical Phase Formation and Growth. An Introduction to the Initial Stages of Metal Deposition* (Wiley, New York, 1996)
41. Y. Zhang et al., Underpotential deposition of copper on electrochemically prepared conductive ruthenium oxide surface, Electrochem. Sol. State Lett. **7**, C107–C110 (2004)
42. K.J. Bachmann, Wet and dry etching: a comparison in the context of solid state electronics applications, in *Semiconductor Micromachining*, ed. by S.A. Campbell, H.J. Lewerenz. Techniques and Industrial Applications, vol. 2, (Wiley, Chichester, New York, 1998)
43. O.J. Glembocki, R.E. Stahlbush, M. Tomkiewicz, Bias-dependent etching of silicon in aqueous KOH, J. Electrochem. Soc. **132**, 145–151 (1985)
44. A.G. Cullis, L.T. Canham, Visible light emission due to quantum size effects in highly porous crystalline silicon, Nature **353**, 335–337 (1991)
45. H.J. Lewerenz, J. Jakubowicz, H. Jungblut, Metastable stage of porous silicon formation: the role of h-terminated low index faces, Electrochem. Commun. **6**, 838–842 (2004)
46. D.J. Monk, D.S. Soane, R.T. Howe, A review of the chemical reaction mechanism and kinetics for hydrofluoric acid etching of silicon dioxide for surface micromachining applications, Thin Solid Films **232**, 1–12 (1993)
47. P. Allongue, V. Costa Kieling, H. Gerischer, Etching of silicon in NaOH solutions. II: Electrochemical studies of n-Si(1 1 1) and (1 0 0) and mechanism of the dissolution, J. Electrochem. Soc. **140**, 1018–1026 (1993)
48. T. Baum, D. Schiffrin, Kinetic isotopic effects in the anisotropic etching of p-Si(1 0 0) in alkaline solutions, J. Electroanal. Chem. **436**, 239–244 (1997)
49. H.J. Lewerenz, M. Aggour, C. Murrell, M. Kanis, H. Jungblut, J. Jakubowicz, P.A. Cox, S.A. Campbell, P. Hoffmann, D. Schmeißer, Initial stages of structure formation on silicon electrodes investigated by photoelectron spectroscopy using synchrotron radiation and in-situ atomic force microscopy, J. Electrochem. Soc. **150**, E185–E189 (2003)
50. P. Allongue, V. Kieling, H. Gerischer, Etching mechanism and atomic structure of H–Si(1 1 1) surfaces prepared in NH4F, Electrochim. Acta. **40**, 1353–1360 (1995)
51. H. Jungblut, J. Jakubowicz, H.J. Lewerenz, Observation of a transitory structure during porous silicon formation: stability of Si (1×1) H-terminated surfaces and facets, Surf. Sci. **597**, 93–101 (2005)
52. K. Siegbahn, Electron spectroscopy for atoms, molecules, and condensed matter, Rev. Mod. Phys. **54**, 709–728 (1982)
53. E. Foca, J. Carstensen, H. Föll, Modelling electrochemical current and potential oscillations at the Si electrode, J. Electroanal. Chem. **603**, 175–202 (2007)
54. R.T. Sanderson, Chemical principles revisited: principles of electronegativity – Part I. General nature, J. Chem. Educat. **65**, 112–118 (1988)
55. H.J. Lewerenz, in Tailoring of Interfaces for the Electrochemical Conversion of Solar Energy, *Advances in Electrochemical Science and Engineering*, vol. 12, ed. by R. Alkire, D. Kolb, P. Ross (Wiley, New York, 2010), pp.61–181
56. A. Dmol, *Density Functional Theory Program with the Insight Molecular Modelling Package* (MSI, San Diego, CA, 1996)
57. H.J. Lewerenz, J. Stumper, L.M. Peter, Deconvolution of charge injection steps in quantum yield multiplication on silicon, Phys. Rev. Lett. **61**, 1989–1991 (1988)

58. J. Stumper, H.J. Lewerenz, C. Pettenkofer, X-ray photoemission spectroscopy analysis of Si(1 1 1) under photocurrent doubling conditions, Phys. Rev. B **41**, 1592–1597 (1990)
59. H.J. Lewerenz, M. Aggour, C. Murrell, J. Jakubowicz, M. Kanis, S.A. Campbell, P.A. Cox, P. Hoffmann, H. Jungblut, D. Schmeißer, High resolution surface analysis of Si roughening in dilute ammonium fluoride solution, J. Electroanal. Chem. **540**, 3–6 (2003)
60. Y. Gassenbauer, A. Knop-Gericke, R. Schlögl et al., Surface potential changes of semiconducting oxides monitored by high-pressure photoelectron spectroscopy: Importance of the electron concentration at the surface, Sol. State Ion. **117**, 3123–3127 (2006)
61. H.J. Lewerenz, J. Jakubowicz, H. Jungblut, Nascent phase of porous silicon, Electrochem. Commun. **6**, 1243–1248 (2004)
62. K. Skorupska, T. Vo-Dinh, H.J. Lewerenz, Scanning probe characterization of enzymes deposited onto step-bunched silicon nanostructures, Phys. Scripta **79**, 065801 (2009)
63. H.J. Lewerenz, Enzyme–semiconductor interactions: routes from fundamental aspects to photoactive devices, Phys. Stat. Sol. (b) **245**, 1884–1898 (2008)
64. K. Skorupska, J. Golbeck, P. Ugarte-Berzel, M. Lublow, H.J. Lewerenz, Immobilization of photosystem II on step-bunched silicon: a combined AFM and Brewster angle reflectometry investigation, unpublished results
65. K. Skorupska, M. Lublow, M. Kanis, H. Jungblut, H.J. Lewerenz, Electrochemical preparation of a stable accumulation layer on Si: a synchrotron radiation photoelectron spectroscopy study, Appl. Phys. Lett. **87**, 262101 (2005)
66. D.K. Avasthi, Developments in nuclear techniques for hydrogen depth profiling, Bull. Mater. Sci. **19**, 4–14 (1996)
67. R. Rizk, P. de Mierry, D. Ballutaud, M. Aucouturier, D. Mathiot, Hydrogen diffusion and passivation processes in p- and n-type silicon, Phys. Rev. B. **44**, 6141–6151 (1991)
68. K.v. Klitzing, G. Dorda, M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantized hall, Phys. Rev. Lett. **45**, 494–497 (1980)
69. D.C. Tsui, H.L. Störmer, A.C. Gossard, Two-dimensional magnetotransport in the extreme quantum limit, Phys. Rev. Lett. **48**, 1559–1562 (1982)
70. S.A. Wolf et al., Spintronics: a spin-based electronics vision for the future, Science **294**, 1488–1495 (2001)
71. P.P. Deimel, in *Semiconductor Micromachining*, ed. by S.A. Campbell, H.J. Lewerenz. Techniques and Industrial Applications, vol. 2 (Wiley, New York, 1998)
72. H.J. Lewerenz, H. Jungblut, *Photovoltaik; Grundlagen und Anwendungen* (Springer, Heidelberg, 1995)
73. J. Zhao, A. Wang, M.A. Green, 24.5% efficiency silicon PERT cells on MCZ substrates and 24.7% efficiency PERL cells on FZ substrates, Progr. Photovolt. Res. Appl. **7**, 471–474 (1999)
74. B. Wu, Photomask plasma etching: a review, J. Vac. Sci. Technol. B **24**, 1–15 (2006)
75. W.P. Maszara, Silicon-on-insulator by wafer bonding: a review, J. Electrochem. Soc. **138**, 341–347 (1991)
76. L. Smith, A. Söderbärg, Electrochemical etch stop obtained by accumulation of free carriers without P–N junction, J. Electrochem. Soc. **140**, 271–275 (1993)
77. E.W. Becker, W. Ehrfeld, P. Hagman, A. Maner, D. Münchmeyer, Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic moulding (LIGA process), Microelectron. Eng. **4**, 35–56 (1986)
78. J. Vollmer, H. Hein, W. Menz, F. Walter, Bistable fluidic elements in LIGA technique for flow control in fluidic microactuators, Sensors Actuat. A **43**, 330–334 (1994)
79. M. Aggour, K. Skorupska, T. Stempel Perreira, H. Jungblut, J. Grzanna, H.J. Lewerenz, Photoactive silicon-based nanostructure by self-organized electrochemical processing, J. Electrochem. Soc. **154**, H794–H797 (2007)
80. H.J. Lewerenz, T. Bitzer, Electrolytic hydrogenation of silicon, J. Electrochem. Soc. **139**, L21–L23 (1992)
81. J.H. Ye, K. Kaji, K. Itaya, Atomic-scale elucidation of the anisotropic etching of (1 1 0) n-Si in aqueous NH_4F: studies by in-situ scanning tunneling microscopy, J. Electrochem. Soc. **143**, 4012–4019 (1996)

82. H.W.B. Skinner, The soft X-ray spectroscopy of solids. I. K- and L-emission spectra from elements of the first two groups, Philos. Trans. Roy. Soc. Lon., Ser. A **239**, 95–134 (1940)
83. S. Eisebitt, W. Eberhardt, Band structure information and resonant inelastic soft X-ray scattering in broad band solids, J. Electron. Spectr. Rel. Phen. **110–111**, 335–358 (2000)
84. H.J. Lewerenz, H. Goslowsky, K.-D. Husemann, S. Fiechter, Efficient solar energy conversion with CuInS2, Nature **321**, 687–688 (1986)
85. S. Menezes, H.J. Lewerenz, K.J. Bachmann, Efficient and stable solar cell by interfacial film formation, Nature **305**, 615–616 (1983)
86. R. Scheer, T. Walther, H.W. Schock, M.L. Fearheiley, H.J. Lewerenz, Development and characterization of a CuInS2 based solar cell with 10.2% efficiency, Appl. Phys. Lett. **63**, 3294–3296 (1993)
87. C.H. Fischer, H.J. Lewerenz, M.C. Lux-Steiner, W. Gudat, F. Karg et al., X-rays shed light on the "Hidden" interfaces of solar cells, Bessy Highlights **2003**, 15–16 (2004)
88. J.E. Jaffe, A. Zunger, Electronic structure of the ternary chalcopyrite semiconductors CuAlS2, CuGaSs, CuInS2, CuAlSe2, CuGaSe2, and CuInSe2, Phys. Rev. B **28**, 5822–5847 (1983)
89. T. Wilhelm, B. Berenguier, M. Aggour, M. Kanis, H.J. Lewerenz, Efficient CuInS2 (CIS) solar cells by photoelectrochemical conditioning, Compt. Rendus Chim. **9**, 294–300 (2006)
90. P. Glatzel et al., The electronic structure of Mn in oxides, coordination complexes, and the oxygen-evolving complex of photosystem II studied by resonant inelastic X-ray scattering, J. Am. Chem. Soc. **126**, 9946–9959 (2004)
91. M. Brüßler, H. Metzner, K.-D. Husemann, H.J. Lewerenz, Phase identification in the Cu–In–S system by perturbed angular correlations, Phys. Rev. B **38**, 9268–9271 (1988)
92. H. Metzner, M. Brüssler, K.-D. Husemann, H.J. Lewerenz, Phase identification in the Cu–In–S system II: a combined study by perturbed angular correlation and X-ray analysis, Phys. Rev. B **44**, 11614–11623 (1991)

Chapter 5
Photon Management: Photonic Crystals, Photosynthesis and Semiconductor–Enzyme Junctions

5.1 Introduction

In the preceding Chap. 4, high energy photons were applied for device fabrication and, also, for the analysis of (photo)electrochemically modified silicon. In addition, the use of soft X-rays in the energetic analysis of metallo-proteins and in materials development was emphasized. The identification of the Gibbs phase triangle for the solar material $CuInS_2$ by a combination of X-ray diffraction and perturbed angular correlations. In the latter, radioactive ^{111} In, incorporated during growth of the material, emits γ-quanta that can be used for identification of non-cubic phases.

This chapter focuses on the features and applications of lower energy photons with energy in the visible and near-visible spectral range. Two systems which have found widespread interest are featured here (1) photonic crystals and (2) photosynthesis. Both are characterized by fascinating properties. Despite their occurrence in nature, the fabrication of photonic crystals and their related structures is presented here by the arrangement of inorganic matter into desired structures. The analysis of photosynthesis has provided science with a plethora of insights into bio-organic chemistry and the related photonic processes. The further deepening of the understanding of these complex structures has the chance to lead the way to advanced solar energy converting systems. The hope is to develop artificial photosynthesis systems that are more robust than the natural ones and that have a considerably higher efficiency.

This chapter begins with a treatment on photonic crystals, their application and shows how Cherenkov radiation can be used to analyse the photonic bandstructure. The second part of Chap. 5 treats the (physico-chemical) processes in photosynthesis, including excitation energy Förster-type transfer and ends with recent results on junctions between semiconductors and enzymes, highlighting the development of a photobiocathode (Table 5.1).

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_5,

Table 5.1 Overview of the content of Chap. 5

5.2 Photonic Crystals

This subsection treats fundamental properties, such as the interrelation between structure, dimensionality and photonic properties, followed by optical and higher photon energy characterization. First, the fundamental properties of photonic crystals are introduced, followed by applications, soliton propagation in photonic crystals. Then, an example of bandstructure determination using Cherenkov radiation is given, based on the electron energy losses due to emission of Cherenkov radiation (see Sect. 2.5) by electrons travelling within the structure of photonic band gap materials (PBGs). The energy loss dependence on direction and primary electron energy allows a mapping of the energy bandstructure of two-dimensional photonic crystals.

5.2.1 Background on Photonic Crystals

The description of the energy bandstructure of electrons in crystalline solids, based on the Schrödinger equation with appropriate potentials that act on the particles [1, 2] resembles that of so-called photonic crystals. In the latter, multiple scattering at periodically arranged structures with modulation dimensions in the wavelength regime of the light results in a bandstructure for photons in a certain analogy to the electronic bandstructure in solids [3]. Based on the intensive research during the last two decades, new devices have been proposed and, in part, already been

realized. Due to the miniaturized feature size and high efficiency, new areas of applications were developed, some of them will be treated below. With the advent of nanophotonics as a research and development field [4], one is led to expect a similarly intense development of photonics as in the history of microelectronics.

Photonic crystals are materials with a periodic variation of the refractory index. The already mentioned analogy with the electronic bandstructure of solids, including the formation of an absolute energy gap as in semiconductors, however, has its limits: the electromagnetic field that describes photons has vectorial character, whereas the electron wave functions in solids are scalar functions. This difference results in differences in the formation of photonic energy gaps and only a few examples exist where nature has realized photonic crystals. Among the best known are the crystallites on butterfly wings and the shimmering appearance of opals. The high demand on miniaturization for preparation of photonic crystals with an energy gap near optical frequencies is an active research area that is also influenced by military aspects [5]. Already, rather early after inception of the concept [6], the fabrication of materials with energy gaps in the IR range in a controlled manner [7, 8] has been demonstrated.

5.2.1.1 Photonic Bandstructures

The properties of the photon field are described by Maxwell's equations. Considering the two curl equations which correlate the electric field E with the magnetic field H for transverse electromagnetic waves (div $\boldsymbol{E}(\mathbf{r}) = \text{div}\, \boldsymbol{H}(\mathbf{r}) = 0$)

$$\nabla x\, \boldsymbol{E}(\mathbf{r}) + \mathrm{i}\omega/c\, \boldsymbol{H}(\mathbf{r}) = 0, \tag{5.1}$$

$$\nabla x\, \boldsymbol{H}(\mathbf{r}) + \mathrm{i}\omega/c\varepsilon(\mathbf{r})\, \boldsymbol{E}(\mathbf{r}) = 0. \tag{5.2}$$

Focussing only on so-called low-loss dielectrics, i.e. neglecting absorptive properties of the structures in the considered frequency range, the dielectric function $\varepsilon(r)$ is not complex but real. In addition, the frequency dependence of ε can be ignored by choosing its value appropriate to the frequency range considered. Equations (5.1) and (5.2) can be decoupled by dividing (5.2) by ε, taking the curl and using (5.1) to eliminate $\boldsymbol{E}(\mathbf{r})$. The resulting equation, entirely written for the magnetic field, is called the *Master equation*:

$$\nabla x\, \{1/\varepsilon(\mathbf{r})\, \nabla \boldsymbol{H}(\mathbf{r})\} = \omega^2/c^2\, \boldsymbol{H}(\mathbf{r}), \tag{5.3}$$

$\varepsilon(\mathbf{r})$ contains the information on the spatial periodic structure of the photonic crystal. Therefore, the bandstructure is obtained by solving the master equation for a given structure, revealed by ε, to find the modes $\boldsymbol{H}(\mathbf{r})$ for a given frequency (energy) and then use (5.4) to determine $\boldsymbol{E}(\mathbf{r})$:

$$\boldsymbol{E}(\mathbf{r}) = \{-\mathrm{i}c/\omega\varepsilon(\mathbf{r})\}\nabla x \boldsymbol{H}(\mathbf{r}). \tag{5.4}$$

The reason for the formulation of the eigenvalue problem in terms of the magnetic field (5.3) relates to the greater difficulty in solving a generalized eigenvalue equation for $\boldsymbol{D} = \varepsilon \boldsymbol{E}$ and also, that, if one reduces the problem to a simple eigenvalue problem, the newly introduced field $\boldsymbol{F}(\mathbf{r}) = 1/\varepsilon^{1/2}(\mathbf{r})\boldsymbol{D}(\mathbf{r})$ is not transverse anymore [9]. It should be noted that the operator in the master equation is hermitian and thus leads to real eigenvalues.

The rather far-reaching analogy of the above considerations with solid-state quantum physics allows a facilitated comprehension of the origin of the photonic bandstructure, based on the longer existing experience with electrons in solids. For electrons, the wavefunction is given by $\psi(\mathbf{r},t) = \psi(\mathbf{r}) \exp i\omega t$, for photons it is described by $\boldsymbol{H}(\mathbf{r},t) = \boldsymbol{H}(\mathbf{r}) \exp \mathrm{i}\omega t$. The eigenvalue problem in quantum mechanics, using the time-independent Schrödinger equation is $\mathbf{H}\psi = E\psi$, where $\mathbf{H}$ and E denote the Hamilton operator $\mathbf{H} = -\mathbf{P}^2/2m + V(r)$($\mathbf{P}$: momentum operator $-ih/2\pi\nabla$) and the energy eigenvalue E, respectively. The analogous eigenvalue problem for a photonic crystalreads $\Theta\boldsymbol{H} = (\omega/c)^2\,\boldsymbol{H}$ where Θ enotes the hermitian operator $\nabla x\{1/\varepsilon\nabla x\}$. The electromagnetic modes of a photonic crystal with discrete translational symmetry in three dimensions can be written as so-called Bloch states composed of a plane wave modulated with a function $\mathbf{u}_\mathrm{k}(\mathbf{r})$ which expresses the periodicity of the real (not the reciprocal) lattice. Hence

$$\boldsymbol{H}_\mathrm{k}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{r}}\mathbf{u}_\mathrm{k}(\mathbf{r}). \tag{5.5}$$

One solves the eigenvalue problem by inserting the Bloch states (5.5) into the master equation (5.3) and obtains the expression

$$\Theta\boldsymbol{H}_\kappa = (\omega(\mathbf{k})/c)^2\boldsymbol{H}_\mathrm{k}, \tag{5.6}$$

resulting in

$$\Theta_\kappa\mathbf{u}_\mathrm{k}(\mathbf{r}) = (\omega(\mathbf{k})/c)^2\mathbf{u}_\mathrm{k}(\mathbf{r}), \tag{5.7}$$

which gives the eigenmodes $\omega_n(\mathbf{k})$, indexed in order of increasing frequency by the band index n. The bandstructure is obtained by solving (5.7) by an iterative minimization technique as described, for instance, in [10]. As in electronic bandstructure representations, the photonic bandstructure can be analogously shown in the first Brillouin zone of momentum space, i.e. the reciprocal lattice. Figure 5.1 shows the bandstructure of a two-dimensional photonic crystal having a hexagonal arrangement of cylindrical pores as displayed in the right-hand part of the figure.

The structure shown in the right-hand side of Fig. 5.1, dielectric pores embedded in a medium with higher dielectric constant, and circular columns, are excellent examples for the description of the physical, i.e. here, optical, properties of photonic crystals. For certain spacings, this dielectric crystal can exhibit a photonic band gap in the x–y plane. Because of mirror reflection symmetry, the modes of two-dimensional photonic crystals can be separated into two polarizations: for the *transverse electrical* field mode (TE), the electrical field is confined in the x–y plane in Fig. 5.1, hence the polarization state can be described by the vector

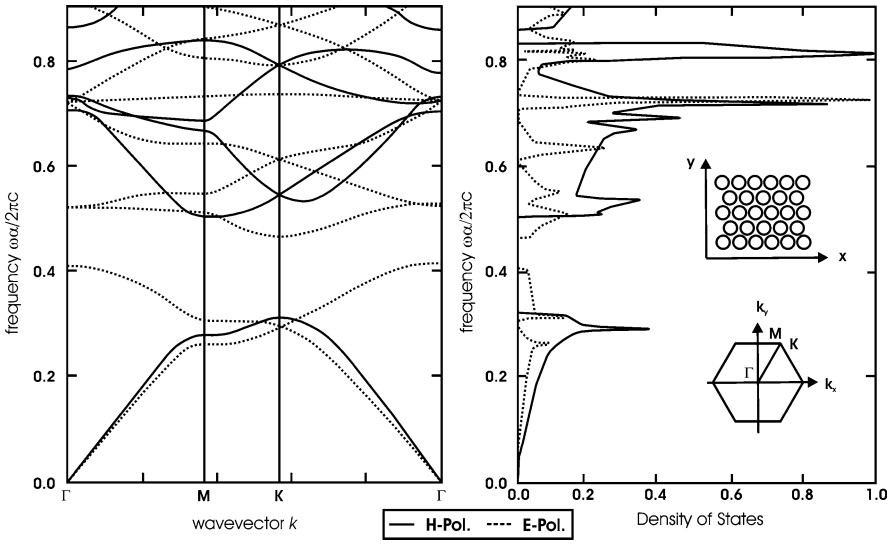


Fig. 5.1 Photonic bandstructure (*left*) and frequency (energy) dependent density of states (*right*) of a two-dimensional photonic crystal; inserts show the assumed hexagonal lattice and the corresponding irreducible part of the first Brillouin zone including main symmetry points in the reciprocal lattice (see text)

($\boldsymbol{E}_x$, $\boldsymbol{E}_y$, $\boldsymbol{H}_z$). The *transverse magnetic* mode (TM) is defined by the vector ($\boldsymbol{H}_x$, $\boldsymbol{H}_y$, $\boldsymbol{E}_z$). This decoupling of the polarizations of the electromagnetic field results in separate bandstructures for the TE and TM modes as shown in the left-hand side of Fig. 5.1. Here, the notation H-pol. stands for a polarization of the electrical field perpendicular to the axes of the pores or rods, i.e. denotes the TE mode. Consequently, the term E-pol. describes the TM mode where the electrical field is parallel to the pores. One sees from the figure that the photonic gap occurs between bands of the TM modes and that a different band gap results from bands attributed to the TE mode. As a general rule, it can be shown that TM band gaps are dominating in a lattice of isolated high ε regions such as in a square lattice of rods, and TE gaps are favoured in a connected lattice [9]. Therefore, a structure consisting of a square lattice of dielectric columns, for instance, possesses an energy gap in the TM modes but not in the TE modes. Oppositely, a square lattice of high dielectric constant veins exhibits a gap in the TE modes but not in the TM modes due to the interconnectedness of the veins (see Fig. 5.2).

For a more quantitative explanation, the so-called *fill factor* is considered. It is defined as the fraction of electrical energy located in the high – ε regions

$$f = \left\{ \int_{V(\text{high}\,\varepsilon)} E^*(r) D(r) \mathrm{d}^3 r \right\} \Big/ \left\{ \int E^*(r) D(r) \mathrm{d}^3 r \right\}. \tag{5.8}$$

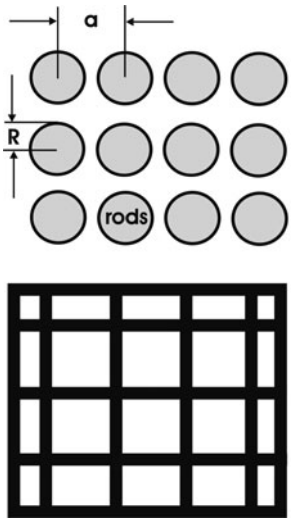


Fig. 5.2 Structures for two-dimensional photonic crystals; (*top*) square lattice of rods, (*bottom*) square lattice of veins; *shaded/darker regions* have high dielectric constant; a: rod distance, R rod radius

Table 5.2 Fill factors for the dielectric and the air band at the X-point of the Brillouin zone of the structures shown in Fig. 5.2

Dielectric band	TM mode	TE mode
Rods	0.83	0.23
Veins	0.89	0.83
Air band		
Rods	0.32	0.09
Veins	0.77	0.14

Historically, one distinguishes the dielectric band (in analogy to the valence band of a semiconductor) and the air band (conduction band) denoting low ε regions, which are often air in pores, etc. The fill factors for the lowest two bands at the X point of the square lattice of rods and veins are given in Table 5.2.

For the rod arrangement, the fill factor difference between the dielectric and the air band in the TM mode is 0.6, whereas it is only 0.23 for the TE mode. This difference of the energy distribution in consecutive modes is responsible for the occurrence of a photonic gap in the TM mode at the X point. In case of the vein structure, the connectedness of the high dielectric constant regions results in a relative gap in the TE mode as can be seen by the fill factor difference between the two bands in the TE mode amounting to 0.69 whereas it is small (0.12) in the TM mode.

An early problem was to design and realize a two-dimensional structure with an absolute gap, thus having a photonic gap in the TM mode as well as in the TE mode. So far, it was shown that TM band gaps are favoured in a lattice of *isolated* high ε regions whereas a TE gap is favoured by *connected* high ε egions. About two decades ago, the first solution to this problem was found [11]. The proposed

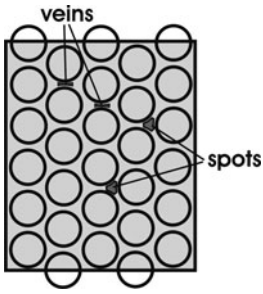


Fig. 5.3 fcc type 2 D-structure with an absolute photonic energy gap; *shaded area*: veins; *dotted area*: spots as indicated; the *circles* indicate air columns

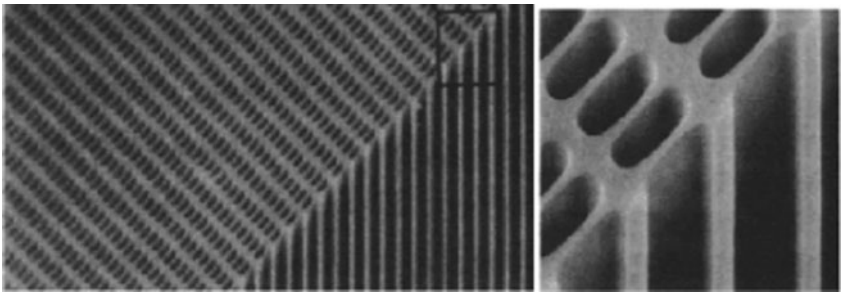


Fig. 5.4 Schematic after an SEM image of a photoelectrochemically produced macropore array in Si (100); *right image* shows magnification; pore width: 6.2 μm, pore pitch: 8 μm (see [8])

structure consisted of a triangular lattice of air columns according to the diamond structure with a two atomic basis as shown in Fig. 5.3.

As can be seen from Fig. 5.3, this structure provides a compromise in the sense that one can imagine it to consist of two types of high ε regions: isolated spots and linked veins as indicated in the figure. A typical result is shown above in Fig. 5.1. The fundamental gap is seen to lie within the TM modes. The value of the energy gap is important for any envisaged application. Due to the linearity of the Maxwell equations, photonic bandstructures are scalable in frequency (energy) space. This is an important property since it means that if one changes the dimensions in $\varepsilon(\mathbf{r})$, the frequency eigenvalues (see (5.6) and (5.7)) scale inversely proportional to the distance of the "dielectric atoms" of a given structure. For photonic band gaps in the IR to visible spectral range and beyond, considerable miniaturization of the structures is needed.

Figure 5.4 shows an arrangement of so-called macropores in a Si (100)-oriented crystal that has been fabricated photoelectrochemically. The structure has an absolute photonic gap of 4.9 μm [8]. The transmission of such a structure has been calculated as well as measured in Γ–M direction according to the insert in the right-hand side of Fig. 5.1. The result is shown in Fig. 5.5. One finds as discussed above, a larger energy gap in the TE mode and the absolute gap, due to the TM mode, is about half of the TE value as indicated by the dotted line in the figure.

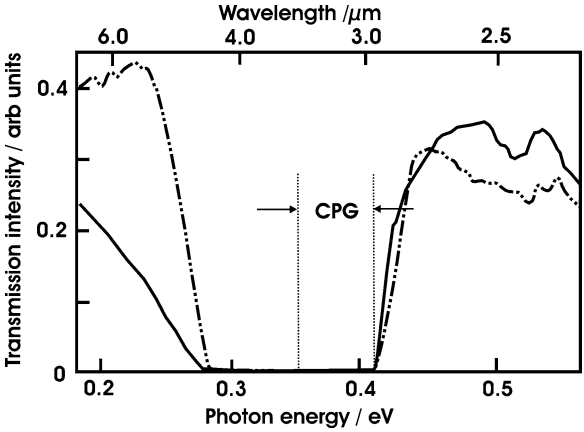


Fig. 5.5 Comparison of calculated (*dashed–dotted line*) and measured FTIR transmission (*full line*) for an fcc 2D photonic crystal with air columns in silicon. The TE mode (H-polarization) in Γ–M direction is analysed; R: radius of the pores, a: distance between centre of two pores; $R/a = 0.438$; $a = 1.55\,\mu$m; *dotted vertical line* indicates the value of the complete photonic gap (CPG)

Using silicon as base material, an important aspect is the non-negligible absorbance of Si above its band gap energy. The preposition of a purely real dielectric constant is then violated. This leads to a reduction of the dielectric contrast between, for instance, air and silicon, resulting in less pronounced photonic gaps.

Finally, we inspect the density of states distribution in Fig. 5.1 (shown on the right). For a given frequency, the DOS is calculated from the summation over all bands and integration over the first Brillouin zone:

$$N(\omega) = \sum\nolimits_n \int_{1\text{stBZ}} \mathrm{d}^3 k \delta(\omega - \omega_n(k)). \tag{5.9}$$

In analogy to electronic energy bandstructure regions with low dispersion (rather flat bands in the $E(k)$ representation of the reciprocal lattice) are characterized by a high DOS. The top of the dielectric band and the bottom of the air band exhibit, related to their free particle-type dispersion, a rather low DOS. One should also note the distinct differences in the TM and TE mode DOS.

Light propagation within the photonic gap resembles electron propagation in solids in relative (metals) and absolute (semiconductors, insulators) energy gaps: for given energy and momentum there are no *extended* states in the photonic gap. Such modes are called evanescent, decaying exponentially as described by their complex wave vector. The term $\mathrm{e}^{-\mathbf{k}z}$ modulates the solution given in (5.5), where $1/\mathbf{k}$ is the decay length given by the imaginary part of the wave vector $\mathbf{k}$. Evanescent modes, although being solutions of the eigenvalue problem, do not satisfy the translational symmetry requirement of the periodic crystal. They cannot be excited in an ideal infinite crystal, but a defect or an edge of an otherwise perfect structure can sustain such a mode. In this case, localized modes within the photonic gap can be created intentionally, for instance.

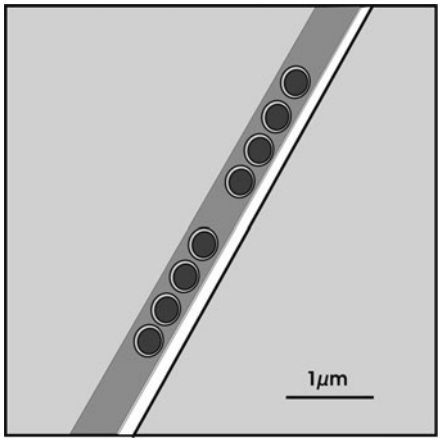


Fig. 5.6 Schematic after an SEM image of a one-dimensional wave guide structure made with silicon on a silicon on insulator (SOI) substrate

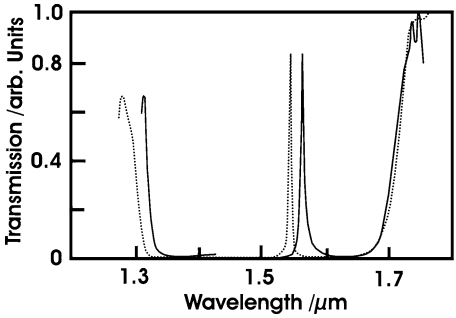


Fig. 5.7 Calculated (*dotted line*) and measured transmission (*full line*, normalized to scale) of the structure shown in Fig. 5.6; note the high transmission and the narrow spectral range within the photonic gap

5.2.2 Applications of Photonic Crystals

Applications using lower dimensional photonic crystals, such as dielectric mirrors, resonant cavities and waveguides encompass stacks of layers, periodic in one dimension (Bragg stacks) or the fabrication of a one-dimensional Si wave guide structure on an SIO (silicon on insulator) substrate [12], for example. The waveguide with a diameter of 0.5 μm has been structured by X-ray lithography such that a periodic arrangement with a well-defined isolated defect has been created (Fig. 5.6). Similar to a Fabry–Perot resonator, the four periodically arranged holes at each side of the defect are each forming a dielectric mirror. At the resonance wavelength $\lambda_R = 1,560\,\text{nm}$, exactly one-half of the wavelength fits in between the two mirrors. The structure can be viewed as a resonator of rather high quality with the resonance frequency given by the chosen dimensions. The measured and calculated transmission is shown in Fig. 5.7 where the resonance function is confirmed.

As shown in Figs 5.6 and 5.7, a point defect leads to localized states within the photonic gap. A wave, resonant to the eigenfrequency of such a defect, cannot leave in the direction perpendicular to the pore or rod structure, it becomes localized. An

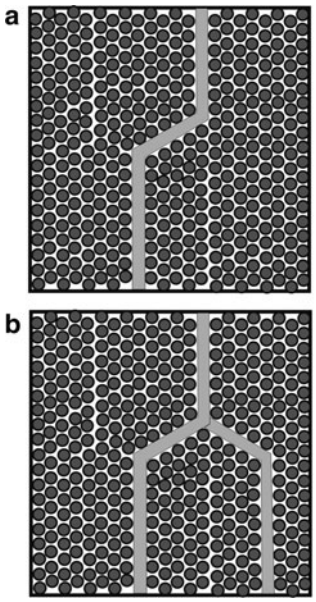


Fig. 5.8 Schematic of SEM images with line defect structures prepared in macroporous Si; (**a**) waveguide, (**b**) integrated beam splitter

arrangement of successive point defects leads to defect bands within the photonic gap and allows one to realize wave guides of almost arbitrary shape. As an example, such structures made using macroporous Si are displayed in Fig. 5.8.

Propagation by excitation of the resonance frequency of these defect bands is restricted to the defect lines. This possibility, to lead photons almost without loss around corners and winds, allows (upon increased miniaturization and with sufficient compatibility to existing technology) an "optical wiring" on a micrometer and submicrometer scale for new computing technologies.

5.2.3 Beyond Classical Electrodynamics

Among the various aspects where the properties of photonic crystals cannot be explained by the Maxwell or Master equation alone (see (5.3)–(5.5)), we select two examples which are (1) of relevance in information technology for optical signal transmission in fibres and (2) relate to the quantum electrodynamics concepts described in Chap. 1, in particular to cavity quantum electrodynamics which allows to manipulate, for instance, the spontaneous emission rate of atoms (compare Sect. 1.4.3).

As carriers of information, solitons have been recognized as well-suited wave packets or pulses. It is their temporal stability that makes them particularly attractive because they allow to transmit signals over large distances at ultrahigh pulse rates. These properties are related to solitons being shape invariant during propagation, i.e. dispersion free. Signals can be transmitted without smearing out of successive pulses.

A soliton, in brief, is a wave packet or pulse that is self-reinforcing during propagation, so that it maintains its shape while travelling at constant velocity. In optics, solitons are solutions of the nonlinear Schrödinger equation. They result from an intricate balancing between group velocity dispersion and Kerr nonlinearity. The former results from the fact that generated pulses have a non-zero bandwidth and that the medium they travel in has a frequency-dependent refractory index $n(\omega)$. This is actually the main problem that limits the transmission bit-rate in optical fibres. The Kerr effect [13] is the change of the refractory index n of a medium due to an electric field. For the AC- or optical Kerr effect, the electric field of the impinging light induces the change in n. This change Δn induces a phase shift that itself results in a frequency shift as shown for a Gaussian pulse profile in Fig. 5.9b where group velocity dispersion (Fig. 5.9a) has been neglected for explanatory reasons. It can be seen that the leading edge of the pulse shifts to lower frequencies, the peak centre shows a linear behaviour where $\omega(t) = \omega_0 + a \cdot t$ and the trailing edge is blue shifted (higher frequencies). This effect, named self-phase modulation, creates extra frequencies thus broadening the frequency spectrum of the original pulse in a symmetric manner. The influence of dispersion in a medium will act differently on the different frequencies: whereas in a region of normal dispersion where n increases as λ decreases (ω ncreases), the red portions of the spectrum have a larger group velocity than the blue ones making the pulse front move faster than the tail thus broadening the pulse in time (see Fig. 5.9a). In case of anomalous dispersion, the wavelength dependency of n is inverted resulting in pulse compression. If the pulse broadening due to self-phase modulation in a region of normal dispersion is balanced by the pulse compression due to anomalous dispersion, a temporal solitary pulse exists that travels without spreading. Since nonlinear effects in optical fibres made of fused silica exist even at low power levels, they can be used to create optical solitons [14].

Soliton propagation in photonic crystals is based on so-called gap-solitons, studied initially for one-dimensional structures. Typically, a pulse whose carrier frequency is located within the energy gap of a photonic crystal is used. If the pulse is intense enough, the change of the refractory index n (or ε) can induce a local change of the photonic bandstructure in such a way that the carrier frequency of the pulse becomes allowed by being energetically located outside the photonic gap for the given pulse time. The pulse is then creating its own transmission channel during its propagation.

In higher dimensional photonic crystals, the peculiarities of the bandstructure offer various possibilities which will be shortly mentioned. In 2D and 3D PBGs, the magnitude of the photonic gap differs for different high symmetry directions and the corresponding modes of the crystal possess different symmetries. This results in principle in a variety of soliton types with different threshold intensities and propagation velocities [15]. It should therefore be in principle possible to switch on a second soliton with a primary one which, due to its intensity, opens up an otherwise forbidden channel defining thus also the direction. Since this “defrosting” and “freezing” takes place on a picosecond time scale, the phenomenon is ultrafast and suggests the principle possibility of optical transistor action.

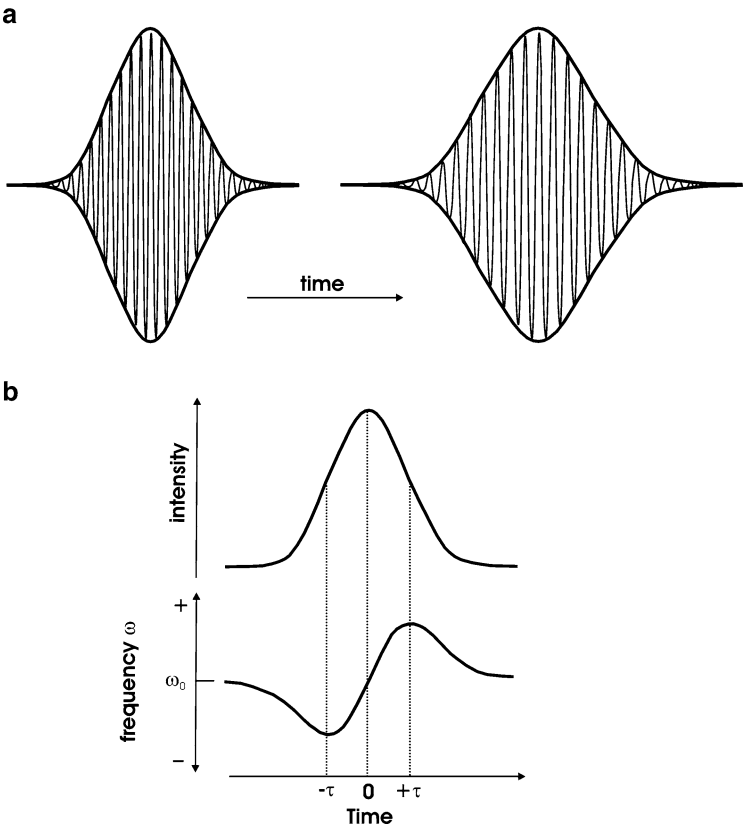


Fig. 5.9 Group velocity dispersion for a Gaussian-shaped pulse (**a**) and self-phase modulation (**b**) for the idealized situation that the medium shows the Kerr nonlinearity but that the refractory index n does not depend on frequency

5.2.4 3D Photonic Crystals

Considering the large number of possibilities to create three-dimensional photonic crystals, those which exhibit an absolute energy gap are of particular interest. As already mentioned, structures that contain connected networks of dielectric spots are promising candidates. From the insights gained by studying the properties of two-dimensional photonic crystals, the design of a three-dimensional fcc photonic lattice, with air spheres (radius r) in a high ε region appeared to be a promising candidate structure. It is, however, not only the so-called dielectric contrast, defined by the ratio of the dielectric constants of the high- and low-ε regions, but also their spatial overlap. In Fig. 5.10, no complete photonic gap exists despite a high dielectric contrast of 13. The occurrence of an absolute gap depends, in this structure, on the ratio of the radius of the air spheres compared to the lattice constant.

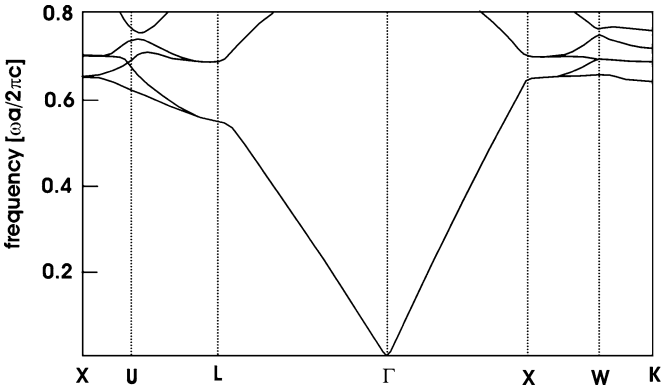


Fig. 5.10 Photonic bandstructure for dielectric contrast $\varepsilon_{\text{high}}/\varepsilon_{\text{low}}$ of 13; note that despite the high contrast, an absolute gap is not observed; shown is the usual irreducible part of the first Brillouin zone for fcc lattices

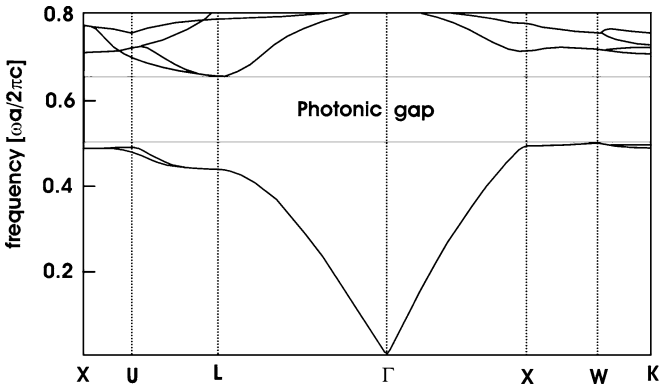


Fig. 5.11 Photonic bandstructure for a three-dimensional fcc arrangement of air spheres in a host of $\varepsilon = 13$; $r/a = 0.325$

Figure 5.11 shows that an absolute gap is obtained for the ratio $r/a = 0.325$ where r denoted the radius of the sphere and a is the lattice constant. The according photonic bandstructure is shown in Fig. 5.11. With the used r/a ratio of 0.325, most of the volume of the structure (81%) is air. Although it should be possible to realize such a structure with colloids, the differences in refractory indices are too small to allow the existence of an absolute energy gap. With metal islands or spheres, one encounters the difficulty of additional absorption. The limiting condition for the ratio of the refractory indices is $\eta = n_{\text{h}}/n_{\text{low}} > 2.9$ [16]. Therefore, the initially fabricated artificial matrix made from infiltration and successive selective removal of an opal with TiO_2, for instance, did not show an absolute gap since, in this case, $\eta = 2.7$ [17].

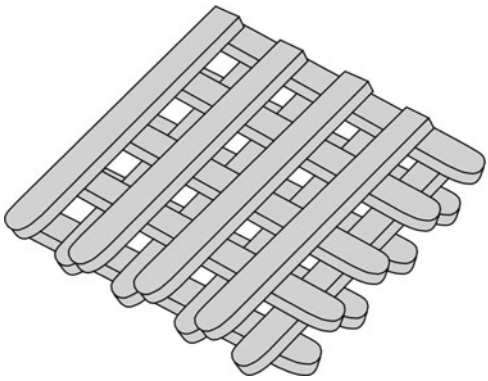


Fig. 5.12 Schematic of a realization of a 3 D photonic crystal with an absolute energy gapin the IR; note that this so-called Lincoln log exhibits a different photonic bandstructure than that shown in the above figure. The stack is made from Si slabs with a width of about 1.5 μm

An early realization of a 3D crystal with an absolute gap used a so-called Lincoln log structure which is shown in Fig. 5.12. The dielectric "logs" form a face-centred tetragonal lattice that can be imagined as an fcc lattice that has been compressed or extended in the [100] direction. In the vertical direction, the unit cell extends over four planes. The photonic gap of this structure lies in the infrared region at a wavelength of ∼12 μm [5]. Taking logs with $\varepsilon = 13$ in air, an absolute photonic gap in the region between $\omega a/2\pi c = 0.47$ and 0.56 occurs with the maximum of the lower band at the X point and that of the higher band at the L-point.

5.2.5 *Cherenkov Radiation Analysis*

In Sect. 2.5, the basic aspects of Cherenkov radiation have already been described. Recapitulating, this radiation is characterized in *conventional* materials by the following properties (1) it occurs for velocities of charged particles that exceed the phase velocity within the medium; (2) the propagation is only in the forward direction; and (3) a conical wavefront exists that points in the forward direction of the moving particle. Deviations from this behaviour are expected in so-called negative index materials that exhibit negative permittivity and permeability that have been introduced in Chap. 3 [18] where Snell's law is reversed near or in periodic structures where Bragg scattering of light can result in radiation without velocity threshold (Smith–Purcell effect [19]). In photonic crystals, complex Bragg scattering processes can occur and the Cherenkov radiation originates from coherent excitation of the crystal's eigenmodes by the moving charged particle. In this case, the radiation results from, both, the transition radiation that occurs upon crossing a boundary or when the charge travels in an inhomogeneous dielectric [20] and from conventional Cherenkov radiation.

The eigenmodes, e.g. the energy bandstructure of photonic crystals and the resulting optical properties, can be at least partly analysed by the energy losses experienced by an ultrafast moving electron due to emission of Cherenkov radiation [21].

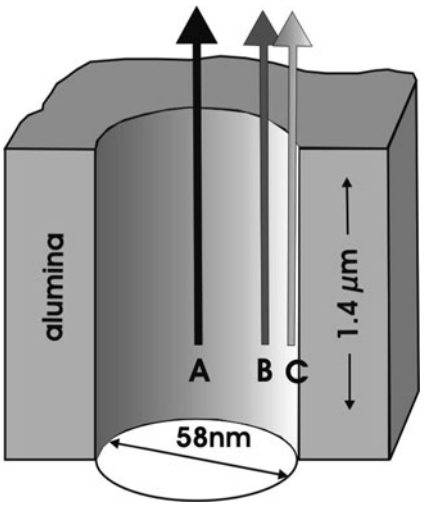


Fig. 5.13 Geometric arrangement for EELS measurements on a porous alumina matrix; pore diameter and depth are indicated in the figure; A, B, C refer to different positions of the electron beam penetrating the pore without propagating in the material; only the geometry of a single pore of the array is shown

Experimentally, optical properties of solids have been routinely investigated by electron energy loss spectroscopy (EELS) [22], performed in electron microscopes by energy analysis of fast electrons that have been traversed a sample. The initial experiments have been performed on a 2D photonic crystal, porous aluminastructures that were prepared by an established anodization process of Al in acidic solutions at rather high anodic potentials. The $1.4-\mu$m-thick Al_2O_3 sample was analysed using 200 keV electrons with a 2-nm beam diameter that travelled inside the pores of about 60-nm diameter as shown in Fig. 5.13. The resulting energy loss spectra for the different trajectories indicated in Fig. 5.13 are displayed in Fig. 5.14. All spectra are characterized by a pronounced maximum at a loss energy of about 7 eV. As this peak is located within an energy range where Al_2O_3 is very weakly absorbing ($\hat{\varepsilon}(\omega) = 3.76 + 0.03i$), the feature must be due to the coupling of radiation modes of the sample to the charge moving along the respective trajectories. As can be seen from the experimental data, the peak is best resolved for a centre beam travelling through the porous alumina. For beam trajectories closer to the inner pore surface, the feature is somewhat blurred by the higher energy surface plasmons [23] as the beam gets closer to the pore walls.

This effect can be described in more detail by consideration of different pore structures as shown in Fig. 5.15. Here, a single pore is contrasted with a pore-shell structure and an array of seven concentric nanopores. For the single pore and the pore–shell structure, the dotted lines show theoretical calculation of the electron energy loss probability. The pore–shell structure has been taken as an approximation of the concentric pore arrangement to describe so-called guided modes where the outer shell radius corresponds to the average interpore separation in the array that is shown in Fig. 5.15. The theoretical model uses the retardation force, acting on the electron which results from the electric field induced by the moving electron and the frequency-dependent dielectric function of alumina. Electron motion along the

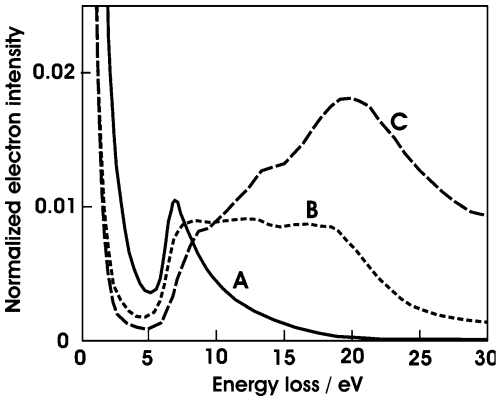


Fig. 5.14 Electron energy loss spectra measured using a scanning transmission electron microscope with 200-keV electron primary energy (see text)

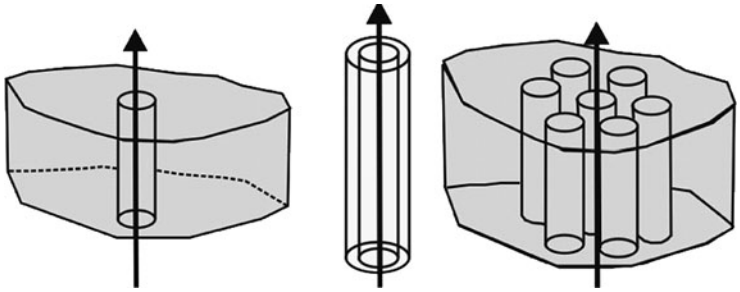


Fig. 5.15 Geometries for Cherenkov radiation experiments and their simulation as shown in Fig. 5.16; the *grey areas* represent the porous alumina host material; *left*: single pore; *middle*: core-shell structure; *right*: concentric pore array with a centre nanopore surrounded by six pores (see text)

single pore results in a rather featureless spectrum because the Cherenkov radiation can leak away from the pore. In the cylindrical pore-shell configuration, guided modes exist that result in the maximum near 7 eV, as measured experimentally (Fig. 5.16, curve A). The corresponding dashed–dotted line shows a resonance at this energy and the solid curve in Fig. 5.16 gives the data for the seven-pore arrangement.

The comparison of the energy bandstructure for localized electron beams in infinite crystals is shown in Fig. 5.17 where the bandstructure and the energy loss function are compared. Electrons moving along the z-axis (parallel to the cylinders that represent the nanopores) can be described as point-like sources:

$$\delta(z - vt) = \sum \exp(\mathrm{i}kz - \mathrm{i}kvt), \tag{5.10}$$

where k denotes the wave vector (see for instance Figs 5.1 and 5.10). The relation between energy (frequency ω) and the z-component of the momentum of the excited

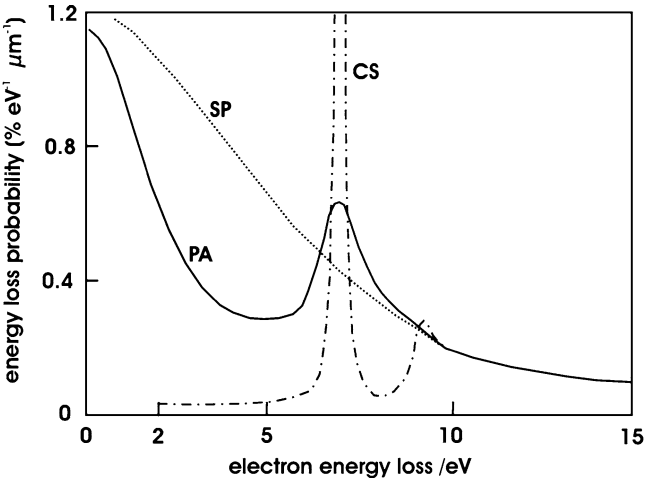


Fig. 5.16 Calculated electron energy loss curves for 200-eV electrons moving at the centre of a pore with different environment (compare also Fig. 5.15); *SP* single pore, *CS* core-shell structure, *PA* pore array with six pores concentrically oriented around the centre pore where the electrons move

radiation that is allowed in the sample, k_z, is then given by $E = k_z\, v$. Although most main features of the calculated DOS are reflected in the spectrum, the agreement is not complete. Whereas the peak of the DOS at 8.3 eV is also found experimentally, the higher energy maximum of the DOS is not well reflected in the EEL spectrum; the structure at 7 eV occurs in the spectrum albeit with more broadening. The reasons might be related to imperfect porous alumina matrices and size and distance dispersion of the nanopores as well as to the electron beam alignment.

5.3 Photonic Processes in Photosynthesis

Photosynthesis is the process on which life on earth depends directly or indirectly. It uses sunlight, water and carbon dioxide to produces the oxygen in the atmosphere (as a by-product) and carbon dioxide fixation results in the formation of carbohydrates. In this section, the focus is on mechanistic aspects of the energy conversion processes in the catalytically active centres, photosystems I and II. A slightly simplified overall reaction scheme is given in (5.11):

$$6\,CO_{2(g)} + 12\,H_2O_{(l)} + 24\,h\nu \rightarrow C_6H_{12}O_{6(l)} + 6\,O_{2(g)} + 6\,H_2O_{(l)}, \qquad (5.11)$$

where the light-induced reaction is

$$12H_2O + 24h\nu \rightarrow 6O_2 + 24H^+ + 24e^-, \qquad (5.12)$$

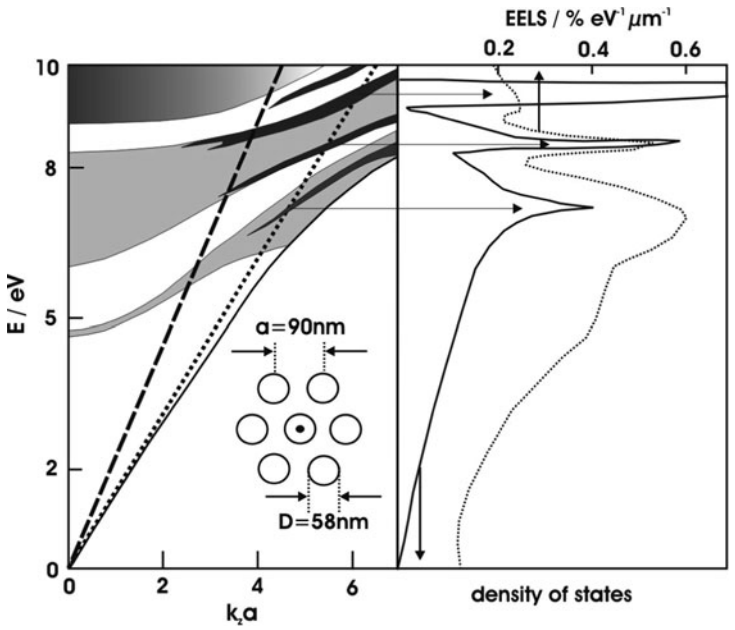


Fig. 5.17 Comparison of the calculated DOS (density of states) of a 2D photonic crystal with geometry as shown in the *inset* on the left image with the resulting projected density of states upon excitation with electrons of kinetic energy 200 keV ($v_{el} = 0.7c$), for which the dispersion relation $E = k_z v$ holds, indicated by the *dotted line* in the left-hand figure; the *horizontal arrows* connect the 2D bandstructure with the DOS (*full line* in the right-hand figure); the DOS (left image) is shown in two grey tones that correspond to low (*light grey*) and high (*darker grey*) DOS; the upper part is not specified because it does not intersect with the electron dispersion; the *dashes line* corresponds to the limiting velocity of the light, c, e.g. $E = k_z c$; the measured electron energy loss spectrum (EELS) is given by the *dotted line* in the right-hand part of the figure

The dark reaction, the carbon dioxide fixation, is

$$24H^+ + 24e^- + 6CO_2 \rightarrow C_6H_{12}O_6 + 6H_2O. \tag{5.13}$$

The energy of the photons from sunlight that are used in the process is defined by the absorbing centres and will be discussed below.

5.3.1 Structural and Energetic Aspects

The primary steps in photosynthesis are the so-called light harvesting and charge separation. For light harvesting, all photosynthetic organisms contain antenna complexes. The function of these pigment–protein complexes is to absorb sunlight and to transport the excitation energy to the photosynthetic reaction centre (RC)

where charge separation occurs. The peripheral components of light-harvesting complexes II and I (LH II, I,respectively) absorb at higher energy, and the core components absorb at lower energy. LH II, which surrounds LH I, absorbs light at a high photon energy by chlorophyll molecules and carotenoids. The excitation energy is then rapidly (time scale: ps) transferred within LH II and subsequently to LH I which surrounds the RC that absorbs at longer wavelength. Thus, the RC can trap the electronic excitation of LH II. The energy cascade described provides a means to funnel energy from the outer regions of the antenna complex to the RC. The RC accepts the excitation energy through its own chlorophyll molecules. Energy transfer from the light-harvesting complexes to the reaction centre occurs within <100 ps and with an efficiency of $\sim 95\%$. To achieve this high transfer efficiency, the thermally relaxed excitations must conserve the photon energy and suppress fluorescence.

The structural arrangement of plants and algae with regard to the photosynthetic reaction has been described extensively [24, 25]. Here, the emphasis is on the structural aspects of the two photosystems and of the reaction centre. In photosystem II, the oxygen-evolving complex and its photonic excitation is of interest. Similarly, the structural properties of photosystem I, where hydrogen can be evolved, and its energetics will be treated. Figure 5.18 shows a schematic overview of the interaction of the photosystems with light and of the basic consecutive processes where photon energy is converted into (storable) chemical energy.

Following photon absorption by the antenna complexes, the energy is transferred to the reaction centre where charge separation occurs. This energy transfer and the charge separation will be treated in more detail below. At the active site of the

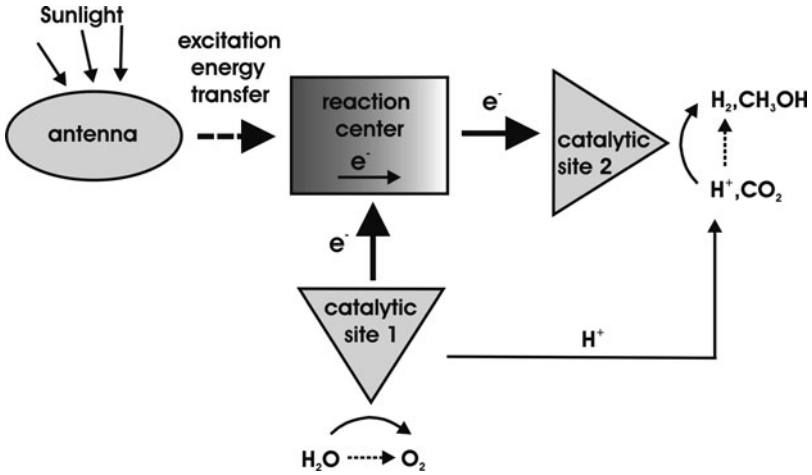


Fig. 5.18 Schematic of photon absorption, excitation energy transfer and electron transfer chain in the photosynthesis process involving photosystem II (catalytic site 1) and photosystem I (catalytic site 2); note that besides carbohydrate formation, proton reduction to molecular hydrogen is indicated at catalytic site 2

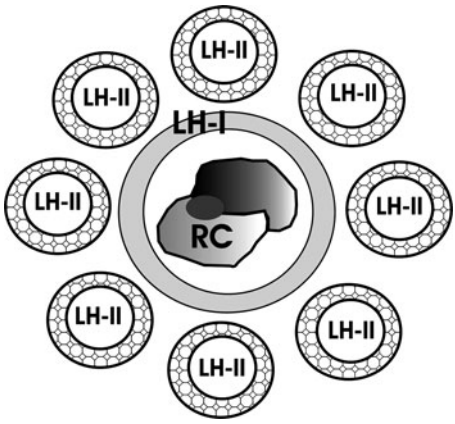


Fig. 5.19 Schematic arrangement of light-harvesting complexes LH I and LH II around the photosynthesis reaction centre RC; the heterodimeric enzyme is shown in two different shadings in the centre of LH I and the active site is visualized by a *light grey* ellipse

catalytic proteins (enzymes), oxidative water splitting (catalytic site 1, PS II) or hydrogen evolution, for instance, (catalytic site 2, PS I) takes place via complex processes that have not yet been fully elucidated. Figure 5.19 shows a schematic of the light-absorbing antenna complexes and their structural arrangement around the reaction centres. The above-mentioned light-harvesting complexes I and II are indicated in the figure. The light is absorbed in the LH II molecular rings and the photon energy is transferred via the LH 1 complex to the reaction centre (see schematic in Fig. 5.18) which consists of a heterodimeric enzyme with its active site.

The structure of green plant LH II is represented in Fig. 5.20. The complex LHC-II is a trimer, consisting of three almost identical subunits. Each subunit contains 14 chlorophylls (eight chlorophyll a, six chlorophyll b), four carotenoids (two luteins, a neoxanthin and a violaxanthin) and two lipids. The chlorophyll molecules are arranged such that their excitation energy can be transferred without loss to another molecule. Carotenoids have an antenna as well as a protection function but their contribution to the overall light harvesting is rather small. The conjugated π-systems of the lutein-molecules, located in the centre of the monomer, are arranged parallel to the neighbouring chlorophylls in such way that the deleterious chlorophyll-triplett-states are avoided.

The pigment molecules are embedded in hydrophobic so-called apoproteins (Fig. 5.20) which results in a change in their chemical properties. As a consequence, the pigment molecules are able to absorb the sunlight over a broader spectral range depending on their interaction with the protein where they are embedded and more of the incident solar spectrum can be used for energy conversion.

Generally, the light-harvesting systems of both photosystems contain distinct pigment–protein complexes that not only maximize the efficiency of absorption but also that of the energy transfer. Figure 5.21 shows absorption spectra for the pigment molecules chlorophyll a, b and for carotenoids.

Figure 5.21 shows that chlorophyll absorbs in the red and blue spectral region, resulting in reflection of green light upon illumination with white (solar) radiation.

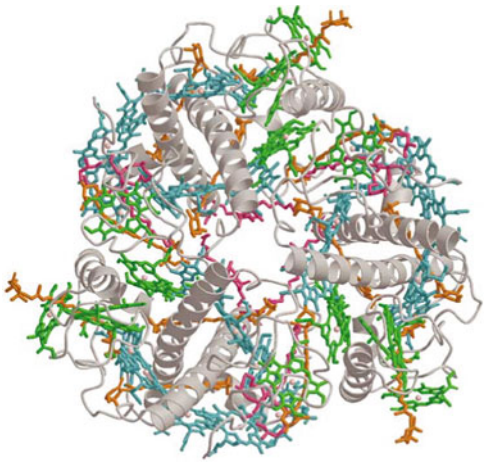


Fig. 5.20 *Top view* from the stromal side of the LH 2; the colour code is *grey*: polypeptides (apoproteins); *cyan*: Chlorophyll a; *green*: Chlorophyll b; *orange*: carotenoids; *pink*: lipids (after [26])

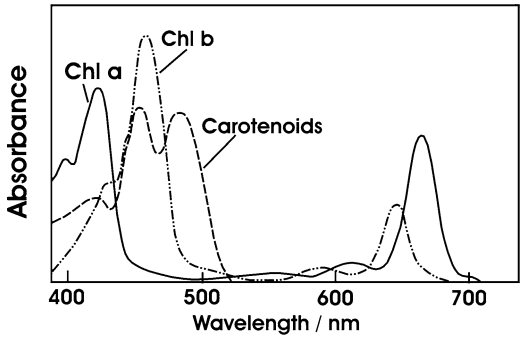


Fig. 5.21 Absorption spectra of chlorophyll a, b and of carotenoids; Chl a is a reaction centre pigment, Chl b an antenna pigment (higher absorption onset)

Chlorophylls consist of a hydrophylic part, e.g. the tetrapyrole ring that contains a Mg atom, and a hydrophobic phytol tail. Accordingly, the photosynthetic pigments are amphipathic which allows their integration into membranes and hygrophobic domains of membrane proteins. Chl a and b are structurally identical except that they differ by the substitution of a CH_3 group by a CHO group at the outer part of the tetrapyrole ring, respectively.

The reaction centre (see Fig. 5.19) comprises an active site which consists of a $Mn_4O_4 \cdots Ca$ cluster in case of PS II and of an Fe–Ni cluster, for example, in PS I. Figure 5.22 shows a schematic of the arrangement for PS II where the Mn–O–Ca cluster and the surrounding protein complex catalyse the oxidation of water (5.12).

The analysis of the functioning of this Mn–O–Ca complex for light-induced water splitting comprises the identification of oxidation states and the assessment of the overall energetics treated below (see Fig. 5.25). Excitation energy transfer occurs from the light-harvesting complexes and results in a chain of excited states that leads to water electrolysis with oxygen evolution as a byproduct. In the so-called Joliot–Kok diagram [27,28], the oxidation states are labelled $S_1 \rightarrow S_4$ and a corresponding

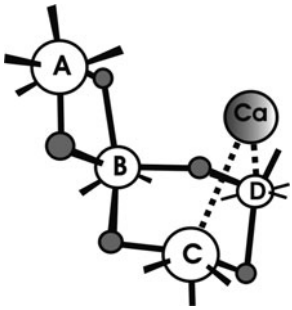


Fig. 5.22 Structural arrangement of the active site of the enzyme water oxidase of photosystem II; *grey circles*: oxygen atoms; A–D: differently positioned Mn atoms

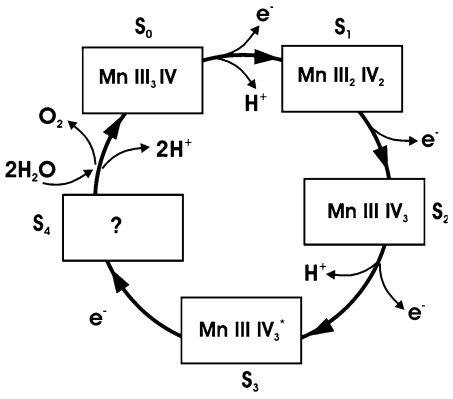


Fig. 5.23 Joliot–Kok scheme for the oxidation states of the Mn–O–Ca cluster of photosystem II; roman numerals indicate Mn oxidation states; S defines collective oxidation states; "note that S_3 is described by an oxidation that is likely stored in the proximity of the Mn cluster (indicated by the asterix) but not by direct oxidation of the remaining Mn (III) atom" and that S_4 is not identified. The steps from S_0 to S_4 are light-induced; the roman numerals refer to the Mn oxidation state (from II to IV) and the subscript to the number of Mn atoms

oxidation cycle is shown in Fig. 5.23. Electron paramagnetic resonance (EPR) and X-ray absorption spectroscopy have provided the basis for the scheme in Fig. 5.23. Upon successive photon absorption by photosystem II, a stepwise progress from the initial state, S_1, towards the more oxidized S_2 and S_3 states takes place until the hypothetical intermediate S_4 is reached which is precursor state for the release of dioxygen. Also, the reduction of the oxidized complex to the state S_0 occurs in this step and the cycle begins again. The experimental evidence supports the view that Mn oxidation states cycle largely, although not necessarily exclusively, between the Mn(III) and Mn(IV) oxidation states. There is the possibility that S_0 consists of Mn II III IV_2.

It should be noted that there is still an ongoing debate about the nature of some of the states S_x. This is partly related to analyses based on X-ray diffraction [29]. In addition, more insights have been obtained using EPR analyses [30].

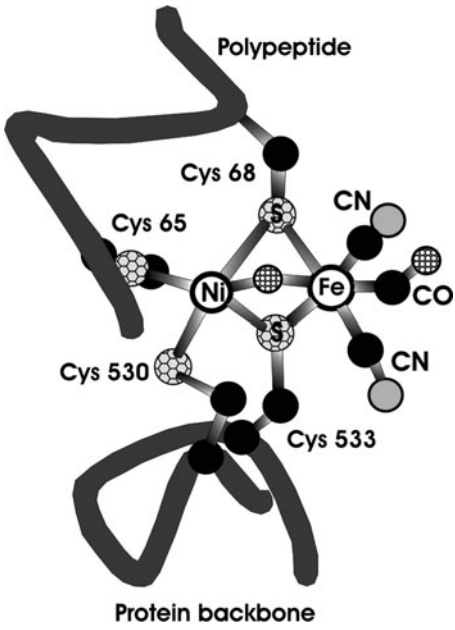


Fig. 5.24 Active site of a Fe–Ni hydrogenase of photosystem I; *cys* cysteine labelled according to position on the peptide; *black circles*: C; *grey circles*: N; *hatched circles*: O; Fe, Ni, S as indicated

The analogous metallo-complex for PS I is displayed in Fig. 5.24. Here, the selected structure is a hydrogenase with an Fe–Ni active site. Hydrogenases catalyse the reversible oxidation of molecular hydrogen and are thus able to reduce protons that have been generated at PS II during water oxidation to H_2. The Ni–Fe heterobinuclear site has an inter-atomic distance that changes from about 0.29 nm in the oxidized, inactive form to ~0.25 nm in the reduced active form. As can be seen from Fig. 5.24, the Ni–Fe core incorporates two terminal thiolates and two bridging thiolates at the Ni site. The bridging thiolates are bound to a $Fe(CN)_2CO$ fragment. For artificial photosynthesis approaches that focus on light-induced hydrogen generation, the synthesis of similarly structured complexes is one of the strategies that are pursued.

The overall energy schematic of photosynthesis is known as the Z-scheme [31] and it usually describes the excitations and products that occur in the photosynthesis process on an electrochemical scale as shown at the ordinate in Fig. 5.25 where potentials are given relative to that of the normal hydrogen electrode.

In PS II, the oxidation of water by the excited Mn–O–Ca complex (Fig. 5.22) is accompanied with an energy increase of the electrons, extracted from water (increasing cathodic potentials are equivalent to electron energy increase). The photonic energy corresponding to 680 nm is 1.82 eV and the corresponding energy increase can be read from the ordinate in Fig. 5.25 where also the excitation of PS I by a slightly lower energy (1.77 eV) is shown. The oxidation of the active complex of PS II produces an excited ground state with successively oxidative charges, labelled h^+ that leads to water splitting as shown in Fig. 5.23 if the energy requirements for photolysis are met. For the system at physiological pH, the thermodynamic

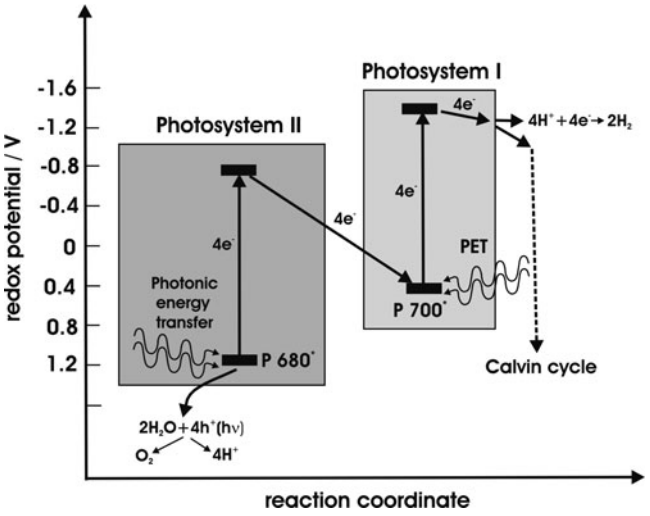


Fig. 5.25 Z-scheme of the energetic conditions of photosystems I and II upon excitation with photonic energy from the pigment molecules of the light harvesting complexes; P 680* and P 700* denote the excited reaction centres that absorb at 680-nm and 700-nm wavelength, respectively; the term "photonic energy" is used here to express that the active sites are not directly excited by light but that photon energy is transferred to them

value for O_2 evolution is $+0.81$ V vs. the normal hydrogen electrode (NHE). The level for the reaction centre of PS II is more positive, thus enabling the consecutive oxidation following the Kok cycle. The excited electrons from PS II are transported to PS I via quinines, a cytochrome b_6f complex and plastocyanin (not explicitly stated at the arrow between the photosystems). The electrons lose energy along this transport chain, partly due to the production of ATP (adenosine triphosphate) from ADP. In PS I, the electrons are again photonically excited, as shown in the above figure, for transfer to the active site where $NADPH^+$ can be reduced to NADPH (nicotinamide adenine dinucleotide phosphate), entering the Calvin cycle where CO_2 fixation occurs [32]. The present interest in the preparation of bioanalogue systems that generate hydrogen is indicated in the figure by the reduction of protons from the process to H_2.

5.3.2 Photonic Excitation Energy Transfer

The principle processes that govern the energy transfer from the absorbing pigments to the reactive sites are shortly reviewed. The structure and the function of the antenna complexes are closely interrelated, forming the so-called funnel for highly optimized photon energy capture by the catalytically active site. The number of the peripherical LH II antenna complexes depends strongly on environmental

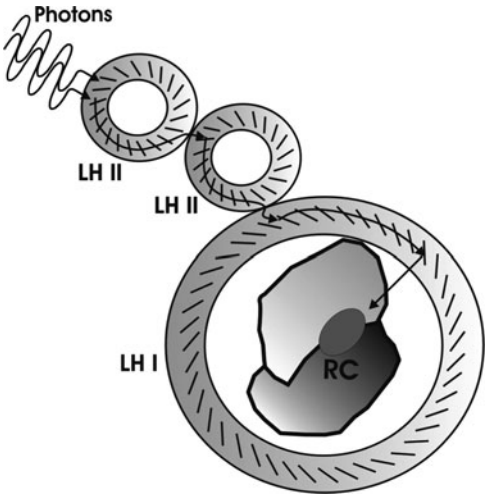


Fig. 5.26 Schematic for the photon energy transfer via light-harvesting complexes to the photosynthesis reaction centre; light is absorbed in the antenna complexes LH II by the chlorophyll–polypeptide complex where the *inclined lines* shall symbolize the chlorophyll molecules and the *grey area* the protein matrix

conditions such as the available light intensity level. Under low illumination levels, a larger number of LH II complexes are found which increases the effective area for light harvesting.

Figure 5.26 shows a rough schematic of the spatial energy transfer chain of the photosynthesis apparatus. As also shown in Fig. 5.19, the core complex LH I that contains the reaction centre (RC) complex is surrounded by the peripherical antenna complexes LH II in a two-dimensional arrangement. Upon photoexcitation of the absorbing complexes in LH II, the excitation energy is transferred via LH Ito the reaction centre. The lifetime of the fluorescence of the LH 2 complexes is in the 1 ns range and the transfer to the LH 1 complexes takes place within 5–10 ps. Therefore, the fluorescence of the LH 2 complexes can only be determined on isolated samples. The energy transfer from LH 1 to the reaction centre has time constants in the range 5–25 ps [33].

Because the nanoscale dimensions of photosynthetic protein pigment complexes result in strong pigment–pigment interactions, optical excitations can be described by a generalized Frenkel exciton model as will be shown below. Excitation processes in macromolecular ensembles that consist of virtually identical absorbing molecules can typically not be assigned to individual molecules. Therefore, localized states do not provide an appropriate basis for the description of the excitation of such aggregates. Rather, the eigenstates of the system are expressed as coherent superpositions of the localized excitations. In analogy to the Bloch states of solid-state crystal lattices, the states of the ensemble of N molecules, where molecule n has been excited, are described by

$$|k\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k \frac{n}{N}} |n\rangle \,. \tag{5.14}$$

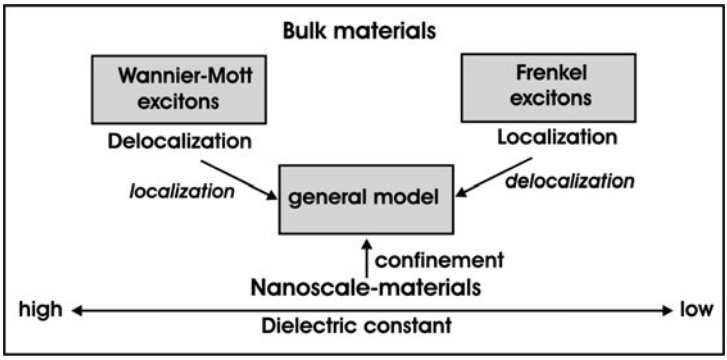


Fig. 5.27 Schematic of the exciton concepts for delocalized and localized electron–hole pairs in bulk materials and the concept transfer to nanoscale materials

The states $|k\rangle$ represent so-called Frenkel excitons and the excitation energy is delocalized over all molecules. In the nearest neighbour approximation where the intermolecular interaction is restricted to adjacent molecules, the exciton energy is given by

$$E(k) = E_{\mathrm{o}} + 2J \cos k \frac{2\pi}{N}, \tag{5.15}$$

where J denotes the interaction energy ($J_{n,n+1} = J_{n,n-1} = J$) and E_o is the photonic excitation energy. In Fig. 5.27, a scheme is shown in which the traditional excitonic models of Wannier and Mott are contrasted to that of Frenkel. Due to the reduced electrostatic interaction in materials with high dielectric constant, Wannier–Mott excitons represent the delocalized mode in crystals, whereas Frenkel excitons are spatially closer bound. In nanosystems, a more general approach should comprise the key features needed to describe the excited states of a molecule as well as that of an extended system as shown in the figure below.

Such more general description in nanosystems can be derived starting from a delocalized description and introducing quantum confinement or by using the Frenkel model of a localized (tight binding) formulation and introducing a certain amount of delocalization as was done above. An important feature of the photosynthesis light-harvesting system is revealed by considering the optical transition probabilities: for absorption of sunlight, electrical dipole transitions must be allowed and the transition dipole moment between a ground state $|g\rangle$ and an excited state $|k\rangle$ must be non-zero. Because of the series development of the Frenkel exciton (see (5.14)), the transition probability differs substantially from that of a single localized state. The transition matrix element M which enters the transition probability as $|M|^2$, has the form

$$M(k) = \langle g|\, P\, |k\rangle = \frac{1}{\sqrt{N}} \sum\nolimits_{n=1}^{N} \mathrm{e}^{\mathrm{i}2\pi k \frac{n}{N}} \langle g|\, P\, |n\rangle\,. \tag{5.16}$$

The sum is non-zero only if the contributions from individual transition moments $m(n) = \langle g|P|n\rangle$ are coherently superimposed such that constructive interference occurs. Besides the individual phases, given by the quantum number k, it is also the spatial arrangement of the absorber molecules within the aggregate that defines the transition probability.

In real systems, the chemically equivalent molecules have somewhat different environments which results in differences in the excitation energy thus widening the absorption bands of the ensemble, resulting in better spectral harvesting of the solar light. This modification is included by the term Δ_n in the value for E_0 in (5.15) and it describes the deviation of the excitation energy from its mean value. Because the basis sets $|n\rangle$ and $|k\rangle$ represent the limits of localized and delocalized states, the ratio (Δ_n/J_{mn}), where J_{mn} denotes the interaction energy of molecule n with molecule m, is taken as measure for the localization. In the B 800 and B 850 bands of the LH II complex of bacteriochlorophyll a (BChl-a) molecules, the above ratio $\Delta/J_{800} = 5$, hence the disorder is rather large and the intermolecular interaction energy is small. Accordingly, these excitations can be described by states that are localized on the respective single BChl-a molecules. For the B 850 band, $\Delta/J_{850} \sim 0.3$ and hence the excitations within this ring system are better described by a delocalized Frenkel exciton model. This delocalization within the B 850 ring system has the effect that the photonic excitation energy exists throughout the ring and can thus be efficiently transferred to a neighbouring ring system (LH II or LH I).

The excitonic coupling that results in excitation energy transfer originates from the interactions of the transition dipoles of the participating chromophores. In the conventional approach, this coupling is described by the point-dipole approximation introduced by Förster for the weak electronic coupling limit [34]. To realistically describe the observed phenomena in photosynthesis, Förster's original model had to be considerably extended and present theories are based on multi-chromophoric Förster processes that include disorder and temperature effects which will be shortly mentioned at the end of this section.

The validity of the Förster approach has been debated due to differences in experimental estimates of the energy transfer rates with that predicted by single chromophoric Förster resonance energy transfer of the LH II complexesof B 800 and B 850 [35]. The observed discrepancies result in a factor of 5–10 [36]. It should be noted, however, that these differences are not those between experiment and theory but between the evaluation of experiments and the predictions of the Förster model. Below, it will be shown that inclusion of disorder and temperature effects into standard FRET (Förster resonance energy transfer) theory yields satisfying agreement between the improved multi-chromophoric FRET theory and accordingly evaluated experiments.

Förster resonance energy transfer describes an efficient excitation energy transfer mode without the actual transfer of electrons. It is, for instance, also of considerable importance in the coupling of quantum dots, quantum wells and macromolecules to photonic absorbers such as semiconductors, nanostructured systems, nanocrystals and macromolecules for advanced applications in photovoltaics and solar fuel

generation. Förster energy transfer occurs with fluorescent molecules such as the pigment molecules in photosynthesis. These molecules absorb light at a specific wavelength interval and emit electromagnetic radiation at lower photon energy resulting in a radiation-less energy transfer. Förster theory was experimentally confirmed in 1967 ([37], and references therein) using a series of RNA fragments of known length and showing that FRET depended on the sixth power of the inverse donor–acceptor distance. Presently, Förster energy transfer is extensively used to analyse the spatial arrangement of proteins and nucleic acids with emphasis on structural changes in proteins such as folding.

Because of their important role in the Förster process, the dipole transition moments, the Franck Condon factors and the decay mechanisms following electronic excitations are reviewed. The oscillator strength of an electronic transition in the dipole approximation (weak perturbation) is given by

$$f = \frac{8\pi^2 m_e v}{3he^2} \cdot \left| M_{fi} \right|^2 . \tag{5.17}$$

The transition matrix element M_{fi} is given by the expression

$$M_{fi} = e \int \psi_f{}^* \, \bar{X} \, \psi_i \, \mathrm{d}^3 r, \tag{5.18}$$

where $\bar{X}$ denotes the space operator and the asterisk indicates the complex conjugate wave functions ψ in the initial (index i) and final state. The inverse prefactor in (5.17) is $7.1 \times 10^{-43}\,\mathrm{C^2\,m^2\,s^{-1}}$ With an assumed absorption maximum at a wavelength of 300 nm, $v = 10^{15}$ Hz and an oscillator strength of 10^{-2}, a value of for the transition dipole M_{fi} of 2.7×10^{-30} Cm is obtained which corresponds to 0.81 D($1\,\mathrm{D} = 3.336 \times 10^{-30}$ Cm). Such transition moments can be envisaged as a redistribution of electron density within the molecule that generates a radiation-induced dipole.

In these molecular excitations, the vertical electronic transition connects to those vibrational levels of the molecule in the upper electronic state for which the nuclei have the same initial separation and for which the overlap of the vibrational wave functions is large as schematically drawn in Fig. 5.28.

Because other vibrational levels have a non-vanishing probability of being at the same nucleus separation, these probabilities have to be integrated to assess the actual transition probability. Accordingly, the wave functions in the initial and final state are written as products of electronic and vibrational contributions, such as $\psi_{\mathrm{el}}(r) \cdot \psi_{\mathrm{vib}}(R)$, and the electronic contribution is then separated from the vibrational term:

$$M_{fi}^{\mathrm{el,vib}} = e \int \psi^*_{\mathrm{el},f}(r) \bar{r} \psi_{\mathrm{el},i}(r) \mathrm{d}^3 r \cdot \int \psi^*_{\mathrm{vib},f}(R) \cdot \psi_{\mathrm{vib},i}(R) \mathrm{d}^3 R. \tag{5.19}$$

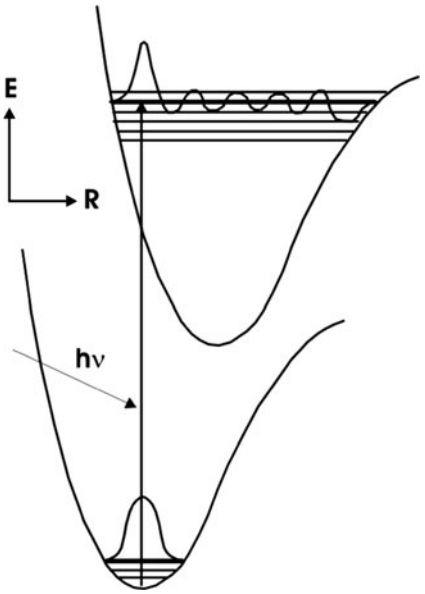


Fig. 5.28 Transition schematic on the influence of Franck–Condon factors in molecular electronic excitations; ground state (lower energy curve) and excited state are shown including vibrational substates; note that R is the nucleus coordinate

The second integral in (5.19) represents an overlap integral between the vibrational initial and final states of the wave function. The square of the expression for the integral that contains the vibrational excitations

$$S^2_{\mathrm{vib},f,i} = \left| \int \psi^*_{\mathrm{vib},f}(R) \cdot \psi_{\mathrm{vib},i}(R) \mathrm{d}^3 R \right|^2 \tag{5.20}$$

is known as *Franck–Condon factor* for the transition shown in (5.19). As a consequence, the overlap of the vibrational substates, the larger is the optical absorption related to that electronic-vibrational transition, as schematically indicated in Fig. 5.28.

The fate of excited states can be divided into radiationless energy dissipation and photon emission as fluorescence and phosphorescence. The corresponding processes such as internal conversion, quenching, intersystem crossing and vibrational relaxation have been intensively reviewed in the literature [38] and are not discussed in more detail here. For the description of the Förster transfer process, however, it is important to derive an expression for the fluorescence yield. The (intrinsic) lifetime of the excited state is given by

$$\tau_0 = \frac{1}{k_\mathrm{F} + k_\mathrm{NR}}, \tag{5.21}$$

where k_F is the fluorescence rate and k_NR the rate given by all non-radiative dissipation processes. The fluorescence decay time is expressed by

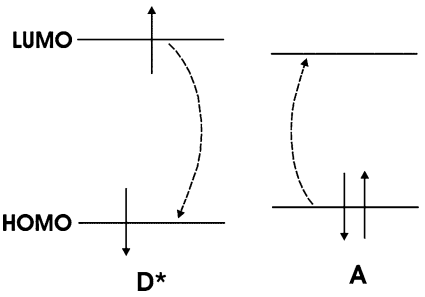


Fig. 5.29 Schematic for excitation energy transfer between donor (D) and acceptor (A) fluorophor

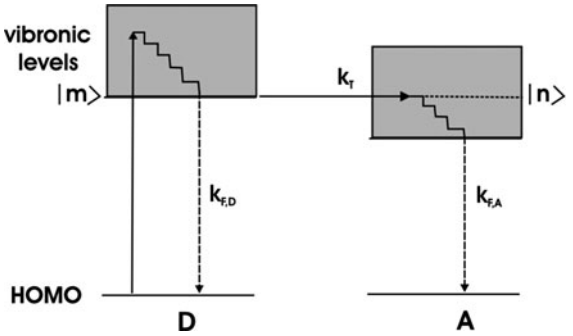


Fig. 5.30 Schematic of resonant excitation energy transfer between donor D (excited state $|m\rangle$ and acceptor A empty state $|n\rangle$ including the vibronic and electronic deexcitation processes within each molecule

$$\tau_{\mathrm{F}} = \frac{1}{k_{\mathrm{F}}}. \tag{5.22}$$

The fluorescence yield of the donor (light-emitting) molecule in the absence of any acceptor molecule is obtained from the ratio of the lifetimes

$$Y_{\mathrm{D}} = \frac{\tau_0}{\tau_{\mathrm{F}}} = \frac{k_{\mathrm{F}}}{k_{\mathrm{F}} + k_{\mathrm{NR}}}. \tag{5.23}$$

We now consider a donor–acceptor molecule pair where the donor emission spectrum overlaps with the acceptor absorption spectrum and the fundamental absorption edge of the donor given by the energy gap between HOMO and LUMO is at higher photon energy than that of the acceptor. Figure 5.29 shows the situation for an optically excited donor molecule and a nearby acceptor in its ground state (D*A). The fluorescence of the donor, indicated by the dashed downwards oriented arrow, results in excitation of the acceptor: D*A $\rightarrow$ DA*. The energy of the excited molecule D* is transferred in a singlet–singlet transition to the acceptor molecule.

The excitation energy transfer is quantum mechanically described by quantum physics perturbation theory assuming weak interaction for which Fermi's golden rule holds [39]. The transfer rate expressed by the interaction matrix element V_{mn} and the energy conservation (energies v_n and v_m) between donor state $|m\rangle$ and acceptor state $|n\rangle$ is represented by the delta function that defines the resonance condition (Fig. 5.30):

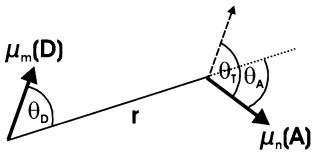


Fig. 5.31 Geometry for the interaction of transition dipoles $\boldsymbol{\mu}_{\mathrm{m}}$ of the donor molecule (D) and $\boldsymbol{\mu}_{\mathrm{n}}$ of the acceptor molecule (A) at distance r in polar coordinates

$$k_{\mathrm{T}} = 4\pi^2 \left|V_{mn}\right|^2 \delta(\nu_n - \nu_m). \tag{5.24}$$

The matrix element V_{mn} contains the dipole–dipole interaction for the transition dipole of the donor molecule and that of the acceptor molecule which will be labelled now according to the resonant interacting states $|m\rangle$ (donor) and $|n\rangle$ (acceptor) and $\boldsymbol{\mu}_{\mathbf{m}}$, $\boldsymbol{\mu}_{\mathrm{n}}$ instead of M_{fi}. This simplifies the indexing of the contributing states and dipole moments. The interaction matrix element is given by the product of the transition dipole moments $\boldsymbol{\mu}_{\mathrm{m}}$ and $\boldsymbol{\mu}_{\mathrm{n}}$ and an orientation factor κ that describes the relative position of the two transition dipoles as shown in Fig. 5.31.

The orientational factor $\boldsymbol{\kappa}$ is given by the geometric expression

$$\kappa = (\cos\theta_{\mathrm{T}} - 3\cos\theta_{\mathrm{D}}\cos\theta_{\mathrm{A}}) \tag{5.25}$$

and the interaction matrix element can be expressed as

$$V_{mn} = \frac{l^2}{4\pi\varepsilon_0\varepsilon r^3}\vec{\mu}_m\vec{\mu}_n\kappa, \tag{5.26}$$

where l is a factor that describes the difference in electromagnetic interaction with respect to vacuum. Each transition dipole moment is given by

$$\vec{\mu}_{fi} = -e\sum_k \left\langle \varepsilon_f \left| \vec{r}_k \right| \varepsilon_i \right\rangle \left\langle \nu_f \mid \nu_i \right\rangle = \vec{\mu}_{\varepsilon}^{f,i}\, S(\nu_f, \nu_i) \tag{5.27}$$

in Dirac's bracket notation (see Chap. 6). The first term on the right-hand side is the electronic transition contribution and the second term represents the overlap integral between the vibrational state in the initial (subscript i) and final (f) electronic states of the molecule. The transition moment μ_{fi} in (5.27) is a general but more detailed expression that applies for both, the donor and acceptor molecule transitions in (5.24). Therefore, one obtains for the respective transition moments of donor and acceptor molecule

$$\vec{\mu}_m = S_m\vec{\mu}_{\mathrm{D}}; \quad \vec{\mu}_n = S_n\vec{\mu}_{\mathrm{A}}. \tag{5.28}$$

The expression for the resonant energy transfer according to (5.24) is then

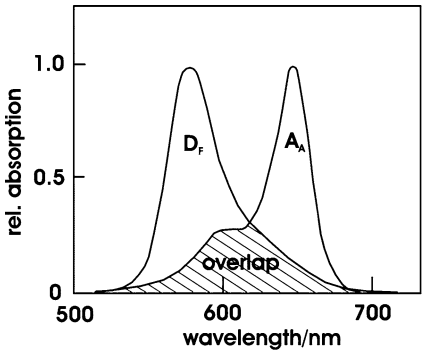


Fig. 5.32 Schematic indicating the spectral overlap of emitter (donor) molecule fluorescence (D_F) and of receiver (absorber) molecule absorption (A_A) spectrum

$$k_T = \frac{l^4\kappa^2}{4\varepsilon_0{}^2 n^4 r^6} \sum_{n,m} \mu_D^2 \mu_A^2 S_m^2 S_n^2 \delta(\nu_m - \nu_n). \tag{5.29}$$

The Förster transfer rate k_T is proportional to r^{-6} which results from the distance dependence of the matrix element in (5.26) and the definition of k_T via V_{mn} in (5.24). In (5.28), the Franck Condon factors S that describe the vibrational overlap have been separated from the electronic contributions, and the energy conservation as well as the labelling of the transition moments relates to the vibrational states. As a result, the Förster transfer rate can be expressed by the product of the fluorescence spectrum of the emitter (donor) molecule, $F(\nu)$, and the absorption spectrum of the receiver (acceptor) molecule, $A(\nu)$:

$$k_T = 8.8 \times 10^{17} \frac{\kappa^2}{n^4 \tau_D r^6} \int \frac{A(\nu) F(\nu)}{\nu^4} \mathrm{d}\nu. \tag{5.30}$$

The integral term in (5.30) is known as the overlap integral

$$J = \int \frac{A(\nu) F(\nu)}{\nu^4} \mathrm{d}\nu \tag{5.31}$$

and Fig. 5.32 shows a schematic graph on this type of interaction.

The so-called Förster radius is defined as the distance for which the efficiency of energy is 50%. In this case, the fluorescence appears partly quenched. This distance dependence provides an excellent measure of intermolecular distances in macromolecular and biological environments and FRET has been employed as a "biological ruler" [40, 41] It can be shown that

$$E_{FRET} = \frac{1}{1 + \left(\frac{R_0}{R}\right)^6} \tag{5.32}$$

and with

$$R_0^6 = 8.8 \times 10^{17} \frac{\kappa^2}{n^4} J, \tag{5.33}$$

one obtains an expression for the Förster rate

$$k_T = k_D \left(\frac{R_0}{R} \right)^6. \tag{5.34}$$

R_0 is in the order of 5–10 nm. Accordingly, this mechanism provides an efficient excitation energy transfer over rather large distances without charge carrier transport. In the analysis of excitation energy transfer in photosynthesis, discrepancies between the Förster transfer rate determined from experiments with the predictions of single chromophoric FRET theory have been observed as mentioned above. The rate for Förster transfer was determined assuming single chromophoric events [36] and a pronounced improvement in the agreement between measured and modelled transfer rates was achieved by postulating a so-called multi-chromophoric (MC) FRET which includes disorder and temperature effects [42, 43]. The distance dependence of MC-FRET is a function also of disorder that affects the spatial and energetic nature of exciton states and of temperature which alters the relative population of such states. It is also emphasized that a simple exponential evaluation of experimental data is not appropriate in these systems. Also, a type of inter-chromophoric term has been introduced that involves quantum mechanical coherence among chromophores resulting in collective donation and acceptance of the excitation energy, leading to increased rates of energy transmission. The advanced model yields convincing agreement with experimental observations and emphasized the role of Förster processes in photosynthetic entities.

5.4 Semiconductor–Enzyme Junctions: Implications for Biomimetic Approaches

The assessment and control of the fundamental properties of enzymes, adsorbed on semiconductor surfaces, has the potential for the design of hybrid inorganic–organic structures that are (photo)catalytically active. In a typical biomimetic approach, the light-induced excess minority charges from an illuminated semiconductor absorber could be transferred to the reactive site of an enzyme, an enzyme mutant or a suited macromolecule. This type of work is yet in its precursor stage [44]. The crucial parameter in this approach is the efficient transfer of the excitation energy from the organic or inorganic absorber to a catalytically active, possibly macromolecular, species residing on the absorber surface.

If one wants to substitute the enzymes in an attempt to arrive at more robust systems, the application of metallo-proteins appears promising. Hitherto, most

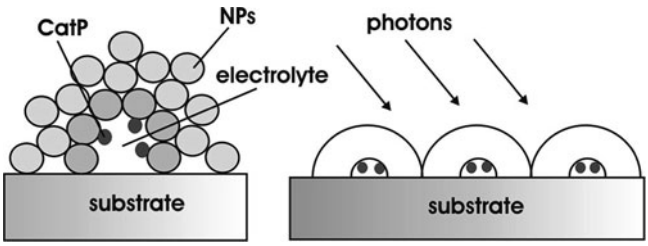


Fig. 5.33 Envisaged catalytically active structure consisting of light-absorbing nanoparticles (NP) that have a semi-ring arrangement and of catalyst particles (CatP) in the inside of the ring or tube structure; the catalyst particles are not excited directly by the light (see text)

attempts involve the "dye sensitization" approach where an excited exciton in a macromolecule transfers its energy into an electrode that acts as excess carrier scavenger [45]. This approach encounters the difficulty of photon absorption with a high absorption coefficient over a wide spectral range and represents one of the limitations of dye sensitization cells [46].

5.4.1 Introduction

Among the problems associated with hybrid inorganic–organic solar energy converting structures is the energy transfer mode: if direct electron transfer via tunnelling or a hopping-type mechanism is considered, the device structure will have to be different from that for a Förster or Dexter excitation energy transfer system. Figures 5.33–5.35 give an overview of envisaged structures. Leaving out the well-established direct electron transfer, the excitation energy transfer based on (1) exciton hopping or (2) a Frenkel exciton–Förster process is displayed. The schemes are tentative and a realization has not yet been demonstrated. In Fig. 5.33, the "traditional" route of photon absorption in nanoparticles that are arranged such that they can absorb most of the solar light is shown. The catalytic sites are located within a tubular arrangement where solution can penetrate and the gaseous products can escape.

The tubular or ring-type structure in Fig. 5.33 ensures that the optical absorption process takes place in the nanoparticles that form the outer ring. Therefore, the catalyst can be made of a material with reduced stability as it is not exposed directly to the sunlight. The inside of the tubes/rings is filled with electrolyte, and hydrogen evolution can take place at the catalyst surfaces. The excitation energy transfer along the nanoparticles is shown in Fig. 5.34 in an energy schematic:

The generation of an exciton in the particle on the left in Fig. 5.34 by illumination is followed by exciton hopping where the excited electron moves to the neigh-

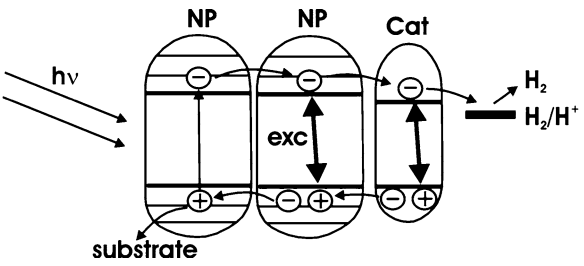


Fig. 5.34 Envisaged scheme for exciton hopping induced catalytic evolution of hydrogen; the process is assumed to occur with inorganic nanoparticles (NP) that are functionalized with catalyst particles (CatP) or molecules

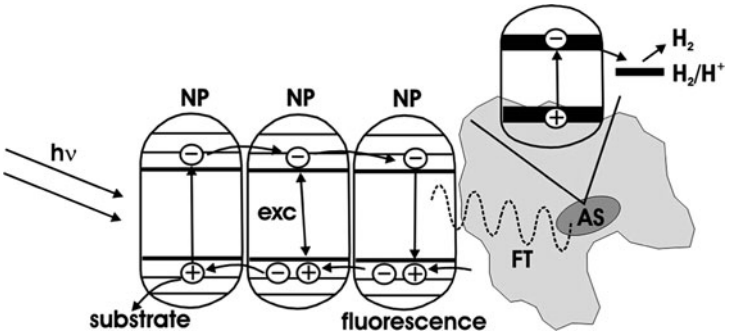


Fig. 5.35 Envisaged biomimetic structure for Förster transfer-based excitation energy transport from nanoparticles to the catalytically active site of an enzyme or a macromolecule with metallic centres (see text); *exc* exciton, *FT* Förster transfer, *AS* catalytically active site within a macromolecule/enzyme (*grey area*); the electronic orbitals at the active site (HOMO, LUMO or extended π-systems) are drawn in bulk *horizontal lines*

bouring particles' excited level and a ground state electron from that neighbouring particle moves towards the unoccupied ground state of the particle that has been optically excited, due to a gradient in chemical potential. The final state of the interaction between the two nanoparticles is that the initially generated exciton on particle 1 has moved to the adjacent particle. This process resembles a Dexter transfer which is efficient only at rather short distances because the transfer rate depends on the sum of the van der Waals radii of the interacting species [47]. In the next step in the excitation energy transfer chain, the exciton is transferred to the catalyst particle (molecule) where, at its active site, fuel generation occurs. Here, the case of hydrogen evolution from an acidic solution has been depicted. In the configuration of tubes or rods, shown in Fig. 5.33, the catalyst itself is not directly illuminated. Instead, the absorption takes place in the spherically shaped nanoparticles and the energy is transferred to the catalysts. The realization of such structures requires a low resistivity contact for the substrate to collect the holes (in the case of H_2 evolution) which can then be used to maintain an anodic catalytic process (Cl_2 or O_2 evolution, for example).

Possible realizations of Förster-transfer-based devices are sketched in Figs 5.35 and 5.36. In Fig. 5.35, the excitation energy is collected by nanoparticle in an analogous manner as is shown in Fig. 5.34, the difference being that the catalyst

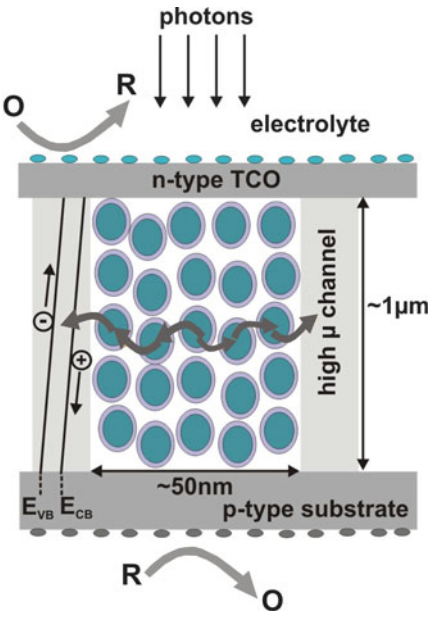


Fig. 5.36 Visualization of a photocatalytic device that operates based on the Förster excitation energy transfer mechanism (see text); *TCO* transparent conductive oxide, E_{CB}, E_{VB} conduction and valence band, respectively, *R*, *O* reduced and oxidized species in solution, respectively, μ mobility, *small arrows* indicate the Förster transfer, the *larger grey arrows* the electrochemical reactions; the *small ellipsoidal objects* indicate catalysts

particle is now specified as either an enzyme with an active site in its interior or a macromolecule where the catalytically active metals are also embedded in the central region. The distance from the contact of the outer parts of the molecule with the nanoparticle is too large, for instance, for a Dexter exciton exchange. Therefore, electron transport mechanisms such as tunnelling, hopping or superexchange [48] could occur for the transfer of the photon energy to the catalyst, or a Förster-type process without actual electron transfer provides the excess energy at the catalytically active site. In the investigation of semiconductor–enzyme structures with STM, strong indications have been obtained that electron transfer processes from the semiconductor surface to the enzyme's catalytic site are difficult to realize due to the protein surrounding bio-water (see below). Therefore, the Fig. 5.35 shows a Förster-transfer-based device where the final step is hydrogen evolution.

As realization, a structure that uses quantum size properties of nanoparticles is shown in Fig. 5.36. It is composed of nanoparticles (quantum dots) with different sizes. Because in these small dimensions, the energy gap is a function of the size, the energy gap of the smaller particle is larger and the corresponding exciton energy is largest, too. With increasing size, the energy gap shrinks, thereby allowing for Förster transfer as indicated by the grey arrows. The excitation energy is transferred over distances in the range of up to 10 nm between neighbour nanoparticles and the device needs a compartmentalization where the subunits have a lateral size of approximately 50–60 nm. Although such a device appears somewhat futuristic, it would enable ultrafast excitation energy transfer in a system that is typically limited by low mobilities in and across the nanoparticles. The conductive channels on each

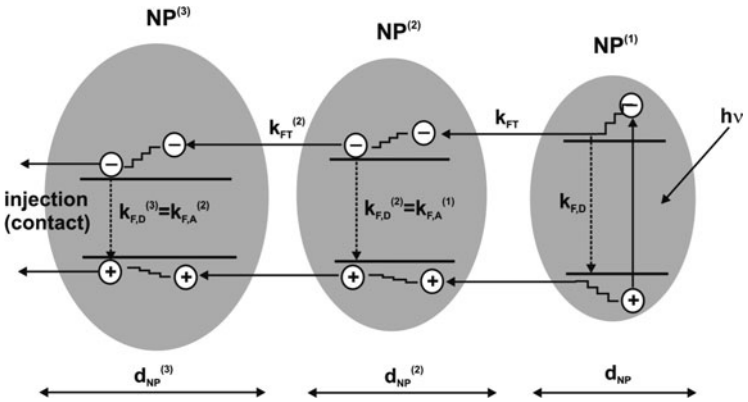


Fig. 5.37 Schematic on a Förster excitation energy transfer between quantum dot sized nanoparticles of different sizes; d_{NP} nanoparticle size (here, an ellipsoidal particle has been chosen for generality and d represents its smallest dimension), $k_{F,D}$ rate of fluorescence of the respective donor, labelled up to (3), according to the labelling of the nanoparticles ($NP^{(1)}$–$NP^{(3)}$, k_{FT} Förster transfer rate

side of the device have the function to separate electron and holes and transport them to the respective contact (front/rear) where they induce redox reactions at the surface of the respective catalyst. For good photoconductivity, the use of an intrinsic semiconductor can provide large drift velocities in the rather constant electric field within the semiconductor.

The excitation energy process according to (5.24–5.34) is visualized in Fig. 5.37. Three quantum-sized nanoparticles have been drawn that differ in size. Upon irradiation, the Förster transfer takes place from the smaller to the larger particles and, in the device shown in Fig. 5.36, the transfer from particle 2 to particle 3 can overlap with the transfer induced by the direct illumination of particle 2 and successive Förster transfer to particle 3. The limitation of the system is given by the spatial range of the energy transfer. Also, the vertical dimension of the device (Fig. 5.36) depends on the absorption properties of the particles; assuming transitions with a large oscillator strength such as direct transitions in bulk semiconductors, the absorption coefficient could easily reach values of $10^5\,cm^{-1}$, corresponding to an absorption length in the range of a few hundred nanometers. Therefore, such devices could act as multi-chromophoric thin film absorbers because the absorption threshold for each nanoparticle is different, depending on its size.

After the consideration of possible architectures that are based on energy-transfer processes mostly realized in nature, we return to systems that catalyse biochemical reactions via enzymatic processes. The investigation of enzymes properties and dynamics has been performed by scattering methods in a non-native state [49, 50] which has yielded considerable insight into their ternary and quaternary structure. For envisaged solar applications, the interaction with semiconductors is of interest

because of possibilities of charge or excitation energy transfer processes between (illuminated) semiconductors and enzymes. The first step in such experiments is the immobilization of the enzymes on semiconductor surfaces thereby maintaining their natural structure and shape which provides the basis for enzymatic function. Related experiments are described below.

5.4.2 Electronics and Charge Transfer Aspects at Protein–Semiconductor Junctions

5.4.2.1 Life Cycle of a Retrovirus and Reverse Transcription

The development of hybrid devices which operate at inorganic (solid)–organic (enzyme) interfaces is based on the aforementioned immobilization. Of particular interest is site-selective immobilization. For light-sensitive structures, either chromophores coupled to solids or macromolecules coupled to photoactive semiconductors can be investigated. The former approach where photonic excess energy is collected in organic macromolecules has already been described to some extent above in the treatment of photosynthesis although it operates without a solid phase. It is pursued intensively in research on artificial photosynthesis [51]. The latter approach, however, will be further described here.

The initial work was based on the first imaging of proteins on semiconductor electrodes by STM [52]. At that time, the motivation behind the work was the attempt to image a well-known tertiary structure of comparably large heterodimeric enzyme whose structure was known due to their medical relevance. Therefore, the work that will firstly be described is related to the STM imaging and to the interpretation of the images of retroviral enzymes, called reverse transcriptases. These enzymes are of particular relevance in the life cycle of retroviruses such as the human immunodeficiency virus (HIV), the avian myeloblastosis virus (AMV) and the murine leukaemia virus (MuLV), for instance. The reader finds an overview on the topic in [53] and Fig. 5.38 shows a schematic of the life cycle of a retrovirus.

A characteristic of retroviruses is that they contain all necessary components for viral reproduction. Instead of undergoing endocytosis, i.e. engulfing the virion with the cell membrane, the virion attaches with its glycoprotein shell to the membrane and fuses directly with the plasma membrane (step 1 in the above figure). Then, following entry by fusion and capsid uncoating, the nucleocapsid enters the cytoplasm of the cell (step 2) and deoxynucleoside triphosphates from the cytosol enter the nucleocapsid. In the nucleocapsid, RT catalyses the synthesis of the proviral DNA from the RNA genome of the virus (step 3). The viral DNA copy is integrated into the nucleus (only one host-cell chromosome is depicted) and spliced into one of many possible sites of the host-cell chromosomal DNA, catalysed by the enzyme integrase (step 4). The integrated viral DNA, referred to

as a provirus, is transcribed by the host-cell RNA polymerase, generating mRNAs (messenger RNA) and genomic RNA molecules. The host-cell machinery translates the viral mRNAs into glycoproteins and nucleocapsid proteins (step 5). In that process, the proteases cleave newly synthesized polyproteins at the appropriate places to create the mature protein components of an infectious HIV virion. The nucleocapsid proteins assemble with genomic RNA to form progeny nucleocapsids, which interact with the membrane-bound viral glycoproteins. Eventually the host-cell membrane buds out and progeny virions are pinched off (step 6).

The process of reverse transcription (step 3 in Fig. 5.38) is displayed in some detail, but abbreviated, in Fig. 5.39. The top graph shows so-called minus strand synthesis beginning near the 5′ end of the plus strand viral RNA where the transfer RNA (tRNA) acts as primer and binds to the primary binding site (PBS) of the host strand. The growth direction of the new strand through the u5 region is indicated by an arrow and synthesis ends at the r region of the viral genome. Below, a more advanced step of reverse transcription is drawn: here, the first jump (template exchange) of the synthesized minus strand genome has taken place, after disintegrating of the initial RNA host part (r, u5) by RNase H. The pbs is unaffected by the process. After passing the polypurine tract (ppt) region of the template (see arrow), the template escapes dissolution by RNase H and it serves itself as primer for plus strand synthesis by a mechanism called DNA-dependent DNA polymerization (DDDP). In the third graph in Fig. 5.39, a situation is shown where the template strand has largely been digested by RNase H. In that situation,

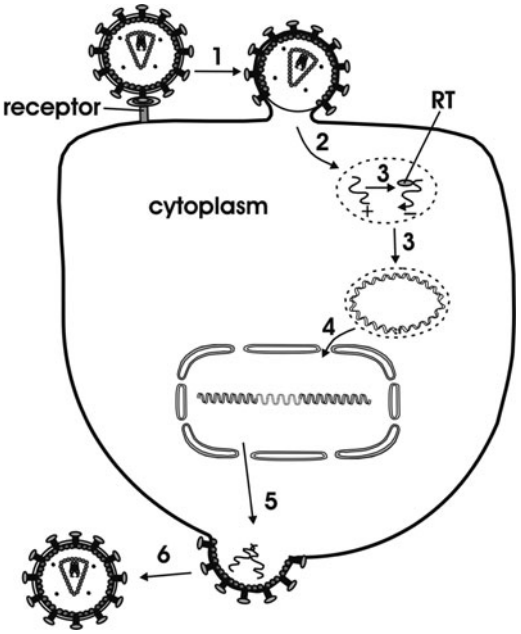


Fig. 5.38 Schematic of the life cycle of a retrovirus; RT: reverse transcriptase (see text)

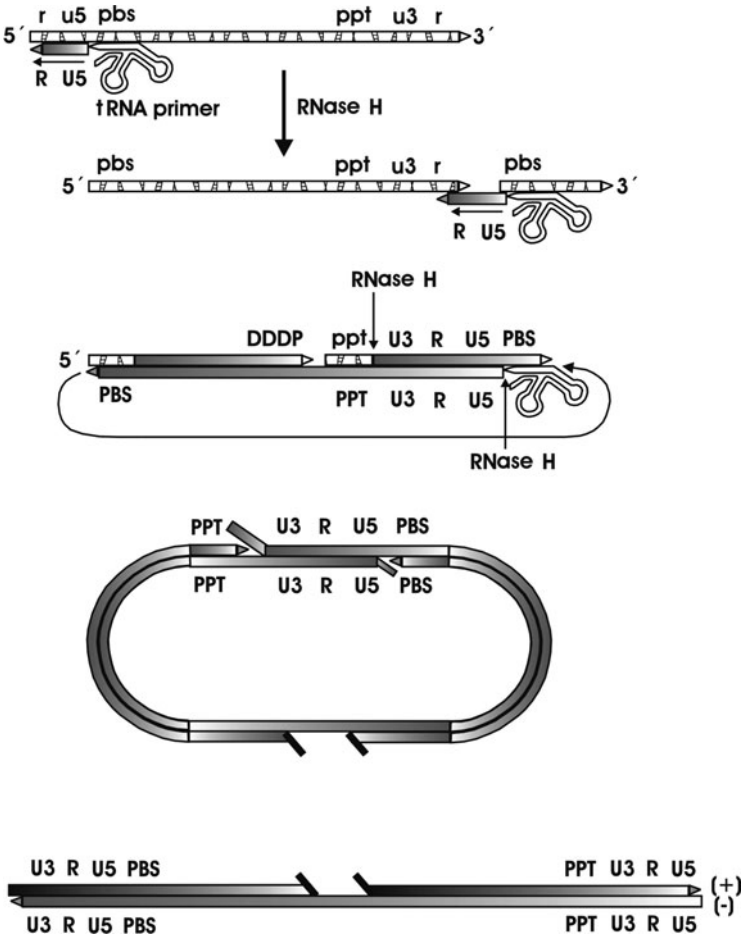


Fig. 5.39 Scheme of selected steps of the process of reverse transcription in the life cycle of a retrovirus (see text)

minus strand synthesis of the first (lower) copied strand extends to the PBS site of the new strand, and the second strand, growing in the opposite direction, is almost completely synthesized. Only the original pbs and ppt sites remain. The arrows, labelled RNase H indicate the removal of tRNA and ppt. The then exposed PBS fuses with the PBS sequence at the 3′ end of the minus strand DNA, resulting in a circular DNA with overlapping 5′ ends.

The overall synthesis of the new proviral DNA strand is terminated at the template strand breaks at the PBS and PPT ends, resulting in a linear molecule with long terminal repeats, shown at the bottom of Fig. 5.39.

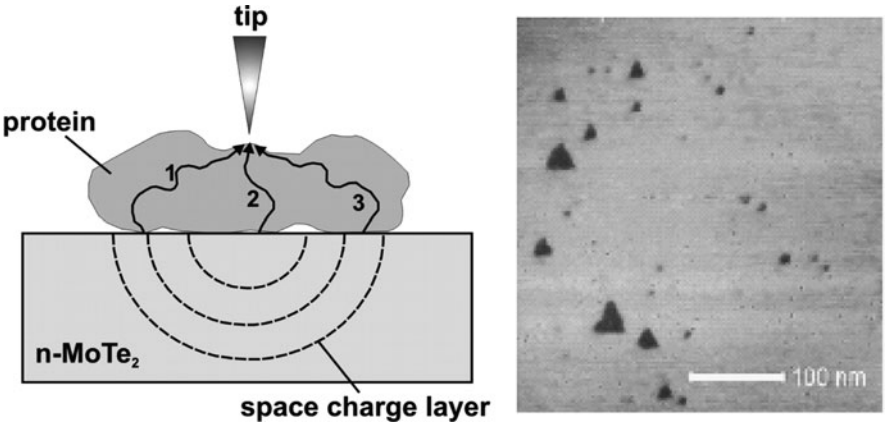


Fig. 5.40 Imaging geometry scheme for STM experiments in ambient atmosphere (*left*) and electron micrograph of a typical part of the $MoTe_2$ surface; also indicated is the space charge region in the semiconductor that exists upon contact formation between the tip and the semiconductor and upon biasing the semiconductor relative to the tip. Pathways 1–3 indicate possible routes for electronic transport through the molecule; an alternative path, not shown here, will be discussed below (see Fig. 5.55)

The origin of the indicated nucleotide sequence gap is presently discussed. This strand discontinuity might constitute a binding site for a protein that mediates nuclear transport or integration of the provirus. Finally, linearization occurs via strand displacement and/or repair and ligation, i.e. the connection of the 3′ hydroxyl end with the 5′ phosphate end via the enzyme ligase. This genome is then integrated into the host genome (step 4 in Fig. 5.38).

5.4.2.2 STM Imaging of Enzymes on Semiconducting Substrates

The STM imaging geometry is shown in Fig. 5.40 together with electron micrographs of the substrate used for the adsorption (immobilization) of the enzymes. A layered semiconductor, $MoTe_2$, has been used as substrate because of its triangular nano-dimensioned defects at the surface which result from the crystal growth procedure [54].

Since the semiconductor is n-type doped, the triangular defects represent negatively charged surface states, and around and underneath them, positively charged space charge regions, as shown in the figure, exist. The isoelectric point (IP) of the RT of HIV 1 is at pH∼9. Accordingly, at the pH of the used buffer solution for deposition of the enzymes ($pH < 7$) the enzyme's overall charge is positive. Therefore, the forces that govern adsorption in the sense of the DLVO theory [55, 56], namely the electrostatic and the van der Waals attraction, add, resulting in an attachment of the molecules to these charged surface sites. A schematic is drawn in Fig. 5.41.

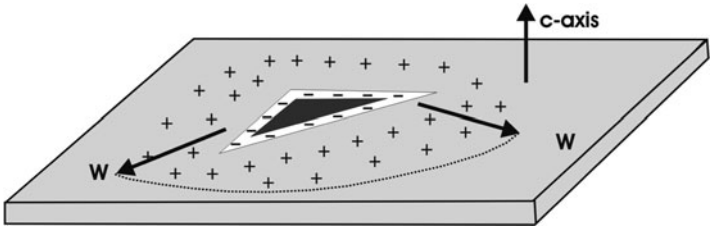


Fig. 5.41 Charge distribution around nanoscopic triangular defects at $MoTe_2$ surfaces; the defects act as surface states that carry negative charge on an n-type semiconductor. The resulting space charge does not only occur perpendicular to the surface but also in the surface plane as indicated by its lateral extension W; the c-axis defines the crystallographic orientation perpendicular to the sandwich planes of the material (see Fig. 5.42)

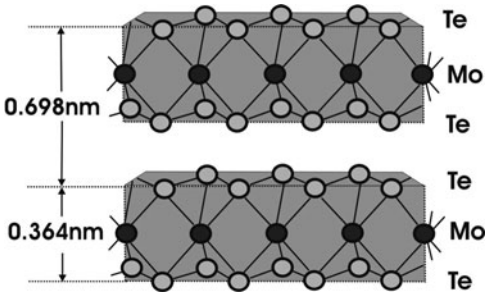


Fig. 5.42 Structural arrangement of layered group VI transition metal dichalcogenides; example: $MoTe_2$ with trigonal prismatic arrangement of the metal-chalcogen bonding, forming sandwich-type layers; the c-axis is the growth axis of the materials and points perpendicular to the layers

The energetic situation for a semiconductor with surface states is displayed in Fig. 5.43 perpendicular to the surface.

The situation depicted in Fig. 5.43 is more general than is actually assumed later in the discussion of the results in order to show the influence of a high density of surface states [57]. The higher this density, the less discharging or charging of surface states occurs when contact with a metal such as the STM tip is formed. For high electronic quality surfaces, the negligible surface state density results in an increase of the band bending upon contact formation and the difference $E_F - E_F^C$ shows up as an increase in eV_{BB}. For strong Fermi-level pinning and an accordingly high surface state density, the band bending remains unaffected and the band edges at the surface shift; in this case, $E_F - E_F^C = \Delta E_{CB,VB}$. The latter is likely to be the case with the rather strongly disturbed surface of $MoTe_2$ that is seen in Fig. 5.40.

The experimental arrangement for STM experiments under ambient conditions is shown in Fig. 5.44. In the constant current mode applied, the tip is scanned across the sample and the tip moves in order to restore the constant current that has been set. The value used in the STM experiments on the reverse transcriptases was 200 nA.

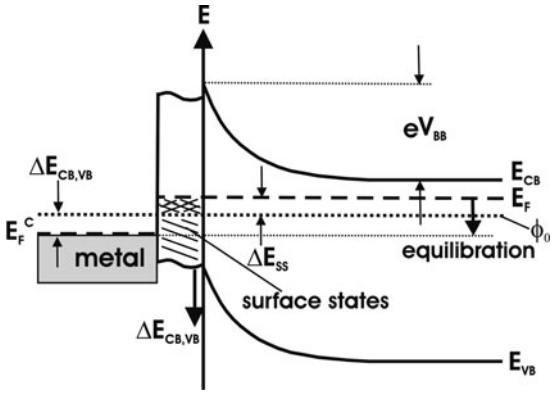


Fig. 5.43 Contact formation between an n-type semiconductor with surface states as indicated and a metal (STM tip), represented by its Fermi level; here, the situation of partial Fermi-level pinning is shown: before contact to the metal, the semiconductor was in equilibrium with its surface states, given by the Fermi-level E_F; upon contact, the new Fermi level is given by E_F^C, i.e. the Fermi level of the metal; the contact potential difference E_F–E_F^C is, in this case of *partial* Fermi-level pinning, split between a part that increases the band bending eV_{BB} by discharging surface states down to the value ϕ_0 and a part $\Delta E_{CB,VB}$ that results in a shift of the band edges towards lower energy

If the conductivity of the samples would change lateral to the surface, then the tip would retract (increased conductivity, larger current) or approach the sample if more insulating areas are scanned. Therefore, this observation mode displays not a true height image of the specimen but an image where electronic properties of the tip-sample system can strongly contribute. In addition, the curvature of the tip apex can substantially alter the size of the imaged adsorbate, due to tunnelling occurring already when the tip approaches the sample laterally. This enhancement factor is dependent on the specific tip and cannot be predicted because a reproducible preparation of high quality tips with extremely low tip apex curvature is not possible at present.

The biomolecule handling is involved and includes storage at −20°C, dialysis of the used carrier solution to remove organic materials and most of the HEPES ($N-2-$hydroxyethylpiperazine-$N'-2-$ethanesulphonic acid) buffer solution and electrophoresis was performed to identify the reverse transcriptases. The RTs of HIV 1 and AMV were supplied in heterodimer form as 66 kD/51 kD unit for HIV 1 and as 95 kD/65 kD unit for AMV. To ensure the quality of the STM experiments, blind experiments on the dialysed buffer solutions without the enzymes were carried out [58]

Figure 5.45 shows two structure schematics of the RT of HIV 1.

The similarity of the tertiary structure with the right human hand has led to an according assignment of the enzyme sub-structure; the grip of the protein around the newly synthesized proviral DNA (compare Figs 5.38 and 5.39) particularly emphasizes the mentioned analogy with a human hand.

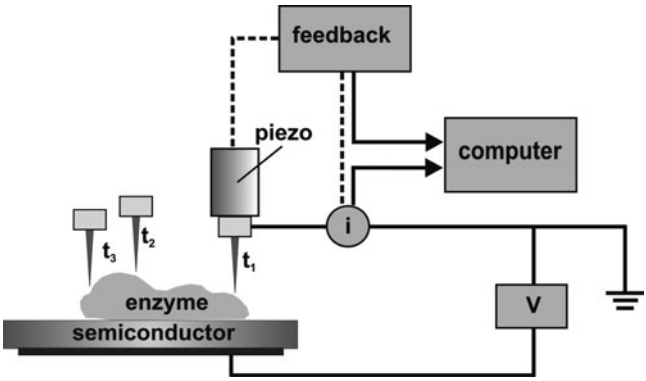


Fig. 5.44 Schematic diagram of a constant current STM experiment; the tip (*grey*) moves across the biomolecule (BM) with a given scan velocity v_{scan}; at the times t_1–t_3, the feedback adjusts the z-direction of the piezo to maintain the set constant current (I), producing an image that contains electronic information, too

Due to the uncertainties of the molecule–surface interactions in the presence of an electrolyte that also comprise the deposition dynamics from a pipette or syringe onto the surface and the unknown factors in deceleration of the solution and enzyme precipitation, the actual shape of the deposits on the surface cannot be predicted. Figure 5.46 shows that obviously, the position proposed in Fig. 5.45 (top) has been assumed by the molecule. Interestingly, all major features can be assigned to the image, including thumb, fingers, palm, active site and RNase H activity and the features are also visible for the lower molecular weight subunit in the background. The top view (Fig. 5.47) also shows the complex tertiary structure of the enzyme. It should be noted that the height given on the axis of Fig. 5.46 is about 2 nm, i.e. too large for through space tunnelling from the substrate to the tip at the preset current. The factors that contribute to the rather undistorted images will be discussed after the presentation of two additional images in Figs 5.48 and 5.49. Whereas Fig. 5.48 appears to show the first evidence of a reverse transcriptase, attached to a synthesized polypeptide strand, Fig. 5.49 is important for the discussion of the conduction mechanism that allows STM imaging of the proteins without destroying their tertiary structure.

Figure 5.48 shows a linear strand that has been "gripped" by a C-shaped entity. Also, 3 smaller not very clearly resolved features are seen along the strand. The C-shaped feature has been tentatively interpreted as a reverse transcriptase acting on an RNA strand for catalysis of proviral DNA [59], whereas the smaller ones could also be a result of the scanning process across the biomolecular system where the tip can pick up less strongly bonded parts resulting in an erosion of the height of initially larger features. Therefore, the smaller seen features are interpreted of remnants of reverse transcriptases.

Figure 5.49 presents a result whose interpretation has raised considerable discussions. It shows DNA strands imaged by STM in the constant current mode [60].

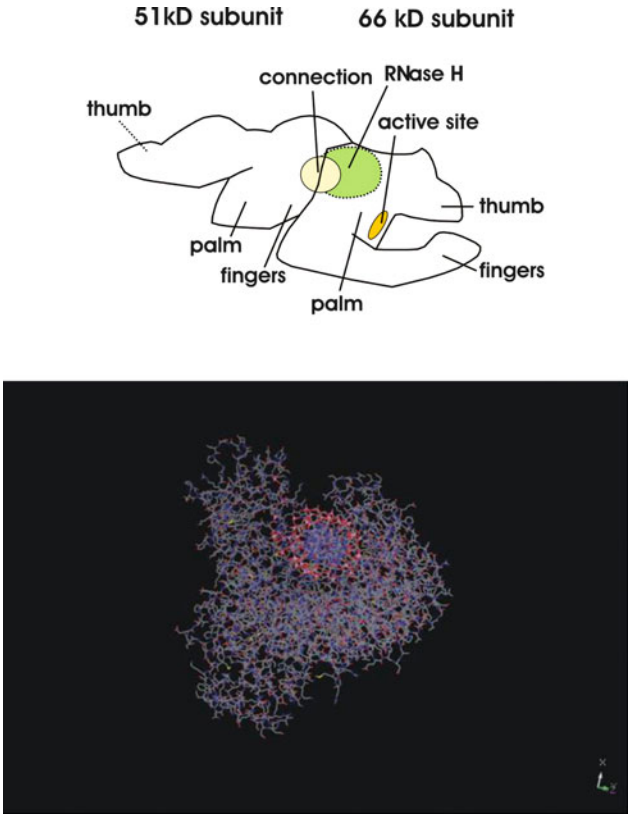


Fig. 5.45 Structure images of the reverse transcriptase of HIV 1; the similarity with the Klenow fragment has led to the assignment of tertiary structure subunits according to the right human hand as fingers (F), thumb (T), palm (P) and includes the connection (C) and the ribonuclease activity RNase H (R) (*upper image*); the image also shows a possible position on the surface with the 66-kD subunit in front and the 51-kD subunit (without the RNase H) behind it; the *lower image* depicts the situation of proviral DNA strand synthesis with the "grip" of the RT that demonstrates the analogy with the human hand

With a customized microscope, imaging at or below the pA range has been possible. The experiment has been performed on DNA strands that were deposited on mica. Mica is known as an insulator as is the polypeptide molecule. In the original work, it was shown that the imaging process depended on the relative humidity (RH) and the applied voltage to the surface contact on the mica as seen in the experimental arrangement schematic in Fig. 5.50. For voltages of –7 V vs. the grounded STM tip and RH $>$ 74%, images could be obtained despite the fact that sample and substrate are insulators. It was claimed [60, 61] that the high conductivity of ultrathin water films was responsible for the effect, with a conductivity increase of "surface water" with regard to bulk values in the order of 10^4. The rather undistorted STM images of the reverse transcriptases (Figs 5.46 and 5.47) demonstrate that a constant current

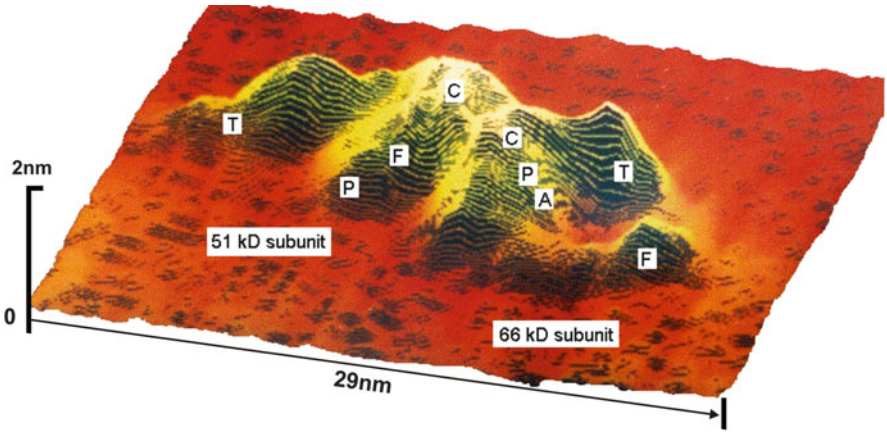


Fig. 5.46 Constant current STM *side view* image of a reverse transcriptase of HIV 1 deposited onto n ─ $MoTe_2$; current: 200 pA; bias: −0.7 V vs tip

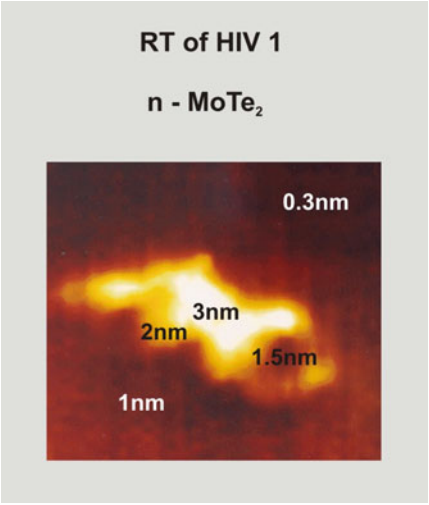


Fig. 5.47 *Top view image* of a reverse transcriptase; experimental parameters as in the *upper figure*; the approximate heights of the major subunit parts are indicated in the figure; the lateral dimension corresponds to that of Fig. 5.46.

of 200 pA can be drawn across the molecule, actually *independent* of the position of the tip above the molecule. If less conductive regions would exist within the protein, the STM tip would closer approach the surface and, due to the scanning movement, disrupt the enzyme. As this appears not to occur, the experimental results, shown in Figs 5.46, 5.47 and 5.49, demand an explanation. In particular, the question arises how and where the electron transfer processes occur across the molecules. For an analysis, the energetic conditions across the tip–air gap–enzyme–semiconductor junction have to be analysed. These junction energetics allow an

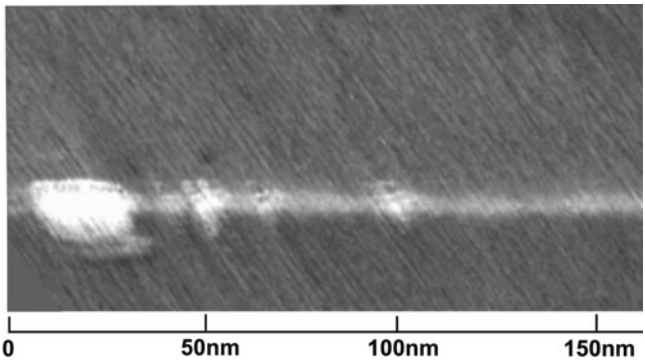


Fig. 5.48 STM image of a hybrid structure recorded on n − $MoTe_2$ under the same experimental measurement conditions as in Figs. 5.46 and 5.47

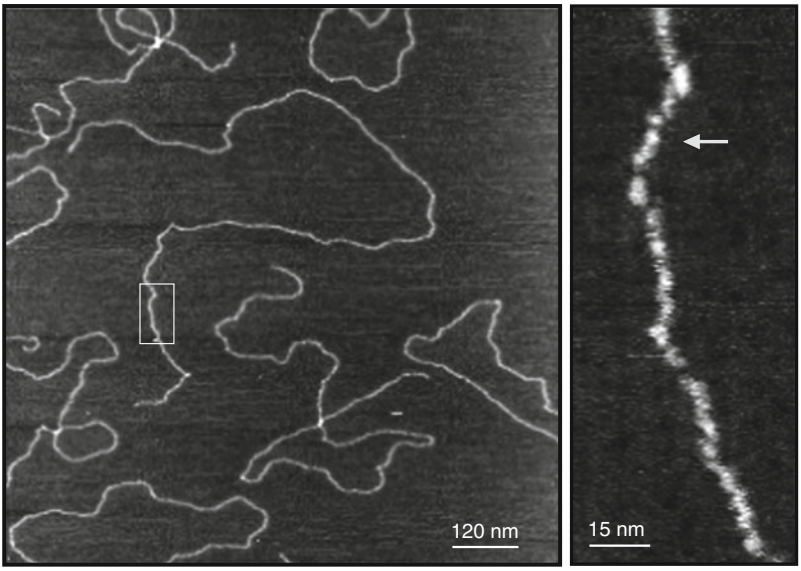


Fig. 5.49 pA-STM image of DNA on mica; bias: −7 V vs. tip, RH = 74%; constant current: 1pA; after ref. [56]; the *right-hand side* shows a magnification of the region within the *white rectangle* of the *left image* (see text)

assessment of the direction and magnitude of the static electric field across this complex junction. Below, the analogy of the system with a planar metal–insulator–semiconductor (MIS) junction will be explored.

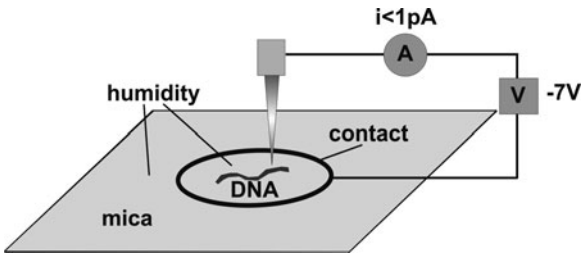


Fig. 5.50 Experimental geometry for the measurement of DNA on mica as shown in Fig. 5.49; relevant parameters are indicated in the figure

5.4.2.3 Junction Energetics

First, the energetic situation at the STM tip–gap–semiconductor junction is considered before contact formation in analogy to a Schottky or MIS (metal–insulator–semiconductor) junction [62]. The energetic situation is displayed in Fig. 5.51. Due to surface states, however, a band bending is already present before contact. The electron affinity of $MoTe_2$ has been determined by ultraviolet photoelectron spectroscopy from the secondary electron cut-off to be 3.9–4 eV. The work function of the Pt tip used in the experiments is somewhat uncertain due to the tip curvature which can reduce the work function to values around 5 eV [63]. Instead of typical values for polycrystalline Pt of about 5.6 eV, a value of 5.2 eV has been taken which, of course, adds a hypothetical aspect to the following considerations. Nevertheless, the analysis, based on the concepts of metal–insulator–semiconductor contacts in the field of applied semiconductor physics, allows a first assessment of the influencing parameters of the direction and strength of the electrical field across the junction and on the voltage dependence of the STM experiment.

The overall contact potential difference of $\Delta E_c \sim 1\,\mathrm{eV}$ from the position of the Fermi level of $n-MoTe_2$ at 4.2 eV below the vacuum level (without band bending) is split between the value of eV_{BB} (0.5 eV) and eV_c (0.5 eV) because $\Delta E_c = e(V_{BB} + V_c)$. Upon contact formation, the Fermi levels equilibrate and the new Fermi level position is identical to that of the grounded tip as can be

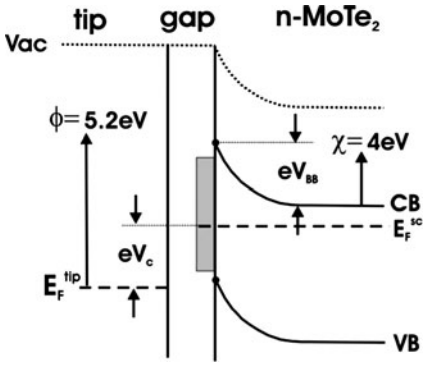


Fig. 5.51 Energy vs. space diagram of the Pt tip–tunnel gap–n–$MoTe_2$ junction; E_F^{tip} and E_F^{sc} denote the Fermi level of the tip and the semiconductor, respectively; eV_c: contact potential difference; *grey box:* surface states; eV_{BB}: semiconductor band bending due to surface states; *CB*, *VB*: conduction band bottom and valence band top, respectively

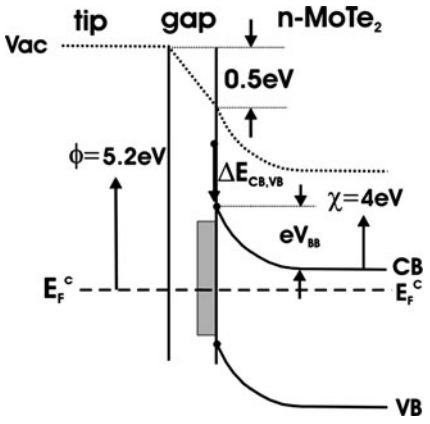


Fig. 5.52 Energetic situation at the STM-tip-gap-n-$MoTe_2$ contact after contact formation; $\Delta E_{CB,VB}$: band edge shift due to potential drop across tunnel gap (see text)

seen in Fig. 5.52. The remaining contact potential difference eV_c of 0.5 eV drops across the interfacial tunnel gap because it has been assumed that strong Fermi level pinning prevails that leaves the charge in the surface states unaffected by the contact formation. This drop in the (electrostatic) Galvani potential results in a downward band edge shift $\Delta E_{CB,VB}$. The energetic distance between the semiconductor Fermi level and the valence band maximum remains unchanged because strong Fermi level pinning exists.

When a negative bias of –0.7 V is applied to the sample, (the potential at which the onset of reliable imaging of the biomolecule occurred), the situation inverses: the electrostatic (Galvani) potential across the tunnel gap increases towards the semiconductor (see Fig. 5.53) which results in an opposite direction of the electric field across the junction. Then, electrons move, according to the definition, in the opposite direction of the electric field, i.e. from the semiconductor to the tip. The Galvani potential across the interlayer where the biomolecule resides is higher at the semiconductor side than on the tip side.

Introducing the reverse transcriptases that are located on the semiconductor surface adds additional complexity to the problem. For well-defined assessment of the overall energetic situation, one would need the ionization potential I_p of such molecules which, however, is presently not known. Also, the knowledge of the energy gap between HOMO and LUMO and information of the energetic width of the polypeptide extended states would be needed. Therefore, only a tentative model can be developed that is based on the investigation of the much smaller amino acids. Photoemission experiments yielded a spread of values for I_p of about 2 eV and absolute values around 8–9 eV [64]. The energy gap of polypeptides has been experimentally determined to lie around 4 eV [65, 66] and the width of the extended states has been calculated for DNA to be in the order of 2 eV [67]. Even with a rather precise knowledge of I_p, the energetic alignment between the enzyme's LUMO and the semiconductor bands can be altered due to surface dipoles that are created upon adsorption. These dipoles can change the relative energetic alignments

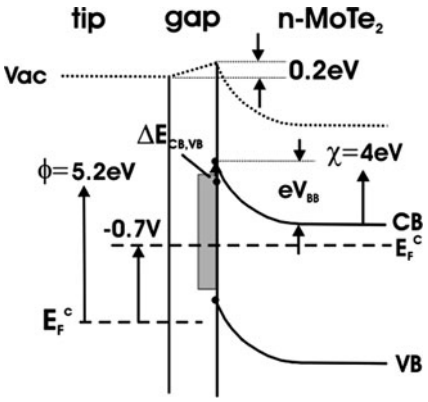


Fig. 5.53 Energetic situation at the STM-tip–tunnel gap $MoTe_2$ junction upon biasing the semiconductor by –0.7 V vs. the grounded tip; note that the negative voltage has been drawn into the figure to show the actual experimental parameter; the according electron energy increases of course by 0.7 eV; also note that the band edge shift is now energetically upwards but that the semiconductor band bending remains unaffected by the external voltage

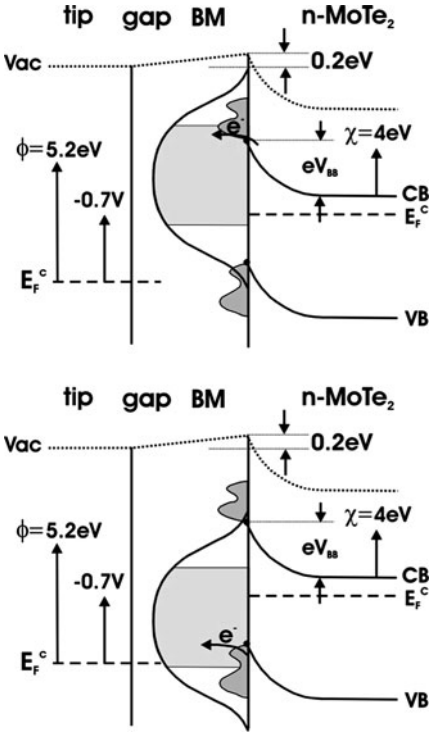


Fig. 5.54 Energy scheme where overlap of the high-density valence and/or conduction band states with the protein extended states would allow charge transport of electrons through the molecule to the tip; *top*: overlap of conduction band with extended LUMO states; *bottom*: overlap of valence band with extended LUMO states; *BM* biomolecule; *dark grey areas*: density of states of $MoTe_2$ at the band edges; *light grey area*: extended biomolecule LUMO states; other notations as in the three figures above

in the range of an eV. Therefore, the following discussion of the energetic origin of the image formation in the STM experiments shown in Figs 5.46 and 5.47 is based on the connection of experimental data regarding the imaging and on the available literature values which gives the following evaluation a hypothetical aspect.

A high density of states at the semiconductor valence band maximum has been determined by UPS and theoretical bandstructure calculations. The latter also show a similarly high density of states at the bottom of the conduction band. It is not known how the surface state density varies with energy. Therefore, Fig. 5.54 shows two energy schemes where both, the valence band and the conduction band, are assumed to overlap with the LUMO extended states of $MoTe_2$, neglecting possible surface state contributions.

The actual process depends on the absolute position of the LUMO with respect to the physics vacuum scale which is unknown. For a lower value, the valence band overlap would result. In addition, one has to make an assumption how the energetic position of the polypeptide changes with applied voltage and how strong it is coupled to the semiconductor. Experiments on the shift of energy states within the Helmholtz layer indicate that the energy shift is about 0.5 of the applied voltage [68, 69].

5.4.2.4 Charge Transport Aspects

The interpretation of the charge transport has to account for basically two general observations: (1) the high resolution of the tertiary structure of the HIV RT in Figs 5.46 and 5.47 and (2) the STM imaging of DNA on mica in Fig. 5.49. First, the geometrical and electronic situation for the reverse transcriptases is redrawn in Fig. 5.55. The difference to the drawing in Fig. 5.40, left image, is the inclusion of a water film that surrounds the protein. The presence of bounded water layers (often referred to as biological water) at the surface of proteins results from the interaction of the water dipoles with hydrophilic and charged parts of the protein [70].

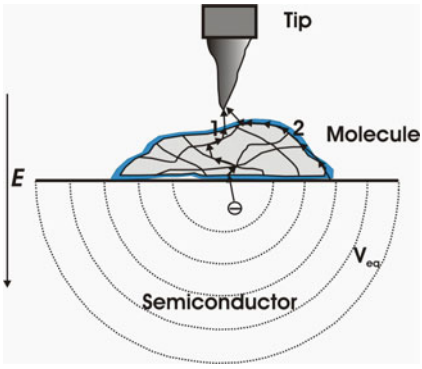


Fig. 5.55 Geometric and electronic situation at the STM-tip/biomolecule/n-$MoTe_2$ contact at a bias of -0.7 V of the semiconductor vs. the tip; E electric field across the junction, V_{eq} equipotential lines of the space charge region in the semiconductor, 1, 2: possible electron pathways with electron transport through or around the molecule (see text)

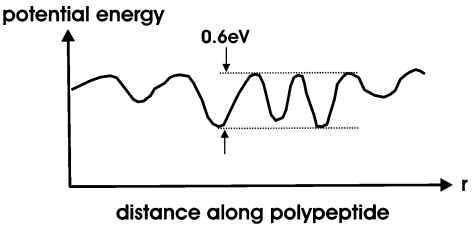


Fig. 5.56 Schematic of the trap distribution with distance r along a polypeptide chain

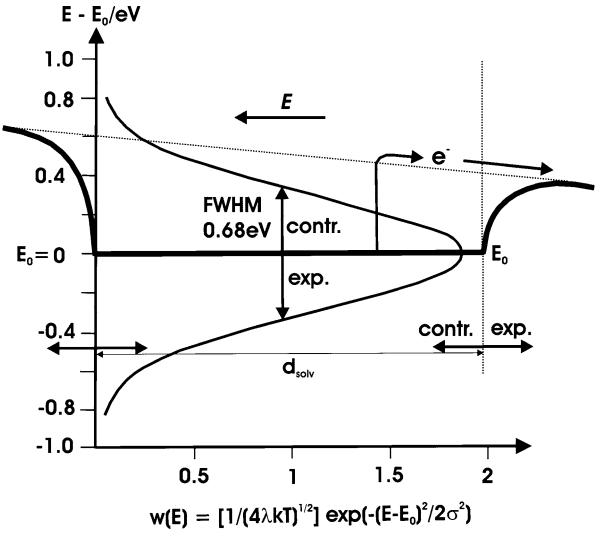


Fig. 5.57 Energy schematic for solvation-assisted trap release of electrons after the Marcus–Gerischer model; E_0: electron trap depth; d_{solv}: spatial extension of the solvation cage in Angström units; contr./exp: contracting and expanding mode, respectively (see text)

Due to the electric field across the junction, electrons motion towards the tip as indicated occurs. The pathway 1 indicates injection into extended LUMO states followed by a subsequent motion that involves hopping-type transport along polypeptide chains and interchain tunnelling. Path 2 includes injection into extended molecule states located near the shell of biological water [70]. The so-called energetic landscape of polypeptides varies by about 0.6 eV [71] (Fig. 5.56).

In the presence of the water shell of biological water which is similarly strongly bound as water in Helmholtz double layers at solid–electrolyte interfaces, an injected electron can become localized by trapping. Immediately, the electron will become solvated in an ultrafast process as indicated in Fig. 5.57. The solvation of electrons in water has been intensively investigated [72] and the results can be transferred to the situation considered here although the solvation shell of the trapped electron on a polypeptide backbone might be only partly solvated due to sterical hindrances. In the process of solvation, a so-called solvation cage is formed. For free electrons in water, the radius of the solvation cage is 0.25 nm and the

trap depth is ∼3 eV [73]. The collective thermal motion of the solvation shell, consisting of oriented water dipoles results in a change of the electron energy that is typically substantially larger than the thermal energy itself. The resulting probability distribution of the electron energies has a Gaussian-type shape in the framework of the Marcus–Gerischer theory [74–76]. The derived expression for the occupation probability $W(E)$ is given by

$$W(E) = \frac{1}{\sqrt{4\pi\lambda kT}} \mathrm{e}^{-\frac{(E_0-E)^2}{4\lambda kT}}, \tag{5.35}$$

where E_0 denotes the equilibrium level of the electron in the solvation cage and λ is the reorganization energy that describes the energy which is needed to cause fluctuations of the polar liquid around the charge from its equilibrium polarization by addition (or, for redox systems also by subtraction) of an electron to the central charge. For an electron in water, $\lambda = 1.6$ eV. Using the Gaussian function

$$f(x) = \mathrm{e}^{\frac{x^2}{2\sigma^2}} \tag{5.36}$$

for the reorganization energy of an electron in water, one obtains for the full width half maximum $E_{1/2} = 2\sigma\sqrt{2\ln 2}$ with $\sigma = \sqrt{2\lambda kT} = 0.29$ eV a value of 0.68 eV. Hence, within a time scale from 50 fs to the low ps range, the electron energy is increased by this amount. The energetic situation is shown in Fig. 5.57. The figure shows the solvation cage by the thicker bold line with an asymmetry that is caused by the electrostatic field across the biomolecule as displayed in the lower part of Fig. 5.54. An average trap depth of 0.6 eV has been assumed that is lowered to about 0.4 eV on the right hand side of the trap by the electric field that has a field strength of the order of $\boldsymbol{E} \sim 5\times10^5$ V cm^{-1}. In the contracting mode of the solvation shell, an electron increase of 0.34 eV results, using the FWHM approximation made above, as indicated in Fig. 5.57. This value is close to that of the trap depth and the probability for escape of the trapped electron depends on the attempt frequency. The Gaussian probability distribution for solvation excitation is drawn in the same direction as the spatial extension of the electron trap site and it can be seen that fast and directed release from the traps becomes possible by the presence of water. Subsequent trapping and detrapping results in a conduction mechanism that has similarities with the Poole–Frenkel transport [77]. Since the model is based on a solvation-assisted carrier transport it explains easily the water conductivity effect observed in the STM experiments on mica (Figs 5.49 and 5.50).

The undistorted imaging of the reverse transcriptases has therefore been attributed to the high conductivity for excess charges in the water shell that surrounds the protein. This conduction mechanism also explains the observation that all relevant features of the tertiary structure are revealed and why the molecules are not disrupted by the STM tip upon approaching (seemingly) less conductive regions. Although this finding in itself is intriguing, it poses serious questions in the

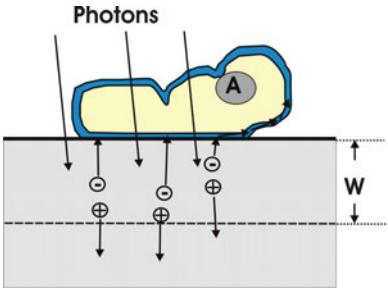


Fig. 5.58 Schematic of charge transport along the biological water shell of a protein adsorbed on a semiconductor where light-induced excess carriers (electrons) are transported to the electrolyte for hydrogen evolution; *A* active site, *W* space charge layer extension (see text)

use of enzymes or mutants of them for artificial photosynthesis: as Fig. 5.58 shows, the electronic charge trajectories would then surround the molecules.

For catalytically active molecules, the active site that is typically embedded into the macromolecular structure will be difficult to reach by, for instance, light-induced transferred excess electrons (compare Fig. 5.58). Therefore, one might have to consider other ways of transport of the photonic excess energy. One alternative, the Förster transfer mechanism, has already been indicated in the beginning of this section (Figs 5.29 and 5.30). Such electron transfer has presently not yet been realized in enzymatic systems. Its implementation demands the preparation of suited nanoscale photon absorber materials that have well-designed exciton lifetimes. In a different approach, one might use extremely thin conductivity material, such as single- or multiwall carbon nanotubes [78] that penetrate into the protein towards the active site and thus allow charge transfer via electron or hole conduction. The involved nanotechnology and the complex photophysics indicate that further development of the research area of artificial photosynthesis will have to be developed from, both, the biological and bio-analogue side and from inorganic chemistry and solid-state expertise. This chapter closes with the development of a protein–semiconductor system that shows catalytic activity upon illumination.

5.4.3 Development of a Photocatalytically Active Bio-cathode

In the research approaches of artificial photosynthesis, the predominantly considered structures resemble those that have been realized by nature: a chromophore (dye) molecule is photo-excited and the light-induced excess charge carrier is injected into a substrate electrode. The remaining carrier, typically a hole in the excited state, induces an oxidation reaction. In the electrochemical photovoltaic mode, this corresponds to the oxidation of iodide to tri-iodide. This is the basic operation of so-called dye sensitization cells [79]. An overall schematic of the functioning is shown in Fig. 5.59.

On a wide-gap semiconductor, such as TiO_2, with an energy gap that is too large to absorb solar photons in sufficient amount for efficient solar energy conversion, a dye is adsorbed that has a considerably lower threshold energy for photoexcitation.

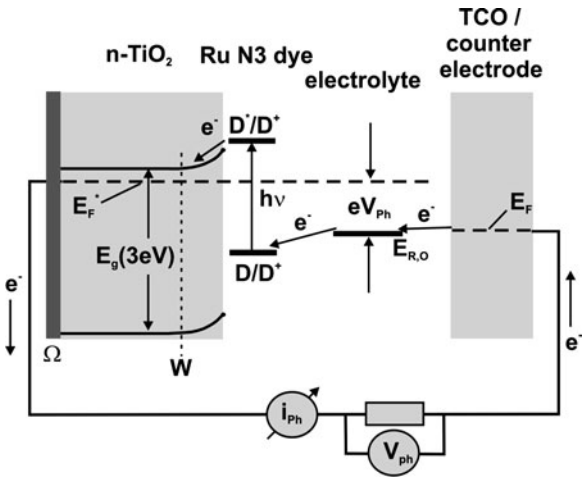


Fig. 5.59 Principle of operation of a dye-sensitized solar cell; D, D*, D* denote the ground state, the oxidized and the excited state of the Ru dye molecule, respectively; E_F, $E_F{}^*$ are the Fermi level of the system in the dark and under illumination, respectively; W space charge layer width, Ω is the ohmic back contact of the semiconductor; $E_{R,O}$ is the energy (potential) of the used redox couple; V_{Ph} denotes the photovoltage obtained under the above condition (see also text)

Upon illumination with visible and near infrared light, an exciton is formed on the excited dye [80]. With the LUMO located energetically above the conduction band of anatase TiO_2, electron injection occurs due to a chemical potential gradient regarding the charges on each side of the dye. The semiconductor, being in a depletion situation with a positively charged space charge layer (width W) as indicated by the downward band bending accepts the photo-induced electron as drawn in Fig. 5.59. The positive charge on the dye molecule is compensated by a reduction reaction with the iodine/iodide electrolyte (in acetonitrile, CH_3CN), and a photocurrent flows in the external circuit if the oxidized redox species becomes reduced at the counter electrode. In Fig. 5.59, neither the so-called short circuit condition ($V = 0$) nor the open circuit situation ($i_{Ph} = 0$) is shown; rather, an intermediate situation where an external load defines the obtained photovoltage and photocurrent.

For a technical realization of the dye sensitization principle, the absorption of one layer of dye molecules has been much too low. Therefore, a sponge-like structure was created by a sol–gel technique that allowed penetration of the dye. With a very high surface to volume ratio, adsorption of the dye at the inner surfaces of the structure and penetration of the electrolytes resulted in efficient absorption of the photons from the sunlight. Solar-to-electrical conversion efficiencies have reached the 10% range [81] but large-scale application is delayed by the search for less poisonous electrolytes and general stability of the system.

The sensitization principle can also be applied for photocatalytic operation if the oxidized (or reduced) dye induces an electrochemical reaction such as water oxidation, for example. Then, at the counter electrode, hydrogen evolution could occur if the energetic conditions for water electrolysis are met [82]. One could, of course, investigate so-called half-cell reactions, i.e. reactions that occur on one of the electrodes for water splitting. This allows detailed analysis of the materials and interface issues and the reaction energetics of the respective photoanode or

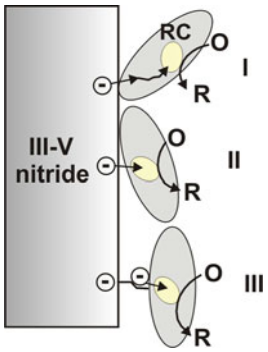


Fig. 5.60 A macromolecule or an enzyme is adsorbed on a semiconductor (here selected from the III–V nitride family); after photoexcitation of the semiconductor electrons are transferred to the reactive site (RC) of the molecule where they induce a reduction reaction (O – R); I: injection by tunnelling and subsequent hopping; II: direct tunnelling to the active site; III: injection via a linker molecule and subsequent tunnelling (see text)

photocathode. This route will be followed below where the properties of a hybrid photocathode are analysed.

A different approach for photocatalytic operation uses photon absorption in the semiconductor substrate with subsequent charge transfer to the reactive site of a catalytically active molecular entity on the semiconductor surface. The concept could be considered as an inversion of the dye sensitization principle and its application becomes possible when considering the above discussed semiconductor–enzyme junctions (Sect. 5.3). The basic idea is visualized in Fig. 5.60.

As absorber, a semiconductor from the III-nitride family can be used. The materials span an energy gap from 6 eV (AlN) to below 1 eV (InN) [83] and thus allows not only the tuning of the photon absorption edge but also that of the conduction band with composition for efficient absorption and charge transfer. Figure 5.60 shows three configurations of the macromolecule or enzyme on the surface which differ by their position of the reactive site with respect to the absorber surface. This is a direct outcome of the studies of the reverse transcriptases with STM (Sect. 5.4.2) where it was found that conduction could occur at the circumference of a protein thereby inhibiting direct transfer to an active site.

Another aspect relates to the energetic coupling of the semiconductor conduction band with the LUMO of the molecule. Its importance has already been discussed in the context of the dye sensitization cell above. The energetic locations of LUMO states of macromolecules and, in particular, of enzymes (proteins) are widely unknown. In addition, interface dipoles can result in changes in the electronic coupling of semiconductor conduction band and macromolecule LUMO that exceeds 1 eV and, the width and location of extended states in the molecules are not defined. For the III-nitrides, the conduction band position changes by about 3 eV with a change in composition from the large gap to the minimum gap material. This can in principle be exploited for efficient charge injection as shown in Fig. 5.61.

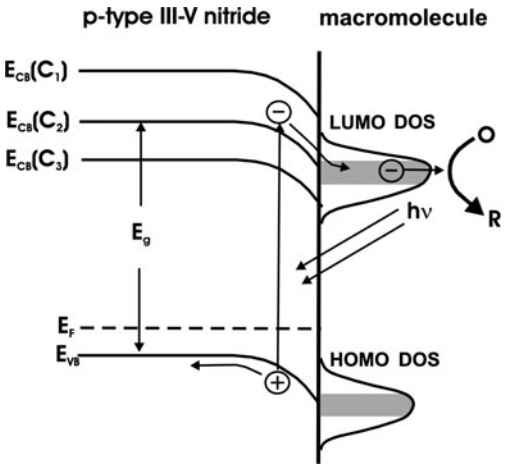


Fig. 5.61 Energy schematic on the injection of photo-generated excess electrons from a group III-nitride compound semiconductor into a catalytically active macromolecule; C_i denote different compositions; *grey-shaded areas* indicate extended states within the macromolecule; *DOS* density of states; the valence band energy position has been left unchanged for simplification of the figure (see text)

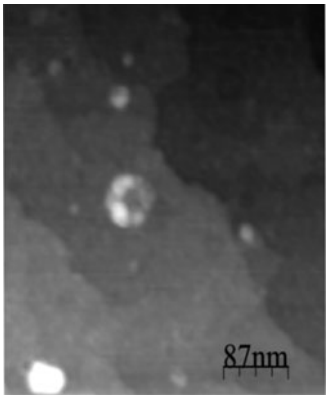


Fig. 5.62 Tapping mode AFM image of the reverse transcriptase of the avian myeloblastosis virus immobilized near and at step edges of step-bunched Si(111)

For a first realization of the concept, an enzyme with considerable stability and well-known structure has to be used in conjunction with a semiconductor that has proven to allow well-defined and site-specific immobilization. Such a system is the laccase/step-bunched Si structure: laccase is an oxygen reduction catalyst and on step bunched Si, enzymes have been immobilized in such a stable manner that they withstood repeated scanning using tapping mode atomic force microscopy (AFM) [84]. Figure 5.62 shows a reverse transcriptase of the avian myeloblastosis virus, imaged on step-bunched Si. The latter is characterized by atomic terraces that are separated by a multiple of the bilayer height of 0.314 nm (see Fig. 4.15). The area around the steps is characterized by a lateral change in the electrostatic potential,

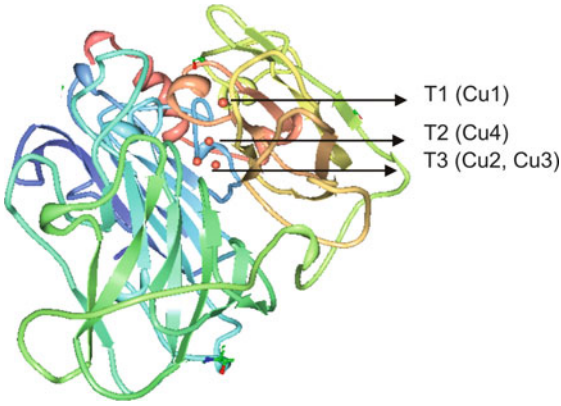


Fig. 5.63 Ribbon representation of the enzyme laccase; *red spheres* indicate the location of the active centres, labelled T1–T3

determined by Kelvin probe AFM [85]. These more negatively charged sites provide adsorption sites due to electrostatic, van der Waals and mechanical interaction.

The enzyme laccase is characterized by three sites that are catalytically active (Fig. 5.63). All centres contain Cu as transition metal. The T1 centre is located towards the exterior of the enzyme and this property increases the chance for a direct charge transfer from the semiconductor substrate into the molecule via a tunnelling process.

Enzyme immobilization has been successful at step edges, similar to the observations in Figs 5.62 and 5.64 shows a photocurrent–voltage characteristic of the system p-Si(111) step-bunched/laccase/electrolyte/Pt counter electrode [86]. At $+0.05$ V vs. the normal hydrogen electrode, a cathodic photocurrent is observed that only exists when the solution is oxygen purged and after deposition of the enzyme-containing carrier solution. The plateau of the current extends to -0.35 V. No effect is seen without illumination, and filtering of the white light used for illumination showed that the effect is not due to the direct photo-excitation of the enzyme. This indicates that the enzyme maintains its catalytic function after adsorption and shows, as seen for the reverse transcriptases, that the immobilization process maintains to a large extent the actual tertiary structure of the protein. The result represents the first bio-photocathode using an enzyme that maintains its activity on a semiconductor surface.

This chapter closes with a recent result where the complete photosystem I has been immobilized on step-bunched Si for the first time [87]. In artificial photosynthesis, efforts are made to largely use the molecular machinery of the photosystems, provided from nature, and to modify them for higher efficiency. This has been done, for instance, by attaching Pt nanoparticles via a linker molecule to photosystem I [88]. For sustained operation and separation of the resulting gases, adsorption on an electrode surface will provide a conducting substrate that supplies the charge carriers from the opposite electrode (which is the anode in the case of photosystem I that reduces protons). Attempts to immobilize photosystem I have

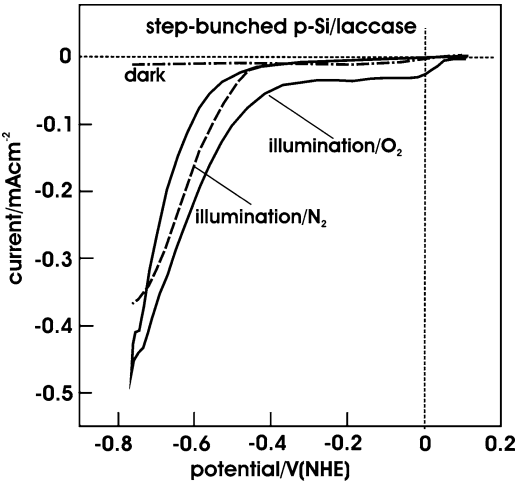


Fig. 5.64 Photo- and dark current–voltage characteristics of the system step-bunched Si(111)/laccase/citric acid with K_2HPO_4 (pH 5.2); illumination: white light W-I lamp, 60 mW cm^{-2} (see text)

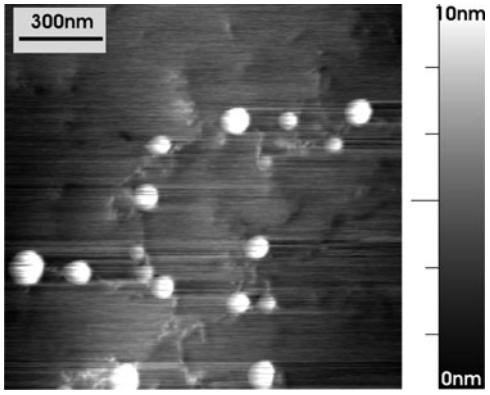


Fig. 5.65 Tapping mode AFM image of photosystem I adsorbed onto step bunched Si; note concentration of the biomolecule at and near the edges of steps

failed so far, and only very recently, shape conserving immobilization could be realized (see Fig. 5.65). Line sections across the molecules show a height between 7 and 8 nm and a width of 100–120 nm. The next step in the development would be the immobilization of the photosystem with attached Pt nanoparticles and the investigation of their photosensitivity with regard to hydrogen generation.

After having reviewed various aspects related to the flow of light in photonically active materials and in photosynthesis in this chapter, a further aspect of photons, e.g. their quantum nature, will be addressed in Chap. 6.

References

1. E. Schrödinger, An undulatory theory of the mechanics of atoms and molecules, Phys. Rev. **28**, 1049–1070 (1926)

2. J.C. Slater, An augmented plane wave method for the periodic potential problem, Phys. Rev. **92**, 603–608 (1953)
3. N.W. Ashcroft, N.D. Mermin, *Solid State Physics* (Holt, Rinehart, and Winston, New York, 1976)
4. T. Vo-Dinh, H.-N. Wang, J. Scaffaldi, Plasmonic nanoprobes for SERS biosensing and bioimaging, J. Biophoton. **3**, 89–102 (2010)
5. S.Y. Lin et al., A three-dimensional photonic crystal operating at infrared wavelengths, Nature **394**, 251–253 (1998)
6. E. Yablonovitch, Inhibited spontaneous emission in solid state physics and electronics, Phys. Rev. Lett. **58**, 2059–2062 (1987)
7. V. Lehmann, The physics of macropore formation in low-doped n-type silicon, J. Electrochem. Soc. **140**, 2836–2843 (1993)
8. U. Grüning, V. Lehmann, Two-dimensional photonic crystal based on macroporous silicon, Thin Solid Films **276**, 151–154 (1996)
9. J.D. Joannopoulos, S.G. Johnson, J.N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, Princeton, NJ, 2008)
10. R.D. Meade, A.M. Rappe, K.D. Brommer, J.D. Joannopoulos, O.L. Alerhand, J. Jin, *The Finite-Element Method in Electromagnetics* (Wiley, New York, 1993)
11. K.M. Ho, C.T. Chan, C.M. Soukoulis, Existence of a photonic gap in periodic dielectric structures, Phys. Rev. Lett. **65**, 3152–3155 (1990)
12. J.S. Foresi et al., Photonic-band gap microcavities in optical waveguides, Nature **390**, 143–145 (1997)
13. J. Kerr, A new relation between electricity and light: dielectric media birefringent, Philos. Mag. **50**, 446–458 (1875)
14. A. Hasegawa, in *Optical Solitons in Fibers*, Springer Tracts in Modern Physics, vol. 116 (Springer, Heidelberg, 1989)
15. N. Aközbek, S. John, Optical solitary waves in two- and three-dimensional nonlinear photonic band-gap structures, Phys. Rev. E **57**, 2287–2319 (1998)
16. H.S. Sözüer, J.W. Haus, R. Inguva, Photonic bands – convergence problems with the plane wave method, Phys. Rev. B **45**, 13962–13972 (1992)
17. J.E.G.J. Wijnhoven, W.L. Vos, Preparation of photonic crystals made of air spheres in titania, Science **281**, 802–804 (1998)
18. J.B. Pendry, A.J. Holden, W.J. Stewart, I. Youngs, Extremely low frequency plasmons in metallic mesostructures, Phys. Rev. Lett. **76**, 4773–4776 (1996)
19. S.J. Smith, E.M. Purcell, Visible light from localized surface charges moving across a grating, Phys. Rev. **92**, 1069(1953)
20. X. Artru, G.B. Yodh, G. Mennessier, Practical theory of the multilayered transition radiation detector, Phys. Rev. D **12**, 1289–1306 (1975)
21. F. J. Garcia de Abajo, A.G. Pattantyus-Abraham, N. Zabala, A. Rivacoba, M.O. Wolf, P.M. Echenique, Cherenkov effect as a probe of photonic nanostructures, Phys. Rev. Lett. **91**, 143902(2003)
22. D. Ugarte, C. Colliex, P. Trebbia, Surface- and interface-plasmon modes on small semiconducting spheres, Phys. Rev. B **45**, 4332–4334 (1992)
23. A.G. Koutsioubas, N. Spiliopoulos, D. Anastassopoulos, A.A. Vradis, G.D. Priftis, Nanoporous alumina enhanced surface plasmon resonance sensors, J. Appl. Phys. **103**, 1–6 (2008)
24. A. Ben-Shem, F. Frolow, N. Nelson, Crystal structure of plant photosystem I, Nature **426**, 630–635 (2003)
25. M. Rögner, E.J. Boekema, J. Barber, How does photosystem 2 split water? The structural basis of efficient energy conversion, Trends Biochem. Sci. **21**, 44–49 (1996)
26. J. Standfuss, A.C. Terwisscha van Scheltinga1, M. Lamborghini, W. Kühlbrandt, Mechanisms of photoprotection and nonphotochemical quenching in pea light harvesting complex at 2.5A resolution, EMBO J. **24**, 919–928 (2005)

27. B. Kok, B. Forbush, M. McGloin, Cooperation of charges in photosynthetic O_2 evolution. 1. A linear 4-step mechanism, Photochem. Photobiol. **11**, 457–475 (1970)
28. T.M. Bricker, A time-resolved vibrational spectroscopy glimpse into the oxygen-evolving complex of photosynthesis, PNAS **103**, 7205–7206 (2006)
29. I.J. Hewitt, J.-K. Tang, N.T. Madhu, R. Clérac, G. Buth, C.E. Ansona, A.K. Powell, A series of new structural models for the OEC in photosystem II, Chem. Commun. **25**, 2650–2652 (2006)
30. W. Lubitz, E.J. Reijerse, J. Messinger, Solar water-splitting into H2 and O2: design principles of photosystem II and hydrogenases, Energy Environ. Sci. **1**, 15–31 (2008)
31. R. Hill, F. Bendall, Function of the two cytochrome components in chloroplasts: a working hypothesis, Nature **186**, 136–137 (1960)
32. M. Calvin, The photosynthetic cycle, Bull. Soc. Chim. Biol. **38**, 1233–1244 (1956)
33. X. Hu, A. Damjanovic, T. Ritz, K. Schulten, Architecture and mechanism of the light-harvesting apparatus of purple bacteria, Proc. Natl. Acad. Sci. USA **95**, 5935–5941 (1998)
34. Th.v. Förster, Intermolecular energy migration and fluorescence, Ann. d. Phys. **2**, 55–75 (1948)
35. G.D. Scholes, Long range resonance energy transfer in molecular systems, Annu. Rev. Phys. Chem. **54**, 57–87 (2003)
36. B. Schuler, E.A. Lipman, P.J. Steinbach, M. Kumke, W.A. Eaton, Polyproline and the "spectroscopic ruler" revisited with single molecule fluorescence, Proc. Natl. Acad. Sci. USA **102**, 2754–2759 (2005)
37. X. Hu, T. Ritz, A. Damjanovic, F. Autenrieth, K. Schulten, Photosynthetic apparatus of purple bacteria, Q. Rev. Biophys. **35**, 1–62 (2002)
38. J.R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Springer, New York, 2006)
39. P.A.M. Dirac, *The Principles of Quantum Mechanics* (Oxford, Clarendon, 1958), p. 180
40. L. Stryer, Fluorescence energy transfer as a spectroscopic ruler, Annu. Rev. Biochem. **47**, 819–846 (1978)
41. J. Zheng, FRET and its application as a molecular ruler, in *Handbook of Modern Biophysics, Biomedical Applications of Biophysics*, vol. 3 (Humana Press, Riverside, New Jersey, 2010), pp. 119–136
42. S. Jang, M.D. Newton, R.J. Silbey, Multichromophoric Förster resonance energy transfer, Phys. Rev. Lett. **92**, 1–4 (2004)
43. G.R. Fleming, G.D. Scholes, Quantum mechanics for plants, Nature **431**, 256–257 (2004)
44. G. Jutz, A. Böker, Bionanoparticles as functional macromolecular building blocks – A new class of nanomaterials, Polymer **52**, 211–232 (2011)
45. M. Grätzel, Dye-sensitized solar cells, J. Photochem. Photobiol. C: Photochem. Rev. **4**, 145–153 (2003)
46. N. Robertson, Optimizing dyes for dye-sensitized solar cells, Angew. Chem. Int. Ed. **45**, 2338–2345 (2006)
47. D.L. Dexter, A theory of sensitized luminescence in solids, J. Chem. Phys. **21**, 836–851 (1953)
48. B.P. Paulson, J.R. Miller, W.-X. Gan, G. Closs, Superexchange and sequential mechanisms in charge transfer with a mediating state between the donor and acceptor, J. Am. Chem. Soc. **127**, 4860–4868 (2005)
49. H. Lederer, O. Schatz, R. May, H. Crespi, J.-L. Darlix, S. F.J. LeGrice, H. Heumann, Domain structure of the human immunodeficiency virus reverse transcriptase, EMBO J. **11**, 1131–1139 (1992)
50. M. Tarek, G.J. Martyna, D.J. Tobias, Amplitudes and frequencies of protein dynamics: analysis of discrepancies between neutron scattering and molecular dynamic simulations, J. Am. Chem. Soc. **122**, 10450–10451 (2000)
51. D. Gust, T.A. Moore, A.L. Moore, Mimicking photosynthetic solar energy transduction, Acc. Chem. Res. **34**, 40–48 (2001)
52. H. Jungblut, S.A. Campbell, M. Giersig, D.J. Müller, H.J. Lewerenz, STM observations of biomolecules on layered materials, Farad. Disc. **94**, 183–198 (1992)
53. H.J. Lewerenz, Enzyme-semiconductor interactions: routes from fundamental aspects to photoactive devices, Phys. Stat. Sol (b) **245**, 1884–1898 (2008)

54. H.J. Lewerenz, H. Jungblut, S.A. Campbell, D.J. Müller, Direct observation of reverse transcriptases by STM, AIDS Res. Hum. Retroviruses **8**, 1663–1667 (1992)
55. B.V. Derjarguin, L. Landau, Acta Physicochimica (URSS) **14**, 633 (1941)
56. E.J. Verwey, J.T.G. Overbeek, "*Theory of the Stability of Lyophobic Colloids*" (Elsevier, Amsterdam, 1948)
57. H.J. Lewerenz, Surface states and Fermi level pinning at semiconductor/electrolyte junctions, J. Electroanal. Chem. **356**, 121–143 (1993)
58. S.A. Campbell, J.R. Smith, H. Jungblut, H.J. Lewerenz, Protein imaging on a semiconducting substrate: a scanning tunnelling microscopy investigation, J. Electroanal. Chem. **599**, 313–322 (2007)
59. H.J. Lewerenz, K. Skorupska, J.R. Smith, S.A. Campbell, Surface chemistry and electronics of semiconductor-nanosystem junctions II: enzyme immobilization, charge transport aspects and scanning probe microscopy imaging, J. Sol. State Electrochem. **13** 195–203 (2009)
60. R. Guckenberger, M. Heim, G. Cevec, H.F. Knapp, W. Wiegräbe, A. Hillebrand, Scanning tunnelling microscopy of insulators and biological specimens based on lateral conductivity of ultrathin water films, Science **266**, 1538–1540 (1994)
61. R. Guckenberger, M. Heim, STM on wet insulators: Electrochemistry or tunnelling? Response to a technical comment, Science **270**, 1851–1852 (1995)
62. S.M. Sze, *Semiconductor Devices* (Wiley, New York, 1980)
63. K. Maturova, R.A. Janssen, M. Kemerink, Connecting scanning tunneling spectroscopy to device performance for polymer:fullerene organic solar cells, ACS Nano **4**, 1385–1392 (2010)
64. H.-W. Jochims, M. Schwell, J.-L. Chotin, M. Clemino, F. Dulieu, H. Baumgärtel, S. Leach, Photoion mass spectrometry of five amino acids in the 6–22 eV photon energy range, Chem. Phys. **298**, 279–297 (2004)
65. R.N. Jones, H.J. Greech, The ultraviolet absorption spectra of protein solutions, JOSA **33**, 209–217 (1943)
66. P. Facci, D. Alliata, L. Andolfi, B. Schnyder, R. Kötz, Formation and characterization of protein monolyers on oxygen-exposing surfaces by multiple-step self-chemisorption, Surf. Sci. **504**, 282–292 (2002)
67. T. Koslowski, Localized and extended electronic eigenstates in proteins, J. Chem. Phys. **110**, 12233–12239 (1999)
68. D.M. Kolb, W. Boeck, K.-M. Ho, S.H. Liu, Observation of surface states on Ag(1 0 0) by infrared and visible electroreflectance spectroscopy, Phys. Rev. Lett. **47**, 1921–1924 (1981)
69. H. Jungblut, J. Jakubowicz, S. Schweizer, H. J. Lewerenz, Mechanism of initial structure formation on highly doped n-Si(111), J. Electroanal. Chem. **527**, 41–46 (2002)
70. S.K. Pal, J. Peon, A.H. Zewail, Biological water at the protein surface: dynamical salvation probed directly with femtosecond resolution, Proc. Natl. Acad. Sci. USA **99**, 1763–1768 (2002)
71. E.W. Schlag, S.-Y. Sheu, D.-Y. Yang, H.L. Selzle, S.H. Lin, Charge conductivity in peptides; dynamic simulations of a bifunctional model supporting experimental data, Proc. Natl. Acad. Sci. USA **97**, 1068–1072 (2000)
72. X. Shi, F.H. Long, H. Lu, K.B. Eisenthal, Femtosecond electron solvation kinetics in water, J. Phys. Chem. **100**, 11903–11906 (1996)
73. P. Kambhampati, D.H. Song, T.W. Kee, P.F. Barbara, Solvation dynamics of the hydrated electron depends on its initial degree of electron delocalization, J. Phys. Chem. A **106**, 2374–2378 (2002)
74. R.A. Marcus, On the theory of electron-transfer reactions VI. Unified treatment for homogeneous and electrode reactions, J. Chem. Phys. **43**, 679–702 (1965)
75. H. Gerischer, Über den Ablauf von Redoxreaktionen an Metallen und an Halbleitern. I Allgemeines zum Elektronenübergang zwischen einem Festkörper und einem Redoxelektrolyten, Z. Phys. Chem. N. F. **26**, 223–247 (1960)
76. H. Gerischer, Über den Ablauf von Redoxreaktionen an Metallen und an Halbleitern. III. Halbleiterelektroden, Z. Phys. Chem. N. F. **27**, 48–79 (1961)
77. M. Zielinski, M. Samoc, An investigation of the Poole-Frenkel effect by thermally stimulated current technique, J. Phys. D: Appl. Phys. **10**, L105–L107 (1977)

78. T.W. Ebbesen, H.J. Lezec, H. Hiura, J.W. Bennett, H.F. Ghaemi, T. Thio, Electrical conductivity of individual carbon nanotubes, Nature **382**, 54–56 (1996)
79. B. O'Reagan, M. Grätzel, A low-cost, high efficiency solar cell based on dye-sensitized colloidal TiO_2 films, Nature **353**, 737–740 (1991)
80. B.A. Gregg, Excitonic solar cells, J. Phys. Chem. B **107**, 4688–4698 (2003)
81. J.M. Kron et al., Nanocrystalline dye-sensitized solar cells having maximum performance, Prog. Photovolt: Res. Appl. **15**, 1–18 (2007)
82. S. Licht, B. Wang, S. Mukerji, T. Soga, M. Umeno, H. Tributsch, Efficient solar water splitting, exemplified by RuO_2-catalyzed AlGaAs/Si photoelectrolysis, J. Phys. Chem. B **104**, 8920–8924 (2000)
83. I. Vurgaftman, J.R. Meyer, Band parameters for nitrogen-containing semiconductors, J. Appl. Phys. **94**, 3675–3696 (2003)
84. K. Skorupska, T. Vo-Dinh, H.J. Lewerenz, Scanning probe characterization of enzymes deposited onto step-bunched silicon nanostructures, Phys. Scripta **79**, 1–4 (2009)
85. K. Skorupska, Ch. Pettenkofer, S. Sadewasser, F. Streicher, W. Haiss, H.J. Lewerenz, Electronic and morphological properties of the electrochemically prepared step bunched silicon (111) surface, Phys. Stat. Sol. B **248**, 361–369 (2011)
86. K. Skorupska, H.J. Lewerenz, P. Ugarte-Berzal, A. Rutkowska, P.J. Kulesza, unpublished results
87. K. Skorupska, P. Ugarte-Berzal, M. Lunlow, H.J. Lewerenz, unpublished results
88. R.A. Grimme, C.E. Lubner, D.A. Bryant, J.H. Golbeck, Photosystem I/molecular wire/metal nanoparticle bioconjugates for the photocatalytic production of H_2, J. Am. Chem. Soc. **130**, 6308–6309 (2008)

Chapter 6
Exploiting the Quantum Nature of Photons

6.1 Introduction

A topical overview on the sections of this chapter is given in Table 6.1 As can be seen, very fundamental physical concepts and modern developments are entangled as is the mutual interaction between advances in basic research and the sometimes surprisingly fast realization of devices, spurred by feasibility and market potential. The chapter begins with a review of the seminal paper by Einstein et al. [1] and with the transformation of the concept discrepancy between EPR and quantum physics into measurable phenomena by Bell's inequality [2]. The use of the non-local character of quantum phenomena in teleportation of quantum states and in quantum cryptography is then presented. The remainder of this chapter is devoted to quantum computing and initial successes in its realization.

6.2 Einstein–Podolski–Rosen Concept and Bell's Inequality

6.2.1 Introduction: Entanglement and Correlations

Quantum computing directly uses quantum physical phenomena such as entanglement of quantum mechanical states and their superposition. Early work in that field goes back to the seminal article of Einstein, Podolski and Rosen (EPR) [1]. It has become popular to physicists to use the analogy of Bertlmann's socks [2] for a daily-life interpretation of EPR correlations: Dr. Bertlmann likes to wear two socks of different colours; if one sees that the first sock is pink, then the second sock will not be pink. Hence, detection of the first and the knowledge about Bertlmann gives immediate information about the second sock (Fig. 6.1).

In the EPR Gedanken experiment, uncharged particles of spin $+1/2$ and $-1/2$ were moving in opposite directions in a Stern–Gerlach magnet where a non-uniform

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_6,

Table 6.1 Overview of the content of Chap. 6

Fig. 6.1 Les chaussettes de Msr. Bertlmann (see text)

magnetic field acts on the particles. In "classical" expectations, because of the angular variations of particle axes with respect to the field, a range of locii where the deflected particles hit a screen should be observed. Instead, two separated groups are found (see Fig. 6.2). That means that by observing one particle, the result on the other, even very remote particle, is immediately predictable. Is there a possibility that the measurement of the first particle influences the outcome on the other remote particle? This thought was rejected by EPR as "spooky action at a distance" or "spukhafte Fernwirkung". If one avoids such action at a distance, it is necessary to postulate that the system possesses real properties that are fixed before observation. Because quantum physics formalism does not contain such properties in advance of observation, EPR termed that theory incomplete and the so-called local hidden variables had to be introduced to restore causality and locality [3].

In this postulate of -no action at a distance-, the principle of *local causality* (often termed *locality*) is maintained by EPR. There have been initially various

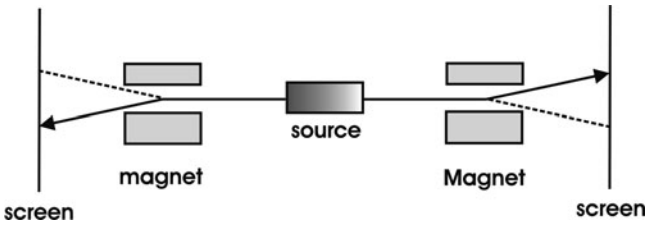


Fig. 6.2 Stern–Gerlach experiment with spin $\pm 1/2$ particles

misunderstandings of the statements made by EPR: the difficulty for them was not that events in different places could not be correlated; the problem was that it could not be accepted that an intervention at one place could immediately affect an event at a (distant) other place. In the considered ideal case, the anti-correlation is perfect in the sense that measurement of spin $+1/2$ results in a remote detection of spin $-1/2$ of the other particle; in realistic experiments, however, source and detector properties will result in a reduced correlation as will be discussed below. Summarizing this early development, quantum theory was left in a rather unsatisfactory situation by EPR: either the theory was incomplete by not accounting for elements of physical reality or it violated the principle of finite propagation speed.

It took 29 years until John S. Bell published his seminal article in which he derived an inequality that represents an upper limit of the strength of correlations of any theory based on locality [4]. Consider now two photons with orthogonalized single particle states in the superposition entangled Bell-state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|+1\rangle|-1\rangle - |-1\rangle|+1\rangle). \tag{6.1}$$

Here, Local hidden variable theory (LHVT) Dirac's bracket notation of quantum states is used again. Entanglement means that $|\psi\rangle$ cannot be written as a product such as $|\psi\rangle = |\varphi\rangle|\zeta\rangle$. The entangled quantum state cannot be factorized. Schrödinger communicated first the unsettling fact that in quantum mechanics, a system can be in an "in-between" mixed state [5–7] and, in order to demonstrate this nature of quantum mechanics, applied it to a macroscopic system, known as the "Schrödinger cat". Here, a cat and an excited atom are both in a box. If the atom decays, it sets of some kind of trigger (phantasy has been rather unlimited here) that kills the cat. The state of the cat and the atom are entangled and the cat in the box exists in a *dead and alive* situation in stark contrast to macroscopic reality. Once Dr. Schrödinger opens the box, the wave function collapses and the cat is found either dead or alive.

Bell investigated which restrictions would occur for the correlation of results from measurements as shown in Fig. 6.3 if one assumes the validity of a local hidden variable theory. We therefore consider a photon pair, generated, for instance, by a non-linear optical process. The polarization of each photon shall be measured by

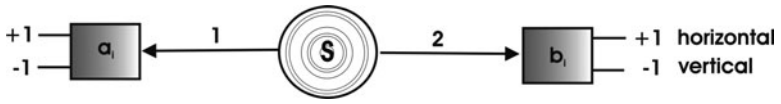


Fig. 6.3 Gedanken experiment for the derivation of Bell's inequality: the source (S) emits correlated photons. Observers A, B measure the polarization of their respective photon. As result of arbitrary switching the polarizers between $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{b}_1$, $\mathbf{b}_2$ either horizontally (coded –1) or vertically (+1) polarized photons are detected

spatially separated observers A, B and, in addition, it is assumed that the observers incidentally switch their polarizers between two orthogonal directions labelled $\mathbf{a}_1$,$\mathbf{a}_2$ and $\mathbf{b}_1$, $\mathbf{b}_2$ (see figure).

Only two results are possible for each polarizer setting: horizontal (coded +1) and perpendicular (–1) polarization. For *local* theories, the following assumptions hold: (1) it can be predicted any time how the polarization vector of the two photons is oriented; (2) results of observers A, B only depend on the local hidden variable λ and each own apparatus and (3) results of one observer are independent of the choice of the other observer on the polarization. Under these suppositions, for an arbitrary state of the photon pair, the following results can occur. A: $a_1 = \pm 1$ or $a_2 = \pm 1$; B: $b_1 = \pm 1$ or $b_2 = \pm 1$. In order to obtain an experimentally measurable value, the expression $|(a_1 + a_2)b_1 + (a_2 - a_1)b_2|$ is evaluated. For all possible results, one obtains

$$|(a_1 + a_2)b_1 + (a_2 - a_1)b_2| = 2. \tag{6.2}$$

For frequent repetitions of the experiment, the hidden parameter λ can assume different values that are weighed by a normalized probability distribution $P(\lambda)$. Averaging over a statistical ensemble of experiments yields for each local hidden variable theory (LHVT):

$$\int_{\Lambda} \mathrm{d}\lambda P(\lambda)|(a_1 + a_2)b_1 + (a_2 - a_1)b_2| = 2 \tag{6.3}$$

and generally

$$\int_{\Lambda} \mathrm{d}\lambda P(\lambda)|(a_1 + a_2)b_1 + (a_2 - a_1)b_2| \leq 2. \tag{6.4}$$

This expression (6.4), Bell's inequality, defines an upper limit for possible correlations of measurement results of spatially separated observers within the framework of LHV theories. A few years later, Clauser, Horne, Shimony and Holt published a variation of Bell's inequality that contains experimentally accessible observables [8].

What are then the predictions of quantum mechanics for the statistical mean values of measurements on entangled states? From (6.1) one obtains, using the eigenstates of the spin operator $|\pm 1\rangle_{\mathrm{A}}, |\pm 1\rangle_{\mathrm{B}}$ with eigenvalues ± 1 which again code the horizontal and vertical polarization direction

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|+1\rangle_A|-1\rangle_B - |-1\rangle_A|+1\rangle_B). \tag{6.5}$$

$|\psi\rangle$ is often referred to as the two particle "Bell" state. It can be shown that the correlations for that state, predicted by quantum mechanics violate the Bell and Clauser–Horn–Shimony–Holt inequality yielding a value of $2\sqrt{2}$ which is by a factor of $\sqrt{2}$ larger than the upper limit for LHV theories. The quantum mechanical correlations are thus stronger than those of any local theory. This result allows to determine experimentally whether it is possible at all to replace quantum mechanics by a LHVT. It is also important that the difference between the upper limit of 2 for local theories and $2\sqrt{2}$ for quantum mechanics provides an experimental "buffer" to account for the (already mentioned above) reduced expected correlations due to realistic experimental conditions. Series of experiments employing photons, atoms and so-called cluster states [9] have been performed that show violation of the Bell-type inequalities [10] representing accumulative evidence for the non-locality in quantum physics. *The* conclusive experiment (and only one is necessary), however, has not yet been performed; in the case of photons, for instance, the problem is the detection probability of correlated photons that is far below 100% and theories based on hidden variables cannot be excluded with absolute certainty.

6.2.2 Quantum Teleportation

The use of entangled quantum physical states for what has become known as *quantum teleportation* [11] and *quantum cryptography* [12] is described here as a further introduction to quantum computing. Teleportation is the wording used by science fiction writers to describe disintegration of an object at one place and its (or a replica's) appearance at a distant place. Here, we consider the destruction of the original and the generation of a distant perfect replica.

Suppose, observer A (the sender) has a particle, with subscript 1, in a given quantum state $|\psi\rangle$ and that it is intended that B, at a distant location, should have this particle in that quantum state. One possibility would be to send classical information to B, so that B is able to reconstruct the state precisely. Since the specification of the polarization state of a photon requires specifying a direction in space, where the value of an angle can vary continuously, A would have to convey a principally unlimited amount of classical information to B, for B being able to reconstruct $|\psi\rangle$ precisely. For classical information transfer, the binary expansion, for instance, of an angle variable represented by a real number is represented by a potentially infinite sequence of 0s and 1s. Then, if a different information transfer path is sought, it is known that no measurement that A can perform on $|\psi\rangle$ will be sufficient to allow B its reconstruction because of the evasiveness of quantum systems that can be in a superposition of several states at the same time. A measurement will force the system – according to the so-called projection postulate – into only one of

these states. It will be seen that it is actually this projection postulate that enables teleportation of $|\psi\rangle$ from A to B. The key for teleportation to occur is the use of an entangled ancillary (photon) pair of photons 2 and 3 that is initially shared between A and B. This entangled pair, written in (6.6) according to (6.5) as

$$|\psi\rangle_{23} = \frac{1}{\sqrt{2}}(|+1\rangle_2|-1\rangle_3 - |-1\rangle_2|+1\rangle_3), \tag{6.6}$$

does not contain information on individual particles; it indicates, however, that the two photons will be in opposite states. The crucial feature of an entangled pair is, that as soon as a measurement is made on one of the photons that projects it, for instance, onto $|+1\rangle$ (horizontal polarization, see Fig. 6.3), the state of the other is determined to be $|-1\rangle$ and vice versa. The teleportation scheme is usually set up as follows: A has particle 1 in its initial state $|\psi\rangle_1$ and particle 2 from the ancillary entangled pair; B has particle 3 from the entangled pair. A then performs a so-called Bell-state measurementon 1 and 2 that projects them onto the entangled states [12]

$$|\psi^-\rangle_{12} = \frac{1}{\sqrt{2}}(|+1\rangle_1|-1\rangle_2 - |-1\rangle_1|+1\rangle_2), \tag{6.7}$$

where the characteristic anti-symmetric state was chosen from the four projection possibilities. A Bell-state measurement is the projection of an arbitrary state onto the basis of the four states of a two-particle system. Quantum mechanics predicts that if particles 1 and 2 are projected into $|\psi^-\rangle_{12}$, particle 3 is immediately projected into $|\psi\rangle_1$, the initial state of particle 1: since particles 1 and 2 are now entangled, in whatever state 1 is, 2 must be in the opposite state, i.e. orthogonal to $|\psi\rangle_1$. Initially, particles 2 and 3 were prepared in the state $|\psi\rangle_{23}$ (6.6), meaning that particles 2 and 3 are orthogonal to each other. This is only possible if particle 3 is in the same state as particle 1 was in, initially. The state of particle 3 is then (a and b complex)

$$|\psi\rangle_3 = a|+1\rangle + b|-1\rangle. \tag{6.8}$$

Figure 6.4 shows a schematic of quantum teleportation.

During the Bell-state measurement, $|\psi\rangle_1$ becomes entangled with particle 2 and the original state $|\psi\rangle_1$ is destroyed on A's side of the teleportation process. This information transfer can, in principle, take place over arbitrary distances. Experimentally, quantum entanglement has been demonstrated rather early to occur over distances in the order of 10 km [13]. Meanwhile, the "transfer velocity" of an experiment with entangled states has resulted in a lower limit of a factor of 10^4 of the velocity of light [14].

It should be realized that a complete Bell-state measurement gives equal probabilities of 25% for the other (than the anti-symmetric) Bell states. Then, particle 3 assumes one of three different states. Classical communication on which of the Bell states results were obtained by A then allows B to bring particle 3 into the original state of particle 1 by an appropriate (unitary) transformation. Also, quantum teleportation follows the verdict of the *no-cloning theorem* [15]. This theorem has

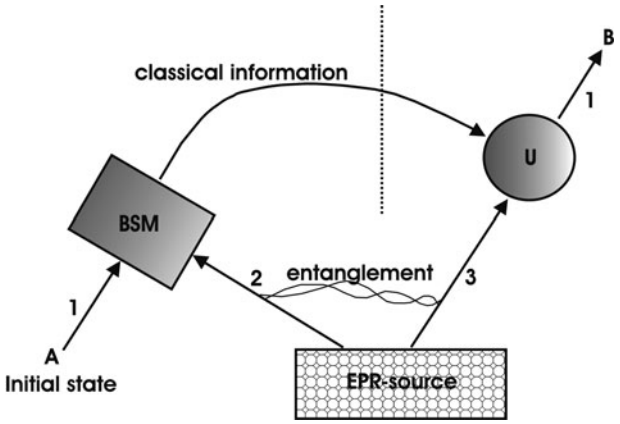


Fig. 6.4 Principles involved in quantum teleportation; quantum state 1 of observer A is sought to be teleported to B. An ancillary entangled particle pair from an Einstein–Podolsky–Rosen source, such as a non-linear optical element, is shared between A and B. A performs a joint Bell-state measurement (BSM) on particle 1 and on one of the entangled particles (here particle 2) thus projecting them onto an entangled state. Then, via a classical information channel, that result is sent to B. A unitary transformation U (for instance rotation of polarizers) on the other ancillary particle (3) transforms it into the original state of particle 1

the profound implications in quantum cryptography and quantum computing, states that an arbitrary unknown quantum state cannot be identically copied. Hence, after successful teleportation, particle 1 is not available in its original state anymore and therefore particle 3 is not a clone but the result of teleportation. It can be shown that the *no-cloning theorem* has very fundamental implications: it (1) correlates with the uncertainty principle of quantum physics and (2) it prevents superluminal communication via quantum entanglement.

Experimental evidence for successful quantum teleportation has accumulated in recent years. Production as well as detection of entangled states is necessary. Entangled photons can be produced by parametric down conversion, where, in a non-linear optical material, an incoming photon decays spontaneously into two photons that are in the state described by (6.6). Projection into a Bell state necessitates that the photons are indistinguishable. This is achieved by superposing at a beam splitter (Fig. 6.5). It can be shown that projection of the photons (1 and 2) onto the anti-symmetric $|\psi^-\rangle_{12}$ state allows to register coincidence detections when detectors are placed in each of the beam splitter outputs. To suppress photon distinction via their time of arrival, generation uses a pulsed (≤ 200 fs) pump beam and transmission to narrow bandpass filters that results in a photon coherence time much longer than the duration of the pulse.

More recently, the experimental realization of the teleportation concept has been extended to quantum state teleportation using, for instance, three calcium atoms in a linear Paul trap [16]. The measured fidelity was in excess of 75%, demonstrating the quantum nature of the process since classical maximal fidelity is 66.7%; to rule

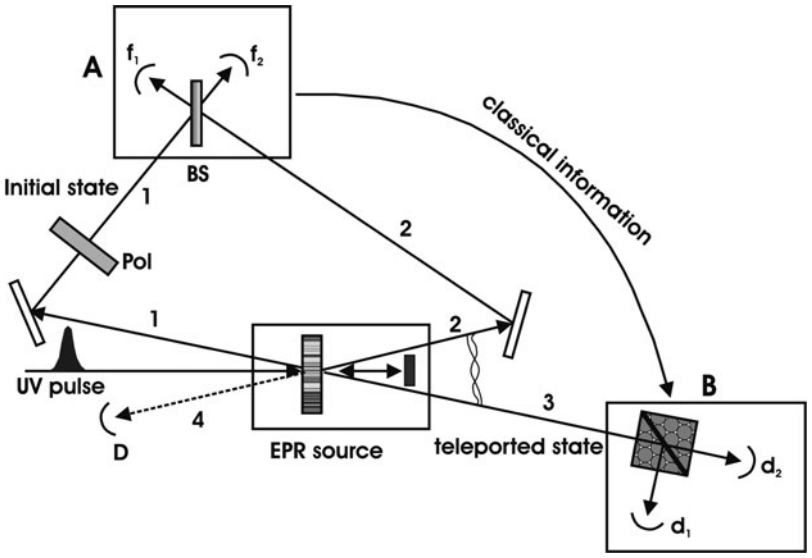


Fig. 6.5 Experimental realization of quantum teleportation:a UV pulse, passing through a nonlinear optical element, generates an ancillary entangled photon pair (photons 2 and 3). Retroflection results in a second passage of the pulse through the element, generating photons 1 and 4; 1 is to be teleported, 4 serves as indicator (detector D) that a photon to be teleported has been created. At A (*above left*), coincidences are counted after the beam splitter BS where photons 1 and one of the ancillaries are superposed. B obtains classical information on a coincidence count of A in detectors f1, f2 that identifies the anti-symmetric $|\psi^-\rangle_{12}$ Bell state and knows that photon 3 is in the initial state of photon 1 which can be tested by polarization analysis using a polarizing beam splitter and detectors d1, d2

out hidden variable theories, however, a fidelity larger than 87% is required ([17], and references therein). Also, teleportation of an atomic ensemble quantum state has been reported [18].

6.2.3 Quantum Cryptography

In quantum cryptography, the use of public and secret keysis distinguished. Public keys are typically based on one-way functions that can be easily calculated in the direction of encoding, but are extremely difficult to determine in the opposite direction of decoding [19]. An example is the factorization of prime numbers. Generally, the security is based on mathematical complexity; hence, an eavesdropper would need (too) much time for the deciphering of a message. As an introduction, consider the early approach that uses a secret key, known as "one-time pad" [20]. Here, the sender A adds to a binary code a modulo 2 key as shown in Table 6.2 If that key consists of an arbitrary sequence of 0's and 1's, has the same length

Table 6.2 "One-time pad" consisting of a message, a secret code (modulo 2), an encoded publicly transmitted text that is decoded using the original secret key into the original message; the modulo 2 key has the mathematical operations: 0 +0 = 1 + 1 = 0; 0 + 1 = 1 + 0 = 0

Sender (A)								
Message	0	1	1	0	1	0	0	1
Key	1	0	0	1	1	0	1	0
Sum (modulo 2) encoded text	**1**	**1**	**1**	**1**	**0**	**0**	**1**	**1**
Data transfer	1	1	1	1	0	0	1	1
Receiver (B)								
Encoded text	1	1	1	1	0	0	1	1
Key	1	0	0	1	1	0	1	0
Sum (modulo 2) decoded text	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**1**
Original message	0	1	1	0	1	0	0	1

as the message and is only used once, then the resulting number sequence can be transmitted publicly because its information content is null.

The problem of using a secret key is hence reduced to the secure distribution of that key. The important aspect of quantum information transfer is, as was shown above, that a measurement alters an *unknown* quantum state. Hence, if bits have been altered during transmission, one knows that the message has been modified by a third person. Although, in quantum cryptography, two classes of coding principles, i.e. based on one- or two-particle systems exist, we choose to present here the first one where transfer based on single photons is considered because of better intelligibility; also, it describes the historical protocol, known as BB 84 [21]. For non-local correlations of entangled particles that can be used to prepare correlated sequences of bits, the interested reader is referred to the work of Ekert [22].

In the BB 84 protocol, A uses polarization filters to send binary information via linearly polarized photons to B in two different incommensurable alphabets: alphabet I codes for horizontal/vertical polarization. Horizontally polarized photons and those with polarization –45° are coded as "1"; in alphabet II, vertically and +45° polarized photons are coded "0". A sends arbitrary sequences of photons in one of these states to B and protocols the polarization state of each particle in an arbitrary selection of its polarizer settings. B uses two analysers: one allows distinction of horizontal and vertical polarization, the other of diagonally polarized states. Before each detection measurement, B chooses one of the analysers, notes that choice and protocols whether or not a photon has been registered. Then, after a sufficient number of measurements, A and B (publicly) compare their respective lists with respect to those events where B's analyser met the state of the photon sent by A. In that case (see Fig. 6.6), both have identical bit-values. All events, where no photon has been detected or where the states of sent photon and analyser setting were not coinciding, are neglected.

The publication whether a photon was horizontally/vertically or diagonally polarized, gives, however, no detailed information on the state of the particle sent. If an eavesdropper, E, tries to measure a photon and, corresponding to the result, sends a substitute photon to B, an error occurs: E will measure 50% of the photons

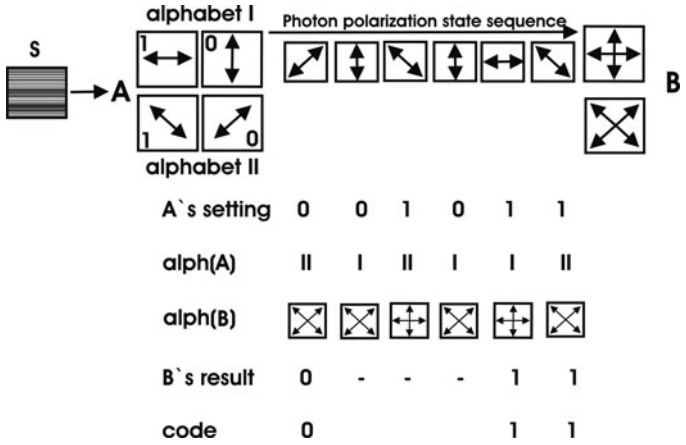


Fig. 6.6 Quantum key distribution following the historical BB 84 protocol where bits are coded as polarization states of photons (see text)

in the wrong basis, hence in 50% of the cases in which the polarization state of the photons sent by A and in which the analyser settings of B are identical, altered photon states are detected in B. For 50% of these altered photons, B obtains a result in contradiction to the originally chosen state by A. In A and B's lists, an error of 25% occurs allowing identification of a third party in the information transfer chain.

6.3 Concept Transfer: Quantum Computing

6.3.1 Brief Background on Quantum Computing

Classical computers, actually introduced by Zuse [23, 24], use physical realizations of bits (the fundamental unit of information, represented as 0 or 1) via a macroscopic system such as the magnetization on a hard disc or the charge of a capacitor. For example, an information consisting of n-characters and stored in a classical computer, is described by a string of $8n$ zeros and ones. Whereas this type of computer obeys classical laws of physics, physical phenomena unique to quantum mechanics are used in a quantum computer. The fundamental difference of both computing schemes is well acknowledged by the introduction of the qubit, the quantum bit in contrast to "standard" bits of present technology. As was shown above, the superposition principle of quantum mechanics leads to a series of behaviour of quantum particles that, if transposed into the macroscopic world as done with Schrödinger cats, results in paradoxa. In quantum computing, the qubit, as a fundamental unit of information, can be considered quaternary in its nature compared to the binary classical bit. A pure qubit state is a linear coherent quantum

superposition of two (entangled) states of a basis vector:

$$|\psi\rangle = a|0\rangle + b|1\rangle, \tag{6.9}$$

where a, b are probability amplitudes and are generally complex with $|a|^2+|b|^2=1$. Equation (6.9) shows that, in general, a qubit can exist as 0, 1 or simultaneously as both 0 and 1 where the coefficients a, b determine the probability of each state. These are the typical quantum mechanical properties described above that govern the deviations from the Bell limit, quantum teleportation and quantum cryptography.

Historically, the idea of quantum computation emerged when scientists were contemplating on fundamental limits in computation based on classical machines. According to Moore's law, which stated in the 1970s that the number of transistors on a chip doubles every two years (surprisingly, this trend has been maintained until the recent time where chips reach close to 1 billion transistors), an eventual limit was foreseen. The continuous size reduction would result in structures where the individual elements reach the size of molecules and/or atoms. At that scale, the physical laws that govern the circuit properties will be clearly quantum mechanical in nature, raising the question whether a limitation was reached in the miniaturization or whether one could use the quantum world behaviour for a new type of computing philosophy. In 1982, Feynman attempted to answer that question by proposing a model that demonstrated how a quantum system could be employed to perform computations. In addition, it was claimed that such a machine would be able to simulate quantum physical processes [25]. Three years later, Deutsch communicated a theory which showed, that, in principle, any physical process could be modelled by a quantum computer [26].[1] This work initiated the search for applications of a quantum computer. In 1994, Shor described a method how quantum computers could crack a particular important task in number theory, i.e. the factorization problem [29, 30]. Difficulty in factorization is used presently as security code on conventional computers: the most popular public key cryptosystem that provides security, for instance, of electronic bank accounts. RSA, named after Rivest, Shamir and Adleman is the most used cryptosystem [31]. It relies on the fact that integer factorization is thought to be infeasible with a classical computer for large numbers which are a product of two prime numbers of approximately equal size, such as, for instance, products of two 300-digit prime numbers. Shor showed that with his algorithm, a quantum computer could factor large integers quickly. This "killer algorithm" could, in principle, break many of today's cryptosystems and led to a tremendous interest in quantum computing schemes and their possible implementation in real machines. In 1998, the first working quantum computer based on two-qubit nuclear magnetic resonance (NMR) (see below) has been

[1] In an essay dedicated to the psychologist Jung, Pauli discussed the topic of the immediate presence of the conscious as well as the subconscious, postulated by Jung [27] from a physicists view, already indicating the implications of a simultaneous presence of complex information in humans [28].

demonstrated, followed by a three-qubit NMR computer and the first execution of what has become known as Grover's algorithm [32]. Subsequently, rapid progress in realization has been made: the first five and seven qubit NMR computer and the first execution of parts of Shor's algorithm occurred in the year 2000; developments in 2005, 2006 until the present time are too numerous to account for all of them; among them are steps towards stable quantum memories, qubit caging in a fullerene particle for isolation (see decoherence below) and the reversible transformation of quantum information from atoms to photons.

6.3.2 *Principles of Quantum Computing*

Upon extending the principle of coherent superposition of quantum states, prepared as superposition of its two logical states 0 and 1 to a three-bit system, the capabilities of quantum computing begin to become visible: a classical register can, at a given time, store one of the eight possible configurations (0 0 0, 1 1 1, 0 0 1, 0 1 1, 1 0 0, 1 1 0, 0 1 0 and 1 0 1). A quantum register consisting of three qubits, however, can store all eight numbers in a quantum superposition. All eight numbers are simultaneously present in the register due to the superposition principle. This storage capacity increases exponentially with the number of qubits: for four qubits, 16 different numbers can be stored at once and for N qubits, the storage capacity is 2^N. Now, consider Fig. 6.7. If one has created a register, for example, of three qubits as shown in the above figure, operations can be performed on them. If the qubits are atoms, then their states can be altered by tuned laser pulses that affect electronic states. Then, the initial superposition of encoded numbers evolves into a new, different one and in this process, each number of the initial

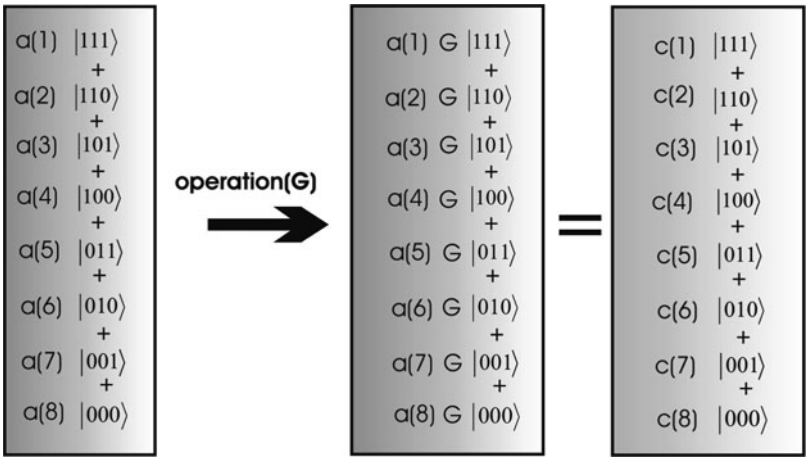


Fig. 6.7 Principle of quantum computing (see text)

superposition is affected. As a consequence, extended *parallel* computation in one element of quantum hardware takes place. A quantum computer, by using only one computational step, can perform the same mathematical operation on 2^L different input numbers that are encoded in different superpositions of the L qubits. In principle, classical computers can perform the same calculation but would just need more time and/or memory. Such computers would have to repeat the same computation 2^L times or would have to use that number of different processors which work in parallel. In classical computers, performing complex operations as those described here, time and memory elapse quickly because of their exponential increase: to solve a specific problem, computers follow a specific set of instructions, called algorithm. Fast algorithms are multiplication, slow ones are, for example, factorization or calculation of chess strategies. To calculate the factors of the number 31,707 takes much longer time than to calculate 117×271. For a computer algorithm, its applicability strongly depends on the fact that processing time does not increase sharply when the same method is used for increasingly larger numbers. Factorization of a 30-digit number takes about 10^{13} times more time/memory than that of a three-digit number (using the simplest trial division method). Increasing the digits of a factorization problem soon results in an enormous increase in processing time and memory on classical systems, limiting its usability for advanced problems and deciphering. It is noteworthy to specify this limitation more clearly: while technological progress increases computational speed by a multiplicative factor, the exponential dependence between input size and processing time limits the usability of conventional systems. The perceived power of quantum computing (where new quantum technology has still to be developed much further) lies in the advancement of quantum algorithms based on quantum superposition that comprise an exponential number of different terms.

6.3.3 Mathematical Aspects of Quantum Computation

Consider the eight-dimensional complex vector in the qubit register in Fig. 6.7: an algorithm for quantum computation must in each step of the algorithm modify this vector by a unitary matrix (unitary transformation of a matrix U: $U^{-1} = U^*$, the complex conjugate). The invertibility of the matrix ensures reversibility of each calculative step. A quantum algorithm is hence implemented by a sequence of unitary operations that typically act on a small number of bits in each step. This is actually an "engineering" of entanglement of quantum states, being closely related to building of gates that act on qubits [33, 34]. On a classical computer, logic gates other than the NOT gate are not reversible (except the AND gate if both input bits are 1). In quantum computation, reversible gates are possible. An n-qubit reversible gate is a unitary mapping U of the space of qubits $L_{\mathrm{qb}}(n) = l^2(\{0,1\}^n)$, where $\{0,1\}^n$ denotes the classical bit space [35]. Among the single and multiple qubit gates, the single qubit NOT gate and the multiple qubit controlled NOT gate (CNOT) are considered here. The former describes the transformation of the state

$$a|0\rangle + b|1\rangle \Rightarrow \text{gate} \Rightarrow b|0\rangle + a|1\rangle, \tag{6.10}$$

where the gate for single qubits are unitary 2×2 matrices such as the one used here

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{6.11}$$

The multiple qubit CNOT gate is based on two inputs: (1) a so-called control qubit and (2) a target qubit. If the control qubit is the state $|0\rangle$, then the target qubit is unchanged; if the control qubit is $|1\rangle$, then the target qubit is flipped. Hence, that gate permutes the four basis states as follows: $|0\,0\rangle \rightarrow |0\,0\rangle; |0\,1\rangle \rightarrow |0\,1\rangle; |1\,0\rangle \rightarrow |1\,1\rangle; |1\,1\rangle \rightarrow |1\,0\rangle$. This operation is shown more explicitly in (6.12)

$$\begin{aligned} \oplus(|0\rangle_c|0\rangle_t) &= |0\rangle_c|0\rangle_t, \\ \oplus(|0\rangle_c|1\rangle_t) &= |0\rangle_c|1\rangle_t, \\ \oplus(|1\rangle_c|0\rangle_t) &= |1\rangle_c|1\rangle_t, \\ \oplus(|1\rangle_c|1\rangle_t) &= |1\rangle_c|0\rangle_t, \end{aligned} \tag{6.12}$$

where the symbol $\oplus$ denotes the CNOT gate operator given by the unitary matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{6.13}$$

and the subscripts c and t refer to the control and target qubit, respectively. In quantum computation, every multiple qubit gate can be constructed from CNOT gates and single qubit gates. Although the logic of the gates appears rather straightforward, their experimental implementation is not easy. The reason is that the implementation of a two-qubit state requires very precise control of the couplings between the qubits. Before turning to experimental realization of quantum computing, another practicality issue is considered: if the quantum states in the computer interact with the environment, even in the slightest possible way, it would cause the system to decohere. Then, the unitary character (and hence the invertibility) of the quantum computation steps are lost. For candidate systems such as those based on NMR and MRI (magnetic resonance imaging) (see below), low temperature so-called dephasing times range between nanoseconds and seconds. Typically, computational error rates are proportional to the ratio of operating time and decoherence time; accordingly any calculation must be performed in a time much shorter than the decoherence time. The scalability condition sets the figure for the required error rate at each gate at 10^{-4}, hence each gate must perform its task 10 000 times faster than the decoherence time of the system.

6.3.4 Realization of Quantum Computing

Selected possible candidates for quantum computation comprise among others:

(1) NMR-based systems in solution and in solid state (Kane computer).
(2) Trapped ion quantum computers.
(3) Quantum optical systems.
(4) ESR-based systems using fullerenes.
(5) Superconductivity-based systems (charge and flux qubits).

The above systems will be treated shortly with regard to their underlying physics, feasibility of implementation and processing properties.

NMR is a well-established technique and, rather early, successes in running quantum algorithms using NMR systems have been reported [36]. The method uses the spin $1/2$ states of atoms in molecules as qubits. Each spin $1/2$ nucleus provides a single qubit and simple logic gates that affect only one qubit of the quantum system, can be realized using rf magnetic fields which interact strongly with the spins thus enabling their precise control. In liquid NMR, the signal from a single molecule is by far too small to be detected and measurements have to be done on an ensemble of molecules. Then, the synthesis of a pure initial state has to be achieved from a thermal ensemble. In pulsed NMR, a transverse rf magnetic field is applied to a sample in a strong longitudinal field $\vec{B}_0$. $\vec{B}_0$ induces Zeeman splitting of the levels of the spin system. In thermal equilibrium and for a two-level system, the ratio of population of the higher energy level to the lower energy level is slightly different from unity. This deviation of about 10^{-6} is nevertheless measurable as macroscopic magnetization of the sample as a precession. Due to differences of the chemical environment in the selected molecule even for identical atoms, selective single-spin rotation can be accomplished using rf transverse pulses in resonance with the atom transition frequencies [37, 38].

An example for implementation of a quantum algorithm using a custom-made molecule, shown in Fig. 6.8, that allowed factorization of $N = 15$ demonstrates the specific demands of quantum computing.

Quantum factorization using this synthesized molecule requires coherent control over seven qubits in the course of a long chain of controlled interactions which poses high demands on the spin–spin coupling network. In the considered molecule, all seven spin states are well separated in their transition frequency $\omega_i/2\pi$ and hence complex operations can be performed. Logic gates and unitary transformations are

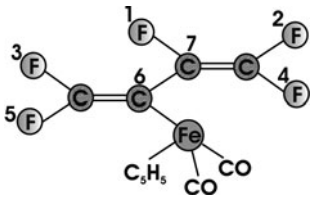


Fig. 6.8 Perfluoro-butadienyl–iron complex quantum computer with 5^{19}F and 2^{13}C spin $1/2$ nuclei as qubits labelled 1–7

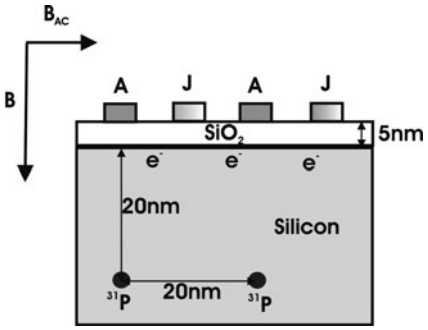


Fig. 6.9 Schematic of a one-dimensional array of (here only two) ^{31}P atoms in a Si host for quantum computing); A-gates control the hyperfine interaction via the nucleus resonance frequency; J-gates allow control of the coupling between adjacent nuclear spins via the exchange energy (function of overlap of electron wave functions); **B**, static magnetic field ($\sim$2 T), $\mathbf{B_{AC}}$, ac. magnetic field ($\sim$1 mT)

constituted by sequences of rf pulses that alter spin orientations and couplings. Liquid NMR quantum computing, however, has encountered scepticism rather early [39] advocating designed quantum computing systems such as those based on regular arrays of quantum dots and ESR-based experiments rather than bulk liquid NMR quantum computers which exhibit quite classical dynamics.

It is in that direction that the subsequent invention of a nuclear spin-based solid quantum computer by Kane [40] was aimed. It is based on an array of separated, isotopically pure ^{31}P atoms with spin $1/2$ embedded as donors in a crystalline Si lattice with the silicon purified to contain only ^{28}Si. From theoretical considerations that neglect the valley degenerate anisotropic band structure of Si, the design of a spin-impurity-host quantum computer based on Si follows as shown in Fig. 6.9. ^{31}P atoms are placed at 20 nm below a SiO_2/Si interface with a horizontal separation of also 20 nm. Above each donor atom, metallic gates are arranged where an applied voltage alters the Larmor frequency of individual donor atoms allowing them to be individually addressed. Nuclear spins, however, do not interact with each other at a distance of 20 nm. Although able to perform single qubit operations, quantum computing requires two qubit operations. Since nuclear spins can interact with electrons, applying a voltage at the A-gate results in externally controlled precession frequency of a single donor nuclear spin-electron system. Spins can be selectively brought into resonance with the oscillating magnetic field $\mathbf{B_{AC}}$ thus enabling arbitrary operations on each nuclear spin. The two-bit operation requires nuclear spin exchange between two donor nuclear spin systems. Such interaction can be realized by electron-mediated interaction brought about by the exchange interaction that, in its simplest form, depends on the donor distance, the semiconductor Bohr radius and its dielectric constant. Spin measurements in this quantum computer are possible because distinct nuclear spin states are converted into different electron spin states and the electron spin is determined by its influence on the exchange-coupled

two-electron system. If a differential voltage is applied to the A gates, electrons in the singlet D^- state (two electrons bound to the same donor) can move between the donors. Using sensitive single-electron capacitance techniques, this motion can be detected, allowing spin measurement.

Summarizing, the solid-state NMR quantum computer works under control of the following external parameters: A-gates above the donors define the strength of the hyperfine interactions and hence the resonance frequency of the nuclear spins; the J-gates between the donors allow for electron-mediated coupling between nuclear spins and the a.c. magnetic field inverts spins at resonance.

This approach has several principle advantages: (1) it uses advances and existing knowledge in silicon technology; (2) it is scalable to an arbitrary number of qubits, due to the fact that qubits can be individually addressed electrically. Although the electron spin relaxation time at low temperature (1.5 K) is in the range of 10^3 s and the nuclear spin relaxation time is about 10 h, the lifetime of the D^- readout state might be too small. Recent realization attempts are based on progress in nanoscience using either low energy ion beam implantation or STM lithography with subsequent epitaxial Si overgrowth [41].

Another variation of the quantum computer is based on the use of trapped ions [42]. The computer is based on the confinement of charged atoms (ions) in free space in an oscillating electromagnetic rf field as has been shown by Paul, initially four decades ago [43]. The Paul-trap allows to store qubits in stable electronic states of the respective ion. Scaling has been demonstrated by using trapped ion systems where a large number of qubits is obtained by shuttling ions into arrays of traps. This, together with the exceptional stability (thousands to millions of years) and the long decoherence time of these so-called hyperfine qubits makes trapped ion systems a particularly promising approach for a scalable universal quantum computer. Recently, a major advance has been communicated that relates to so-called open system quantum computers [44]. Open quantum systems include interaction of the quantum mechanical system with ambient/environment. Usually, this interaction with the environment has been considered deleterious as it leads to quantum decoherence of entangled states and, in general, to dissipation (loss in coherence and superposition properties of quantum systems). Contrary to that experience, it was shown that an additional ion of a quantum system that contained four trapped ions can be used for a controlled interaction between the quantum system and the environment. Rather surprisingly, it was found that by using dissipation via the properties of the environment ion, it became possible to enhance quantum effects such as entanglement [44].

A linear Paul trap for ion storage is schematically drawn in Fig. 6.10. It consists of four rods, two of which are each separated into three parts where a constant bias can be applied. The outer parts are used as caps where the bias provides linear confinement. The rf voltage $V = V_0 \cos \omega t$ is applied between the dark grey electrodes. Ion confinement in the x–y plane results from the ponderomotive (pseudo)potential of the oscillating electric field $\bar{E}$ and is given by (see Fig. 6.10)

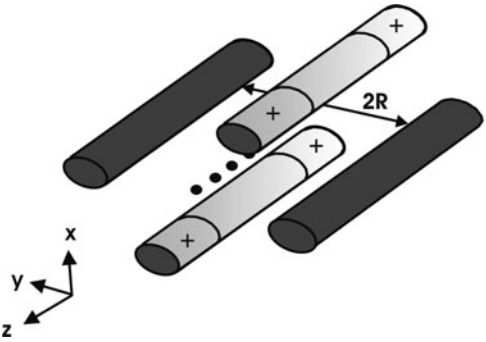


Fig. 6.10 Schematic drawing of a linear Paul trap; the oscillatory field is applied between the dark electrodes that are separated by 2R; the lighter electrodes are dc biased, the positive charging (+) results in confinement of ions in z-direction; the *black spheres* represent a linear bead-type ion arrangement (see text)

$$U(\bar{r}) = \frac{q^2}{2\omega_T^2}\langle \bar{E}^2(\bar{r})\rangle, \tag{6.14}$$

where the oscillation frequency of an ion in the trap in the x–y plane is

$$\omega_{xy} \cong \frac{qV_0}{\sqrt{2}\omega_T m_i R^2}, \tag{6.15}$$

E is the rf electric field, q and m are the ion charge and mass, respectively. According to (6.14), ions seek the position of the minimum of $|E(r)|$. For several ions, their mutual Coulomb repulsion results in an equilibrium configuration in the form of a linear array, similar to beads on a string as schematically shown in Fig. 6.10 by the black spheres. For typical voltages and frequencies ($V_0$500 V; $\omega_T/2\pi$100 MHz; $R = 200\,\mu$m; $q =$ 1e) the distance between two $^9Be^+$ ions is about 3 μm.

A viable quantum computer architecture must comprise the highest possible number of qubits. For various reasons given in the respective literature, realistic scalable systems must use arrays of interconnected traps which themselves contain only a small number of ions. As information carriers between traps, photons or ions can be used (see below).

Within each trap, qubits, separated by optical frequencies, that are stored in metastable hyperfine levels or qubits stored in ground state configurations, separated by microwave energies can be used. The latter has several important advantages, such as a very long radiative lifetime, and is shortly described here. The example in Fig. 6.11 shows the lowest $^{111}Cd^+$ ion levels.

Detection of the internal hyperfine state is accomplished by irradiation with a circularly polarized laser beam close to the $P_{3/2}$–$S_{1/2}$ transition wavelength of 214.5 nm. If the ion is in the $|\downarrow\rangle$ spin state ($F = 1$), the beam scatters many photons but virtually none if the $|\uparrow\rangle$ state is occupied. The resulting qubit detection selectivity reaches 99.7% for this ion type.

Single qubit gates, realized by hyperfine frequency qubit rotations can be based either on resonant microwave radiation (where the radiation is tuned to the transition

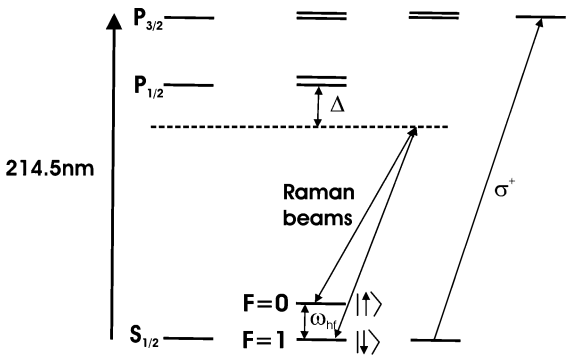


Fig. 6.11 Schematic of the internal electronic levels of the $^{111}Cd^{+}$ ion; $F = 1$, $F = 0$, hyperfine ground states; σ^{+} circular polarized so-called detection beam; also shown are Raman beams used for realization of quantum logic gates (see text)

frequency ω_{hf}, see Fig. 6.11) or on Raman transitions with two photon fields. In this scenario, two laser beams with photon energies that are detuned by the energy Δ (Fig. 6.11) from an excited state differ in photon energy by $\Delta E_{\mathrm{Photon}} = \hbar\omega_{\mathrm{hf}}$. Due to difficulties of addressing individual qubits using microwave technology, the optical method which uses stimulated Raman transitions is considered more advantageous for the purpose of quantum computing.

Qubit gate entanglement can be achieved by the Coulomb interaction between the trapped ions. An external "control field" is used for the entanglement of trapped qubits that is based on correlation of the ion positions. Consider the example where the coupling of a qubit, stored in the $|\downarrow\rangle$ and $|\uparrow\rangle$ $S_{1/2}$ hyperfine ground state, is made via the ion's position $\vec{X}$ by using the electric field $\vec{E}(\vec{X})$ of a laser. The coupling of qubit states proceeds due to dipole coupling to excited P states (Fig. 6.11) $|e_{xc}\rangle$ and is described by the interaction Hamiltonian

$$H_{\mathrm{I}} = -(\vec{\mu}_{\uparrow e} + \vec{\mu}_{\downarrow e})\vec{E}(\vec{X}), \tag{6.16}$$

where $\vec{\mu}$ denotes the electric dipole operator between the states $|s\rangle$ and $|e\rangle$. Basically, simultaneous as well as sequential interaction, both generating entanglement, can be employed. The most common methods used are: (1) motion-sensitive stimulated Raman transitions and (2) the spin-dependent optical force technique. Due to different possibilities in the realization of gates (CNOT for the Raman transition method, phase gates for method (3)), both methods are considered important for implementation. Several gate schemes based on trapped ions have been demonstrated including a four-qubit gate [45] and a π-phase gate [46] that showed 97% fidelity. Recently, an advanced positioning system for Cs atoms (instead of ions) has been realized that allows ordering and shifting of the neutral particles [47].

In quantum optical systems, basically the equivalence between (traditional) linear optical elements (beam splitters, mirrors, etc.) and single bit quantum gates can be used for optical quantum computation. Such an approach uses the fact that several qubits (n) can be represented by a *single* photon in an interferometric

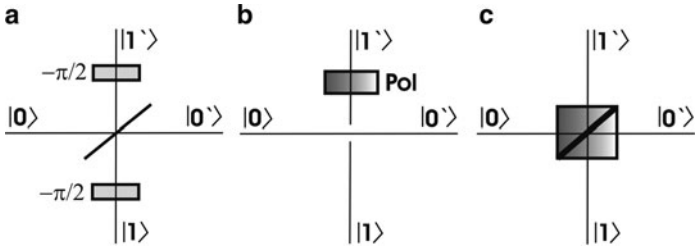


Fig. 6.12 Optical simulation of the basic quantum logic gates; (**a**) Hadamard gate (using a symmetric lossless beam splitter); (**b**) CNOT gate (polarization rotation element); (**c**) CNOT gate with switched control and target qubits (polarizing beam splitter element) (see text)

arrangement with 2^n photon paths. Then, the appropriate arrangement and cascading of the linear optical elements allows simulation of networks of one- and two-qubit gates. This, however, leads to an exponential growth of optical paths that limits the applicability of this principle to simple circuits. The method is described here due to the intelligibility of the simulation of the most fundamental quantum logic gates which are shown in Fig. 6.12.

Figure 6.12 shows single photon experiments using a Mach–Zehnder interferometer that allows optical simulation of quantum gates. In Fig. 6.12a, the two input modes entering the beam splitter are $|0\rangle$ and $|1\rangle$, the exit modes are $|0'\rangle + \mathrm{i}|1'\rangle$ or $|1'\rangle + \mathrm{i}|0'\rangle$, where the complex values result from the phase shift of $\pi/2$ between transmitted and reflected photons. With the phase shifters inserted into the interferometer arms as shown in Fig. 6.12a, the lossless beam splitter performs a Hadamard transformation [48]:

$$\begin{pmatrix} |0'\rangle \\ |1'\rangle \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} |0\rangle \\ |1\rangle \end{pmatrix}. \tag{6.17}$$

In Fig. 6.12b, it is assumed that the two optical modes carry polarization; the spatial degree of freedom carries the control qubit, polarization the target qubit. If the photon is in the vertical spatial mode, it will, according to the figure, undergo a polarization rotation, flipping the polarization from horizontal ($|0\rangle$) to vertical ($|1\rangle$), a CNOT gate operation has been realized (see (6.12, 6.13)). The broken line in the figure indicates that there is no interaction between the crossing modes [49]. The polarizing beam splitter inserted in Fig. 6.12c acts on the location qubit (flipping it or not, equivalent to reflection or transmission, respectively) in dependence of the photon polarization state, thus achieving CNOT operation.

The early optical quantum computers, based on linear optical elements, were believed to be non-scalable. At that time it was widely accepted that scalable systems needed a non-linear component, such as a Kerr medium (exhibiting a non-linear light-intensity-dependent component to the refractory index). It has been demonstrated, however, that universal quantum computers based on linear optics,

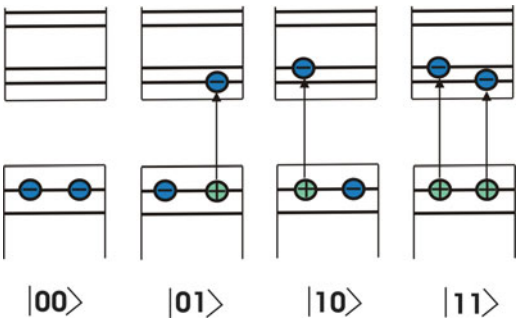


Fig. 6.13 Low photon energy exciton transitions in a single quantum dot with the crystal ground state $|0\,0\rangle$, two single exciton states $|0\,1\rangle$ and $|1\,0\rangle$ and the biexciton state $|1\,1\rangle$ where 0 refers to no exciton and 1 to an exciton state

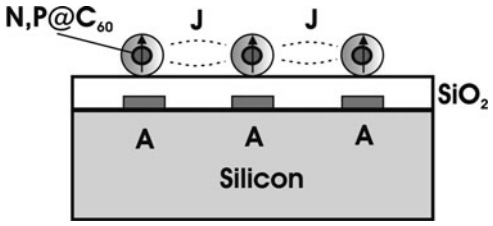


Fig. 6.14 Concept for an ESR solid-state quantum computer; qubits based on electron spins (*arrows*) of endohedral fullerenes (*spheres*) are used; the system is based on a Si/SiO_2 structure; A- and J-gates are as in Fig. 6.9

single photons and photon detectors can be realized, leading to a new paradigm for optical quantum computing [50, 51].

Another development involves the use of optically excited excitons and biexcitons in a quantum dot system that behaves like a CNOT gate [52]. Figure 6.13 shows the qubit-state assignment within the quantum dots that were fabricated by sandwiching a 4.2 nm GaAs layer between two 25 nm AlGaAs layers.

Rabi oscillations (a periodical time evolution of the population difference in a two-level system that occurs at the electrical dipole transition frequency (see [53], and references therein) between the ground state and one-exciton states and between one-exciton states and the two-exciton state can be achieved with lasers. The above quantum dot system was shown to operate like a CNOT gate providing a novel option for a solid-state quantum computer.

Electron spin resonance (ESR)-based solid state quantum computers use, for example, endohedral fullerenes, i.e. fullerenes with inserted atoms (N, P), as spin system. Using doped fullerenes as information-carrying system has the advantage that the carbon cage protects the spin system from the environment while simultaneously providing a handle for spin manipulation. Qubits are encoded in electron spins of ^{31}P (nuclear spin $I = 1/2$) and $^{14}N(I = 1$, three hyperfine lines), inserted into the fullerenes. Quantum logic operations are performed using electron spin resonance. Qubit interaction between the encapsulated atoms occurs by magnetic

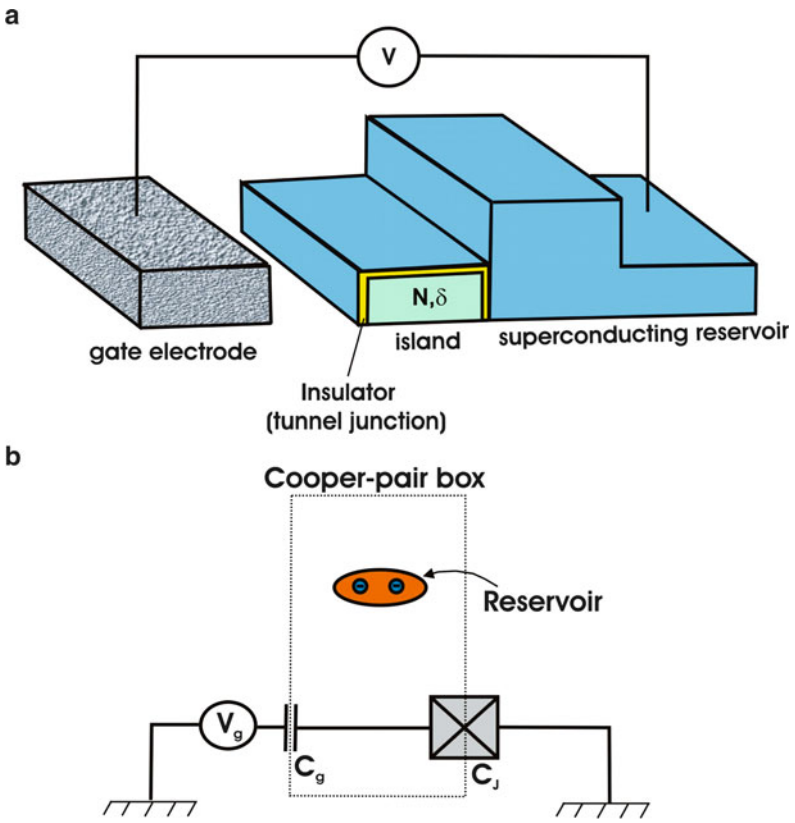


Fig. 6.15 Schematic of a Cooper-pair box arrangement (**a**) and its electric circuit (**b**); N, number of extra Cooper-pairs stored in the box; δ, phase difference between the superconducting island and the superconducting reservoir

dipole interaction. The conceptual realization resembles that of solid-state NMR quantum computer systems as shown in Fig. 6.9 and is given in Fig. 6.14:

A-gates and J-gates for spin addressing and qubit coupling, respectively, are used as shown in the figure above where the dopant atoms of the NMR system are replaced by the "doped" fullerenes. Qubits are set and read out via ESR pulses [54, 55].

An electrically controlled quantum computer system involves so-called Cooper-pair and flux qubits, i.e. micrometer-sized loops of superconducting metals that are separated by Josephson junctions. Josephson junctions consist of two superconducting materials, separated by a thin insulating layer that allows intermetallic current flow via tunnelling of Cooper pairs. The coherent tunnelling of Cooper-pairs in low capacitance Josephson junctions and Coulomb blockade effects provide the physical realization of qubits where logical states differ by a Cooper-pair charge. A sequence of gate voltages allows single- and two qubit operations and nanoscale Josephson

junction technology, based on what has become known as Cooper-pair box, is presently considered for realization [56]. The latter consists of a nanometer-scale superconducting electrode that is connected to an electron reservoir via a Josephson junction. It has been shown that such a system constitutes an artificial two-level electronic system where the levels are charge states of multiples of 2e [57]. The system, where coupling occurs by Cooper-pair tunnelling, operates if the conditions $\Delta > E_C$(Δ superconductivity gap, E_C single electron charging energy of the box) and $E_C > E_J$(E_J Josephson energy), $E_C > kT$ are met. A Cooper-pair box is visualized in Fig. 6.15 where also the corresponding electric circuit is drawn (Fig. 6.15b). N is the number of excess Cooper pairs stored in the superconducting island, separated by a thin oxide layer from the reservoir. The two lowest lying energy levels of a Cooper pair are well separated from each other and from the other levels and they can be used to implement the $|0\rangle$ and $|1\rangle$ qubit states. The charge qubits are typically read out with an rf single electron transistor, i.e. an extremely sensitive electrometer. The phase qubit and the flux qubit are further types of superconducting qubits; the flux qubits are based on the persistent flow of Cooper pairs in micrometer-sized superconducting loops that are interrupted by a number of Josephson junctions. The basis states of the qubits are defined by the direction of the circulating currents (clockwise or counter-clockwise). For the former, the magnetic moment is pointing downwards ($|\downarrow\rangle$) and for the current moving left, it is pointing upwards ($|\uparrow\rangle$). The characteristics of the superconducting qubits reduce their applicability in large-scale computational systems. In particular, the decoherence time has been too short by an orders of magnitude. Therefore, research on the mechanism of decoherence (compare also [44], and references therein) has been intensively pursued [58].

References

1. A. Einstein, B. Podolsky, N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, Phys. Rev. **47**, 777–780(1935)
2. J.S. Bell, Bertlmann's socks and the nature of reality, J. de Physique, Colloque **C2**, C2-41–C2-62 (1981)
3. P.A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist* (Library of the Living Philosophers, Evanston, 1949).
4. J.S. Bell, On the Einstein–Podolsky–Rosen paradox, Physics **1**, 195–200 (1964)
5. E. Schrödinger, Die gegenwärtige Situation in der Quantenmechanik, Naturwissenschaften **23**, 807–812 (1935)
6. E. Schrödinger, Die gegenwärtige Situation in der Quantenmechanik, Naturwissenschaften **23**, 823–828 (1935)
7. E. Schrödinger, Die gegenwärtige Situation in der Quantenmechanik, Naturwissenschaften **23**, 844–849 (1935)
8. J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, Proposed experiment to test local hidden-variable theories, Phys. Rev. Lett. **23**, 880–884 (1969)
9. P. Walther, M. Aspelmeyer, K.J. Resch, A. Zeilinger, Experimental violation of a cluster state bell inequality, Phys. Rev. Lett. **95**, 1–4 (2005) (020403)

10. C.C. Gerry, J. Mimih, A. Benmoussa, Maximally entangled coherent states and strong violations of Bell-type inequalities, Phys. Rev. A **80**, 1–11 (2009) (022111)
11. D. Bouwmeester, J.-W Pan, K. Mattle, M. Eibl, H. Weinfurter, A. Zeilinger, Experimental quantum teleportation, Nature **390**, 575–579 (1997)
12. C.H. Bennett, D.P. DiVicenzo, Quantum information and computation, Nature **404**, 247–255 (2000)
13. W. Tittel, J. Brendel, B. Gisin, T. Herzog, H. Zbinden, N. Gisin, Experimental demonstration of quantum correlations over more than 10 km, Phys. Rev. A **57**, 3229–3232 (1998)
14. D. Salart, A. Baas, C. Branciard N. Gisin, H. Zbinden, Testing spooky action at a distance, Nature **454**, 861–864 (2008)
15. W.K. Wootters, W.H. Zurek, A single quantum cannot be cloned, Nature **299**, 802–803 (1982)
16. M. Riebe et al., Deterministic quantum teleportation with atoms, Nature **429**, 734–737 (2004)
17. N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145–195 (2002)
18. A. Dantan, N. Treps, A. Bramati, M. Pinard, Teleportation of an atomic ensemble quantum state, Phys. Rev. Lett. **94**, 1–4 (2005) (050502)
19. R.L. Rivest, A. Shamir, L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Commun. ACM **21**, 120–126 (1978)
20. G. Vernam, Cipher printing telegraph systems for secret wire and radio telegraphic communications, J. Am. Inst. Electr. Eng. **45**, 1–115 (1926)
21. C.H. Bennett, G. Brassard, Quantumcryptography: public key distribution and coin tossing, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE Press, 1984), p. 175
22. A.K. Ekert, Quantum cryptography based on Bell's theorem, Phys. Rev. Lett. **67**, 661–663 (1991)
23. K. Zuse, *Der Computer-Mein Lebenswerk* (Springer, Berlin, New York, 1884)
24. K. Zuse, Patent Application Z-391 (Deutsches Patentamt, Berlin, 1941)
25. R. P. Feynman, Quantum mechanical computers, Found. Phys. **16**, 507–531 (1986)
26. D. Deutsch, Quantum theory, the Church–Turing principle and the universal quantum computer, Proc. Roy. Soc. London A **400**, 97–117 (1985)
27. C.G. Jung, Der Geist der Psychologie, Eranos Jahrbuch **XIV** (1946)
28. W. Pauli, *Leben = Physik. Chemie?* (Vieweg, Braunschweig, 1984)
29. P. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science* (1994), pp. 124–134
30. P. Shor, Polynomial-time algorithms for prime factorization and discrete logarithm problems, SIAM J. Comput. **26**, 1484–1509 (1997)
31. R.L. Rivest, A. Shamir, L. Adleman (eds.), On digital signatures and public key cryptosystems, Techn. Memorandum MIT Laboratory of Computer Science (1977)
32. L.K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing* (1996), p. 212
33. E. Hagley et al., Generation of Einstein–Podolsky–Rosen pairs of atoms, Phys. Rev. Lett. **71**, 1–5 (1007)
34. K.M. Ghen, C. Saavedra, P. Törmä, J.I. Cirac, P. Zoller, Entanglement engineering of one-photon wave packets using a single-atom source, Phys. Rev. A **58**, R2627–R2630 (1998)
35. M. Hirvensalo, *Quantum Computing* (Springer, Berlin, 2001)
36. L.M.K. Vandersypen, M. Steffen, G. Breyta, C.S. Yannoni, M.H. Sherwood, I.L. Chuang, Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance, Nature **414**, 883–887 (2001)
37. N.A. Gershenfeld, I.L. Chang, Bulk spin-resonance quantum computation, Science **275**, 350–356 (1997)
38. H.K. Cummins et al., Approximate quantum cloning with nuclear magnetic resonance, Phys. Rev. Lett. **88**, 1–4 (2002) (187901)
39. W.S. Warren, The usefulness of NMR quantum computing, Science **277**, 1688–1689 (1997)
40. B.E. Kane, A silicon-based nuclear spin quantum computer, Nature **393**, 133–137 (1998)

41. R.G. Clark et al., Progress in silicon-based quantum computing, Philos. Trans. Roy. Soc. London A **361**, 1451–1471 (2003)
42. J.I. Cirac, P. Zoller, Quantum computations with cold trapped ions, Phys. Rev. Lett. **74**, 4091–4094 (1995)
43. W. Paul, Elektromagnetische Käfige für geladene und neutrale Teilchen (Nobel-Vortag), Angew. Chem. **102**, 780–789 (1990)
44. J.T. Barreiro et al., An open-system quantum simulator with trapped ions, Nature **470**, 486–491 (2011)
45. C.A. Sackett et al., Experimental entanglement of four particles, Nature **404**, 256–259 (2000)
46. D. Leibfried et al., Experimental demonstration of a robust, high fidelity geometric two-ion qubit phase gate, Nature **422**, 412–415 (2003)
47. Y. Miroshnychenko et al., Quantum engineering: an atom-sorting machine, Nature **442**, 151 (2006)
48. H. Kharaghani, B. Tayfeh-Rezaie, A Hadamard matrix of order 428, J. Combinat. Des. **13**, 435–440 (2005)
49. N.J. Cerf, C. Adami, P.G. Kwiat, Optical simulation of quantum logic, Phys. Rev. A **57**, R1477–R1480 (1998)
50. E. Knill, R. Laflamme, G.J. Milburn, A scheme for efficient quantum computation with linear optics, Nature **409**, 46–52 (2001)
51. P. Kok, W.J. Munro, K. Nemeto, T.C. Ralph, J.P. Dowling, G.J. Milburn, Linear optical quantum computing with photonic qubits, Rev. Mod. Phys. **79**, 135–174 (2007)
52. X. Li et al., An all-optical quantum gate in a semiconductor quantum dot, Science **301**, 809–811 (2003)
53. T.H. Stievater et al., Rabi oscillations of excitons in single quantum dots, Phys. Rev. Lett. **87**, 1–4 (2001) (133603)
54. W. Harneit, Fullerene-based electron-spin quantum computer, Phys. Rev. A **65**, 1–6 (2002) (032322)
55. S.C. Benjamin, Quantum computing without local control of qubit–qubit interactions, Phys. Rev. Lett. **88**, 1–4 (2001) (017904)
56. Y. Makhlin, G. Schön, A. Shnirman, Quantum-state engineering with Josephson-junction devices, Rev. Mod. Phys. **73**, 357–400 (2001)
57. Y. Nakumura, Y.A. Pashkin, J.S. Tai, Coherent control of macroscopic quantum states in a single-Cooper-pair box, Nature **398**, 786–788 (1999)
58. I. Chiorescu, Y. Nakamura, C.J.P.M. Harmans, J.E. Mooij, Coherent quantum dynamics of a superconducting flux qubit, Science **299**, 1869–1871 (2003)

Chapter 7
Spintronics

7.1 Introduction

Table 7.1 gives the overview of the Sects. 7.1–7.4. First, more traditional ferromagnetic spin injector systems are reviewed, including the giant magneto-resistive effect, followed by an extended overview on the basic aspects of magnetism and spin (Sect. 7.3). The more widespread relevance of these concepts, reaching into quantum electrodynamics and eventually to string- and M-theory is emphasized to show the connectivity of modern physics and device development. The chapter ends with semiconductor spintronics that holds the promise of several rather revolutionary advances such as the development of spin quantum devices (spin FETs and LEDs), quantum computing at room temperature and the availability of ultrafast, very dense memory and logic at extremely low power.

As a novel branch of electronics, spintronics – a neologism for spin-based electronics – uses spin manipulation in addition to charge to produce a desired outcome. In principle, spin manipulation has advantages with respect to processing speed and required power compared to conventional electronics. Spin quantization has been proven in the famous experiment of Gerlach and Stern on Ag atoms in a magnetic field [1] followed a few years later by the formulation of the Dirac equation (see below) in 1928 [2]. Another major step towards spintronics was the discovery of the giant magneto-resistive effect (GMR), considered as the starting point of spin-based electronics [3]. The effect has been used to provide new ways to store and read information on hard discs, tapes or magnetic random access memory (MRAM).

The realization and implementation of spintronic devices needs solving of the following aspects (1) efficient injection of spin-polarized electrons; (2) transport of electron spins across interfaces and (3) within materials; (4) manipulation of electron (and nuclear) spins in a sufficiently short time interval and (5) detection of spin polarization and spin-polarized currents. Although most of this section will predominantly address semiconductor spintronics, a short introduction to more general aspects and initial work will be provided by introducing devices based on ferromagnetic spin injectors. Spin-based quantum computing has already

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2_7,

Table 7.1 Overview of content of Chap. 7

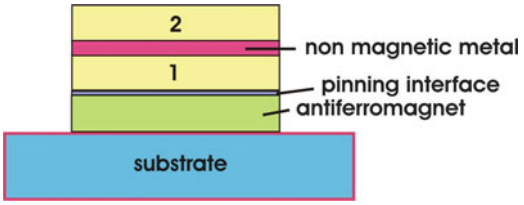


Fig. 7.1 Schematic of a spin valve, a term, coined by IBM, made by two ferromagnetic and an anti-ferromagnetic layer (see text); (1) magnetically pinned layer by direct contact to anti-ferromagnet; (2) soft, magnetically polarizable "free" layer (see text)

been described in Sect. 6.3.4. (realization of quantum computing) and will not be extended here.

7.2 Ferromagnetic Spin Injector Systems

First, consider a spin valve, consisting of two ferromagnetic layers, typically made of alloys of Fe, Ni and Co, that are separated by a thin non-magnetic layer such as a Cu film. One of the outer ferromagnetic layers is magnetically "pinned" in the sense that its magnetization is comparably independent to an applied magnetic field of moderate strength. The pinning has been realized by contact to an anti-ferromagnetic layer that consists of mutual anti-parallel spin domains with net overall magnetization (see Fig. 7.1). The interface between the anti-ferromagnet and the ferromagnet acts as inhibitor to changes of the ferromagnetic layer (1) by external magnetization. The ferromagnet pinning, discovered by Meiklejohn and Bean [4], is due to an exchange bias, i.e. a shift of the centre of the magnetization loop of the ferromagnetic layer in contact with an anti-ferromagnet away from zero external magnetic field. The loop shift has been termed exchange bias field. The magnetization of the second layer (2) can be altered by a comparably small external magnetic field. This layer is also called the free layer.

A spin valve uses the GMR effect, discovered independently by Fert and Grünberg [3, 5]. The GMR effect produces pronounced changes in the electrical resistance in dependence of the relative magnetization of two ferromagnetic layers, separated by a thin metallic but nonmagnetic layer. Although magneto-resistance

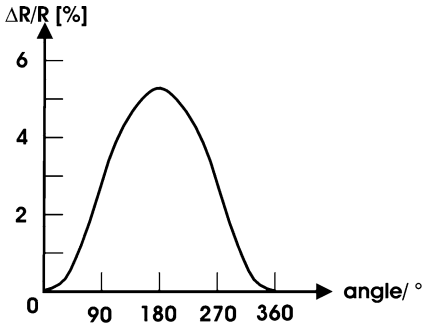


Fig. 7.2 Change of electrical resistance R with relative magnetization of ferromagnetic layers w.r.t. to each other; parallel magnetization: 0° and 360°; anti-parallel: 180° (see text)

(MR), the change in resistance of a metal or a semiconductor in a magnetic field, is an established phenomenon and has been used in heads of hard disk drives, the change in electrical resistance by alternating ferromagnetic layer structures similar to those in Fig. 6.1 exceeded the ∼2% changes in $\Delta R_\Omega / R_\Omega$ by the MR effect substantially. The influence of spin rotation from parallel to anti-parallel in the ferromagnetic layers on electrical resistance is shown in Fig. 7.2. This increase or decrease of the resistance depending on the relative spin orientation in adjacent ferromagnetic layers has resulted in the expression "spin valve": relative spin positions in the layers modulate the flow of current via electrical resistance.

GMR is based on spin-dependent scattering [6] of conduction electrons in a material. Depending on its magnetic orientation, a single-domain magnetic material will scatter electrons with spin up or down differently. When magnetic layers in GMR structures have anti-parallel magnetization, electrical resistance will be high, because electrons with spin up, for instance, that are not scattered in one layer will be scattered in the other. With increasing layers of parallel magnetization, R_Ω will be reduced. Preservation of the spin upon passing the metallic interlayer is mandatory for the functioning of the device and implies that the interlayer thickness is smaller than the so-called spin-scattering length. Typical thicknesses are in the low nm range.

The functioning of a GMR sensor, for application in computer hard disk drives (as shown in Fig. 7.3), is based on the influence of the magnetic field from a bit, that passes beneath the soft magnetic ("free") layer, which induces a rotation of the magnetic orientation of this layer. The magnetization rotation relative to that of the magnetically pinned layer produces a significant change in the electrical resistance of the device that can be read out as a voltage.

A similar device, the magnetic tunnel junction (MTJ), uses an insulating layer between the pinned and the free ferromagnetic layers. The insulator is typically made of alumina (Al_2O_3). Here, the tunnelling resistance is modulated by a magnetic field and shows MR changes up to 40%. Because of the small tunnelling currents, MTJ devices are characterized by rather high resistances but recent work has shown good MR effects at resistances as low as $10\,\Omega\mu\mathrm{m}^{-2}$. Figure 7.4 shows a schematic for an MTJ.

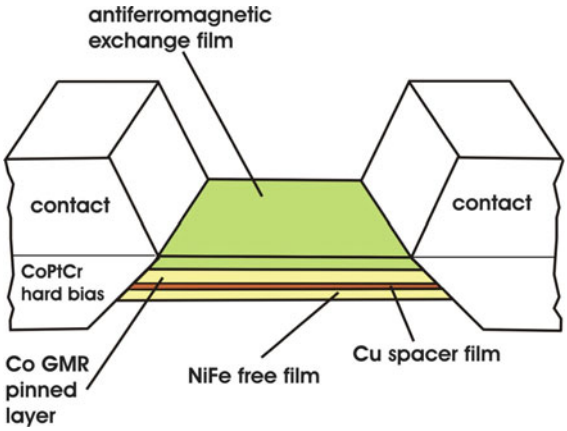


Fig. 7.3 Schematic of a GMR-based magnetic sensor for hard disk drives; during operation, the bits on the disk move under the head

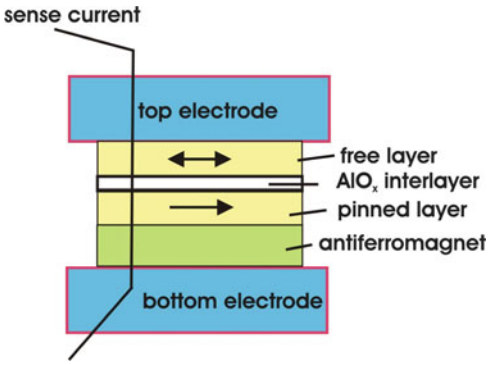


Fig. 7.4 Sketch of a magnetic tunnel junction using an insulating aluminium oxide interlayer (thickness ∼2 nm) and a free layer of ∼5-nm thickness

The functioning of the MTJ can be comprehended by a two-band model for magnetic d-band materials: d-bands are split in each material in a spin up and spin down band with different density of states at the Fermi level. For parallel layer magnetization, majority band electrons tunnel across the interlayer to the majority band of the opposing layer and minority electrons tunnel to their respective counterpart. In anti-parallel configuration, majority electrons are forced to tunnel into the minority band and vice versa. The reduced number of states, induced by the coupling with the minority band, results in an increase of the tunnel resistance.

Compared to metallic spin devices, semiconductor-based spintronics could combine storage, detection, logic and communication capabilities on a chip, replacing hitherto separated components by a single multifunctional device. Envisaged applications include integration of MTJs with Si-based electronics and transformation of magnetic information into an optical signal. Shortly after the discovery of this effect, the spin-field effect transistor (spin-FET) has been introduced [7], leading to a wealth of device proposals that articulated spin-based analogues of conventional electronic devices such as field effect transistors and bipolar junction transistors.

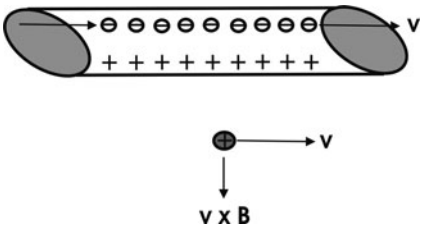


Fig. 7.5 Geometry for considering equivalency between electrostatic and magnetic interactions (see text)

In the spin-FET (see below), transistor current between ferromagnetic source and drain is modulated by a gate voltage that controls the spin precession in the transistor channel via Rashba spin–orbit coupling [8,9]. The latter is a lifting of the degeneracy of electronic quantum states by spin–orbit interaction in the presence of an electric field (that acts as a magnetic field in the rest frame of the moving electron) in conjunction with a structural inversion asymmetry. Such asymmetry exists, for instance, at semiconductor surfaces and hetero-interfaces. The Rashba effect hence allows to extract spin-polarized currents from semiconductor devices via resonant quantum mechanical tunnelling.

7.3 Notes on Magnetism and Spin

As an introduction, also to the understanding of the Rashba effect, the equivalency of magnetic interaction with electrostatics, readily described by special relativity theory is considered: for a positive charge $+Q$, moving with velocity v above a metallic wire through which current flows, a repulsing Lorentz force is observed. Figure 7.5 shows the arrangement; there, the usual picture of electrons moving along the wire with the same velocity v as Q and positive metal cores that are at rest is shown. In an even simpler form, one could consider the positive charge as a smeared out into a homogeneous background, where the metal is then referred to as jellium [10].

The charge Q then experiences a deflection according to the Lorentz force $F = q\vec{v} \times \vec{B}$, where B is obtained from $\oint \vec{B} \mathrm{d}l = \mu_0 \vec{I}$ (Maxwell equation) and $\vec{I} = \lambda \vec{v}$ where λ denotes the electron charge density, resulting in a magnetic force (using $\vec{B} = (\mu_0 \vec{I})/(2\pi r)$) that points directly away from the wire with magnitude

$$F = \frac{Q\mu_0 \lambda v^2}{2\pi r}. \tag{7.1}$$

In that situation, the charge will accelerate away from the wire. Changing now the frame of reference means keeping the electrons and Q at rest and the positive charge is moving relative to them with $-\vec{v}$ (Fig. 7.6).

In this reference frame, no Lorentz force is experienced because Q is at rest. The positive charges with density λ move at velocity v relative to the electrons at rest.

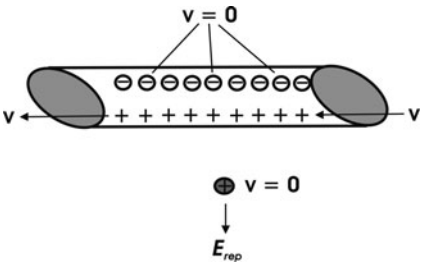


Fig. 7.6 Visualization of the change of the charge motion reference frame; $\boldsymbol{E}_{\text{rep}}$, electrostatic (repulsing) field (see text)

Their charge density in this frame is increased slightly by the Lorentz contraction $\lambda* = \lambda/\sqrt{1-(v^2/c^2)}$. Because of the very small value of v, the surplus density can be approximated by $(\lambda/2)(v^2/c^2)$ and since the electron density in their rest frame is reduced by the same amount, a positive excess electrostatic charge density of $\lambda v^2/c^2$ is obtained. It can be readily shown that the electrostatic force acting on the charge Q is given by

$$F = \frac{Q\lambda v^2}{2\pi r}\frac{1}{\varepsilon_0 c^2}. \tag{7.2}$$

Using $1/c^2 = \mu_0\varepsilon_0$, one obtains the result of (7.1) that has been calculated for the Lorentz force. Hence, whether a particular force on a particular charged system is magnetic or electric depends on the chosen frame of reference.

The *electron spin* has initially been introduced into quantum physics in a phenomenological manner, based on experiments such as that of Stern and Gerlach (see Fig. 7.2) and on the anomalous Zeeman effect. Theoretically, it was established in a relativistic version of the Schrödinger equation that overcame certain shortcomings of the Klein–Gordon equation [2]. As in the Klein–Gordon Ansatz, a relativistic expression for the electron kinetic energy for a free particle was used:

$$H^2 = c^2\vec{p}^2 + m^2c^4. \tag{7.3}$$

Here, $H = \mathrm{i}\hbar(\partial/\partial t)$ and $\vec{p} = -\mathrm{i}\hbar\nabla$ (momentum operator). Rewriting (7.3) and introducing the components of the momentum operator yields $(H^2 - c^2\sum_j \vec{p}_j^2 - m^2c^4)\psi = 0$. This equation can be expressed in a product form whose first part, then linearized, gives already the solution of the complete equation, named after Paul Dirac,who derived it in 1928 [2]:

$$\left(H - c\sum_i \alpha_i\vec{p}_i - \beta mc^2\right)\psi = 0. \tag{7.4}$$

The linearization becomes possible if the constant coefficients α_i and β satisfy the relations (1) $\alpha_i\alpha_j + \alpha_j\alpha_i = 2\delta_{i,j}$ (Kronecker symbol); (2) $\alpha_i\beta + \beta\alpha_i = 0$ and (3)$\beta^2 = 1$. The Dirac equation is first order in the time derivative, as is the Schrödinger equation, is invariant against Lorentz transformation and thus governs the dynamics of quantum waves that describe electrons while satisfying special

relativity. The α's are linear operators that act on the wave function. The condition (1) above is an anti-commutation relation. The conditions (1)–(3) can be satisfied in the simplest way by using 4×4 hermitian matrices for the α's. The necessity to use 4×4 matrices turns out to have considerable physical significance, the first being that the electron wave function must have four components as it is acted on by 4×4 matrices (see below).

The hermiticity of the α's results in a hermitian Hamiltonian with real eigenvalues. There is more than one possible representation of the Dirac matrices with the choice of representation affecting the physical meaning of the individual components of the wave function. Here, we consider the matrices

$$\alpha_0 = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \qquad \alpha_i = \begin{bmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{bmatrix}, \tag{7.5}$$

where 0 and I are 2×2 0 and 1 matrices. The original matrix for α_0 was

$$\alpha_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{7.6}$$

The σ_i are named Pauli matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{7.7}$$

One can show that, for a spin 1/2 particle, the spin angular momentum operator $\vec{S}_i$ can be identified with the Pauli matrices $\vec{S}_i = (\hbar/2)\sigma_i$. As basis set, the eigenvectors of σ_3 are used, i.e. $\varsigma_\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (spin up) and $\varsigma_\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ (spin down). The electron spin is not the only physical observable that is contained in the Dirac equation. Since the matrices are 4×4, the according wave function is a four-component vector or "spinor" $\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$ written as a column matrix here.

In the non-relativistic limit, where the energy is approximately given by $E \sim p^2/2m \ll pc$, the components of the spinor can be interpreted. The first two components, ψ_1 and ψ_2, describe spin up and down, respectively, for the positive energy solution (electron) and ψ_3, ψ_4 describe the spin states for the negative energy solution. The latter troubled Dirac initially because of the apparent unphysical result. Five years after Dirac published his equation, however, Anderson communicated the finding of a particle (in a cloud chamber experiment) with the mass of an electron but with positive charge, the positron [11]. It possessed the properties

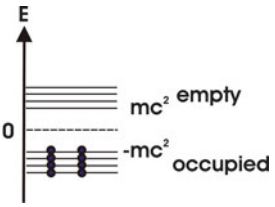


Fig. 7.7 Dirac sea of negative energy states (continuum) and empty positive energy states (see text)

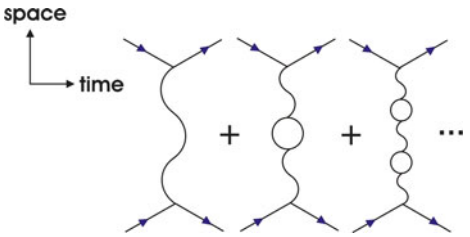


Fig. 7.8 Feynman diagrams for scattering of two electrons; (*straight full black lines*) electrons; (*wavy black lines*) photons (electromagnetic field between the electrons that makes them repel each other); (*full line loops*) photon creating an electron and a positron that annihilate each other, creating a photon in this virtual process

of the "hole" states of Dirac's "negative energy sea" (see Fig. 7.7), thus solving the mystery of the negative energy solutions.

Mathematically, the solutions of the Dirac (and Klein–Gordon) equations have instabilities that describe the generation and annihilation of virtual particles from empty space. Such behaviour was later described by Feynman's formalism using the graphical form of diagrams [12, 13] in the framework of quantum electrodynamics. Figure 7.8 shows a set of Feynman diagrams (in the typical space–time coordinate system) for the scattering of two electrons where the virtual particle generation and annihilation is represented by the respective loops that connect to photons.

After it could be shown that quantum electrodynamics was a renormalizable theory where infinities are avoided, it was considered an established theory in the late 1940s. Three decades later, in the theoretical description of the strong nuclear force, (holding protons and neutrons in the atomic core) so-called dual models [14] led to the renormalizable quantum chromodynamics [15]. Around 1970, it was understood that dual models were theories describing the quantum behaviour of relativistic vibrating strings, leading to the labelling "string theories". After establishing the electroweak theory (uniting weak nuclear forces and electromagnetism), only gravity was left as a force that could not be described by a renormalized field theory. A considerable problem of such theory was the consequences of the value for the spin $J = 2$ of the graviton, the quantum particle of the gravitational force field: the loop integral for a particle of spin J in D-dimensional space is approximated by

$$I_{\text{loop}} \sim \int p^{4J-8} \mathrm{d}^D p, \tag{7.8}$$

(p, momentum). For $4J + D–8 \geq 0$, the integral assumes infinite values. Hence, for the gravitational field, with $J = 0$ and $D = 4$, the loop integral would become infinite with the fourth power of **p**. String theory, however, originally introduced for the description of the strong interaction, contains a mass-less particle with spin $J = 2$. Therefore, string theory was considered an advantageous alternative theory of quantum gravity, in particular as it avoids the zero-distance, infinite-momentum problems of integrals of particle loops (Fig. 7.8). In string theory, distance and momentum are roughly related by $\Delta L \sim (\hbar/p) + \alpha'(p/\hbar)$, α' being the string tension parameter, T, with $T = 1/2\pi\alpha'$. The expected minimum length scale is in the order of the Planck length L_p (gravitational constant G_N)

$$L_\mathrm{p} = \sqrt{\frac{\hbar G_\mathrm{N}}{c^3}} = 1.6 \times 10^{-33}\,\mathrm{cm}. \tag{7.9}$$

More recently, unification of the different existing string theories has been attempted by introduction of the M-theory [15].

Another unexpected feature of the Dirac equation is the so-called *Zitterbewegung* due to interference between the positive and negative energy components of the wave packet. It should result in ultrahigh frequency oscillations ($\sim 10^{21}$ Hz) of the particle over distances in the order of a Compton wavelength for electrons [16]:

$$\lambda_\mathrm{COMPT} = \frac{\hbar}{m_\mathrm{e} c} \sim 3.9 \times 10^{-11}\,\mathrm{cm}. \tag{7.10}$$

Although this Zitterbewegung has not yet been observed directly, an analogue of it has recently been studied in semiconductor quantum wells where the spin–orbit coupling of the electronic wave packets is larger than in vacuo [17].

7.4 Semiconductor Spintronics

7.4.1 Hybrid Ferromagnetic Metal–Semiconductor Structures

Whereas metallic spin devices provide new ways to store and read information in hard discs, tapes or MRAMs, semiconductor-based spintronics could combine storage, detection, logic and communication capabilities on a chip. Such multifunctional devices could replace several hitherto separated components. Examples include the improved integration of MTJs with Si-based electronics and the transformation of magnetic into optical signals. Various further applications have been envisaged [18].

Currently, two basic approaches are pursued in semiconductor spintronics: (1) the development of hybrid structures where ferromagnetic spin injectors and detectors are combined with semiconductors and (2) the development of devices and structures that are exclusively using semiconducting materials. We will treat these

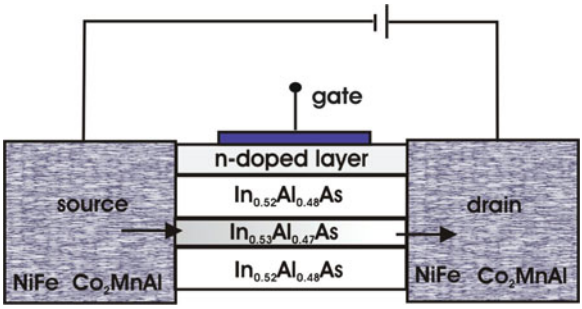


Fig. 7.9 Spin field effect transistor with a semiconductor hetero-structure consisting of, for instance, an asymmetric single quantum well with InAlAs as shown and ferromagnetic spin-injecting and -detecting contacts; L_D (not shown in the figure) is the distance between the edges of source and drain that contact the conductance channel (see text)

approaches separately, beginning with hybrid structures, where the proposition of an electronic analogue of the electro-optical modulator, introduced shortly after the discovery of the GMR, has found widespread attention. This so-called spin field effect transistor (coined spin-FET) design [7] is based on a HEMT (high electron mobility transistor) structure where highly doped source/drain regions are replaced by ferromagnetic contacts. At the Fermi level of these contacts, the concentration of one type of spin is largely exceeding the other, allowing to preferably inject (and detect at the drain) one type of spin polarization. The drain current depends on the gate-voltage-controlled coherent spin of electrons that reach the contact with respect to the drain magnetic moment. Although later somewhat criticized [19], the structure represents an excellent example for describing how spin might be controlled by electrical fields although the device has not yet been realized. A schematic of a spin-FET, based on a semiconductor heterostructure, is shown in Fig. 7.9. The original device was thought to operate in the ballistic regime where the device channel length L_D is smaller than the spin scattering length L_{ss}.

For description of the operational principle of the device, first, the spin–orbit coupling and the effect of a built-in and of an externally applied electric field on this coupling is considered. Electrons, orbiting around nuclei, produce a magnetic field that alters their magnetic moment. This interaction is known as spin–orbit coupling. The Hamilton operator, H_{S-O}, describing spin–orbit coupling, is derived from the Dirac equation (7.4) and has the form

$$H_{S-O} = \frac{\hbar^2}{(2mc)^2}(\nabla V \times \boldsymbol{p})\varsigma, \tag{7.11}$$

∇V is the gradient of the potential energy V that results from the electric field, induced by the hetero-structure formed, which acts on the electrons; $\boldsymbol{p}$ the momentum operator and ς represents a vector of the Pauli spin matrices (compare (7.6)) and the according text).

The wave function of a spin-polarized electron, injected into the 2DEG (two-dimensional electron gas) conduction channel of the device (see *arrows* in Fig. 7.9), can be described as superposition of two spin states (7.6ff) that have slightly different quantum mechanical momenta $\hbar k$. This difference results from the different influence of the electrical field (that becomes a magnetic field, the Rashba field, in the rest frame of the electrons) on the two spin states in the 2DEG, resulting in slightly different free-electron parabolae as shown in Fig. 7.10. Upon coherent movement of the two spin states along the conduction channel, the relative phase shift ξ is given by the electron wave vector difference and the length of the channel,

$$\xi = \Delta k L_{\mathrm{D}} = 2 m_{\mathrm{e}}^{*} \alpha L_{\mathrm{D}} / \hbar^{2}, \tag{7.12}$$

where m_{e}^{*} is the electron effective mass and α the Rashba parameter. The dispersion relation for parabolic bands is given by

$$E^{\uparrow\downarrow}(k_{\mathrm{II}}) = E_0 + \frac{\hbar^2 k_{\mathrm{II}}^2}{2 m_{\mathrm{e}}^{*}} \pm \alpha |k_{\mathrm{II}}|, \tag{7.13}$$

here, k_{II} is the electron wave vector in the 2D plane, E_0 the edge of the ground sub-band and α the parameter for Rashba spin–orbit splitting; the sub-band splitting depends on the size of α where the spin up ($\uparrow$) electrons populate the energetically higher lying sub-band as can be seen from (7.13). The phase shift can be expressed by considering the single-spin components in (7.13) where the electrons, travelling in the x-direction, experience a magnetic field perpendicular to their motion as indicated in Fig. 7.10.

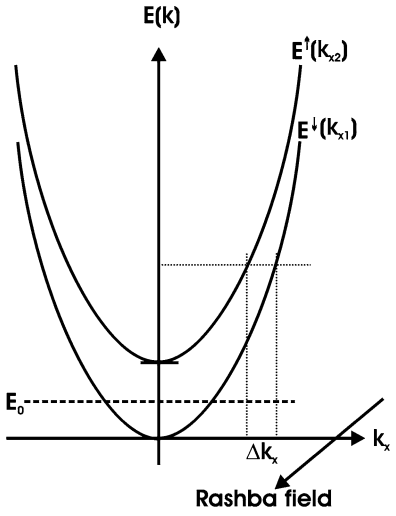


Fig. 7.10 Simplified schematic for momentum phase shift induced by Rashba field sub-band splitting; the 2DEG is assumed to reside in the x–y plane (see text)

$$E^{\uparrow}(k_{x2}) = E_0 + \frac{\hbar^2 k_{x2}^2}{2m_e^*} + \alpha k_{x2},$$
$$E^{\downarrow}(k_{x1}) = E_0 + \frac{\hbar^2 k_{x1}^2}{2m_e^*} - \alpha k_{x1}. \tag{7.14}$$

For equal energy, the phase shift is then $k_{x2} - k_{x1} = 2m_e^*\alpha/\hbar^2$, resulting in the phase shift given in (7.12). As a result, the spin of the electron, moving ballistically along the conduction channel, precesses and can have a distinctively different orientation near the end of that channel than upon injection. The gate voltage, in addition to the electric field of the hetero-junction, can alter this spin orientation and also influence the Rashba parameter via gate voltage-induced sub-bandgap splitting: from Shubnikov–de Haas beating patterns [20], evidence for such gate voltage-induced influence on the Rashba parameter

$$\alpha = \frac{\sqrt{\pi}\hbar^2}{m_e^*}(n^{\uparrow} - n^{\downarrow})(2N_e - 2(n^{\uparrow} - n^{\downarrow}))^{-1/2}, \tag{7.15}$$

(N_e, total carrier density, typically $\sim 10^{12}\,\mathrm{cm}^{-2}$, n, spin-polarized electron concentration in the 2DEG) has been reported. It was found that a rather small variation of the front gate voltage ($\sim$0.15 V) can change α by a factor of 2, demonstrating the possibility of tuning the spin splitting without influencing the charge in the 2DEG channel. Other experiments have shown, however, that the field contribution to the spin–orbit splitting can be smaller than the effect of penetration of the wave function perpendicular to the quantum well [21]. The positioning of the wave function perpendicular to the 2D channel is thus the third parameter that can be used to alter α.

The search for materials combination of ferromagnetic injector and spin-transporting semiconductor has rather recently resulted in spin injection with a degree of polarization of 30% [22]. These systems, however, are characterized by two disadvantages (1) the large conductivity differences between metal and semiconductor demand separation by a highly resistive tunnel junction, except for spin injectors with a polarization degree of close to 100%. It will be difficult to achieve such high polarization degrees because the bulk magnetic properties typically deviate substantially from the according interfacial ones; (2) the spin relaxation during transport can destroy the spin orientation due to a too large distance to the metallic spin injector.

7.4.2 Ferromagnetic Semiconductors

Due to the above-described limitations, the use of ferromagnetic semiconductors has found considerable attention. Advantages include a rather easy integration into semiconductor devices, the possibility to adjust the ferromagnetism by the charge

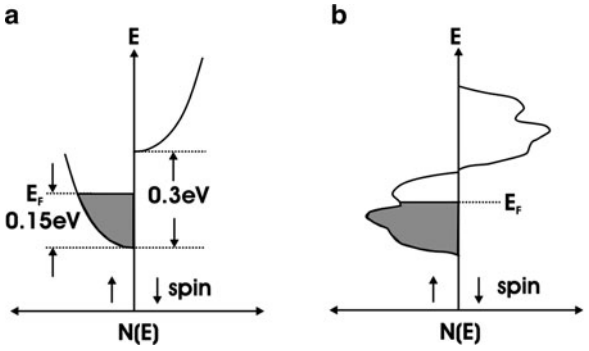


Fig. 7.11 Schematic representation of the density of states $N(E)$ for magnetic semiconductors below the Curie temperature; (**a**) Mn-doped EuS with a Zeeman splitting of $\sim$0.3 eV and with $E_F \sim 0.15$ eV (for 10^{20} cm^{-3} free carrier concentration); (**b**) $N(E)$ for the half-metallic ferromagnet CrO_2; the energy scale in (**b**) is about a factor of 10 larger than in (**a**)

carrier density and the absence of the already mentioned problems with metallic injectors. The demands on the material properties for semiconductor spintronics include a Curie temperature above room temperature, incorporating n- as well as p-type doping, a Zeeman splitting of the conduction- (valence-) band that exceeds the Fermi energy of the electrons (holes) (Fig. 7.11a) and as high spin polarization as possible at room temperature at E_F. The latter occurs in so-called half metallic materials where the Fermi level intersects only one of two spin bands whereas E_F lies in a band gap for the other band (Fig. 7.11b).

The selection of promising spin injection materials for semiconductor spintronics is based on two main criteria (1) the ferromagnetic state of the materials should be preserved at temperatures at or above 300 K; (2) the fabrication technology for the respective material class should already exist. Past work has focused on (Ga,Mn)As and (In,Mn)As where markets exist for their host materials in infrared LEDs and lasers as well as in high-speed digital electronics (GaAs) and magnetic sensors (InAs). The hitherto achieved highest Curie temperatures reported for these materials of 173 K, however, are still too low for most practical applications (see below).

The search for suited semiconductor materials is partly characterized by a rational screening approach, also often used in the development of new pharmaceuticals where, based on experience and intuition, materials are developed and investigated with regard to their magnetic properties and, in particular, with regard to a Curie temperature T_C at or above room temperature and a large carrier spin polarization. Examples encompass a wide range of material classes. Among them are ferromagnetic oxides such as CrO_2 (see Fig. 7.11b) and Fe_2O_3, the latter being characterized by a very high degree of spin polarization at room temperature. Mn-doped II–IV–V_2 ternary chalcopyrites such as $CdGeAs_2$ and $CdGeP_2$ show T_C at about 350 K. The Mn content can reach 20% in $CdGeP_2$ but the materials can exhibit magnetic inhomogeneity, i.e. the simultaneous presence of ferromagnetic and paramagnetic-type phases. The dilute magnetic semiconductor class $Ga_{1-x}Mn_xAs$

Table 7.2 Overview of semiconductors that exhibit room temperature ferromagnetism

$Cd_{1-x}Mn_xGeP_2$	1.7 eV	Mn evaporation	>300 K
(Ga,Mn)N	3.4 eV	MBE Mn incorporation	>300 K
(Ga,Cr)N	3.4 eV	MBE Cr incorporation	>400 K
(Ga,Mn)P:C	2.2 eV	Mn incorporation by MBE or implantation	>330 *K*
$(Zn_{1-x}Mn_x)GeP_2$	1.8–2.8 eV	Sealed ampoule growth	312 K
$ZnSiGeN_2$	3.5 eV	Epilayers on Al_2O_3	∼300 K
$ZnSnAs_2$	0.65 eV	Bridgman growth	329 K
ZnO:Co	3.1–3.6 eV	PLD Co incorporation	>300 K
TiO_2 : Co	3.2 eV	PLD Co incorporation	>380 K
CrO_2	∼2 eV	CVD on TiO_2	∼390 K
$\varepsilon - Fe_2O_3$	n.a.(2 eV)	Sol–gel (ferrimagn.)	>490 K

MBE molecular beam epitaxy, *PLD* physical vapour deposition, epilayers: epitaxially grown material; *CVD* chemical vapour deposition, *na* not available; for spray pyrolysis films, where the phase of Fe_2O_3 has not been identified, the energy gap is 2.05 eV

has been extensively investigated showing ferromagnetism, however, below room temperature only. Table 7.2 gives an overview of the semiconductors that show $T_C > 300$ K including their energy gap and preparation methods.

Room temperature ferromagnetism was also observed in laser-ablated or PVD-prepared Co-doped TiO_2 (see Table 7.2). As this material is transparent in the visible spectral range, having an energy gap around 3.2 eV, it could be relevant in optoelectronic applications. Epitaxial layers of CrSb/GaAs/CrSb show also room temperature ferromagnetism (not shown in the above table) and $Mn_{11}Ge_8$ crystals have T_C near room temperature. The so-called Heusler alloys [23], made of metals which have their spins aligned at room temperature, (NiMnSb with $T_C = 728$ K) are not considered here in the context of *semiconductor* spintronics.

The coherent spin transport through semiconductors and across interfaces is essential for device development where spin coherence times and mobilities define device quality. Figure 7.12 shows an example of interlayer spin transport under externally applied bias at the ZnSe/GaAs hetero-interface. The junctions are made of n-GaAs substrates and n-ZnSe epitaxial layers, grown by MBE with, for example, 100-nm thickness. A 100-fs laser pulse with circular polarized photons and 1.5-eV photon energy produces spin-polarized conduction band electrons selectively in the GaAs substrate. This process is described in Sect. 7.4.3 in more detail. The spin transport is measured at a delay time Δt in ZnSe using a linearly polarized photon probe pulse with $h\nu = 2.8$ eV, i.e. above the band gap energy of ZnSe. Due to the Kerr effect, its polarization rotates proportional to the spin magnetization. It has been shown that the spin signature of materials can be obtained by sweeping a magnetic field $\boldsymbol{B}$ at a fixed pump–probe delay [24]. Without bias, during delay times of $\Delta t = 100$–400 ps, the spontaneous spin transfer from GaAs to ZnSe is in the range of a few percent of the spins present in GaAs. The according signature in the magnetic field response experiment is virtually identical to that of the ZnSe as can be seen in Fig. 7.12b. This so-called two-colour pump–probe measurement [25] where the magnetic field response S of transferred spins has been measured in resonance

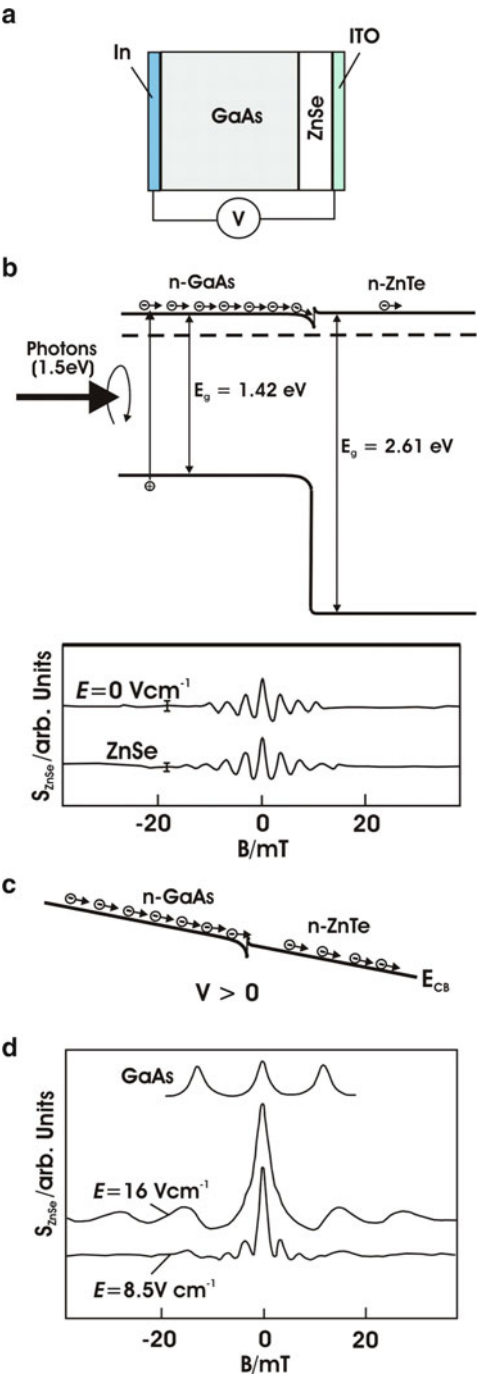


Fig. 7.12 Schematic of bias-induced spin transfer at a GaAs/ZnSe hetero-junction; (**a**) hetero-junction arrangement and energy band scheme $E(x)$ including optical excitation of spin-polarized conduction band electrons; (**b**) magnetic response signature of ZnSe and of the hetero-junction without external bias ($\boldsymbol{E}$= 0); (**c**) energy schematic of the influence of a forward bias voltage on the conduction band spin reservoir in GaAs; (**d**) electric field-induced control of the magnetic response of the ZnSe layer during spin injection (see text); $T = 5\,\mathrm{K}$

scans when the applied magnetic field **B** is swept at a fixed delay $\Delta t = -10\,\mathrm{s}$ shows in the bottom signal that the signature of the ZnSe spins is almost unaffected by the few spins from the GaAs substrate (which acts as spin reservoir due to the long spin lifetime in this material) that crossed the interfacial barrier (compare also Fig. 7.12a). Figure 7.12c shows an energy scheme of the influence of the applied bias on the spin transport across the spike in the conduction band where the forward bias applied to the spin reservoir (see Fig. 7.12a) facilitates transport.

The resulting magnetic response is shown in Fig. 7.12d and compared to that of GaAs. The response of the spins injected into ZnSe to the applied magnetic field ***B*** changes from that of ZnSe towards that of GaAs. With increased bias, the periodicity of the signature resembles the signal of GaAs, showing that one can control the magnetic response of the ZnSe layer by an electric field during spin injection. Thus, sustained spin currents from the GaAs reservoir can be generated by biasing the hetero-junction allowing, in principle, extended functionality of magneto-electronic devices.

Theoretical analyses have shown that the mobility in low electric fields and the diffusion of spin packets depend sensitively on the doping of the semiconductor. The introduction of ferromagnetic semiconductors that are compatible with III–V semiconductor epitaxy is presently viewed as promising contribution in this direction. Such semiconductor devices provide means to manipulate their magnetic properties by external factors such as doping and gating, for instance, due to the carrier-related nature of the ferromagnetism in these materials [26]. As already mentioned above, the class of ferromagnetic semiconductors based on the elements (Ga,Mn)As provides promising candidates for coherent spin transport because the good lattice match with GaAs and AlAs allows preparation of high-quality magnetization-modulated (Ga,Mn)As/III–V hetero-structures for potential spintronic applications. Advances in MBE growth processes and improvement of the growth conditions have resulted in an increase of the Curie temperature to now 173 K [27]. The construction of functional spintronic devices as that shown above in Fig. 7.12 depends on the realization of efficient spin injection from the ferromagnetic regions into the paramagnetic ones. Also, the tunnelling magneto-resistance effect (TMR) [28] allows applications in, for instance, RAMs and magnetic field sensors. Both aspects have been recently realized in structures that are based on the (Ga,Mn)As family, where about 80% spin-polarized electrons from a p-(Ga,Mn)As/n-GaAs Zener diode and a TMR effect larger than 200% has been reported on (Ga,Mn)As/AlAs/(Ga,Mn)As triple layer structures [29].

7.4.3 Spin-Photonics

The spin generation in the GaAs reservoir as described above and schematically shown in Fig. 7.12a has occurred by the use of polarized light. In this subsection, the interaction of photons with electron spins, optical spin generation and resulting devices will be treated in somewhat more detail. The field of *spin-photonics* aims at using photons for the storage and processing of information in magnetic media.

Photons can be used, for instance, to pattern magnetic media as will be shown below and also for the study of fundamental properties of the interaction of light with magnetic materials. The industrial goal of using photons for writing and erasing elementary bits of information in magnetic media presents several advantages that are specific to lasers or synchrotron radiation (1) an adjustable wavelength, (2) short pulses are available in ultrafast optics and (3) the use of the $\pm h$ spin momentum of photons, carried by circularly polarized light which has been used above (Fig. 7.12a). Particularly promising is the extrapolation of the present all-magnetic technology to frequencies well above the GigaHertz range by use of femtosecond laser pulses. In addition, applying linearly or circularly polarized light allows selection of specific optical transitions in semiconductors and can induce all-optical switching in ferrimagnetic materials where opposing magnetic moments are unequal and result in spontaneous magnetization. Although at the present time, *spin-photonics* is an emerging field that is still far from outperforming the well-established all-magnetic technology which is used, for example, in the commercially available computer hard drives of a few hundred gigabytes rotating at a few thousands rd/minutes, it has nevertheless led already to several promising achievements in the recent past. Realized devices comprise spin LEDs [30] that emit circularly polarized light, semiconducting random access memories [31, 32] using a ferromagnetic (Ga,Mn)As nanostructure and spin lasers [33] where the selective coupling of spin-polarized electrons with photons of defined helicity results in a reduction of the threshold laser current, for example.

The interaction of the polarization state of photons and an electron system in an external magnetic field is known from the Zeeman effect [34]. The quantum numbers for electrons in an atom are n, l, m and s which refer to the main-, the orbital-, the magnetic- and the spin quantum number, respectively. In Russell–Saunders coupling, the individual quantum numbers l and s can be added separately due to rather weak individual spin–orbit interactions. The total angular momentum can then be described by $J = L + S$ and not by the individual quantum numbers l_1, $l_2 \ldots$ and s_1, $s_2 \ldots$, thus describing the system in a single-particle approach. In a magnetic field, the splitting of the valence electron levels is given by $\Delta E = \mu_\mathrm{m} \boldsymbol{B}$, where μ_m is the total magnetic moment of the atom which depends on J, L and S. Introducing the magnetic quantum number M_J that depends on the total angular momentum J (with $M_\mathrm{J} = -J, \ldots, 0, \ldots J$), the energetic splitting is given by the product of M_J with the Bohr magnetron μ_B, the Landé factor g and the magnetic field, neglecting the hyperfine interaction which would include the magnetic moment of the nucleus:

$$\Delta E = M_\mathrm{J} \mu_\mathrm{B} g \boldsymbol{B}, \tag{7.16}$$

where the Landé factor expresses the spin interaction in the Hamiltonian

$$H = H_0 + (\boldsymbol{L} + \mathrm{g_s} \boldsymbol{S}) \boldsymbol{B}. \tag{7.17}$$

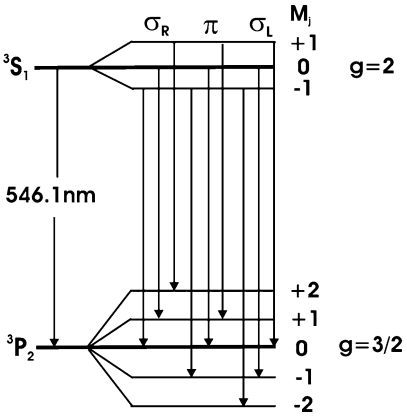


Fig. 7.13 Term scheme for the splitting of the Hg line with $\lambda = 546.1$ nm in the presence of a magnetic field; M_J, magnetic quantum number, σ_R, σ_L and π denote transitions with $\Delta M_J = +1$, –1 and 0, respectively and indicate left- and right-hand photon polarization; the spectral notation of the levels is $^{2S+1}X_J$ where $X = S, P, D\ldots$ denotes $L = 0, 1, 2,$ etc., hence $^3P_2$ is a state with $S = 1$, $L = 1$, $J = 2$

With $g_s \approx 2$; the Landé factor g and M_J depend on the particular state in Russell–Saunders coupling. g is given by

$$g = 1 + \frac{J(J+1) + L(L+1) - S(S+1)}{2J(J+1)}. \tag{7.18}$$

We now consider the term splitting of the mercury line at 546.1 nm in the presence of an external magnetic field as shown in Fig. 7.13. The fundamental transition is from a $^3S_1$ ($J = 1$, $L = 0$, $S = 1$, spin-only) to a $^3P_2$ ($J = 2$, $L = 1$, $S = 1$, mixed spin–orbital momentum) state as indicated in the figure. The term multiplicity in the magnetic field is $2J + 1$, resulting in three levels for the $^3S_1$ term and in five levels for $^3P_2$. The optical transition selection rules for electric dipole transitions are $\Delta l = \pm 1$, $\Delta L = 0, \pm 1$, $\Delta S = 0$, $\Delta J = 0, \pm 1$ and $\Delta M_J = 0, \pm 1$. The light that is emitted in the three triplets is linearly polarized for $\Delta M_J = 0$ transitions, with polarization along the magnetic field, and circularly polarized for $\Delta M_J = \pm 1$ transitions. The photon is a particle with spin 1 with spin components of $\pm m\hbar$. A particle with $m = +1(-1)$ corresponds to right (left) circularly polarized light. In optical excitation transitions, circularly polarized light can therefore excite electrons from one spin state to another one because of spin transfer to the electron system. This is used in modern applications of semiconductor spintronics and has been already used in the experiment shown in Fig. 7.12.

Spin-polarized electrons are injected in the conduction bands of semiconductors by the absorption of circularly polarized light. This process, known as *optical orientation*, provides a powerful tool in the field of spintronics, where it is used to generate a spin-polarized electron density in semiconductors. The principle is shown in Fig. 7.14.

The degree of spin polarization after excitation is determined by the expression

$$P(t) = \frac{I_\uparrow(t) - I_\downarrow(t)}{I_\uparrow(t) + I_\downarrow(t)}. \tag{7.19}$$

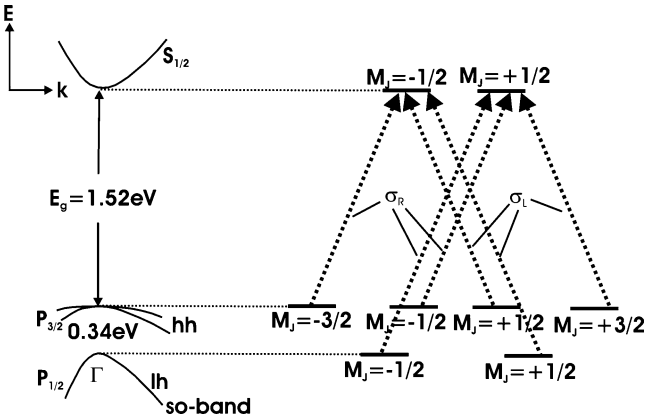


Fig. 7.14 (*Left*): $E(\mathbf{k})$ diagram of the GaAs band structure near the Γ-point at $k = 0$ with the energy gap of 1.52 eV at low temperature ($E_g = 1.42$ eV at room temperature) and a spin–orbit splitting of 0.34 eV; the *right-hand side* shows optical excitation from the sub-bands using left- and right-handed circularly polarized light that generates different electron spin populations in the conduction band; lh and ll refer to light and heavy holes, respectively, defining the according effective mass $m^* \propto 1/(\mathrm{d}^2E/\mathrm{d}k^2)$ which is inversely proportional to the curvature of the $E(\mathbf{k})$ band structure. The effective masses of the light and heavy holes can be described by the Kohn–Luttinger parameters γ_1 and γ_2 using $m^*_{\mathrm{hh}} = (\gamma_1 – 2\gamma_2)^{-1} \ldots m_0$ and $m^*_{\mathrm{lh}} = (\gamma_1 + 2\gamma_2)^{-1} \ldots m_0$ (m_0, electron mass at rest); the spin-off band (so-band) is related to the spin–orbit splitting of the Γ_7 band from the Γ_8 valence band. Below, in case of the spin laser, it is shown that quantum confinement in a quantum well further lifts the degeneracy between the hh and lh valence bands at the band edge

A schematic example is shown in Fig. 7.15. Typically, the initial spin polarization is taken for evaluation as is shown in Fig. 7.16 for a GaAs/AlAs multiple quantum well grown on a GaAs substrate by MBE with a thickness of 15.2 nm.

In these two-dimensional quantum systems, the degeneracy of light and heavy holes at the Γ-point is lifted for the upper valence band (compare Fig. 7.14, left-hand side, where the degeneracy exists for the three-dimensional case) with the light holes lying energetically $\sim$ 5–10 meV below the valence band edge at Γ_8, depending on the actual quantum well dimensions. It can be seen that the spin polarization changes with photon energy from a high positive value (> 0.7) to $\sim$ –0.3($h\nu = 1.54$ eV). The optical transition associated with the change in spin polarization is due to excitation from a light hole state into the conduction band. The observed maxima in P(t) are related to various heavy hole–conduction band transitions [35].

Applications that have already been demonstrated include the spin laser and spin LEDs. They will be shortly reviewed here. A schematic of the spin laser structural arrangement is shown in Fig. 7.17. The laser consists of a top and a bottom $\lambda/4$ (so-called) Bragg reflector which serve as laser mirrors. The Bragg reflectors on top and at the bottom confine a $3/2\lambda$ optical wave. Two InGaAs quantum wells in the optical antinodes of the laser cavity provide the optical gain medium. Such a structure is called a vertical cavity surface-emitting laser (VCSEL), where, in contrast to typical

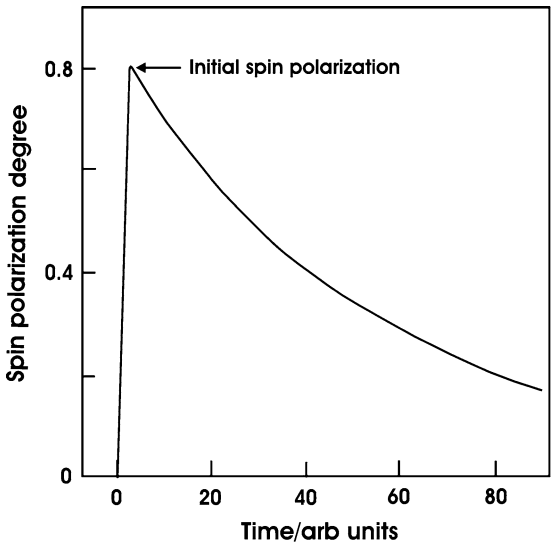


Fig. 7.15 Spin polarization vs. time behaviour after optical spin excitation using circularly polarized light with the value for the initial spin polarization as indicated

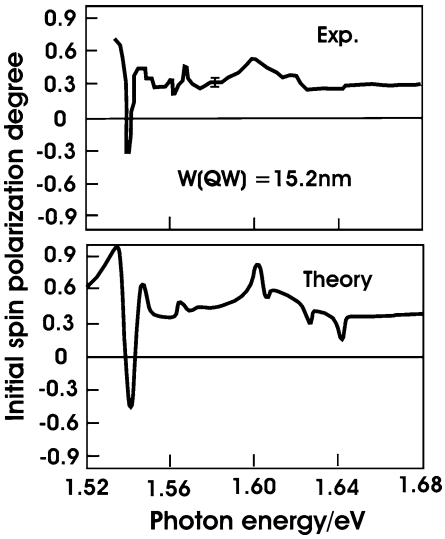


Fig. 7.16 Comparison of experimentally determined and calculated spin polarization degree of a GaAs quantum well (QW) structure with 12 QWs of thickness of 15.2 nm; excitation: Ti:sapphire laser, $T = 4.2$ K; *upper part* shows experimental result (exp) and the *lower part* the calculated response

so-called edge emitting lasers, the laser light is emitted perpendicular to the plane of the active region as indicated in Fig. 7.17. Optical electron and hole excitation ($\lambda = 780$ nm) is done in the growth direction by fs pulses of a Ti:Sapphire laser. Both, the surrounding GaAs barriers and the $Ga_xIn_{1-x}As(x = 0.96)$ quantum wells are optically excited with light of photon energy 1.59 eV (780-nm wavelength) where the reflectivity of the Bragg reflectors has its minimum. This photon energy is both, above the band gap and below the transition from the split-off band (shown in the left side of Fig. 7.14) to the conduction band minimum.

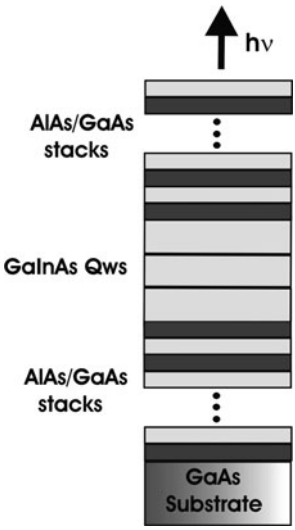


Fig. 7.17 Schematic for the structural arrangement of a VCSE laser, consisting of two GaInAs quantum wells (QWs) with 8 nm thickness

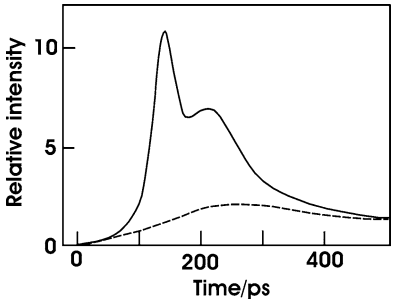


Fig. 7.18 Transient laser intensities for a GaInAs-based VCSE laser after generation of aligned spins (*full line*) and of random spins (*dashed line*)

After injection of spin-aligned electrons by a pump laser with circularly polarized light, coherent radiation of $\lambda = 835$ nm is emitted about 150 ps after injection. In Fig. 7.18, the transient laser intensity after injection is shown for two cases (1) for excitation with linearly polarized light that results in random spin orientation and (2) for excitation with circularly polarized light. The pump power of 6.5 mW was chosen to yield a pump current of 650 mA cm^{-2} which is slightly below the laser threshold for random spin alignment. The increase in emitted light intensity in the latter case is attributed to the fact that the spin-polarized electrons contribute mostly to one circularly polarized lasing mode whereas for random electron spins, more modes are supported.

For aligned spins, the conversion into a single lasing mode drives the laser above the lasing threshold. The laser threshold current, which defines the minimum power consumption needed to achieve lasing, is shown in Fig. 7.19 for the two cases. It is determined from plots of emission power vs. pump power. For aligned spins, this threshold is 0.5 A cm^{-2} and for random spins it is at 0.65 A cm^{-2} demonstrating a

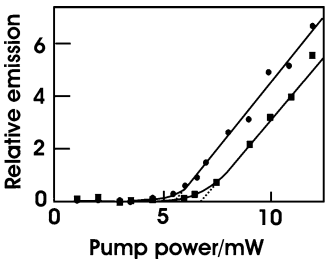


Fig. 7.19 Laser threshold for the VCSE laser of Fig. 7.17; (*circles*) lasing after injection of spin-polarized electrons; (*squares*) lasing for random electron spins (see text)

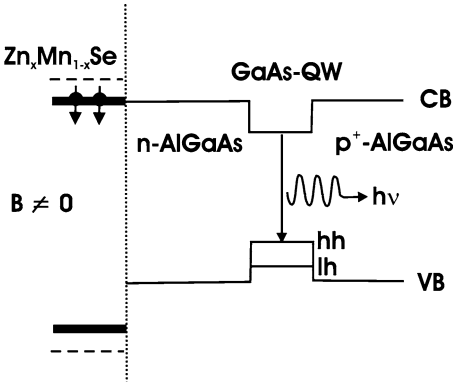


Fig. 7.20 Energy vs. space diagram for a spin-LED based on a GaAs quantum well in flat-band situation; the band offsets are enlarged for intelligibility; in the GaAs QW, the light and heavy hole sub-band lifting of the degeneracy due to the reduced dimensionality is exaggerated (see text)

substantial decrease in the laser threshold current in the former case. In the above considerations, it has been tacitly assumed throughout that the hole states show fast spin relaxation and do not carry a spin polarization.

Whereas for lasing, optical injection of electron spins that results in emission of polarized photons has been described above with a light-to-electron and electron-to-light energy and momentum conversion, the according light-emitting diodes (spin-LEDs) typically use electronic spin injection and conversion to polarized light. An example that demonstrates the operation principle (compare also Fig. 7.12) is displayed in Fig. 7.20.

The spin injector material is a $Zn_{0..94}Mn_{0.06}Se$ semi-magnetic semiconductor (see, for instance, Table 7.2). In an applied magnetic field ($B \neq 0$ in the above figure), this material shows a large band edge spin splitting, also known as giant Zeeman effect, forming spin-polarized carrier populations as indicated in Fig. 7.20 at the conduction band edge. Representative electroluminescence spectra from the structure, obtained for left and right circularly polarized light, are shown for selected values of the applied magnetic field at 4.2 K. At $\boldsymbol{B} = 0$, no polarization is observed, as expected. The emission maximum occurs at a photon energy of 1.524 eV, which confirms that the radiative recombination occurs via the heavy hole excitonic states in the quantum well. Because ZnMnSe is a paramagnet that acquires no net magnetization at $\boldsymbol{B} = 0$, optical polarization is not observed. For increased magnetic field strength, a spin population is generated in ZnMnSe and injected into

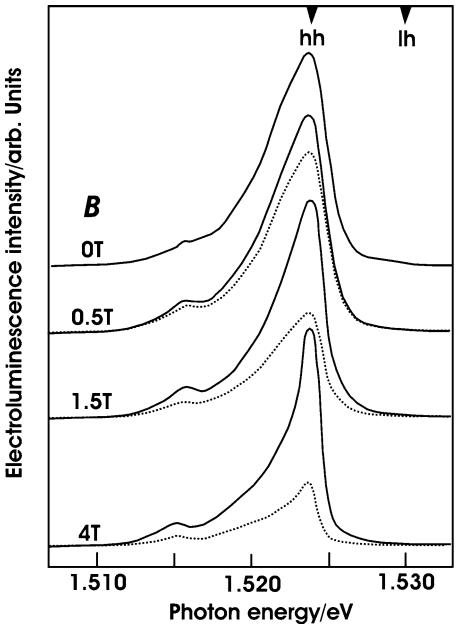


Fig. 7.21 Electroluminescence spectra for a spin-LED for different magnetic field strengths, recorded at 4.2 K; (*full lines*) left circular polarization σ_L; (*dotted line*) σ_R; the energetic position of the heavy and light holes is indicated

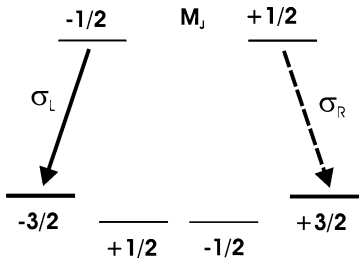


Fig. 7.22 Optical selection rules and occupancy in a GaAs quantum well spin-LED with the two polarizations according to the labelling in Fig. 7.21 (see text)

the LED structure. The corresponding spectra in Fig. 7.21 show a distinct increase in polarization as indicated by the differences between the full and the dotted lines. The shoulder at 1.515 eV is attributed to recombination in the GaAs bulk which indicates by its growing polarization degree that a part of the spin polarized electrons retain their spin until recombination in the GaAs buffer layer.

The missing signal from the light hole band is due to its energetic location below the heavy hole band ($\sim$5 meV for this structure). This split is larger than the thermal energy at 4.2 K, 0.36 meV, so that the light hole states which are at lower electron but higher hole energy are not occupied. Therefore, radiative transitions from the two-fold degenerate conduction with $M_J = \pm 1/2$ to the heavy hole band with $M_J = \pm 3/2$ occurs only. These transitions are schematically drawn in Fig. 7.22 [36].

References

1. W. Gerlach, O. Stern, *Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld*, Zeitschrift f. Physik **9**, 349–352 (1922)
2. P.A.M. Dirac, The quantum theory of the electron, Proc. Roy. Soc. A **117**, 610–624 (1928)
3. M.N. Baibich et al., Giant magnetoresistance of (0 0 1)Fe/(0 0 1)Cr magnetic superlattices, Phys. Rev. Lett. **61**, 2472–2475 (1988)
4. W.H. Meiklejohn, C.P. Bean, New magnetic anisotropy, Phys. Rev. **102**, 1413–1414 (1956)
5. J.A. Barnas, A. Fuss, R. Camley, P. Grünberg, W. Zinn, Novel magnetoresistance effect in layered magnetic structures: Theory and experiment, Phys. Rev. B **42**, 8110–8120 (1990)
6. S.S.P. Parkin, Origin of enhanced magnetoresistance of magnetic multilayers: Spin-dependent scattering from magnetic interface states, Phys. Rev. Lett. **71**, 1641–1644 (1993)
7. S. Datta, B. Das, Electronic analog of the electrooptic modulator, Appl. Phys. Lett. **56**, 665–667 (1990)
8. E.I. Rashba, Fiz. Tverd. Tela (Leningrad) **2**, 1224 (1960)
9. 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, Sov. Phys. Solid State **2**, 1109–1122 (1960)
10. N.W. Ashcroft, N.D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976)
11. C.D. Anderson, The positive electron, Phys. Rev. **43**, 491–494 (1933)
12. R.P. Feynman, Spacetime approach to non-relativistic quantum mechanics, Rev. Mod. Phys. **20**, 367–387 (1948)
13. R.P.Feynman, 1949 spacetime approach to quantum electrodynamics, Phys. Rev. **76**, 769–789 (1949)
14. Y. Nambu, in *Proceedings of the International Conference on Symmetries and Quark Models*, Wayne State University, Detroit, Michigan, USA, 1969, ed. by R. Chand (Gordon and Breach, New York, 1970), p. 269
15. E. Witten, String theory dynamics in various dimensions, Nucl. Phys. B **443**, 85–126 (1995)
16. G. David, J. Cserti, General theory of Zitterbewegung, Phys. Rev. B **81**, 121417(R)1–4 (2010)
17. J. Schliemann, D. Loss, R.M. Westervelt, Zitterbewegung of electrons and holes in III–V semiconductor quantum wells, Phys. Rev. B **73**, 085323 (2006)
18. I. Zutic, J. Fabian, S.C. Erwin, Bipolar spintronics: fundamentals and applications, IBM J. Res. Dev. **50**, 121–139 (2006)
19. S. Bandhyopadhyay, M. Cahay, Alternate spintronic analog of the electro-optic modulator, Appl. Phys. Lett. **85**, 1814–1816 (2004)
20. S.J. Papadakis, E.P. De Poortere, H.C. Manoharan, M. Shayegan, R. Winkler, The effect of spin splitting on the metallic behavior of a two-dimensional system, Science **283**, 2056–2058 (1999)
21. D. Grundler, Large Rashba splitting in InAs quantum wells due to electron wave function penetration into the barrier layers, Phys. Rev. Lett. **84**, 6074–6077 (2000)
22. X. Jiang, R. Wang, R.M. Shelby, R.M. McFarlane, S.R. Bank, J.S. Harris, S.S.P. Parkin, Highly spin-polarized room-temperature tunnel injector for semiconductor spintronics using MgO(1 0 0), Phys. Rev. Lett. **94**, 216601–4 (2005)
23. P.J. Webster, Heusler alloys, Contemp. Phys. **10**, 559–577 (1969)
24. J.M. Kikkawa, I.P. Smorchkowa, N. Damarth, D.D. Awschalom, Room-temperature spin memory in two-dimensional electron gases, Science **277**, 1284–1287 (1997)
25. I. Malajovich, J.J. Berry, N. Samarth, D.D. Awschalom, Persistent sourcing of coherent spins for multifunctional semiconductor spintronics Nature **411**, 770–772 (2001)
26. T. Dietl, H. Ohno, F. Matsukura, Hole-mediated ferromagnetism in tetrahedrally coordinated semiconductors, Phys. Rev. B **63**, 1952051–21 (2001)
27. K.W. Edmonds, K.Y. Wang, R.P. Campion et al., Hall effect and hole densities in $Ga_{1-x}Mn_xAs$, Appl. Phys. Lett. **81**, 3010–3012 (2002)

28. S.S.P. Parkin et al., Giant tunnelling magentoresistance at room temperature with MgO (1 0 0) tunnel barriers, Nature Mater. **3**, 862–867 (2004)
29. P. Sankowski, R. Oszwaldowski, P. Kacman, J.A. Majewski, T. Dietl, Coherent spin transport in magnetization modulated semiconductor heterostructures, Phys. Stat. Sol. (b) **244**, 2391–2398 (2007)
30. A.F. Isakovic, *Spin Transport in Ferromagnet-Semiconductor Heterostructures: Spintronics with Schottky Contact Spin-LEDs and Photodiodes: Materials, Spin Diodes, Experiments and Models* (Lambert Academic Publications, Koeln, 2010)
31. Y. Wu, Nano spintronics for data storage, Encyclop. Nanosci. Technol. **X**, 1–50 (2003)
32. T. Wosinski et al., Magnetoresistive memory in ferromagnetic (Ga, Mn)As nanostructures, Mater. Sci. Poland **26**, 1097–1104 (2008)
33. M. Holub, B.T. Jonker, Threshold current reduction in spin-polarized lasers: Role of strain and valence-band mixing, Phys. Rev. B **83**, 1253091–5 (2011)
34. P. Zeeman, Philos. Mag. **43**, 226 (1897)
35. S. Pfalz, R. Winkler, T. Nowitzki, D. Reuter, A.D. Wieck, D. Hägele, M. Oestreich, Optical orientation of electron spins in GaAs quantum wells, Phys. Rev. B **71**, 1653051–11 (2005)
36. B.T. Jonker, Y.D. Park, B.R. Bennett, H.D. Cheong, G. Kioseoglou, A. Petrou Robust electrical spin injection into a semiconductor heterostructure, Phys. Rev. B **62**, 8180–8183 (2000)

Index

H.-J. Lewerenz, *Photons in Natural and Life Sciences*, Springer Series in Optical Sciences 157, DOI 10.1007/978-3-642-23749-2,