Henry H. Radamson
Anders Hallén
Ilya Sychugov
Alexander Azarov

Analytical Methods and Instruments for Micro- and Nanomaterials

Springer

Lecture Notes in Nanoscale Science and Technology

Volume 23

Lecture Notes in Nanoscale Science and Technology (LNNST) aims to report latest developments in nanoscale science and technology research and teaching – quickly, informally and at a high level. Through publication, LNNST commits to serve the open communication of scientific and technological advances in the creation and use of objects at the nanometer scale, crossing the boundaries of physics, materials science, biology, chemistry, and engineering. Certainly, while historically the mysteries in each of the sciences have been very different, they have all required a relentless step-by-step pursuit to uncover the answer to a challenging scientific question, but recently many of the answers have brought questions that lie at the boundaries between the life sciences and the physical sciences and between what is fundamental and what is application. This is no accident since recent research in the physical and life sciences have each independently cut a path to the edge of their disciplines. As both paths intersect one may ask if transport of material in a cell is biology or is it physics? This intersection of curiosity makes us realize that nanoscience and technology crosses many if not all disciplines. It is this market that the proposed series of lecture notes targets.

Henry H. Radamson · Anders Hallén ·
Ilya Sychugov · Alexander Azarov

Analytical Methods and Instruments for Micro- and Nanomaterials

Henry H. Radamson
Guangdong Greater Bay Area Institute
of Integrated Circuit and System
R&D Center of Optoelectronic Hybrid IC
Guangzhou, China

Institute of Microelectronics
University of Chinese Academy
of Sciences,
Beijing, China

Mid Sweden University
Sundsvall, Sweden

Ilya Sychugov
Department of Applied Physics
KTH—Royal Institute of Technology
Stockholm, Sweden

Anders Hallén
School of Electrical Engineering
and Computer Science
KTH—Royal Institute of Technology
Stockholm, Sweden

Alexander Azarov
Centre for Materials Science
and Nanotechnology
University of Oslo
Oslo, Norway

© Springer Nature Switzerland AG 2023

Preface

The continuous development of semiconductors is resulting in more complicated device structures with an increase of device density in our integrated chips (ICs). We pace towards more advanced technology where complicated multi-layer structures for lasers and detectors are processed, driving transistors in new hybrid designs. Both electronics and photonics are being merged together in a single chip with a 3D configuration. In order to fulfil the industry's demands in this path, the state-of-the-art material characterization has to be implemented to ensure the material quality. As examples, for transistors, new gate oxide material, channel and source/strain engineering, and for lasers and detectors, new types of quantum wells with special bandgap alignment are demanded. The performance of these devices is strongly dependent on their interfacial quality, defect density and surface morphology. Therefore, the controlled processes on an atomic level and characterization of the thin layers have become increasingly crucial for the semiconductor industry.

This book presents the basics of most common analysis methods for characterizing the materials and devices. The text is prepared for the scholars, lab engineers and other interested persons with some background in physics and semiconductor materials. The aim of writing this book is to provide a good knowledge about analysis methods but also highlighting the difficulties, benefits and drawbacks for an applied method. The chapters begin with the history of the technical development, describes the set-up, underlying theory and applications and, finally, give reference to several research examples. The content is gradually expanded from basics to higher levels in order to help the readers with different scientific backgrounds to follow the content.

Guangzhou, China   Henry H. Radamson
Stockholm, Sweden   Anders Hallén
Stockholm, Sweden   Ilya Sychugov
Oslo, Norway   Alexander Azarov

Contents

Symbols and Acronyms

2DEG	Two-dimensional electron gas (definition)
4PP	Four-point probe (measurement technique)
ABF	Annular bright-field (TEM)
ACSEM	Aberration-corrected SEM (definition)
ADF	Annular dark field (TEM)
AES	Auger electron spectroscopy (measurement technique)
AFM	Atom force microscope (AFM)
AL	Aspheric lens (TEM)
APD	Avalanche photodiode (device)
BCC	Body-centered cubic
BE	Backscattered electrons (SEM and TEM)
BEEM	Ballistic electron microscopy (AFM)
BEI	Backscattered electron image (SEM and TEM)
BS	Beam-splitter
BTS	Bias-temperature stress (measurement)
BZ	Brillouin zone (definition)
CCD	Charge-coupled device
CdSe	Cadmium selenide (material)
CMOS	Complementary metal-oxide-semiconductor (device)
CPD	Contact potential difference (definition)
CV	Capacitance–voltage
DLTS	Deep-level transient spectroscopy (measurement technique)
DOF	Depth of field (SEM)
DOS	Density of states (definition)
DSM	Diffused scattered maps (X-ray)
E-beam	Electron beam (lithography)
EDX	Energy-dispersive X-ray (measurement technique)
EELS	Electron energy loss spectroscopy (TEM and SEM)
Ehps	Electron–hole pairs (definition)
EL	Electroluminescence (measurement technique)
EM	Electron multiplication

ESA	Electrostatic sector analyser (SIMS)
ESCA	Electron spectroscopy for chemical analysis (XPS)
EXAFS	Extended X-ray absorption fine structure (EELS measurement)
ExELFS	Extended energy loss fine structure (EELS measurement)
FCC	Face-centered cubic (definition)
FE	Field emission (TEM and SEM)
FIB	Focused ion beam (TEM, SEM, material process)
FinFET	Fin field-effect transistor (device)
FLA	Folded longitude acoustic (phonon)
FLO	Folded longitude optical (phonon)
FTA	Folded transverse acoustic (phonon)
FTIR	Fourier transform infrared spectroscopy
FTO	Folded transverse optical (phonon)
FWHM	Full-width at half-maximum (definition)
GaAs	Gallium arsenide (material)
GFIS	Gas field ion source
GISANS	Grazing-incidence small-angle neutron scattering (measurement technique)
HAADF	High-angle annular dark field (TEM)
HCP	Hexagonal close-packed (definition)
HDO/Si	Hydrodeoxygenation/Si
HeCd	Helium cadmium (material)
HF	High frequency (measurement)
HF	Hydrofluoric acid (chemical)
HPOT	High potential terminal (CV measurement set-up)
HRRLM	High-resolution reciprocal lattice mapping (X-ray)
HX-PES	Hard X-ray photoelectron spectroscopy
iXPS	Imaging XPS (measurement)
KB	Kikuchi band (TEM)
KLs	Kikuchi lines (TEM)
KPFM	Kelvin probe force microscopy (AFM)
LA	Longitudinal acoustic (phonon)
LAADF	Low-angle annular dark field (TEM)
LAB6	Lanthanum hexaboride (TEM)
LDOS	Local density of optical states (modes)
LED	Light-emitting diode (device)
LEIS	Low energy ion scattering (RBS)
LFM	Lateral force microscopy (AFM)
LO	Longitudinal optical (phonon)
LOPC	Longitudinal optical phonon–plasmon coupled (mode) (Raman)
LP	Lorentz polarization factor (X-ray)
LPOT	Low potential terminal (in CV measurement setup)
MB	Matthews–Blakeslee (critical thickness theory)
MBE	Molecular beam epitaxy (deposition technique)
MCA	Multi-channel analyser

MCP	Micro-channel plates
MCT	HgTeCd (compound material)
MEIS	Medium energy ion scattering (RBS)
MIGS	Metal-induced gap states
MIS	Metal–insulator–semiconductor (CV measurement)
MOCVD	Metal organic chemical vapor deposition
MOSFET	Metal-oxide-semiconductor field effect transistor
MSA	Magnetic sector analyser (SIMS)
NA	Numerical aperture (SEM and TEM)
NBD	Nano-beam diffraction (TEM)
NIR	Near infrared (definition)
NRA	Nuclear reaction analysis (RBS)
PL	Photoluminescence (measurement technique)
PLE	Photoluminescence excitation (PL)
PMT	Photomultiplier tube
QD	Quantum dot (0D material)
QHE	Quantum Hall effect (measurement technique)
QW	Quantum well (device structure)
QWP	Quarter wavelength plate
RBS	Rutherford backscattering analysis (measurement technique)
RC	Rocking curve (X-ray)
RIE	Reactive ion etching (etch process technique)
RSF	Relative sensitivity factor
SAXPS	Small area XPS
SCC	Single cubic crystal (definition)
SE	Secondary electrons (SEM and TEM)
SE	Spectroscopic ellipsometry (measurement technique)
SEI	Secondary electron image (SEM and TEM)
SEM	Scanning electron microscopy (measurement technique)
SIL	Solid immersion with aspheric lens
SIMS	Secondary ion mass spectrometry (measurement technique)
SMF	Scanning moiré fringe (TEM)
SNOM	Scanning near-field optical microscope (optical measurement)
SPM	Scanning probe microscopy (measurement technique)
SRP	Spreading resistance profiling (AFM)
SSRM	Scanning spread resistance microscopy (AFM)
STEM	Scanning TEM
STM	Scanning tunneling microscopy (atomic level measurement)
SWL	Shortest wavelength line (X-ray)
TA	Transverse acoustic (phonon)
TDS	Thermal diffuse scattering (TEM)
TEM	Transmission electron microscopy (measurement technique)
TFE	Thermionic-field emission (TEMandSEM)
TLM	Transfer length method
TO	Transverse optical (phonon)

ToF-SIMS	Time-of-flight SIMS (instrument)
UHV	Ultra-high vacuum (TEM and STEM)
VGAA	Vertical gate all around (device design)
VLSI	Very large-scale integration (device related)
W	Tungsten (material)
WD	Working distance (SEM and TEM)
W-H	Williamson–Hall (method)
XPS	X-ray photoelectron spectroscopy (measurement technique)
XRR	X-ray reflectivity (X-ray technique)
Z	Denotes "nucleus charge"
ZLP	Zero loss peak (EELS technique)
a_B	Denotes "exciton Bohr radius"
E_g	Denotes "bandgap energy"
E_T	Denotes "trap bandgap level"
k_B	Denotes "Boltzmann constant ($1.38 \cdot 10^{23}$ J/K)"
k_b	Denotes "Boltzmann constant"
N_C	Effective density of states for the conduction (definition)
N_V	Denotes "effective density of states for valence bands" (definition)
R_H	Denotes "Hall coefficient"
r_H	Denotes "Hall factor"
χ_S	Denotes" Electron affinity"
μ_H	Denotes "Hall mobility"
μ-PL	Micro-photoluminescence (measurement technique)

Part I
Material Characterization Using Photons and Electrons

Chapter 1
X-Ray Techniques

Origin of X-ray

The discovery of x-ray extends to 1895 and experiments carried out on a cathode ray tube by the German scientist Wilhem Conrad Roentgen. His experiment was about an electron beam generated between a cathode and an anode electrode inside a vacuum tube. Although the tube was covered by a thick black paper, he observed a kind of radiation emitted from the tube which could be seen by a green colored fluorescent even reaching some distance away from the tube. The nature of this "invisible light," was unknown and therefore it was called x-ray. Later, Roentgen used x-ray for the first time to take a picture of his wife's hand with a ring on her finger. The picture could show the details of the bone structure in darker contrast, while the ring appeared almost black. It was clear that the different structures had different transparency to the mysterious radiation. Roentgens discovery created the first great opportunity of using x-ray for medical purposes.

At about the same time, crystallographers like Ewald predicted x-ray beam as an electromagnetic wave can create interference (later was called diffraction) in crystalline materials, since the spacing between the planes of crystals is compatible with x-ray wavelength.

About two decades later, Max von Laue and his research group used the x-ray beam on copper sulfate crystal and showed the first x-ray diffraction experiment. The x-ray diffraction phenomenon illustrated the nature of x-ray beams as electromagnetic radiation and its property to make interference with regular array of scatterers in a crystal. It was understood that the x-ray beams could generate spherical waves which interfere constructively or destructively, providing bright and dark spots.

Shortly afterwards, Bragg and his son could formulate Bragg's law, which relates the wavelength of x-rays to the incident angle to the crystal. After a big effort, the first x-ray spectrometer was built and used to study x-ray spectral distributions. By using Bragg's law, it was now possible to could determine the crystal structures and

this new x-ray diffraction spectroscopy (XRD) was developed as a powerful tool to study the materials [1, 2].

1.1   Basic Definitions

In order to present a detailed description about x-ray radiation and its applications in material science, a series of definitions in crystallography and defects are given in this part.

1.1.1   Real and Reciprocal Lattice Space

Crystalline materials consist of well-ordered atoms with electron-density spatially distributed as lattice spots, as shown in Fig. 1.1. In such materials, different atomic planes can be defined as mirrors which can reflect x-ray beams. These crystal planes in real space can be considered as points in a lattice in reciprocal-space, also called as inverse- or Fourier-space. In this case, the reciprocal is a Fourier transform of the actual structure of a crystalline material. Mathematically, a vector, $T = xa_1 + ya_2 + za_3$ and $G = v_1b_1 + v_2b_2 + v_3b_3$ may generate the points in the real space (or direct space) and in the reciprocal space with unit vectors of "a_i" and "b_i", respectively.

The reciprocal space is, furthermore, termed **k**-space, where **k** is the wave vector of, for instance electrons. The **k**-space is closely related to the momentum space through the relation $\boldsymbol{p} = \frac{h}{2\pi}\boldsymbol{k}$, where **p** is the momentum and h is Planck's constant.

The relationship between "b_i" and "a_i" unit vectors are given as following:

$$b_1 = 2\pi \frac{a_2 \times a_3}{a_1 \cdot a_2 \times a_3}, \quad b_2 = 2\pi \frac{a_3 \times a_1}{a_1 \cdot a_2 \times a_3}, \quad b_3 = 2\pi \frac{a_1 \times a_2}{a_1 \cdot a_2 \times a_3} \qquad (1.1)$$

The crystals have structures which are categorized according to the symmetry of the unit cells. There are 7 types of crystal systems: Triclinic, Monoclinic, Orthorhombic, Tetragonal, Trigonal, Hexagonal, and Cubic. The symmetry of each

Real lattice points   **Reciprocal lattice points**

$T = xa_1 + ya_2 + za_3$

Each plane group is a point in reciprocal space

G

1 or more atoms

$G = v_1b_1 + v_2b_2 + v_3b_3$
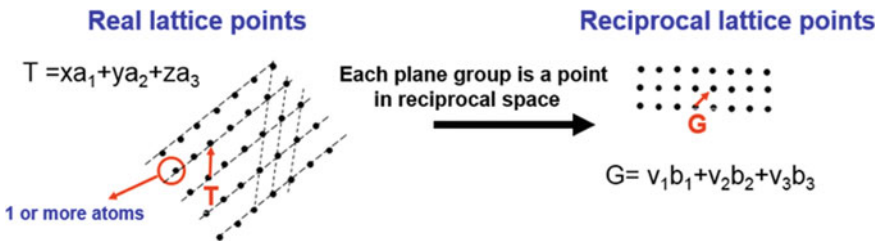
Fig. 1.1 An illustration of real and reciprocal space for a single crystal material

type crystal system is linked to the relationship between the lattice sides a, b, and c
and angles α, β and γ, as shown in the Fig. 1.2.

The unit cell is defined as the smallest group of atoms which are repeated at
regular intervals to form the whole lattice. The "lattice parameter" is considered as
the distance between two points on the corners of a unit cell.

For a crystal system, there are, furthermore, 4 different kinds of arrangements
(Primitive, Base-centered, Body-centered, Face-centered), although not all of these
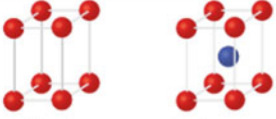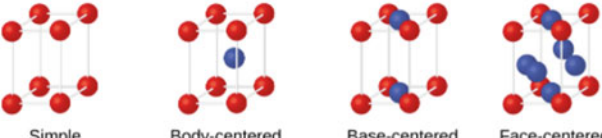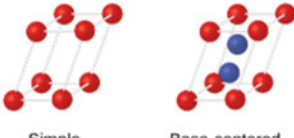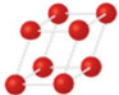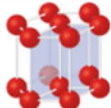
System/Axes/Angles	Unit Cells			
Cubic $a = b = c$ $\alpha = \beta = \gamma = 90°$	Simple	Face-centered	Body-centered	
Tetragonal $a = b \neq c$ $\alpha = \beta = \gamma = 90°$	Simple	Body-centered		
Orthorhombic $a \neq b \neq c$ $\alpha = \beta = \gamma = 90°$	Simple	Body-centered	Base-centered	Face-centered
Monoclinic $a \neq b \neq c$ $\alpha = \gamma = 90°; \beta \neq 90°$	Simple	Base-centered		
Triclinic $a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90°$				
Hexagonal $a = b \neq c$ $\alpha = \beta = 90°; \gamma = 120°$				
Rhombohedral $a = b = c$ $\alpha = \beta = \gamma \neq 90°$				

Fig. 1.2 Different lattice systems and unit cells

arrangements are unique and some of them are not possible because of symmetry reasons. A few examples of unit cells which appear in common crystal materials are: Single Cubic Crystal (SCC), Face-Centered Cubic (FCC), Body-Centered Cubic (BCC), and hexagonal close-packed (HCP).

Now, we define the concept Bravais lattice as the basic building block from which the crystals can be created. This definition has rooted from finding a topological way to make an identical set of points where this set point is indistinguishable from the other sets. Due to symmetry reasons, it turns out that there are only 14 possibilities to arrange arrays of atoms in three dimensions using the translation vector **T**, described above. We can then arrange 14 different groups of points, or unit cells, known as Bravais lattices (after mathematician Auguste Bravais). These lattices in 7 crystal systems, are distinguished by the angles between sides of the "unit cell" and the distance between points in the unit cell. For two dimensional lattices, there are five Bravais lattices.

Si and Ge have diamond structure which consists of two inter-penetrating FCC lattices. In case when the two basis atoms are different, the structure is known as zinc-blende structure. Many III–V semiconductors such as GaAs, AlAs, InAs, or InP have zinc-blende type while GaN have hexagonal structure. Other examples of materials with hexagonal structure are Zn, Co and NiAs. In all these cases, the atoms are in 3D arrangements meanwhile, there is another arrangement of atoms in an atomic sheet form to create 2D crystals.

Most of the metals in our environment are in polycrystalline phase, while ceramic and most organic materials are in amorphous phase. Crystalline materials have the lowest atomic arrangement energy while amorphous materials are on the opposite side.

A Brillouin zone (BZ) is defined as a Wigner–Seitz primitive cell in the reciprocal lattice. The concept of first Brillouin zone is mostly important in the discussion of the electronic and photonic properties of solids. The Brillouin zone is considered as the smallest volume completely enclosed by planes which are the perpendicular bisectors of the reciprocal lattice vectors drawn from the origin. A schematic of Wigner–Seitz primitive cell for reciprocal lattice points in two dimensions and for different primitive cells are shown in Fig. 1.3. The **k** vectors and reciprocal lattice vector, G are also marked in the figure.

The important issues in Wigner–Seitz are the symmetric points (Γ, L, and X) and symmetric lines (Δ, Λ, Σ). For example, in the drawing of bandgap structure of semiconductors such as Si (GaAs or Ge), the maximum point of the valence band locates at Γ point while the minimum point in the conduction depending on the properties of the semiconductor may lie for example for Si near the X point and for Ge near the L point.
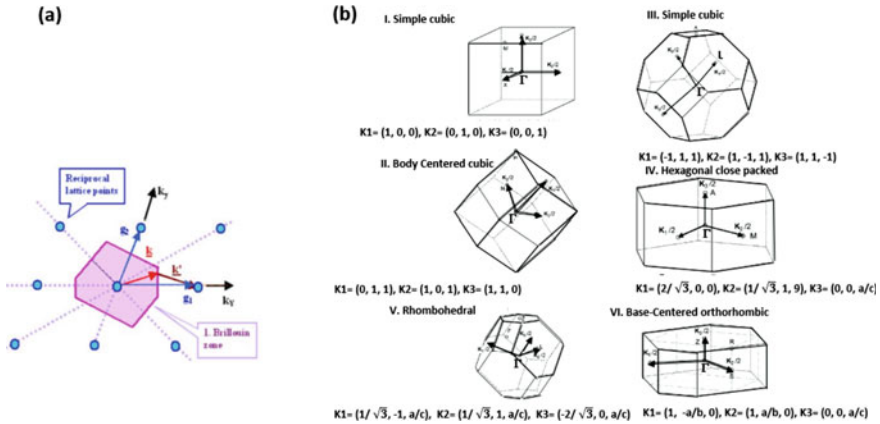
Fig. 1.3 **a**, **b** Different Wigner–Seitz primitive cells

1.1.2 Crystal Structures

So far, we have only discussed ideal crystal structures, but in reality, the crystals contain defects and impurities of various kinds, and they are not infinite, but have surfaces and interfaces to other materials. In recent years, mismatched heterostructures have also been integrated in different types of electronic and photonic devices. These structures are integrated due to the strain which improves the carrier transport and tailors the bandgap for the detecting or lasing of light. In all these cases, to quantify the strain in-plane or normal-to-plane, as well as to determine the layer profile and defect density, are essential tasks in material analysis. X-ray diffraction is a method to provide the above information and can be applied for research purpose, or even in semiconductor industry as in-line measurements.

1.1.2.1 Crystal Defects

Figure 1.4 shows a general view of common defects in the crystalline materials: vacancies, interstitial, precipitates, dislocations and impurity atoms.

The defects in crystal material are divided into three groups:

I. *Local defects*

1. Point defects: these defects appear when one, or few adjacent lattice sites are absent. In principle, point defects are always present in the crystals. The simplest types of point defects are substitutional impurities, vacant lattice sites (vacancies) and interstitial atoms. Point defects are often found in pairs and maintain stoichiometry, or charge neutrality. One example is Frenkel defects, an atom displaced from its lattice site forming a vacancy and an interstitial, or Schottky defects, for instance consisting of a two vacancies

Fig. 1.4 Defects are
denoted as: **a** interstitial
impurity atom, **b** edge
dislocation, **c** Self interstitial,
d precipitate (or cluster),
e small dislocation loop
formed by agglomeration of
interstitials, **f** impurity atom
in substitutional site with
tensile strain, **g** vacancy,
h small dislocation loop
formed by agglomeration of
vacancies, and **i** impurity
atom at substitutional site
with compressive strain
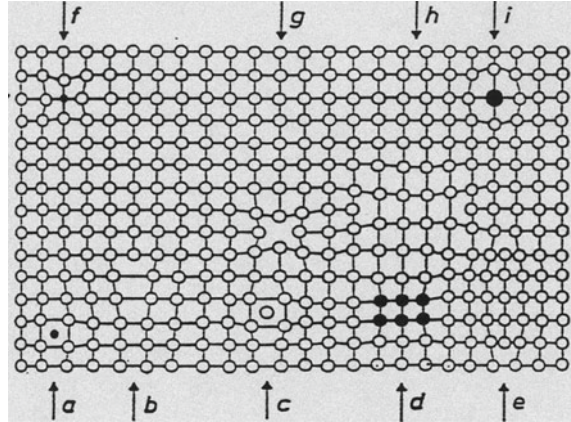
with opposite charge. When the concentration of point defects is sufficient,
then a diffused scattering, similar to substitutional disorder can be emerged in
the x-ray analysis. There is also another type of diffused scattering in presence
of point defects which is known as Huang scattering. This occurs from the
static displacement field due to the elastic deformation of the lattice around the
defect. Point defects are spontaneously created by increasing temperature and
can be quenched by rapid cooling of a material from high temperatures. They
can also be introduced by material processing, for instance by ion implan-
tation, oxidation, or etching. In a larger scale when a small region contains
no atoms at all (or a cluster of vacancies) a void can be formed. Void defects
may occur for example in silicide process when the consumption of Si atoms
is not controlled

2. Precipitates (or clusters): This type of defects can appear in highly doped,
   or post annealed materials as well as segregating during the growth. The
   precipitates are mostly electrically inactive and exist in an amorphous phase.

II. *Linear defects*

These defects mainly appear when groups of atoms are located in irregular positions.
Dislocations belong to linear (or extended) defects that commonly form when the
semiconductors experience strain relaxation. As shown in Fig. 1.5, there are two
main types of dislocations: edge dislocations and screw dislocation, but there is also
intermediate case between these two types so-called mixed dislocations.

Dislocations are characterized by two-direction vectors: the dislocation direction
line, and the Burgers vector. The latter one is a vector which shows the direction and
magnitude of the lattice distortion in presence of a dislocation in a crystal lattice.
The angle between dislocation direction line and Burger vector (α) flags the type of
dislocation.
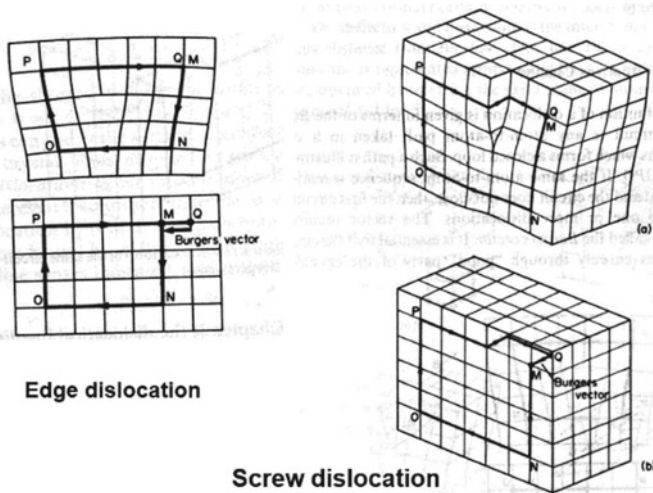
Fig. 1.5 A schematic drawing of edge and screw dislocations

An edge dislocation occurs when an atomic plane is missing in the crystal. The displacement field around the edge dislocation is defined by its Burgers vector, and, is perpendicular to the dislocation line ($\alpha = 90°$).

A screw dislocation is formed by slicing a plane within the crystal up to a certain line, and by removing the atoms on one side of the plane by one unit cell along the direction of the line. For screw dislocations the deformation is in 3-dimensions and, the Burgers vector is parallel to the dislocation line ($\alpha = 0$) [3].

In principle, the dislocations propagate in the direction with lowest required energy, where covalent bonds are weaker, or seriously disturbed. A potential energy is still required before a dislocation is able to move in this direction. Peierls-Nabarro model describes the connection between the movement of dislocations in a certain direction and the number of bonds have to be broken.

The Peierls-Naborro model considers the forces exerted on a dislocation in the crystal structure. Then the tangential stress τ, which explains the role of the atomic layers in different sides of the dislocation in a slip plane (dislocation moves in that plane) is defined as a periodic function of atomic displacement (δ), Burger vector (b) and shear modulus (G is reciprocal space vector) as: $\tau = G \sin [2\pi(\delta/b)]$. The Peierls stress, σ_p is defined as required stress for the motion of a dislocation in the slip plane and can be written as: $\sigma_p = G \exp(- 2\pi\omega/b)$ where ω is the width of dislocation. When a dislocation moves in a slip plane, its energy varies periodically while a minimum value represents the situation that no internal or external forces are acted on the dislocation [4].

In cubic and Zinc-blende crystals, the Burger vector is usually the shortest translation vector of the lattice; $b = a/2 <110>$ and the dislocation direction line is $<110>$ then $\alpha = 60°$ and these dislocations are called 60° dislocations.

III. *Planar defects*

This family of defects stem from a planar locus in the crystal and they are listed as stacking faults (twin planes), grain boundaries, and external surfaces.

Stacking faults are disturbance in the stacking sequence of atomic planes, either missing or adding an atomic plane in the crystal. Stacking faults are often described as a displacement (slip) of an atomic plane perpendicular to the fault plane. Then two vectors are usually discussed: the stacking vector and the slip vector, indicating the displacement direction of atomic planes. The slip vector has a similar role like Burgers vectors in the imaging by transmission electron microscopy analysis [5].

There are other types of planar defects e.g. grain boundaries and antiphase boundaries.

Grain boundaries are a sudden change of crystallographic direction. This type of defect usually forms when two crystals have grown separately and later meet each other.

Antiphase boundaries are built up when the crystallographic direction remains the same, meanwhile an opposite phase occurs at each side of the boundary: As an example, a crystal order ABABABAB could have an antiphase boundary by interrupting the stacking sequence: ABABBABA. Such defects can be observed during the growth of III–V material on Ge buffer layers [6, 7].

1.1.2.2 Strain in Crystalline Materials

Strain is a distortion in the well-arranged crystal lattice. The strain can be induced by lattice-mismatched heterostructures, or in presence of defects and impurities in the crystal. Strain can also be induced by material (or device) processing for example by deposition of stressor nitride layers or thermal cycling treatment when there is a mismatch in thermal coefficients between materials [8]. In crystalline materials, two types of strain can be induced: compressive and tensile as shown in Fig. 1.6. In a cubic crystal with x, y, z coordinates, a compressive force may exert an inwards force in x- and y-directions in which case the cubic symmetry is disturbed and the cube becomes elongated in z-direction. In contrast, if the direction of the force is outwards in x- and y-directions, the crystal shrinks in z-direction and the strain is called tensile. Both types of strain are created in hetero-structures and some examples could be $Si_{1-x}Ge_x/Si$, $Ge_{1-x}Sn_x/Ge$ and GaAs/InGaAs structures for compressive strain, and $Si_{1-y}C_y/Si$, and strained Si/strain-relaxed $Si_{1-x}Ge_x$ buffer layer or strained GaAs/strain-relaxed InGaAs buffer layer for tensile strain.

In general, strain can be revolved in vertical and in-plane directions ($\varepsilon_\perp$ and $\varepsilon_{//}$, respectively) according to:
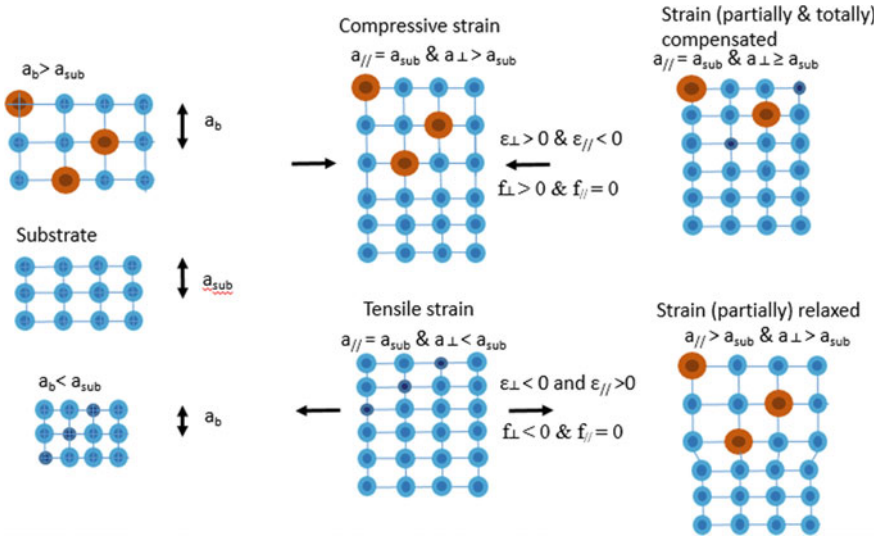
Fig. 1.6 A schematic drawing of different strain types in cubic semiconductor alloys as a result of substitution of lattice atoms with larger (brown) or smaller (dark blue) atoms. In the figure, $a_{//}$ and $a_\perp$ are lattice constant in parallel and perpendicular, respectively

$$\varepsilon_\perp = \frac{a_\perp - a_b}{a_b} \tag{1.1a}$$

and

$$\varepsilon_{//} = \frac{a_{//} - a_b}{a_b} \tag{1.1b}$$

where a_x are the lattice constants in a certain direction or condition as marked in Fig. 1.6. As it was mentioned before, the sign of strain depends on type and direction. In Fig. 1.6, the sign for compressive strain $\varepsilon_\perp > 0$ and $\varepsilon_{//} < 0$ whereas for tensile strain are $\varepsilon_\perp < 0$ and $\varepsilon_{//} > 0$.

The mismatch components in vertical and in-plane directions for a crystalline material are given from these equations:

$$f_\perp = \frac{a_\perp - a_b}{a_{sub}} \tag{1.2a}$$

and

$$f_{//} = \frac{a_{//} - a_b}{a_{sub}} \tag{1.2b}$$

1.1.2.3 Strain and Critical Thickness

The growth of strained alloys results in a stored mechanical energy which can be naturally released when atoms moves according to the forces exerted on them after the layer thickness exceeds a critical value. This critical thickness is directly related to the amount of lattice mismatch between the alloy and its substrate. In order to avoid any strain relaxation, the layer thickness should be grown below critical thickness. Strain relaxation leads to formation of misfit dislocations which deteriorates both the electrical and optical properties of the material. Therefore, there is a need to establish a model to estimate the critical thickness for different alloys.

One of the earliest models to predict the critical thickness was proposed for SiGe alloy grown on Si wafer by Matthews-Blakeslee (MB) already in the 1970'ies [9]. The MB model is known as a "kind of" equilibrium model since it considers a balance between two forces acting during epitaxial growth: a tension force along the dislocation (F_T) and a force caused by stress of misfit (F_a) as shown in Fig. 1.7. The latter force can be written as:

$$F_a = \frac{2G(1+v)}{(1-v)} bh\varepsilon \cos\lambda \qquad (1.3)$$

where G, v, b, and ε are shear modulus, Poisson ratio, burger vector and induced strain, respectively. Parameter λ is the slip angle at the layer interface. The tension force is also formulated as following:

$$F_T = \frac{Gb_2}{4\pi(1-v)} (1 - v\cos^2\alpha)\left(\ln\frac{h}{b} + 1\right) \qquad (1.4)$$

where α represents the angle between misfit dislocation line and Burger vector. Figure 1.7 demonstrates three cases where the strain amount moves towards the maximum level: (I) $F_a < 2F_T$ then the layer interface appears coherently and strain relaxation is impeded, then threading dislocations are located in the layer as shown in (a) and, (II) $F_a = 2F_T$, threading dislocations are formed but at critical stage as shown in (b), and (III) $F_a > 2F_T$, formation of dislocations is promoted as indicated in (c). The formed dislocations disturb the coherence of the layer interface. Finally, the critical thickness (h_c) can be determined through the balance of the forces in case II [9]:

Fig. 1.7 Appearance of dislocations under the acting forces on a strained layer with two interfaces and when the heterostructure is **a** coherent, **b** critical and **c** incoherent
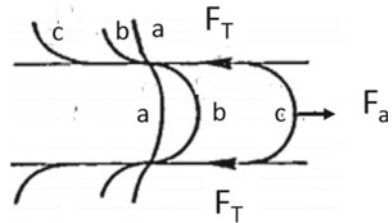
$$h_c = \frac{b(1 - \nu \cos^2 \alpha)}{2\pi f (1 + \nu) \cos \lambda} (\ln \frac{h_c}{b} + 1) \tag{1.5}$$

After almost a decade, when advanced epitaxy technology was developed, it was demonstrated that strained SiGe layers could be deposited with thicknesses far beyond the h_c predicted by Eq. 1.5 in MB model. The large discrepancy between theoretical and experimental data was due to the simplicity of the MB theory which is only based on a balance between acting forces to form misfit dislocations and no other reactions e.g., nucleation, interaction and propagation of misfit dislocations. As a result, the calculated critical thickness for $Si_{1-x}Ge_x$ layers were under-estimated compared to the experimental results.

A few years later, an empirical model with new assumptions was proposed for h_c and the revised equation was as following [10]:

$$h_c = \frac{(1 - \nu)\sqrt{2}b^2}{(1 + \nu)32\pi a f^2} \left(\ln \frac{h_c}{b} \right) \tag{1.6}$$

This equation estimates h_c values with entirely assumption compared to its original version in MB. In the new approach, the initial dislocations are formed in a stochastic fashion which takes into consideration a threshold energy to overcome an energy barrier to generate dislocations in contrast with the MB where the initial dislocations emerge in a simple form in the crystal matrix.

Further attempts to improve the theory for h_c estimation led to a model from Dodson and Tsao (DT) which substitute the MB equilibrium with a new theory which based on an excess stress σ_{ex}. The new idea stems from the difference between the rival forces in Eqs. 1.3 and 1.4 [10, 11]. Hence, the strain energy is released when σ_{ex} becomes greater than zero. The DT equation expresses σ_{ex} as following:

$$\frac{\sigma_{ex}}{\mu} = 2\varepsilon \frac{(1 - \nu)}{(1 + \nu)} - \frac{b}{2\pi h} \frac{1 - \nu \cos^2 \theta}{1 - \nu} \ln \frac{4h}{b} \tag{1.7}$$

where ε stands for the elastic strain and μ for the shear modulus.

In many laboratories, x-ray tools are available to measure the strain and the profile of compound (or composed) heterostructures. Therefore, to determine the critical thickness of a strained material is important to avoid defect formation. As an example, the critical thickness of SiGe on Si heterostructure has studied by different groups. SiGe material is mostly integrated in transistor structures as channel material or as stressor material in source/drain regions.

Figure 1.8 illustrates the calculated h_c-values for Si_xGe_x layers on Si and compares them with the experimental data. The region where the layer thicknesses are above estimated h_c values by MB but below the DT theory's values are known as meta-stable. In fact, the meta-stable region is a result of merging both equilibrium and non-equilibrium theories to handle with stress in epitaxial layers.

Fig. 1.8 Calculated critical
thickness curves for SiGe
alloys vs Ge contents derived
by DT [10, 11] and MB
models [9]. The
experimental data for SiGe
are also marked by (Δ) and
(×) symbols deposited at
different growth
temperatures [10, 11]
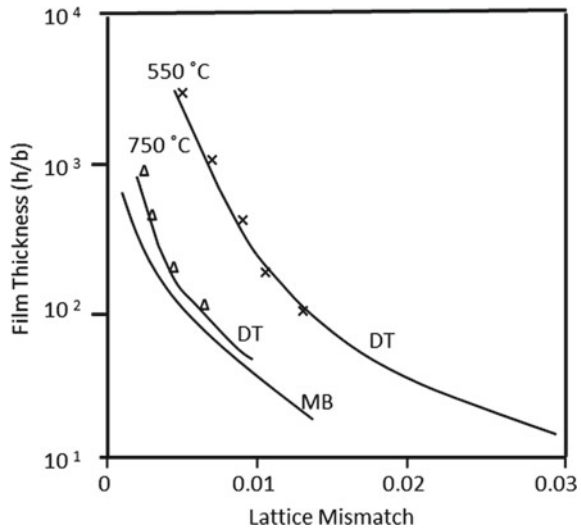
Fig. 1.9 Experimental
curves for critical thickness
of SiGe layers for different
Ge contents and growth
temperatures [12]
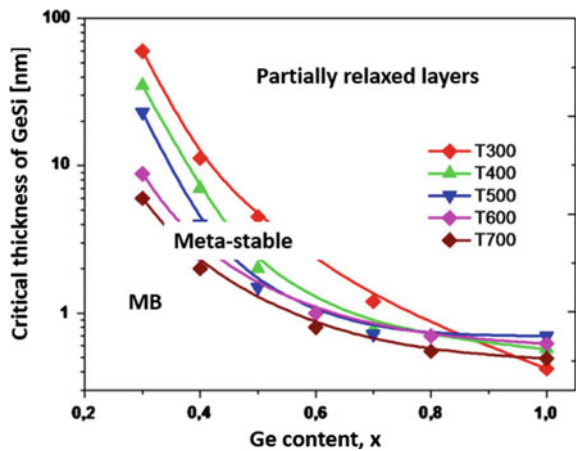
More experimental works have demonstrated the dependence of h_c to the growth
temperature. In fact, the stress relaxation occurs faster by increasing growth temper-
ature due to the decrease of energy barrier for stress relaxation. Therefore, the
meta-stable area shrinks at high growth temperatures as shown in Fig. 1.9.

1.1.2.4 Strain Relaxation on Patterned Substrates

Different studies have demonstrated that epitaxy of strained materials on the patterned
substrates is very different from the global growth on the whole wafer [13]. Therefore,
the critical thickness of the strained materials may differ from the h_c curves discussed
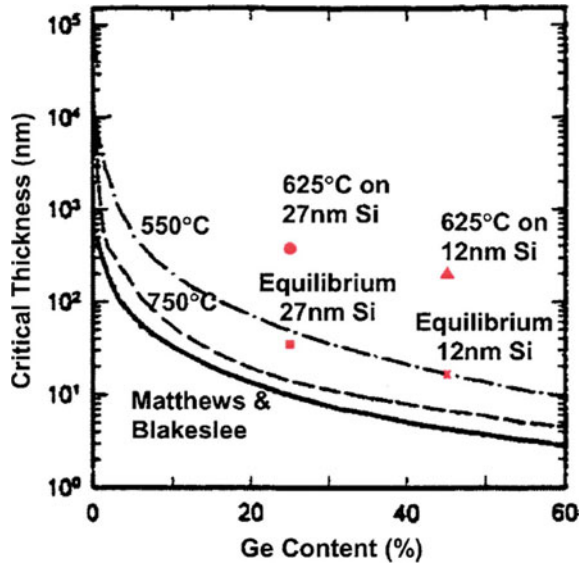
by above theories. As an example, SiGe alloys have been used as stressor materials in recessed source and drain regions [14]. The induced strain by SiGe is non-uniformly distributed inside source/drain openings. In the center area of oxide openings SiGe layer is fully strained but in the regions in vicinity of to the oxide boundaries is partially relaxed. This strain profile creates the situation that the dislocations are terminated into the relaxed area and are not able to move along the whole layer. The outcome is that the SiGe layers can be selectively deposited with thickness above the estimated critical thickness discussed for bulk material as shown in Fig. 1.10.

In advanced FinFET structures, SiGe layers are selectively grown on the Si fins to raise the S/D regions. The relaxation of SiGe layers appears to be far above the critical thickness values for bulk material. The distribution of strain over the Si fins is dependent on shape of Si fin. Since the dislocations glide in the {111} planes, such planes are not formed in order not to promote the dislocation movement. Thus, the maximum induced strain relates to the thickness of fins and not directly to SiGe layer thickness [14].

One of the main techniques to measure strain (and critical thickness) in nano-scale devices in a chip is using x-ray diffraction (XRD). Due to a limited x-ray spot size (1–10 mm) the strain measurements provide a mean value over a series of device arrays and not individual device compared to HRTEM analysis that individual device structure can be investigated. The XRD measurement is based on scanning of the incident beam (ω or θ angle) and the detector moves on the path to collect the diffracted beam (angle 2θ) in a scan type so-called $\omega - 2\theta$.

1.1.3 X-Ray Measurement for Material Analysis

1.1.3.1 X-Ray Emission

X-ray radiation can be generated in laboratory, for instance, by a beam of particles emitted at a target material, as illustrated in Fig. 1.11. For crystallographic studies, an energetic beam of electrons in the range 20–60 kV is commonly used. Electrons are generated by a filament heated by the flow of a current given in mA. The emitted electrons are accelerated towards the positive target, where the x-rays are produced. Therefore, the x-ray emission is often expressed in mA × kV.

The fast deceleration of electrons on the target material results in a large amount of heat (99%) and in the emission of x-ray radiation (1%). The generated heat in all x-ray tubes has to be cooled by water cooling. The x-ray tubes typically have four windows where two are for point focus and the other two for line focus optics. The choice of point or line focus optics depends on which optics are used in the x-ray set-up.

In general, there are two types of atomic processes which may generate x-ray radiation: Bremsstrahlung (German for breaking radiation) and K-shell emission, as displayed in Fig. 1.12. Both these ways of producing x-rays are based on a change of electron states. In Bremsstrahlung the x-ray radiation is emitted when negatively
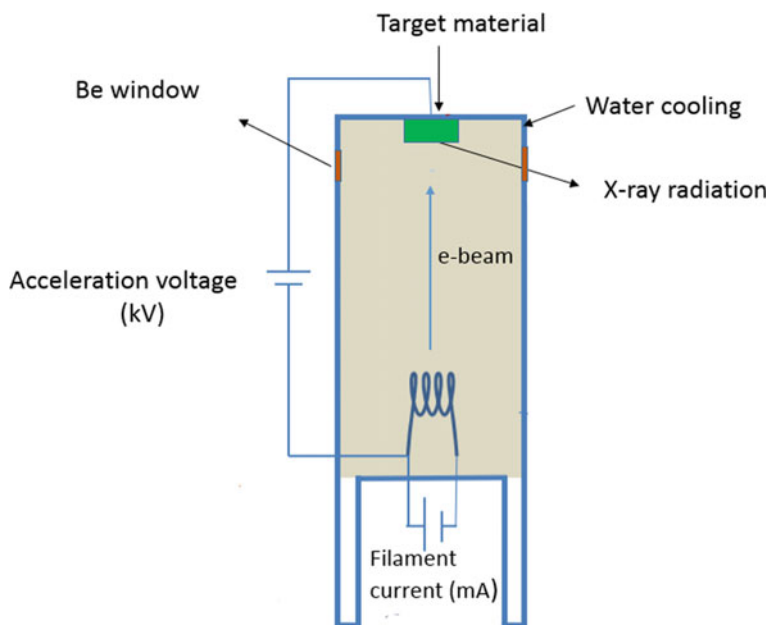
Fig. 1.11 A schematic drawing of an x-ray tube with generator power of kV × mA
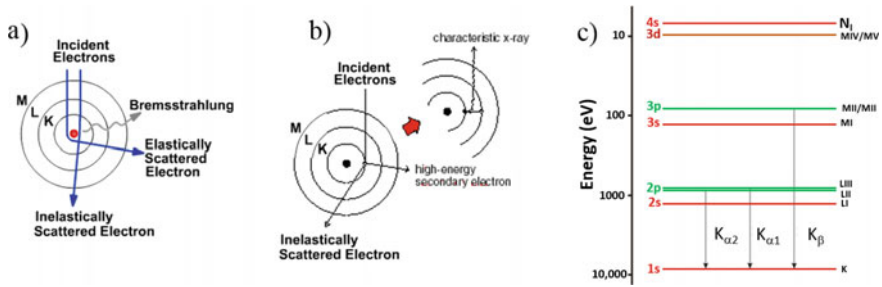
Fig. 1.12 Schematic illustration of elastic and inelastic electrons scattering of an atom when **a** Bremsstrahlung process, **b** K-shell generation of x-ray, and **c** the transmission possibilities in the shells occur

charged electron de-accelerate during interacting via Coulomb force with a nucleus of target atoms. This inelastic energy loss generates x-ray radiation.

In the Bremsstrahlung process, photons with continuously varying wavelengths are generated, but none of these x-ray photons have more energy than the initial electron. Electrons can also be elastically scattered by the nucleus. The elastically scattered electrons are usually scattered at larger angles and even backscattered.

In K-shell emission radiation, the x-ray is generated when incident electrons knock out target electrons from their K-shell, creating a vacant electron state. Then, electrons from outer shells may fill out this vacancy. When this happens, a quantized photon which is characteristic of the target element is emitted. In general, K-shell emission generates x-rays with a single wavelength and higher-intensity than the Bremsstrahlung.

The energies of the characteristic x-rays of an atom depend weakly on the chemical structure, which may shift the energy levels of the outer shell, while and non-bonding, inner, shells of the atom act as main source for the x-ray radiation. A typical ionization time of an atom is only $\sim 10^{-14}$ s and an atom can rapidly be ionized again by a new electron beam which arrives has electrons arriving every 10^{-12} s.

The characteristic lines in this type of spectrum are called K, L, M, … and they correspond to transitions to orbitals with principal quantum numbers 1, 2, 3, … When the two orbitals involved in the transition are adjacent (e.g., $2 \rightarrow 1$), the line is called α. When the two orbitals are separated by another shell (e.g., $3 \rightarrow 1$), the line is called β. Since the transition for β is bigger than for α, i.e., $\Delta E_\beta > \Delta E_\alpha$, then $\lambda_\beta < \lambda_\alpha$. As an example: the wavelengths of K_α and K_β of Cu, an often used target material, are 0.15418 nm and 0.13922 nm, respectively.

The emitted x-rays have always a continuous spectrum of white radiation, but if the electron energy is above a threshold level (excitation potential) then a series of characteristic lines from the filling of vacant electron inner states of the target material can be observed as shown in Fig. 1.13. The excitation potential is considered as the minimum energy required to remove the core electrons. For this excitation process, the maximum energy lost (E_{max}) marks the shortest wavelength line (SWL), which can be calculated from the accelerating voltage (V) equation:

Fig. 1.13 Curves of
blackbody radiation for
different applied acceleration
voltages. The characteristic
lines merge when the unique
threshold voltage of target
material is reached [15]
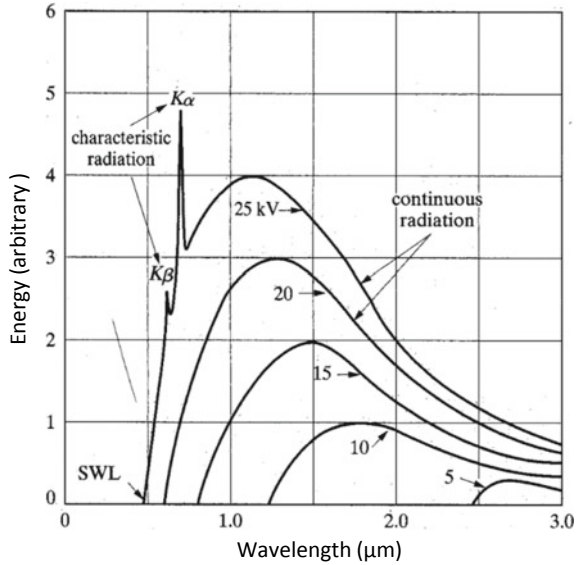
$$E_{max} = eV = hc/\lambda \qquad (1.8)$$

where h, e and c are Planck's constant, electron charge, and speed of light. Then,
by increasing the accelerating voltage of the x-ray tube, SWL becomes shorter but
also the wavelength of the maximum intensity (λ_{max}) shifts to the shorter wave-
lengths. Experiments show that the wavelength of the maximum intensity of the
white radiation lies roughly at $1.5 \times \lambda_{min}$.

Suitable targets for x-ray tubes are often found among the metals since they are
electrically conductive, which avoids charge buildup in the target, and have high
melting points as well as high thermal conductivity to withstand the heat developed
in the target. As an example, an x-ray tube with operating inputs of 40 kV at 30 mA
can produce heat of 1.2 kW, which is more than a typical kitchen toaster. Therefore,
target materials with good thermal stability are limited to chromium, iron, cobalt,
copper, molybdenum to manage the heat energy. The wavelength, λ, of the emitted
characteristic x-ray line, is linked to the atomic number of target material, Z by
Moseley's Law [15]:

$$1/\lambda = c(Z - \sigma)^2 \qquad (1.9)$$

where σ and c are constants. This equation indicates that the wavelength of the
characteristic line for heavier materials is shorted. The wavelength of K_α radiation
for typical target materials e.g., Cr, Cu and Mo are 0.229, 0.154 and 0.071 nm.

1.1.3.2 X-Rays from Synchrotron Electron Accelerators

Circular electron accelerators, so called synchrotrons, provides an alternative technique to generate useful x-rays for material analysis. The synchrotron technique yields high intensity radiation with well-defined linewidths over a large range of wavelengths starting from infrared, through the visible part of the photon spectrum, to hard x-rays. The high intensity of synchrotron facilities have resulted in new dimensions in the study of nanoscale crystals, semiconductor devices and also molecules and biological materials. Figure 1.14a, b show the schematic view of a synchrotron facility.

The technique is based on high energy electrons accelerated by a linear accelerator (or Linac) and then injected into to the booster ring, where they are increased in energy from a few hundred MeV to a few GeV. In a third stage, the electrons are transferred to the outer storage ring and are circulated around this ring by using a series of bending magnets. The electron beam in the ring creates electromagnetic radiation by the bremsstrahlung process at each bending magnet. The generated electromagnetic radiation is collected in a narrow cone tangentially to the electron's orbit and directed towards the sample.

The intensity of synchrotron radiation can be hundred thousands of times higher than the conventional x-ray tubes intensity and is highly collimated. Furthermore, it is also possible to select a certain wavelength for research as shown in Fig. 1.14a [16].

A synchrotron facility is equipped with magnets in pairs to deflect the e-beam and different configuration of these magnets (e.g. Wigglers and Undulators) intensify the radiation as illustrated in Fig. 1.14b.
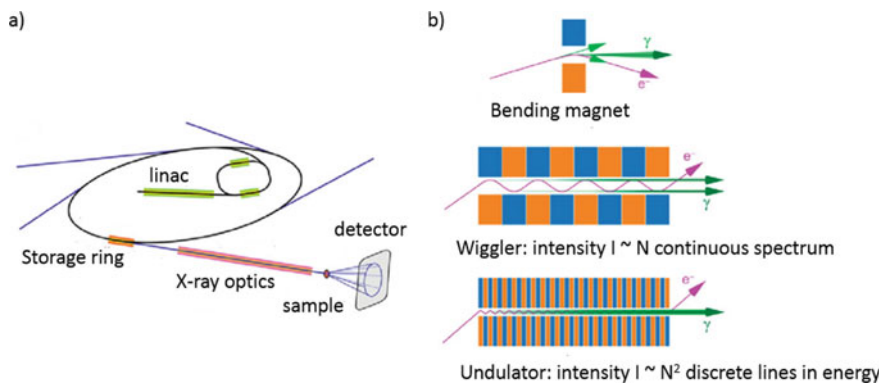
Fig. 1.14 Schematic drawing of a synchrotron technique, **a** the facility and **b** a single magnet which is integrated in wiggler and undulator configuration to intensify the radiation intensity
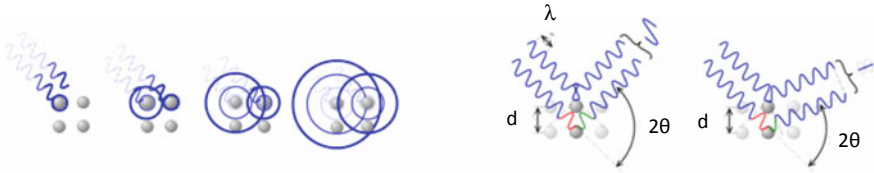
Fig. 1.15   X-ray scattering over an array of atoms in a crystalline material

1.1.4   X-Ray Scattering

The x-ray beam is an electromagnetic wave which may interact with the electron cloud of an atom in a crystal structure. In this case, the electron cloud elastically scatters the x-ray beams as shown in Fig. 1.15. The scattered photons can be viewed as a re-emitted wave of the same wavelength as the incoming beam, which can interfere with photons scattered from other atoms. The outgoing beam may then contain constructive or destructive interference, making the scattered beam stronger or weaker. In this scattering process, each atom is considered as a coherent point scatterer which its strength is proportional to the electron density around the atom.

1.1.5   Powder Diffraction Measurements

1.1.5.1   X-Ray Tool Set-Up

A Bragg–Brentano diffractometer is a common tool for the x-ray measurement. The tool consists basically of an x-ray tube and a slit in the primary optics, or incident side. The x-rays then hit the sample and are scattered. On the secondary optics, or outgoing side, the x-rays are limited by a slit before entering the detector. An advanced Bragg–Brentano configuration offers a programmable divergence slit in the primary optics while, the secondary optics could be high-speed line detector with scatter slit and Ni filter. In principle, a diffractometer has an excellent resolution in the Bragg–Brentano geometry, where the distance between source and sample is the same as the distance from sample to detector.

A common mode of a diffractometer is when the x-ray source is fixed and the sample rotates by the angle θ, as the detector moves by 2θ, a so called $\theta - 2\theta$ measurement. This type of set-up can be modified to an arrangement where the sample is kept in a fixed position and both the source and the detector each can rotate by $-\theta$ and θ, respectively. The x-ray optics for such a set-up are shown schematically in Fig. 1.16 [17].

In Fig. 1.16, the dotted larger circles with sample in the center position represent the goniometer circle (mechanical part) on which the divergent x-rays are focused and diffracted by the sample. In the figures, smaller circles are also associated with
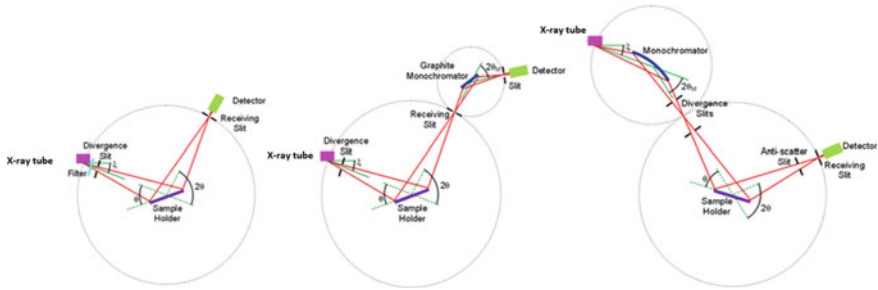
Fig. 1.16 Set-up of different diffractometers for $\theta - 2\theta$ measurements. In such a set-up, all the optical components dealing with the incident beam to the sample is called primary optics. Meanwhile, all the optical components dealing with the diffracted beam is called secondary optics. The left figure shows the simplest design of a diffractometer where two slits in primary and secondary optics decreases the divergence of the x-ray beam. The middle figure shows a diffractometer where a graphite monochromator in secondary optics reduces the beam divergency as well as the intensity of the undesired low intensity wavelengths. The right figure presents a high-resolution measurement where the $K_{\alpha II}$ peak has been removed by the monochromator in the primary optics

the position of pre-sample, or post sample monochromators, used for obtaining better linewidth of the incoming, or outgoing beam, respectively. These types of monochromator crystals are curved to ensure focusing of the x-ray beams. In some diffractometers, a graphite monochromator is placed on the detector arm. The x-ray tube is aligned in a way that the beam divergence on the sample makes angle ξ to the x-ray tube surface (typically $6°$) and the divergence of the beam can be monitored by applying slits. In a diffractometer a suitable filter has to be applied if there is no monochromator. The choice of material as a filter depends on the target material in the x-ray tube. As a common rule, material prior to target material in the periodic table could serve as an excellent K_β filter. This is due to that the wavelength corresponding to the absorption edge of these materials is longer than the K_β wavelength [15]. For example, for an x-ray tube with target metal of Cu, Co is as filter material, and for Fe is Mn as filter material, as shown in Table 1.1.

Table 1.1 K-series wavelengths of most common target materials for x-tubes with suitable filter materials and their absorption edge wavelength for k_β wavelength [15]

Target materials	$K_{\alpha 1}$ (nm)	$K_{\alpha 2}$ (nm)	K_β (nm)	K_β filter	Absorption edge (nm)
Fe	0.193631	0.1940002	0.17567	Mn	0.1896
Co	0.178900	0.179289	0.16208	Fe	0.1744
Ni	0.165794	0.166178	0.15002	Co	0.1608
Cu	0.154059	0.154443	0.13923	No	0.1488

1.1.5.2   Powder Diffraction Measurements

These measurements are performed to determine the material phase and to identify the different crystalline (or polycrystalline) materials in powder form, or bulk [15]. An unknown material is identified through a wide range scan, so called x-ray diffraction pattern, or diffractogram. The measured data are then compared to a large database containing reference patterns [15]. As an example, NaCl analysis has been shown in Fig. 1.17a, b.

The intensity of the peaks (I) in a diffractogram is dependent of different factors, e.g. structure factor (F), multiplicity factor (p), Lorenz factor (L(θ)), absorption factor (A(θ)), where θ is the incident angle, and temperature factor (T) as following [15]:

$$I = F^2 P L(\theta) A(\theta) T \tag{1.10}$$

The structure factor is a fundamental parameter for x-rays scattered by atoms. Mathematically, the scattering amplitude, F_G, over N cells from a diffraction experiment is obtained by:

$$F_G = \sum_N \int dV n(r) e^{-iGr} = N S_G \tag{1.11}$$

where n(r) denotes electron density, dV is the volume element, G is the reciprocal lattice vector and $\mathbf{r_j}$ is the real lattice vector. The S_G in structure factor is then defined as

$$S_{G(hkl)} = \sum_j f_j e^{-iGr_j} = \sum_j f_j e^{-i2\pi \left(hx_j + ky_j + lz_j \right)} \tag{1.12}$$
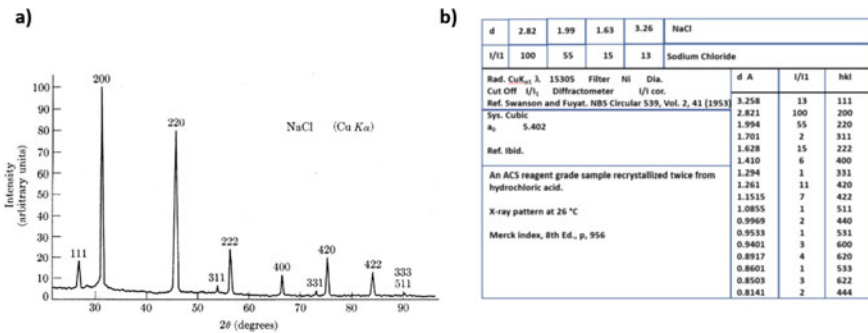
Fig. 1.17 a Diffractogram of NaCl and **b** its powder diffraction file. The file contains the intensity of possible peaks of the material (the numbers associated to the different peaks is the Miller index of the reflection)

The coefficients f_j are defined as atomic form factor. The condition for S_G differs from zero value depends on the unit cell and those conditions are listed for different crystal structures as:

BCC

$$F_{hkl} = \begin{cases} 2f, h+k+l = even \\ 0, h+k+l = odd \end{cases}$$

FCC

$$F_{hkl} = \begin{cases} 2f, \ h,k,l : all\ even\ or\ all\ odd \\ 0, \ \ h,k,l : otherwise \end{cases}$$

Diamond and Zincblende

$$F_{hkl} = \begin{cases} 8f, \ \ \ \ \ \ \ \ h+k+l = 4N \\ 4(1 \pm i)f, \ h+k+l = 2N+1 \\ 0, \ \ \ \ \ \ \ \ \ \ h+k+l = 4N+2 \end{cases}$$

HCP

$$|F_{hkl}|^2 = \begin{cases} 0, \ \ \ \ h+2k = 3N\ and\ l\ is\ odd \\ 4f^2, \ h+2k = 3N\ and\ l\ is\ even \\ 3f^2, \ h+2k = 3N \pm 1\ and\ l\ is\ odd \\ f^2, \ \ \ h+2k = 3N \pm 1\ and\ l\ is\ even \end{cases}$$

The structure factor decides the fate of a scattered x-ray's peak intensity, as discussed previously. The multiplicity factor, M, is related to the number of planes which could be observed by x-ray scan. For example, for (100) plane different combinations are (010), (001), (– 100), (0– 10), (00– 1), but for (111) plane, these are (– 111), (1– 11), (11– 1), (– 1– 11), (1– 1– 1), (– 11–1), (– 1– 1– 1) which makes the intensity of (111) peak is 8/6 higher than (100). The periodicity factor for several crystal types and planes are listed in the Table 1.2.

Table 1.2 Periodicity factor for different planes of crystal materials

System	hkl	hhl	hh0	hhh	hk0	h00
Cubic	48	24	12	8	24	6
Tetragonal	16	8	4	(8)	8	4
Hexagonal	24	12	6	(12)	12	6
Orthorhombic	8	(8)	(8)	(8)	4	2
Monoclinic	4	(4)	(4)	(4)	(4)	2
Triclinic	2	(2)	(2)	(2)	(2)	(2)

In x-ray analysis, the Bragg–Brentano system is used to scan the sample as we discussed before. Therefore, a parameter called the Lorenz factor which is related to the time that the diffraction condition for a crystal is still valid. Since the diffracting planes have different geometry relative to the x-ray tool, the Lorentz factor is dependent on the incident angle as well as the diffraction geometry. The other related factor is the Lorentz-polarization factor (L_P) for x-ray measurement. When a monochromator is placed in the primary optics, the scattered beam is coplanar and Lorentz factor is $L = 1/\sin(2\theta)$. Then p and L_p are given by:

$$p = \frac{1 + \cos^2 2\theta}{2} \tag{1.13}$$

$$L_p(\theta) = \frac{1 + \cos^2 2\theta}{2 \sin 2\theta} \tag{1.14}$$

In general, $L_p(\theta)$ can be written as:

$$L_p(\theta) = \frac{1 + A \cos^2 2\theta}{(1 + A) \sin 2\theta} \tag{1.15}$$

where $A = 1$ when no monochromator is used. Otherwise 'A' becomes more complicated and is obtained from $A = \cos^2 2\theta_m$ where θ_m is the monochromator angle.

A XRD measurement (using the set-up in Fig. 1.16) provides the following information [15].

(a) present phases in the material from the peak positions
(b) phase concentrations from peak intensities
(c) an average of crystallite size or strain by determining the peak widths
(d) background hump in presence of amorphous material.

The peak widths are measured by Full Width Half Maximum (FWHM) and the crystallite size (D) is obtained by using the Scherrer's relationship as shown in Fig. 1.18 [18]:

$$D = \frac{K\lambda}{\beta \cos \theta_B} \tag{1.16}$$

where β is the FWHM of the peak, λ the wavelength of Cu $K_{\alpha I}$ (0.154056 nm), K is the shape factor ($= 0.9$) and θ_B is the incident angle.

The strain in the grain lattice, ε is calculated by the Williamson–Hall (W–H) method. The W–H relation requires the FWHM and grain size to derive the strain value according to:

$$\beta \cos \theta = 4\varepsilon \sin \theta + \frac{k\lambda}{D} \tag{1.17}$$

Fig. 1.18 A typical peak
with its feature signatures
(I_{max}, B and θ) in a
diffractogram
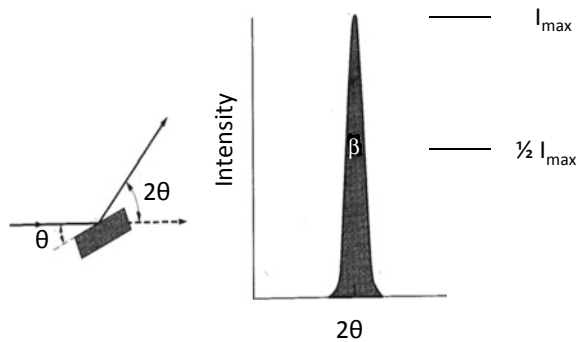
The strain value, ε can be calculated from the slope of the linear fit of the plot of $\beta \cos \theta$ and $\sin \theta$.

1.1.6 High Resolution X-Ray Measurements

1.1.6.1 Set-Up

The high-resolution x-ray set-up is established to remove $K_{\alpha 2}$ peak and preserve $K_{\alpha 1}$. This can be achieved by using a monochromator in front of the x-ray tube as shown in Fig. 1.19. There are different types of monochromators depending on the intensity and the divergence of the x-ray beam. Both these points are crucial for the acquisition of high-resolution data and quality of the measurements. Typical types of monochromators are: LiF, Quartz, or SiO_2, InSb, Ge, Si, in concern to the target materials in the x-ray tube and required divergence of λ.

Two common materials for monochromators in research laboratories are single crystalline Si and Ge, which can be mounted in a two-bounce or four-bounce configuration. For example, for a four-bounce configuration, four Ge(220),or Ge(440), crystals are used. In these monochromators, the first and third crystal are diffractors to remove the $K_{\alpha 2}$ peak but the second and fourth ones are reflectors in order to bring the beam to the original path. The divergence of the x-ray beam after a four-bounce Ge(220) and Ge(440) monochromator is 12 arc-second and 8 arc-second (1 arc sec is 1/3600 degree), respectively. The intensity of the x-ray after the monochromator is significantly lower than the original intensity. Therefore, a mirror is placed prior to the monochromator on the beam path in order to compensate the intensity loss [19].

There is also a hybrid monochromator which consists of an x-ray mirror (parabolic graded multilayer mirror) combined with and a channel-cut Ge crystal. The hybrid module provides high intensity $K_{\alpha 1}$ and $K_{\alpha 2}$ peak is suppressed to < 0.1%, but the beam divergence is ~19 arc sec.

In the secondary optics, there is a receiving slit detector, or a triple axis one. The latter type contains two Ge (220) crystals to reduce the divergence of the
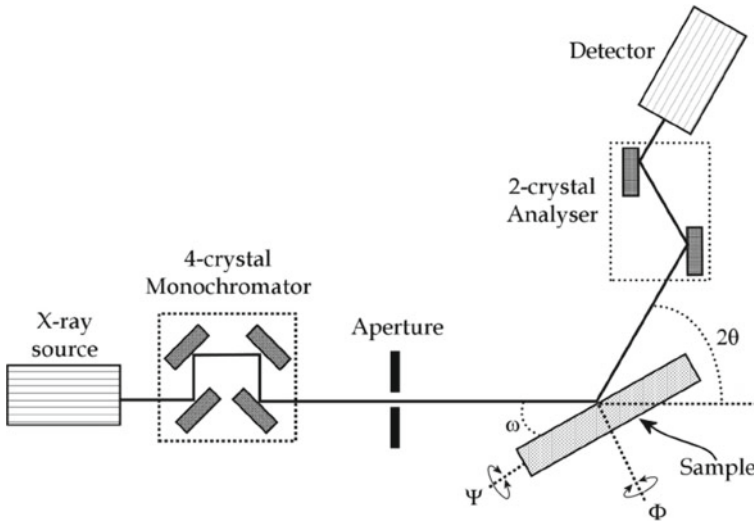
Fig. 1.19 A high-resolution x-ray set-up containing an x-ray tube, monochromator in primary optics, sample goniometer with different rotational motions, and a two-crystal box and detector in the secondary optics part

diffracted beam. The goniometer is a mechanical device that provides movement for the following scans (see Fig. 1.19):

- ω (or θ) scan: Incident beam angle changes and the detector stays at a certain angle
- 2θ scan; detector angle changes but the sample stays at a certain angle
- ω-2θ (or θ-2θ) scan: both the detector and sample angle changes
- ϕ scan: in-plane angle changes (in-plane rotation)
- ψ scan: off-plane angle changes

The above scans can provide one- or two-dimensional x-ray measurements which are dealt in this chapter.

1.1.6.2   X-Ray Diffraction

As we discussed before, the constructive interference of x-ray in a material is defined as diffraction. Since the diffraction in our case is an elastic interaction between an electron density and x-ray photons with wavelength of λ, the absolute value of the wavevector for the incoming and diffracted waves will be preserved; $|K|$ the absolute value of the wavevector for the incoming and diffracted waves will be preserved; $|K_{inc}| = |K_{diff}| = 2\pi/\lambda$. For diffraction with a lattice, it then follows that $K_{diff} = G - K_{inc}$ or $\Delta K = G$ is the condition for diffraction where G is the basis vector in reciprocal space ($G = 2\pi/d$ where d is interplanar distances). This condition leads
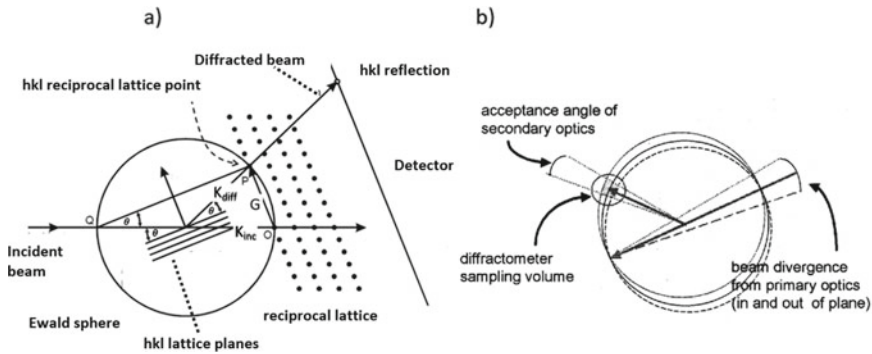
Fig. 1.20 A schematic drawing in 2D of Ewald Sphere in **a** ideal case, and **b** Ewald Sphere after considering the incident x-ray beam divergence

to a well-known relationship calls Bragg's law:

$$2d \sin \theta = \lambda \quad \text{where } d = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \tag{1.18}$$

where θ, a, and (hkl) are the incident angle in relation to a crystal plane, lattice constant and the Miller index for a particular plane, respectively.

Another way to describe the diffraction condition is by using the Ewald sphere method, which was introduced by a German crystallographer called Paul Ewald. If the K_{inc} points at an assumed origin of the reciprocal space, O, the Ewald sphere can be constructed having a radius of $|K_{inc}|$. If the Ewald sphere intersects with other reciprocal lattice points, the condition of $\Delta K = G$ is satisfied, as shown in Fig. 1.20a.

In all x-ray measurements, the x-ray beam has a divergence depending on the optics and the set-up in the instrument. This means that the Ewald sphere actually becomes as shown in Fig. 1.20b. A high-resolution x-ray measurement can be performed around a specific reflection plane with a determined scanning range of the incident angle which is characterized by step size and time per step parameters, by using the set-up in Fig. 1.20. Such a scan is called $\omega - 2\theta$, or rocking curve, RC, where ω is the incident angle and 2θ is the detector angle ($\omega = \theta$ for a symmetric reflection e.g. (004) and ω differs from θ for an asymmetric reflection e.g. (224)). RC measurement is widely used in many fabs and laboratories. A RC is usually performed fast and provides information about the composition, strain, layer quality and layer thickness. During the scan, the Bragg condition may be satisfied for different layers in a multilayer structure, creating distinguished peaks in the RC. The split of layer peak in respect to (real or virtual) substrate peak can be converted to the strain amount in the layer and the composition can be determined by assuming no strain relaxation in the layer.

The x-ray as an electromagnetic wave may refract into the layer if the incident beam is below a certain critical angle. If the interface of the layer is smooth, which is an indicator of a defect-free materials, the x-ray can show interference resulting

$$2d_{layer} \sin\theta_{layer} = \lambda$$

$$2d_{sub} \sin\theta_{sub} = \lambda$$
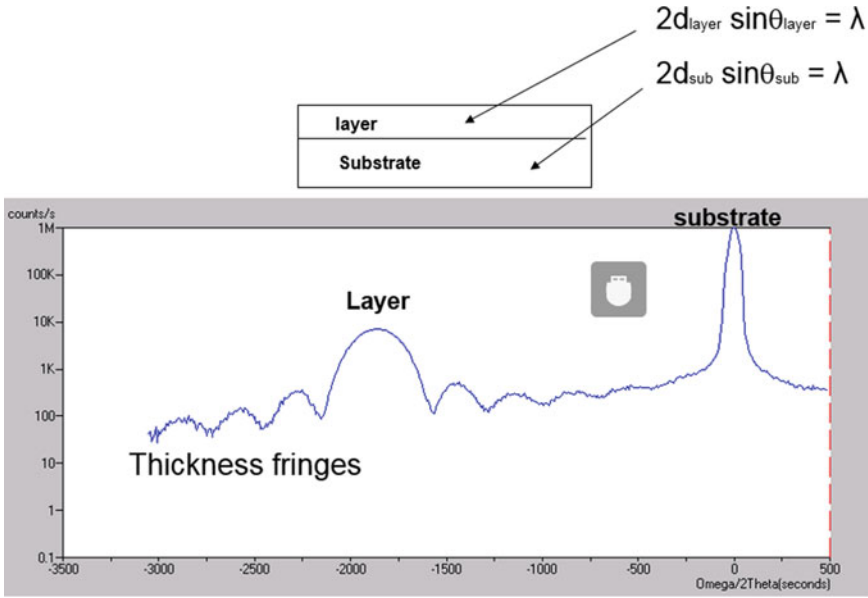
Fig. 1.21   A rocking curve $\omega - 2\theta$ scan around (004) reflection from a heterostructure of SiGe/Si

satellite peaks, or thickness fringes in the RC. The layer thickness can be determined from the two adjacent fringes. The number of fringes is an indication of a smooth interface and for example, it flags for poor epitaxial quality when only a few fringes appear in the RC. The broadening of the layer peaks is expressed in FWHM and it is a signal for defect density in the layers, but it could also be the result of thin layers [20, 21]. Figure 1.21 illustrates a RC of SiGe/Si hetero-structure around (004) reflection plane. The SiGe and Si peaks, as well thickness fringes are marked in the figure.

The x-ray peaks in RCs have a shape which can be approximately considered as a Gaussian profile. The broadening, β, of an RC can be due to several different reasons: the intrinsic broadening due to a crystal structure (β_0(hkl)), the instrumental optics (β_d(hkl)), angular rotation in presence of dislocations (β_α(hkl)), induced strain by dislocations (β_ε(hkl)), curvature of the wafer (β_r(hkl)), and crystal size (poly-crystalline materials) (β_L(hkl)). Thus, the total broadening (β_m(hkl)) is expressed as following:

$$\beta_m^2(hkl) = \beta_0^2(hkl) + \beta_d^2(hkl) + \beta_\alpha^2(hkl) + \beta_\varepsilon^2(hkl) + \beta_L^2(hkl) + \beta_\tau^2(hkl) \quad (1.19)$$

Each term in Eq. 1.19 can be independently calculated. The angular broadening is given by

$$\beta_\alpha^2 = 2\pi \ln 2b^2 D = K_\alpha \quad (1.20)$$

If a peak in a RC is Gaussian distributed then the strain term is expressed according to references [20, 22]:

$$\beta_\varepsilon^2 = \left[8 \ln 2(\varepsilon_N^2)\right] \tan^2 \theta = K_\varepsilon \tan^2 \theta \tag{1.21}$$

where the strain is expressed for different types of dislocations:
Edge dislocations [22, 23]:

$$\varepsilon_N^2 = \left[5b^2/(64\pi^2 r^2)\right] \ln(r/r_0) \times (2.45 \cos^2 \Delta + 0.45 \cos^2 \psi) \tag{1.22}$$

Screw dislocations [22, 24]:

$$\varepsilon_N^2 = \left[b^2/(4\pi^3 r^2)\right] \ln(r/r_0)(\sin^2 \psi) \tag{1.23}$$

60° dislocations:

$$\varepsilon_N^2 = \left[b^2/(\pi^2 r^2)\right] \ln(r/r_0) f(\Delta, \psi) \tag{1.24}$$

Here, Δ is the angle between the normal of dislocation glide plane and the normal to diffraction plane. ψ is the angle between Burgers vector and the normal to the diffraction plane, and r is the radial distance from a dislocation with core r_0 in the direction of strain field.

For 60° dislocation, $f(\Delta, \psi) = 0.070$ and $r_0 = 10^{-7}$ [22] and $r = 0.5\sqrt{D}$. Then the component for strain broadening can be written as:

$$\beta_\varepsilon^2 = 0.160b^2 D \left|\ln(2 \times 10^{-7} \, \text{cm}\sqrt{D}\right| \tan^2 \theta = K_\varepsilon \tan^2 \theta \quad \text{(60 degree dislocations)} \tag{1.25}$$

The broadening related to the crystal size is overshadowed by layer thickness. For GaAs layer, the broadening of 350″, 70″ and 35″ for 0.1, 0.5, and 1 μm layer thickness. Practically, this parameter can be neglected for layers thicker than 1 μm. An expression for this parameter was provided as following [25]:

$$\beta_L^2 = \left[4 \ln 2/(\pi h^2)\right](\lambda^2/\cos^2 \theta) \tag{1.26}$$

where h is the layer thickness. The curvature impact has been formulated as [23, 24]:

$$\beta_\tau^2 = w^2/(r^2 \sin^2 \theta) = K_r/\sin^2 \theta \tag{1.27}$$

where w is the x-ray beam width and r is the curvature radius. In a more practical way, both the intrinsic broadening of a crystal structure and the instrumental optics components can be neglected. In many cases, both the crystal size and the curvature contribution can be also neglected. Then the Eq. 1.19 will be approximately written as:

Table 1.3 X-ray rocking curve data for different reflections for 3 μm GaAs layer grown on Si(100) substrate

(hkl)	Θ (deg)	$\tan^2\theta$	β_m (arc sec)
(002)	15.8	0.080	220
(113)	26.9	0.257	205
(004)	33.0	0.422	220
(224)	41.9	0.805	210
(115)	45.1	1.007	210
(044)	50.4	1.461	235
(006)	54.9	2.020	220
(026)	59.5	2.880	235
(444)	70.7	8.150	295
(117)	76.5	17.350	385

$$\beta_m^2(hkl) = K_\alpha + K_\varepsilon \tan^2\theta \tag{1.28}$$

where the defect density, D is related to K_α and K_ε as follows:

$$\text{Angular broadening}: D = K_\alpha/4.36b^2 \tag{1.29}$$

$$\text{Strain broadening}: D = K_\varepsilon/0.090b^2 D\left|\ln(2 \times 10^{-7}\text{cm}\sqrt{D}\right| \tag{1.30}$$

As an example, FWHM analysis for GaAs grown on Si to show how the above theory works. Table 1.3 illustrates the x-ray experimental data and calculated data according to ref [21].

The table's data has been illustrated and linearly fitted in Fig. 1.22. The slope of the curve and intersection with Y-axis will provide independently the defect density in GaAs/Si structure. The fitting as a straight line is given by:

$$\beta_m^2 = 41.6(arc\ sec)^2 + 5850(arc\ sec)^2(\tan^2\theta) \tag{1.31}$$

$K_\alpha = 41.6$ (arc sec) ^{2} corresponds to 9.78×10^{-7} (rad^2) which gives an estimation for defect density of 1.4×10^8 cm^2. Meanwhile, the other component K_ε is 5850 (arc sec)^2 or 1.375×10^{-7}(rad^2) and the defect density is calculated to 1.5×10^8 cm^2.

The calculated defect density from K_ε and K_α can have a discrepancy of a factor of two (or less). This is still a good agreement with experimental data and a better result can be obtained by having a better approximation.

1.1.6.3   Analysis of Superlattices

XRD technique has been used for analysis of superlattices of ternaries compounds of III–V and II–VI materials [26, 27]. A superlattice is a larger periodicity than

Fig. 1.22 Relationship of
x-ray data versus $\tan^2(\theta)$
extracted from Table 1.3
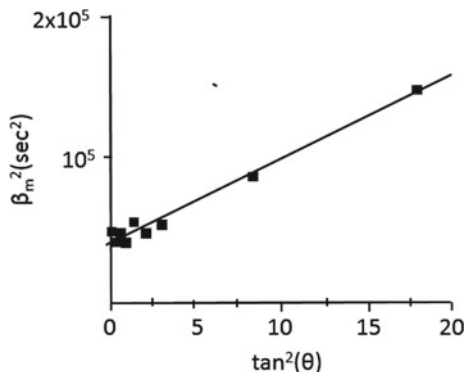
found in a crystal in one or several dimensions. It can, for instance, consist of many
thin crystal layers of different materials, which is typical arrangement for quantum
well structures used for many optoelectronic semiconductor devices. Therefore, the
composition of ternaries, layer thickness, mismatch of quaternaries, tilt, relaxation,
strain, and curvature all have an impact on the performance of devices. For example,
the composition of compound semiconductors relates to the bandgap, whereas the
layer thickness is linked to the quantum size effect.

Rocking curve from a superlattice structure contains two main peaks, from the
substrate and from the layer, but due to interference of x-ray beams within different
layers (between different interfaces) many smaller satellite peaks will also appear. For
RC analysis of cubic crystals, (004) reflection is widely applied. Using the Bragg's
law (from Eq. 1.18)

$$\lambda = \frac{a}{2} \sin \theta \qquad (1.32)$$

the lattice constant 'a' can be determined which provides information about
mismatch. The composition of superlattice can be calculated by using Vegard's Law,
an empirical law which states that the lattice parameter for a compound is roughly
equal to the weighted mean of the lattice parameters of each constituent. For a ternary
compound of InGaAs, where x percent of the Ga is replaced by In, the lattice constant
will be

$$a_{\mathrm{In_xGa_{1-x}As}} = x a_{\mathrm{InAs}} + (1-x) a_{\mathrm{GaAs}} \qquad (1.33)$$

The second important point is to extract the thickness of each layer in the
superlattice structure. By considering Bragg's law, two neighboring peaks can be
formulated:

$$N\lambda = 2\Lambda \sin \theta_n \text{ and } (N-1)\lambda = 2\Lambda \sin \theta_{n-1} \qquad (1.34)$$

where Λ is the thickness of one period of superlattice often consisting of two layers, $\lambda(CuK_\alpha) = 0.15405$ nm, θ_n and θ_{n-1} are the nth-order and n − 1 th-order peak in RC. By combining these equations:

$$\frac{2\sin\theta_n - 2\sin\theta_{n-1}}{\lambda} = \frac{1}{\Lambda} \tag{1.35}$$

The positions of satellite peaks of order n:

$$\frac{2\sin\theta_n - 2\sin\theta_{SL}}{\lambda} = \pm\frac{n}{\Lambda} \tag{1.36}$$

where θ_{SL} stands for the superlattice peak or the zero-order peak.

As an example, analysis of two types of superlattices of III–V material are presented in Fig. 1.23a, b. The figures illustrate the structure of two superlattices with InAlAs buffer/InGaAs/InP, and InAlAs buffer/InGaAs/InAlAs grown on InP substrate. The signature of superlattice structures is satellites peaks which are marked as 0, 1, 2, … to n'th-order, but also a distinguishable peak from InAlAs buffer layer with thickness fringes. The first task in the analysis is to identify the InAlAs peak and zero-order order peak in the rocking curve. One problem is the thickness fringes of buffer layer appears in low intensity part due to thick thickness and they are among the superlattice satellite peaks (blue-marked) and it can be usually difficult to be recognized. One may claim to find these peaks easily otherwise to be sure about the peak information, a reference sample containing only the buffer layer can be grown and be analyzed by XRD.

The zero-order peak may be located to the right or left side of the substrate peak depending on the strain type of material. As an example, the sample in Fig. 1.23a has compressive strain, while the sample in Fig. 1.23b has tensile strain compared to InP substrate [28].

The interference fringes for InAlAs buffer layer have been marked in the Fig. 1.22a, b by red arrows. In the figures, $\theta_{sub} = 31.69545$ while theoretical value for (004) reflection for InP is $\theta_{InP} = 31.6680$. The thickness of a period in superlattice (Λ) is determined by using input angles of $\theta_4 = 32.29795$, and $\theta_{SL} = 31.67345$, which are blue-marked by arrows in Fig. 1.23a, and applying Eq. 1.36.

$$\frac{2\sin\theta_4 - 2\sin\theta_{SL}}{\lambda} = \pm\frac{4}{\Lambda} \tag{1.37}$$

where $\lambda = 0.15405$ nm which gives $\Lambda \approx 34.2$ nm.

In Fig. 1.23a, b, the even peaks have remarkably weaker intensity than the odd order peaks. This shows that the thicknesses of the two layers in SL are quite similar and they are actually $\Lambda/2 = 17.1$ nm. The offset angle of the substrate can be calculated from the difference between the theoretical value of Bragg's angle for InP (θ_{InP}) and the measured value by XRD (θ_s):

$$\text{offset} = \theta_s - \theta_{InP} \tag{1.38}$$
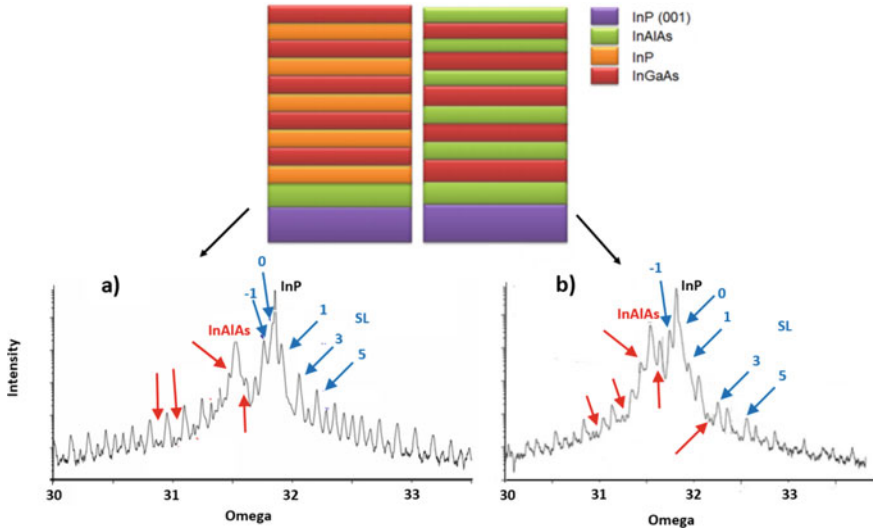
Fig. 1.23 RCs and structures of superlattices with 5 periods of III–V materials with InAlAs buffer layer grown on InP substrate when **a** InGaAs/InP, and **b** InGaAs/InAlAs

where $\theta_s = 31.69545$ and theoretical value for $\theta_{InP} = 31.668$. Then by inserting θ_s and offset angle the following equation

$$\frac{a}{2} \sin(\theta_{SL} - offset) = \lambda \tag{1.39}$$

The lattice constant $a_{InGaAs} = 5.874$ is determined which by inserting lattice constant $a_{InAs} = 6.058$ and $a_{GaAs} = 5.623$, the composition of InGaAs in the SL is estimated to be $x = 0.577$ [28].

1.1.6.4 High-Resolution Reciprocal Lattice Mapping

High-resolution reciprocal lattice mapping (HRRLM) is a series of $\omega - 2\theta$ RCs around a reflection plane, when in each scan the incident angle, ω is varied by a $\Delta\omega$ following an iteration: $\omega_{i+1} = \omega_i \pm \Delta\omega$. This series of RC's are compiled and due to intensity variations of successive ω_{i+1}, contours are formed as illustrated Fig. 1.24. The presence of defects in the investigated material creates distortion in the crystal matrix which acts as scattering centers for the diffracted beam. As a result, Bragg diffraction could be observed for an interval of incident beam angles depending on the amount of defect type and density. The idea behind this method is to find out how much the lattice distortion has affected Bragg diffraction. HRRLMs of defected material display a broadening around the layer and substrate peaks [19].

Fig. 1.24 A schematic view of how a HRRLM is performed. The rocking curves are compiled and the contours with different intensities are seen from top side

Out of the Bragg angle

ω-2θ direction
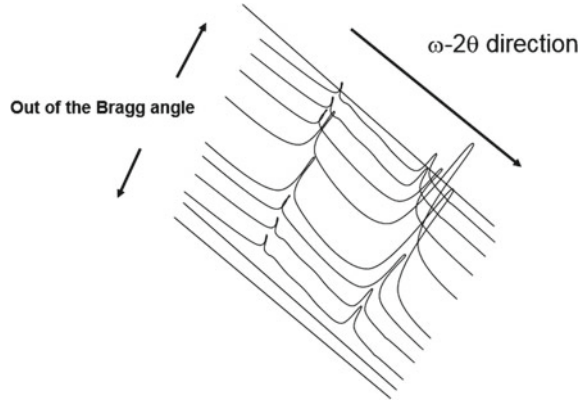
Plotted in this manner, the position in the map of the layer peak relative to substrate peak provides the parallel and perpendicular mismatch parameters, while the feature of the peaks illustrates the defect density. The lattice mismatches can be derived from four input parameters of ω_{sub}, ω_{lay}, θ_{sub} and θ_{lay} for substrate and layer peaks obtained from HRRLMs as following:

$$f_z = \frac{\Delta a_z}{a_{sub}} = \frac{a_{lay}^z - a_{sub}}{a_{sub}} = \frac{\sin\theta_{sub}\cos(\omega_{sub} - \theta_{sub})}{\sin\theta_{lay}\cos(\omega_{lay} - \theta_{lay})} - 1 \tag{1.40a}$$

$$f_{xy} = \frac{\Delta a_{xy}}{a_{sub}} = \frac{a_{lay}^{xy} - a_{sub}}{a_{sub}} = \frac{\sin\theta_{sub}\sin(\omega_{sub} - \theta_{sub})}{\sin\theta_{lay}\sin(\omega_{lay} - \theta_{lay})} - 1 \tag{1.40b}$$

where in these equations the indices "sub" and "lay" denotes for the substrate and the layer peak, respectively. The symbols 'z' and 'xy' refer to perpendicular and parallel to the plane components, respectively.

For totally strained materials, the in-plane mismatch is aligned to the substrate. Therefore, $f_{xy} = 0$, while $f_z > 0$ for compressively strained crystals and $f_z < 0$ for tensile strained materials. The total mismatch, f can be written in terms of f_z and f_{xy} components [29–32]:

$$f = (f_z - f_{xy})\frac{1 - v}{1 + v} + f_{xy} \tag{1.41}$$

where v stands for Poisson ratio of the material. In Eq. 1.41, v describes how a crystal elastically responses when a force is applied. This v can be written in terms of elastic constants (see also Table 1.4) [33].

$$v = \frac{c_{12}}{c_{12} + c_{11}} \tag{1.42}$$

where the elastic constants, C_{ij} has to be determined in order to calculate the Poison ratio. As an example, the elastic constant for an alloy e.g. $Ge_{1-x}Si_x$ is written as:

$$C_{ij}(Ge_{1-x}Si_x) = (1-x)C_{ij}(Ge) + xC_{ij}(Si) \tag{1.43}$$

The strain relaxation (R) is an important value which indicates how much strain has been released. Relaxation is expressed in percentage and it can be obtained as:

$$R = \frac{f_{x,y}}{f} \tag{1.44}$$

For alloy materials, the composition is commonly determined by applying Vegard's law:

$$f = \frac{x \times a_B}{a_A} \tag{1.45}$$

where x is the composition and a_A and a_B are the lattice constants of materials in the alloy. For more accurate values a parabolic equation can be introduced, instead of the linear Vegard's law. As an example, for $Si_{1-x}Ge_x$ alloy, the Ge content is obtained from total mismatch value according to the following equation [38]:

$$f(x) = 3.675 \times 10^{-2}x - 5.01 \times 10^{-3}x^2 \tag{1.46}$$

The compressive/tensile strain in an alloy can be compensated when atoms with different radial size are introduced into the lattice to counteract the induced strain. Strain compensation is realized either in ternary or quaternary materials, e.g. $Si_{1-x-y}Ge_xC_y$ and $Ge_{1-x-y-z}Sn_xSi_yC_z$, or by high doping, for instance B-doping in $Si_{1-x}Ge_x$ layers.

There are also different proposed formulae to calculate the lattice constant for binary or ternary material systems in group IV. As an example, the lattice constant of $Ge_{1-x-y}Sn_ySi_x$ can be derived from the following equation:

$$a_{GeSi}(x) = xa_{Si} + a_{Ge}(1-x) + \theta_{SiGe}x(1-x) \tag{1.47a}$$

$$a_{GeSn}(y) = ya_{Sn} + a_{Ge}(1-y) + \theta_{SnGe}y(1-y) \quad y < 0.20 \tag{1.47b}$$

$$a_{GeSnSi}(x,y) = a_{Ge} + \Delta_{SiGe}x + \theta_{SiGe}y + \Delta_{SnGe}y + \theta_{SnGe}y(1-y) \tag{1.47c}$$

Table 1.4 Elastic constants of the group IV materials [33]

Elastic constant	Ge [34]	Sn [35]	Si [36]	C [37]
c_{11} (Mbar)	1.26	0.69	1.67	10.79
c_{12} (Mbar)	0.44	0.29	0.65	1.24

where $\theta_{SiGe} = -0.0026$ nm, $\theta_{GeSn} = +0.0166$ nm, $\Delta_{SiGe} = a_{Si} - a_{Ge}$ and $\Delta_{SnGe} = a_{Sn} - a_{Ge}$ [39].

The Fig. 1.25 describes qualitatively the various signatures of an HRRLM plot.

As an example, Fig. 1.26a–c illustrate three HRRLMs from a SiGe/Si sample where the compressively strained SiGe layer with high epitaxial quality (Fig. 1.26a) has been thermally treated at 800 °C and 1000 °C (Fig. 1.26b, c, respectively) to increase the strain relaxation. As a result, the threading dislocation density (TDD) increases with increasing annealing temperature. The following points have to be noticed:
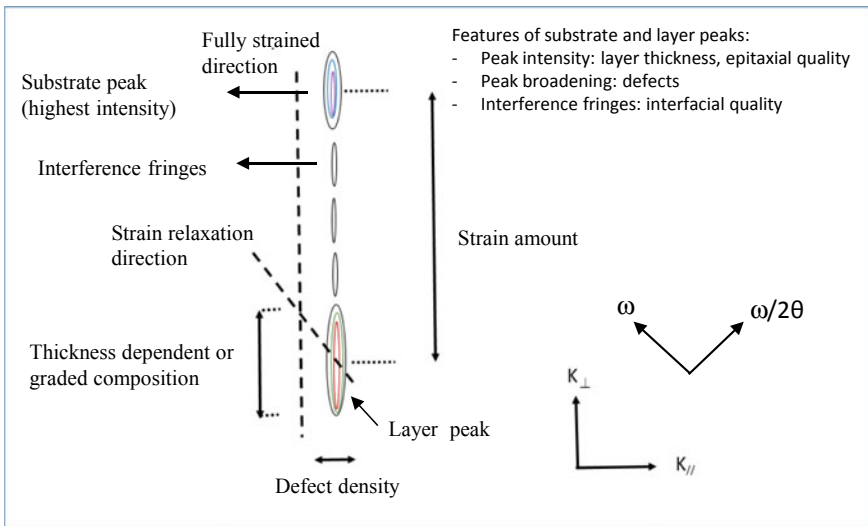
Fig. 1.25 A schematic drawing of a HRRLM with important indicators for defect density, strain amount, layer thickness and material composition. In the figure, the unit vectors $K_{//}$ and $K_{\perp}$ as well as the scanning directions off ω and $\omega/2\theta$ are marked
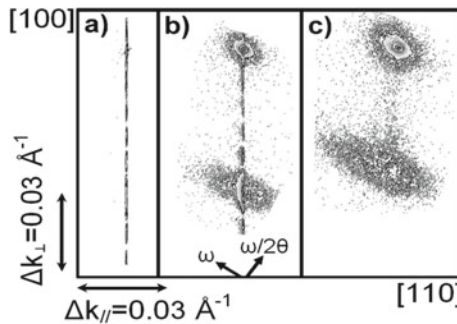
Fig. 1.26 HRRLMs from a SiGe/Si sample, where **a** is as-grown, and **b** and **c** are samples annealed at 800 °C 1000 °C, respectively [30]

1. The alignment of Si and SiGe peaks along $K_\perp$ in Fig. 1.26a is an indicator of strained SiGe with minor strain relaxation
2. The broadening of Si and SiGe peaks in Fig. 1.26b, c shows the presence of threading dislocations. The number of interference fringes between the Si and SiGe peaks is decreasing by increasing the annealing temperature showing interface quality of SiGe/Ge is deteriorating
3. The split of SiGe and Si peaks is an indicator of the amount strain and it decreases by annealing treatment in Fig. 1.26b, c.

According to the points 1–3 above, the figure indicates that SiGe in Fig. 1.26a is totally strained with low defect density. Meanwhile SiGe layers in Fig. 1.26b, c are defected where the layer in Fig. 1.26c is mostly strain relaxed.

1.1.6.5 Application of High-Resolution XRD

A precise strain measurement can be performed by using HRRLM when the misfit parameters in-plane and perpendicular direction (f_{xy} and f_z, respectively) can be measured around a reflection (or reciprocal point) in reciprocal space. The choice of reflection is important to consider, since the incident angle is different causing the penetration length of the x-ray beam, as well as the acquisition of information, will be different. For example, a (113) reflection of Si with incident angle of 2.8° is more sensitive to the defects located closer to the surface compared to (224) and (115) reflections with incident angle of 8.7° and 31.7° [30].

As an example, Fig. 1.27 a-c illustrates HRRLMs of a $Si_{1-x-y}Ge_xC_y$/Si multi-layer around (113), (224) and (115) reflections. TEM analysis of this multi-layer structure is illustrated in Fig. 1.27d. In this structure, the germanium composition has increased successively from 17 to 28% from the substrate, meanwhile the carbon content was kept constant at 1%. Since the incorporation of carbon is dependent on the Ge content, finally the layer with high carbon content, $Si_{0.73}Ge_{0.26}C_{0.01}$ and the cap layer became defected in the multilayer structure.

HRRLM in Fig. 1.27a, shows a large intensity scattering around the Si peak which is mainly generated from the defected cap layer and it is overlapped with other peaks from Si layers in the structure and the Si substrate. The scattering around the Si peak increases in Fig. 1.27b but decreased remarkably for Fig. 1.27c. A similar behavior is observed for $Si_{0.73}Ge_{0.26}C_{0.01}$ among other SiGeC peaks. These results show that HRRLM around (113) is most sensitive to the defects close to the surface compared to maps around (224) and (115) [30].

1.1.6.6 Strain Design in Nanoscale Devices

One of the major trends for semiconductor industry is to follow Moore's law and to constantly increase the number of transistors per chip area. One important part to follow this law has been to control biaxial and uniaxial strain in the channel region
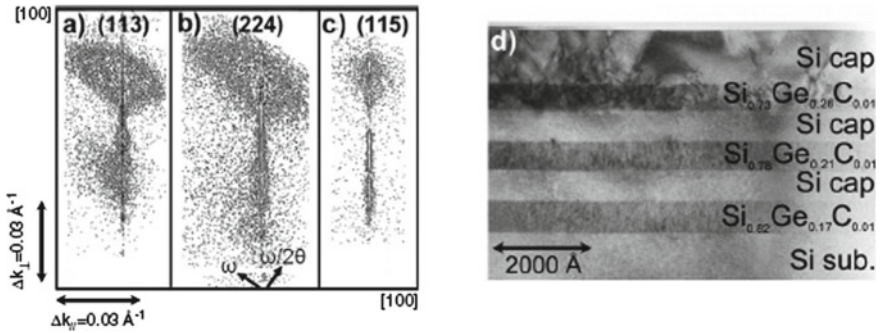
Fig. 1.27 HRRLMs of $Si_{1-x-y}Ge_xC_y$/Si multi-layer obtained at reflections: **a** (113), **b** (224) and **c** (115) and **d** TEM cross-section image of the structure. The Ge and C contents are written in the TEM micrograph. The HRRLMs highlight the sensitivity of measurements to defect density in the epi-layers [30]

[40]. In uniaxial strain, a strained material is selectively deposited in the source/drain areas as stressor source for the channel region, whereas biaxial strain can be induced by global deposition of stressor material over the entire wafer as channel material. High-resolution reciprocal space maps can be applied to directly measure the strain and strain relaxation during different process steps, such as thermal treatment for dopant activation and silicide formation over the transistor chip. By focusing the x-ray beam over a chip, signals from many nanoscale transistors are collected in the detector and, recently, several reports have demonstrated XRD as powerful tool to measure strain in sub 10 nm technology node [41].

High-resolution reciprocal space maps is also a characterization method to determine the substitutional dopant concentration in Si and group IV alloys. For these cases, the strain compensation in presence of dopants can be determined and interpreted into the substitutional dopant concentration.

As an example, the strain analysis has been applied to determine the SiGe profile in source/drain regions for forming silicides to reduce contact resistance [42]. The main concern relates to the point defects which are generated during the silicide formation. In this process the Si atoms are consumed and the Ge atoms are pushed out of the growing silicide into the remaining SiGe layer, resulting in an increasing Ge composition. The resulting SiGe below the formed silicide layer then becomes strain relaxed. The best solution to this problem is to introduce a sacrificial SiGe layer with low Ge content on the top layer which can be consumed during the silicide formation and the active SiGe is untouched [42].

As an example, Fig. 1.28a–c show three HRRLMs around (113) reflection measured on a 22 nm pMOS chip with **a** intrinsic $Si_{0.65}Ge_{0.35}$, **b** B-doped $Si_{0.65}Ge_{0.35}$ and **c** B-doped $Si_{0.65}Ge_{0.35}$/NiSiGe in S/D areas [43]. In this case, a sacrificial layer consisting of a $Si_{0.80}Ge_{0.20}$ cap layer was introduced by using rapid thermal annealing at 500 °C for 30 s in N_2 ambient.
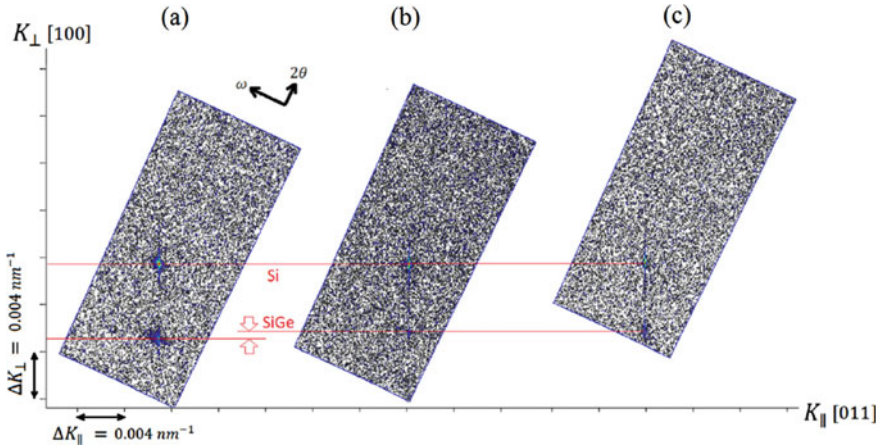
Fig. 1.28 HRRLMs from 22 nm pMOS with $Si_{0.65}Ge_{0.35}$ stressor material in source/drain regions when **a** SiGe intrinsic layer and **b** B-doped SiGe layer **c** as b but with a $Si_{0.80}Ge_{0.20}$ sacrificial cap layer used for Ni silicide formation [43]

The SiGe peaks in Fig. 1.28a, b are aligned to Si along $K_\perp$ direction showing a negligible strain relaxation in the SiGe layer. Meanwhile, in Fig. 1.28b, the position of the B-doped SiGe peak is closer to the Si peak as a result of strain compensation caused by B atoms in SiGe matrix. The difference of the intrinsic and doped SiGe positions along $K_\perp$ direction is converted to a boron concentration of 2×10^{20} cm^{-3} by using the contraction coefficient of boron in Si ((6.3 ± 0.1) $\times 10^{-24}$ cm^3/atom).

Finally, Fig. 1.28c illustrates that the position of the SiGe peak is not altered during the NiSiGe formation and the strain is kept constant [43]. The thickness of the sacrificial layer is an important parameter to preserve the strain.

The strain analysis in transistors with smaller size than 22 nm, e.g. for the 14 nm technology node, synchrotron facilities has been used for strain analysis of transistor chips, as shown by the HRRLM in Fig. 1.29 [44]. The high intensity and well defined energy of the synchrotron x-ray source gives the possibility for acquisition of sufficient signal levels to determine the strain in very small SiGe crystals.

For sub 10 nm nodes, the XRD technique produce too weak signal and cannot provide information for strain analysis.

1.1.7 Reflectivity Measurement

In x-ray reflectivity (XRR) measurement, the x-ray is used as an electromagnetic wave to provide information of the surface, or interface, roughness, as well as the layer thickness, or the density of multilayer [45–49]. In XRR analysis, the sample can be single or poly crystalline, or even amorphous. The principle of XRR is based on the basic optical laws of refraction and reflection of light, thus it does not need to obey

Fig. 1.29 HRRLM at (− 115) reflection from 14 nm pMOS chip with SiGe material in source/drain regions [44]. The SiGe peak is aligned with the Si peak along $K_\perp$ showing a minor amount of strain relaxation
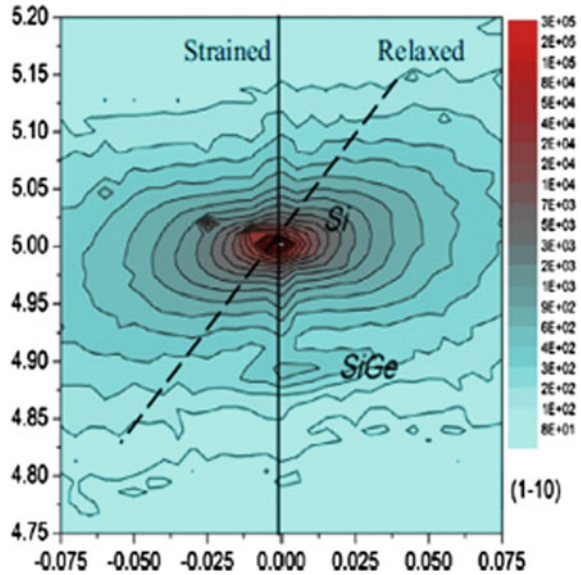
Bragg's law for diffraction. Furthermore, by using a very small incident angle, i.e. the angle between the surface and the direction of the incident x-rays in XRR, thin films with thicknesses of only a few nanometers can be measured. For Bragg reflections of crystal planes, such gracing, or glancing, incident beams results in very weak Bragg reflection peaks. The set-up for XRR measurement has to be carefully chosen. The primary optics consists of mirror, monochromator, an automatic attenuator, while the secondary optics consists of a parallel plate collimator and the detector. The reason for the choice of the components in the primary optics is the need of sufficient high intensity beams, but the automatic attenuator cuts extremely high intensity beam to ensures that no damage occurs to the detector. In the secondary optics part, the diffracted beam has to be collimated by a parallel plate collimator in order to obtain decent results in XRR measurements.

Figure 1.30 illustrates typical XRR curves when different material parameters are varied. The following features can be observed from the XRR curves:

1. The x-ray beam refracts and reflects in the material, creating an interference pattern (fringes) due to the different refractive indices at an interface.
2. The materials have electron density contrast in the sample resulting in larger amplitude signals in the fringes. The density of the material is also distinguished from the critical angle where the material composition is estimated.
3. The layer thickness is estimated from two neighboring fringes.
4. The surface roughness is obtained from the rate of signal decay.

There are also simulation programs which can be applied and fitted to the experimental data to quantify the material composition, thickness and interface roughness.

Fig. 1.30 Typical XRR curves and their informative features
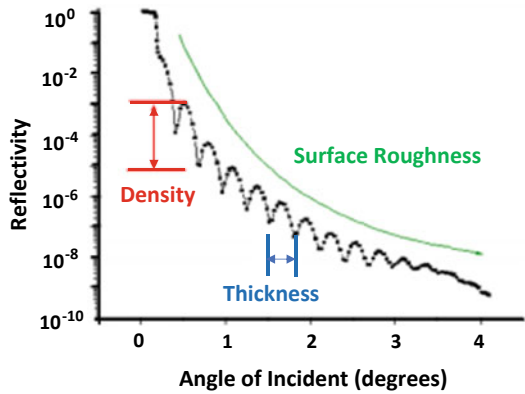
However, the best use of such simulation programs is obtained when several of the material parameters are known.

In reflectivity measurements, the surface reflection for a certain wavelength, i.e. how mirror like surface or smooth interface, has a central role in the features of XRR curves. The most dramatic changes appear when the surface roughness causes diffused scattering wings which are known as Yoneda wings. These diffused Yoneda wings appear when the incidence or exit angles are equal to the critical angle for total external reflection.

In reflectivity measurement the incident beam is around or below 1 degree which makes the x-ray beam shinning a large area of the surface. An estimation of the x-ray beam projection 'x_{proj}' with incident angle of 'θ' when a slit with a width of 'd' is d/sinθ. Furthermore, in reflectivity measurements, if the incident beam is less than the critical angle for the total reflection, d then refraction into the layer occurs which creates interference beams. The refractive index is expressed in complex numbers and is obtained from the following equations [45]:

$$n = 1 - \delta - i\beta \tag{1.48}$$

$$\delta = \left(\frac{r_e\lambda^2}{2\pi}\right) N_0\rho \sum_i x_i(Z_i + f_i') \bigg/ \sum_i x_i M_i \tag{1.49}$$

$$\beta = \left(\frac{r_e\lambda^2}{2\pi}\right) N_0\rho \sum_i x_i(Z_i + f_i'') \bigg/ \sum_i x_i M_i \tag{1.50}$$

where r_e, N_0, λ and ρ denote radius of an electron (2.818×10^{-9} m), Avogadro number, wavelength of x-ray, and density of the layer, respectively. In the above equations, crystal parameters z_i, M_i, x_i, f_i', f_i'' stand for atomic number, atomic weight, atomic ration or molar ratio and atomic scattering factors of an atom at the i-th position, respectively. In principle, in the real part, δ is dependent on wavelength, density and composition of the material and it lies in a typical range of values of 10^{-5}

to 10^{-6} for x-ray wavelength of 0.1 nm. The imaginary part, β is related to absorption and is written by the linear absorption coefficient μ as following [45]:

$$\beta = \lambda\mu/4\pi \tag{1.51}$$

The critical angle (θ_{cri}) is defined as $\theta_{cri} = \sqrt{2\delta}$ and it lies typically in a range of 0.2 to 0.5 degrees and for total reflection. This indicates that the layer density can be obtained from θ_{cri}. This parameter increases when the wavelength of the x-ray beam, or the density of the material increases.

Reflectivity measurements is often linked to grazing-incidence, small-angle x-ray scattering (GISAXS) technique for studying nanomaterials and thin films [46, 47].

The application of 2D measurement using the XRR diffused scattered maps (XRR-DSMs) extends the technique to provide information about surface roughness in the presence of nano materials, island formation within the layer and defect formation. As it was mentioned above, high-resolution reciprocal lattice map has been commonly used as an efficient tool to measure the mismatch (or strain) in directions parallel and perpendicular to the growth direction. The broadening around the layer and the substrate peak could indicate the quality of epitaxial layer. As previously discussed, another way to qualify the layer quality is from the FWHM value of the layer peak in rocking curves. Although such measurements provide very important data, still surface roughness and interface quality, which are indicators for the presence of defects or atomic segregation, cannot be properly observed.

Recently, XRR-DSM is being further developed as an alternative method to analyze the epitaxial layer quality. The primarily application of grazing angle measurements is to study quantum dots where the x-ray shines at a grazing angle on the surface of the sample and is scattered by such dots before being collected in the detector. A view of this type of measurements is illustrated in Fig. 1.31 [48].

GISAXS technique gives information about ordering in lateral and normal direction on a surface, or inside of thin layers. As an example, GISAXS technique has been applied to study lamellar films like copolymers. In a block copolymer, there are two immiscible polymer chains which are chemically linked to each other.

There are different types of arrangements of the blocks depending on the energies between the blocks as shown in Fig. 1.32a–c. For example, the lamellae are formed
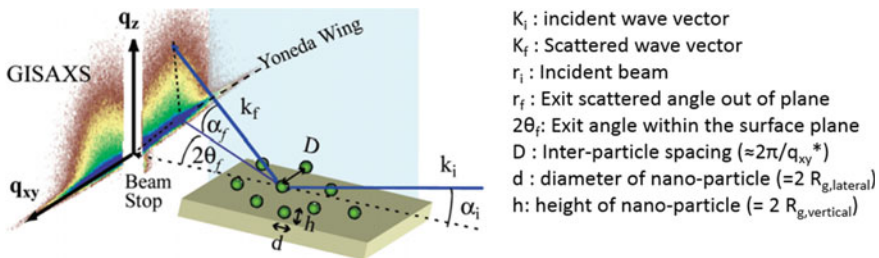
Fig. 1.31 Principle of GISAXS maps for nano particles on a surface [48]
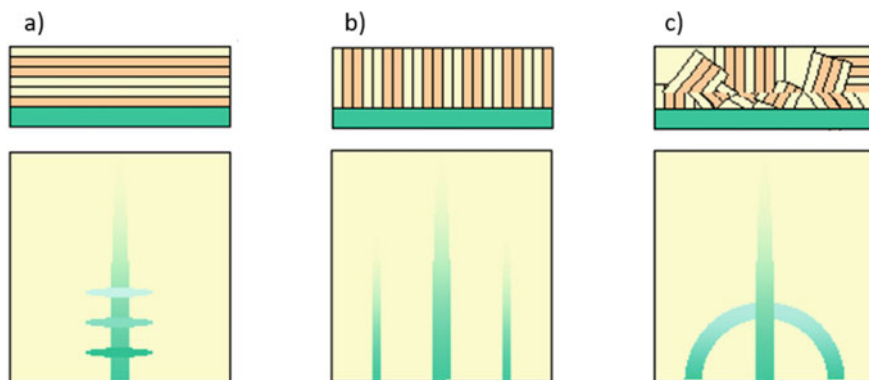
Fig. 1.32 GIAXS maps of **a** horizontal, **b** vertical and **c** randomly ordered lamellae structures. The main peak strongly appears in the center of the map, whereas the signature peaks appear in parallel or horizontal features, depending on the lamellae's orientation. A typical ring pattern merges for disorder or partial ordered lamellar in picture (c) which can be thought of as a combination of signatures in pictures (a) and (b)

in parallel shapes when one of the blocks energetically accommodate one of the two interfaces, or even both, to the substrate. A parallel lamellae configuration with its GISAXS signature is shown schematically in Fig. 1.32a. The map contains stripes with regular spacings along the q_z direction (see Fig. 1.31). These stripes in the diffused reflectivity measurement are known as Bragg sheets. This type of GISAXS pattern is obtained when both incident and scattering angles are well exceeding the critical angles of film and substrate.

Figure 1.32a–c illustrate when blocks have comparable interface energies, then chain stretching at the interface becomes important. Chain stretching takes place at the link between the individual blocks of the polymer. A transition from parallel to vertical lamellae can also occur in the polymers since the interactions are scaled by the degree of polymerization. GIAXS signature for perpendicular lamellae appears through parallel peaks as shown in Fig. 1.32b [49–51].

In thick films, the ordering induced at the interfaces can be unfavorable in the film, and 3D powder bulk structure can be formed. The GIAXS maps signature appears as rings, full or partial, indicating a distribution of domains from complete disordered to partial ordered (or tilted) lamellar as shown in Fig. 1.32c.

As an example, GIAXS technique has been used to study the growth of iron oxide on quartz substrates by using different amount of ferric solutions $Fe(NO_3)_3$, with $NaNO_3$ ionic strength (IS) as shown in Fig. 1.33. In these experiments, heterogeneous nucleation of iron oxide nanoparticles on quartz could be studied in the real time. The exposed quartz substrate showed no visible particle scattering during the first 2–3 min. but after 3 min, two identical scattering lobes appeared in the GISAXS measurements. These lobes reveal oxide nanoparticles of both a well-defined spacing and size formed on the quartz surface. The results showed, furthermore, that the

growth of nanoparticles occurred rapidly laterally but not vertically. Finally, the nanoparticles coalesce, although they did not create a continuous film.

Another example of the grazing angle measurements can be linked to the XRR DSM technique, which correlated to the surface roughness in presence of defects. Figure 1.34a, b illustrates XRR-DSMs of two samples of $Ge_{0.88}Sn_{0.12}$/Ge layers grown on a (100)Si substrate with different interfacial quality. The AFM results (not shown) have revealed that the surface RMS roughness of the sample in Fig. 1.34a is 2.5 nm, where for sample in Fig. 1.33b it is 1.8 nm. The surface roughness is linked to the epitaxial quality in these samples. As a result of lower surface roughness, the diagonal line, which contains interference fringes, is significantly longer in sample in Fig. 1.34b compared to the one in Fig. 1.34a. The x-ray beam at grazing angle is very sensitive to the interface quality. Surface roughness affects the interference
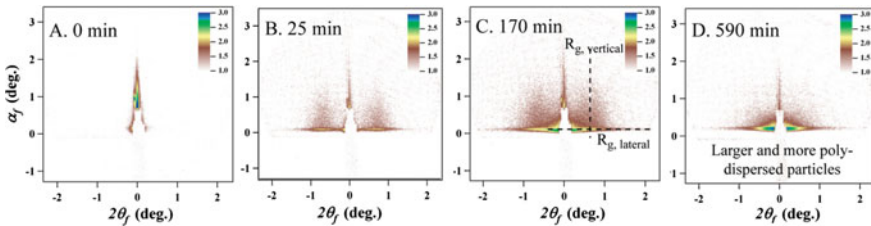
Fig. 1.33 GISAXS images from time series of exposure of quartz substrate to ferric solutions ($[Fe^{3+}]$) 0.1 mM, ($[NaNO_3]$) 10 mM, and pH 3.6. There is a shift of the lobe location toward a smaller $2\theta_f$ showing that the nanoparticles grow larger in size, while the vanishing of the distinct lobes shows the particles have coalesced. Iron oxide particles also start to grow with well-defined inter-particle spacing. Later, the nuclei coalesce, forming larger particles and exhibit a more polydisperse distribution [49]
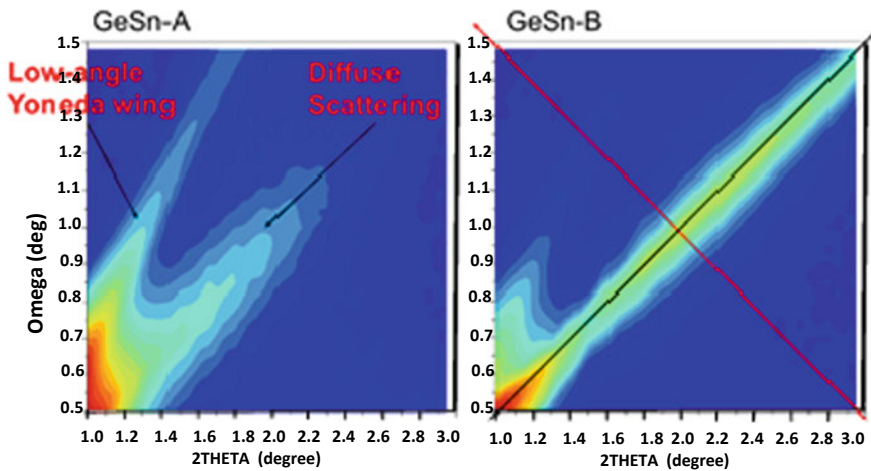
Fig. 1.34 XRR-DSMs obtained from $Ge_{0.88}Sn_{0.12}$/Ge (buffer layer) grown on (100)Si substrate with surface roughness of **a** 2.5 nm, and **b** 1.8 nm

and the low-angle Yoneda wings appear broader. The example shows that the XRR-DSM technique is a new way to investigate the interface quality, which can be further developed in the future.

1.1.8 Analysis of Texture Materials

A single crystal consists of periodic arrays of atoms. Although this perfect crystal arrangement exists, a real single crystal may also display anisotropic properties in the crystallographic directions, which dependent on the crystal symmetry. Many materials are polycrystalline, meaning that they consist of a number of smaller crystals. Such materials are, in fact, very common in our environment and include most metals, many ceramics and minerals. A crystallite (also known as a grain) is one such small crystal in a polycrystalline material where the crystallite orientation can have any direction in relation to other crystallites depending on, for instance, growth and processing conditions.

The word "texture" is defined as the distribution of crystallographic orientations of a polycrystalline material. The material could be either fully randomly oriented (no distinct texture) or some preferred orientation with weak, moderate or strong texture depending on the percentage of preferred oriented crystallites. The degree of crystallinity becomes larger when more crystallites are involved in the material. It is important to point out that the degree of crystallinity determines much of the material properties, such as electrical transport, optical response, and mechanical properties of materials. When a material is polycrystalline, it is therefore important to determine the distribution of the crystallite orientations, i.e. the texture.

Before we elaborate further on the material analysis, the difference between the crystal symmetry and sample symmetry in textured samples have to be explained. In fact, the crystal symmetry always exists, but can be disturbed, for instance, by the presence of high concentrations of defects. Meanwhile, the sample symmetry relies on a statistical basis with data from many crystallites of a polycrystalline material with texture character.

Until now, this chapter has dealt with different x-ray analyses: power diffraction, high-resolution XRD and reflectivity with grazing angle. In all these measurements, the scans are performed through changing stepwise the incident beam (θ or ω), while the detector angle, or the diffracted beam angle, 2θ is used to record the intensity of the x-ray in a Bragg Brentano geometry. In order to study materials with different degree of crystallinity and texture, we need a more advanced goniometer than used up to now, shown in the x-ray instrument set-up in Fig. 1.19. In such an advanced goniometer, there are two further angles: in-plane rotation angle (ϕ) and tilt angle (ψ) making a 4-circles diffractometer. In this way, x-ray technique can be used to create pole figures (PFs) for finding out the distribution of crystallite orientations in a polycrystalline sample. PFs are established when the sample is rotated around axes (ψ and φ) while the θ and its 2θ corresponding to the Bragg reflection of interest can be kept constant. Then, the orientation distribution (OD) in a 3D volume is expressed

$\phi1=0$ $\phi1=45°$ $\phi1=45°$ $\phi1=45°$
$\psi=0$ $\psi=0$ $\psi=17°$ $\psi=17°$
$\phi2=0$ $\phi2=0$ $\phi2=0$ $\phi2=10°$
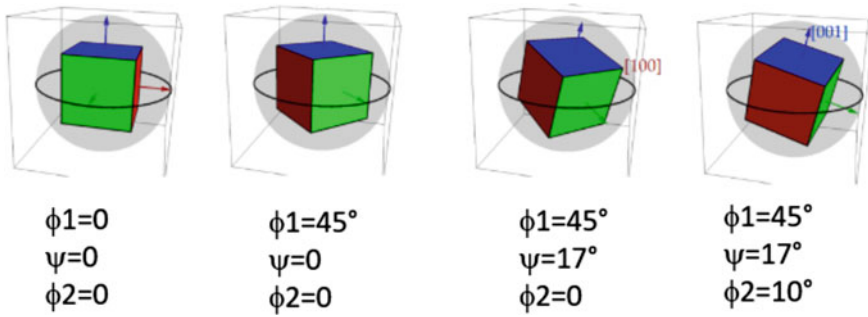
Fig. 1.35 An illustration of a unit cell orientation in terms of Euler's angles

in terms of Euler angles, which relates the three space angles of a body in relation to a reference system (x, y, z). $\{\phi_1, \psi, \phi_2\}$. Figure 1.35 shows some examples of unit cells with different orientations in the space.

Pole figures are defined as a 2D graphical representation of orientation, presenting the orientation of a certain plane normal (so-called a pole) with regards to the sample reference frame [52]. The principle of forming of a PF has been presented in Fig. 1.36a, b using the following steps:

(i)   In a in polycrystalline material, consider a unit sphere in a grain.
(ii) A projection plane can be considered at the equator position of the sphere where a unit cell of crystallite sample is placed at the origin. Draw the normal vectors from each side of the unit cell and determine the intersection points of the normal vectors to the unit sphere in the upper hemisphere.
(iii) Connect the intersection points to the south pole in lower hemisphere.
(iv) The intersection points to the projection plane are in the 3D space and to be mapped on to a 2D plane to form a PF.

There are special notations to describe the planes and orientations specially for the texture materials. For example, (hkl) or {hkl} denotes a plane or set of planes
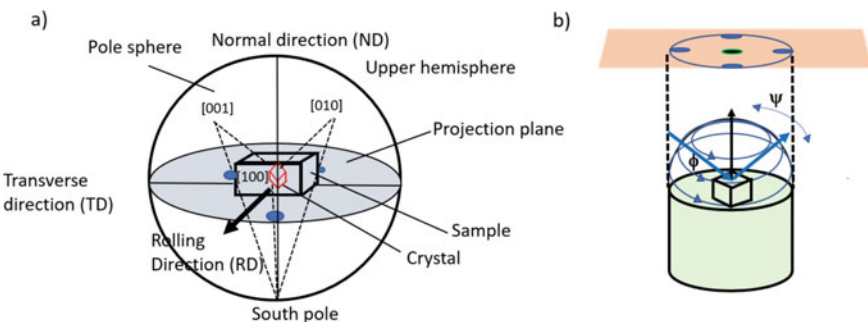
Fig. 1.36 a, b Principles of how a PF is drawn
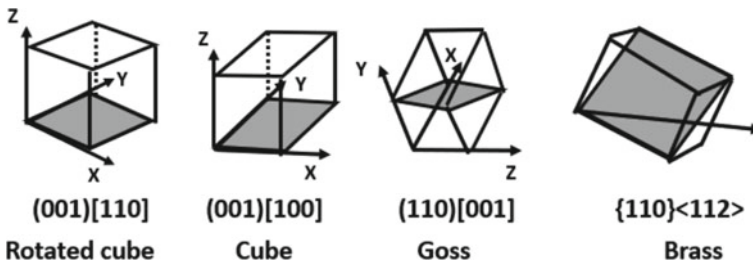
Fig. 1.37 An illustration of the cubes with different orientations and planes

and [uvw] or <uvw> stand for the vector direction or set of symmetrically equivalent directions. In the PF analysis, some special words e.g. "Goss, Brass, cube, … elements" or simply "(uvw) [u′v′w′] elements" are often mentioned for the polycrystalline material as displayed in Fig. 1.37.

Figure 1.38 shows different examples of PFs providing an understanding about the effect of the tilt of crystallites. As an example, to create a PF of (111) planes from a Si sample: First, the detector angle (2θ) has to be set for Si (111) planes (for the commonly used CuK$_\alpha$ sources, this is ≈ 28.4°). The scan occurs by rotating φ, and incrementally increase ψ. This measurement provides a map over all directions where Si (111) crystallites in the layer are oriented. Due to the crystal symmetry, the data at four points at ψ = 45° can be observed. The background is glass and it is amorphous.

There is also a complementary method to the pole figure described so far, the so-called inverse pole figure (IPF). The IPFs demonstrate the sample direction relationship to the crystal reference system. The difference between these measurements is that the plane of interest is chosen for PFs while for IPFs the sample direction of interest is chosen. The axes in PFs are the sample directions but for IPFs, the axes are crystallographic directions as shown in Fig. 1.39.

It could be more intuitive to establish an IPF as showing which plane normal in the crystal is parallel to the selected sample direction [53]. As an example, an IPF is suitable to be plotted instead of a PF for the <111> directions in connection to the rolling direction (RD) and normal direction (ND). In this case, an IPF can be plotted to demonstrate the rolling direction in regards to common directions [100], [110] and [111].

In general, the form of an IPF unit triangle depends on which Laue group the crystal belongs. As shown earlier in Fig. 1.2, there are following crystal systems: Triclinic, Monoclinic, Orthorhombic, Tetragonal, Trigonal, Hexagonal and Cubic and they belong to Laue classes. In the Laue classes, low-symmetry crystal systems (e.g., Triclinic, Monoclinic and Orthorhombic) have a single corresponding symmetry for the diffraction data, meanwhile, for the high-symmetry crystal systems (e.g., Hexagonal, Tetragonal, Trigonal, Hexagonal and Cubic), there are two symmetry options [54].
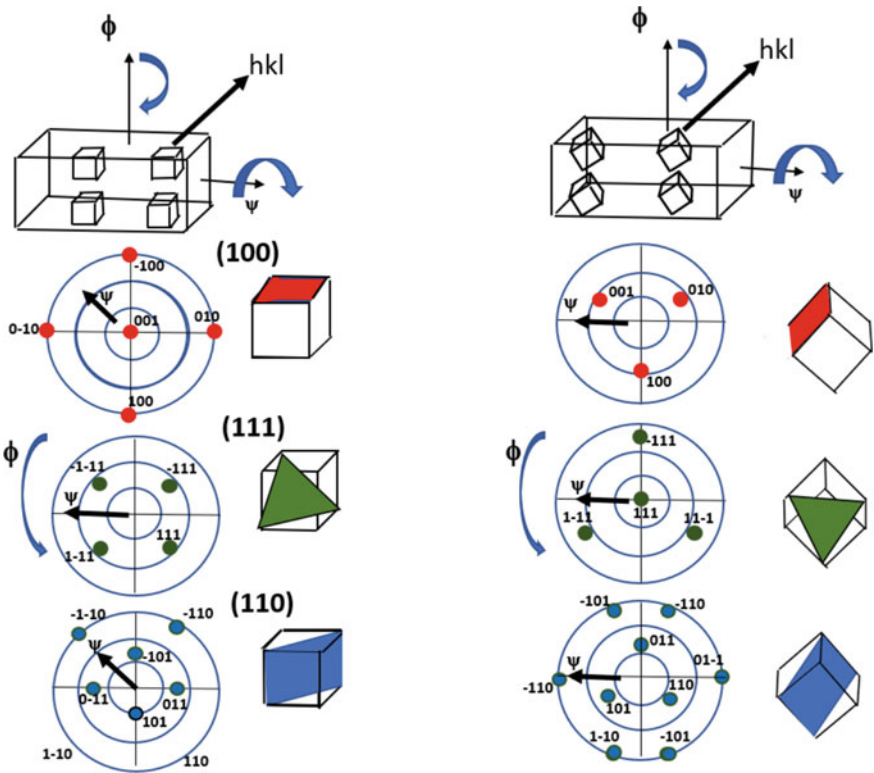
Fig. 1.38 Examples of PFs for different planes and orientations
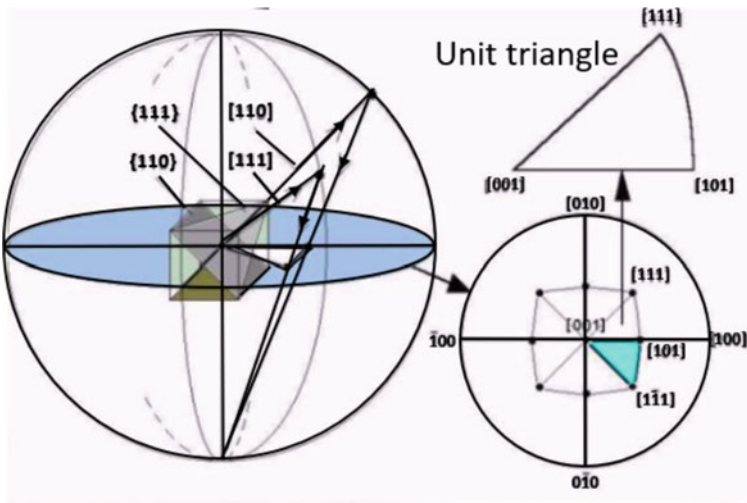
Fig. 1.39 An Illustration how an IPF is established

The Laue class will decide the size of the "unit triangle" in the entire stereographic projection. As an example, a triclinic crystal belongs to Laue group 1 and has no symmetry. Then, a complete PF is required to demonstrate all the possible crystal directions. As a result, for triclinic structures, the IPF is similar to a PF and is a full circle.

But as structure symmetry is increased, the size of the symmetrically equivalent region is decreased. For example, this is the case for Laue group 11 (cubic high symmetry) where 24 symmetrically equivalent regions exist in the full project circle.

One of the analysis techniques for studying of polycrystalline materials is electron backscattered diffraction (EBSD) technique, built on scanning electron microscopy (SEM), further described in Chap. 4 EBSD technique is based on electron backscatter diffraction patterns formed of the Kikuchi bands [55]. A devoted EBSD detector is able to collect the back scattered electrons within a large solid angle in a field emission scanning electron microscope. EBSD provides micro texture analysis in terms of the crystal orientation, size and morphology of the grains. It is important to mention here that there are some issues which may introduce artefacts in EBSD analysis:

  i. How the statistical sampling has been done since the grain size estimation can be affected.
 ii. The quantified phase fraction varies with the magnification since this decides the extent of statistical sampling
iii. The specimen preparation is a mechanical process when electro-polishing method is usually used and this can damage the pattern quality.

As an example, a study of Ti alloys is presented here [56]. There is a large interest to study the phase transformations and microstructural characteristics of Ti-alloys. In this example, the specific microstructural features are α (hexagonal) phase and β phase (BCC) in Ti–6Al–2Sn–4Zr–6Mo alloy (or Ti-6246). The Ti-alloys are classified into five main classes; α, near-α, α + β, near-β, and β-alloys, depending on the relative α/β fraction.

Figure 1.40a–c show a detailed study of Ti-6246 layer by using EBSD and PFs. Ti-6246 material consists of different phases and the fraction of formed phases is important for understanding the properties of the material. Here, x-rays were provided by a synchrotron facility in order to obtain a sufficient intensity for accurate analysis. The EBSD map is based on the backscattered electrons in a SEM. The number of elastically backscattered electrons is sensitive to the target atoms atomic number (Z) and is higher for a heavier material, creating a so-called Z-contrast (see more details in Chap. 4). However, other types of contrast can also be established when the number of the atoms in the grains, as well as when the crystal structure in the grains is different. Therefore, a distinguishable contrast due to the phase change over the sample is observed in Fig. 1.40a and in this way the fraction of the different phases can be estimated.

In EBSD analysis, the quantified phase fraction varies with the magnification of the electron microscope since this decides the extent of statistical sampling. For example, in Fig. 1.40a, the indexed β fraction at low magnification (250×) is 23.3% meanwhile at high magnification (1000×), the β fraction was 11.8%. An overall β
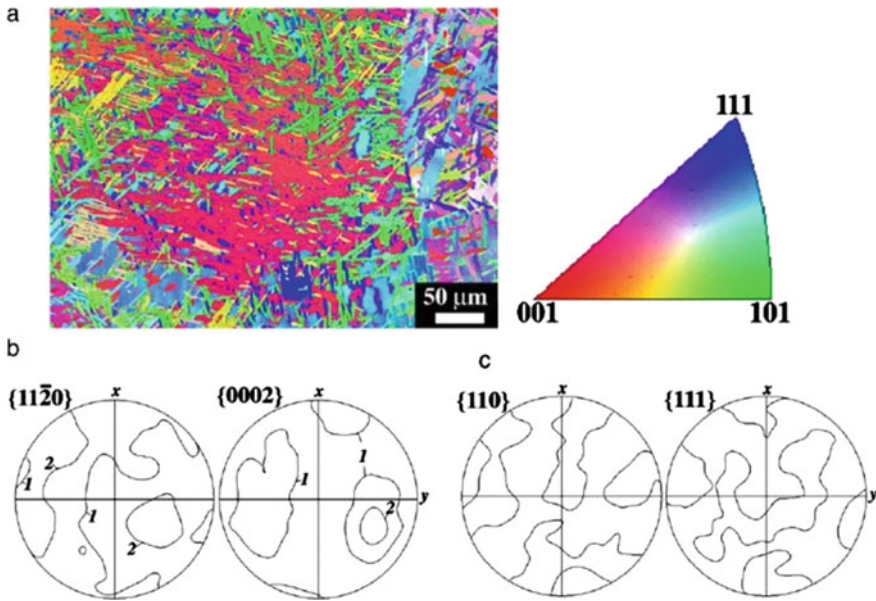
Fig. 1.40 **a** EBSD map for Ti-6246 and IPF, **b** PFs for the α-phase, and **c** β-phase, demonstrating weak textures in both phases. The PFs were formed from an area of 5×5 cm^2

fraction average was ~21.8 $\pm$ 11.1%. It is important to notify that EBSD analysis has to be carefully performed since, as we mentioned earlier, artifacts related to sample preparation and statistical sampling can appear. In the figure, IPF shows the distribution of the phases from EBSD map.

In Fig. 1.40b, c, PFs of both α and β phase samples show a kind of pattern which indicates weak texture in the material. Quantitative phase analysis PF data was done by applying Rietveld refinement [57]. This task can be accomplished by using different software programs e.g. TOPAS software [58] using parameters for α-Ti and β-Ti phases separately.

The orientation distribution function, or orientation density function (ODF), is a function which gives information about the orientation of a volume percentage of crystallites in a polycrystalline sample. ODFs demonstrate the three Euler angles and a density value indicating the amount of the contribution of a specific orientation. ODFs is shown as a ratio to that expected for a completely random distribution of orientations. A mostly used method to calculate ODFs is to apply a "Series Expansion Method". This is a two-step approach which initially calculates the texture coefficients and afterwards uses those coefficients to generate the ODF. The ODF densities can later be shown in 3D Euler space, but commonly ODF are plotted in a series of 2D sections by holding one Euler angle constant. There are even textures where the ODF can be displayed in 1D by keeping two Euler angles fixed. There are also reports showing the difficulties for reproducing ODF from the PFs. In general, the content of PFs is not enough for reproducing a complete ODF. The problem roots

from the series expansion method which is commonly used to reproduce ODF from PFs from data from texture is in form of Fourier coefficients. Therefore, ODF is typically obtained from both x-ray diffraction PFs and EBSD [53].

In summary, x-ray technique is considered as one of the most important characterization tools in material science. The technique is non-destructive, fast and cost effective and mainly used for crystalline and polycrystalline materials. The applications of x-ray technique is constantly increased and it will be one of the most popular material analysis tools in future.

References

1. Berger, H. (2012). The mystery of a new kind of rays: The story of Wilhelm Conrad Roentgen and his discovery of x-rays. ISBN: 1475239971.
2. Roentgen, W. C. (1896). On a new kind of rays. Translation from German to English of Roentgen's first x-ray publication by A. Stanton. *Science, 3*, 227–231.
3. Föll, H., & Kolbesen, B. (1976). Aggllomerate von Zwischengitteratomen (Swirl-Defekte) in Silizium—ihre Bedeutung für Grundlagenforschung und Technologie Jahrbuch der Akademie der Wissenschaft in Göttingen, 27.
4. Nabarro, F. R. N., & Hirth, J. P. (2007). Dislocations in solids, Vol. 13, ISSN-13: 978-0-444-51888-0.
5. Benedek, G. Point and extended defects in semiconductors. Springer, Nato Science Series of B, ISBN 978-1-4684-5709-4.
6. Sumino, K. (1990). Defect control in semiconductors. ISBN 978-0-444-88429-9.
7. Bioud, Y. A., Boucherif, A., Myronov, M., Soltani, A., Patriarche, G., Braidy, N., Jellite, M., Drouin, D., & Arès, R. (2019). Uprooting defects to enable high-performance III–V optoelectronic devices on silicon. *Nature Communication, 10*, Article number: 4322.
8. Radamson, H. H., He, X., Zhang, Q., et al. (2019). Miniaturization of CMOS. *Micromachines-Basel, 10*, 293.
9. People, R., & Bean, J. C. (1985). Calculation of critical layer thickness versus lattice mismatch for GexSi1-x/Si strained-layer heterostructures. *Applied Physics Letter, 47*, 229.
10. Dodson, B. W., & Tsao, J. Y. (1988). Stress dependence of dislocation glide activation energy in singlecrystal silicon-germanium alloys up to 2.6 GPa. *Physics Review B, 38*, 12383.
11. Dodson, B. W., & Tsao, J. Y. (1989). Scaling relations for strained-layer relaxation. *Applied Physics Letters, 55*, 1345.
12. Yue, L., Nix, W. D., Griffin, P. B., & Plummer, J. D. (2005). Critical thickness enhancement of epitaxial SiGe films grown on small structures. *Journal of Applied Physics, 97*, 43519.
13. Radamson, H. H., Bentzen, A., Menon, C., & Landgren, G. (2002). Observed critical thickness in selectively and non-selectively grown Si1_xGex layers on patterned substrates. *Physica Scripta, T101*, 42.
14. Wang, G., Luo, J., Liu, J., Yang, T., Xu, Y., Li, J., Yin, H., Yan, J., Zhu, H., Zhao, C., et al. (2017). pMOSFETs featuring ALD W filling metal using SiH4 and B2H6 precursors in 22 nm node CMOS technology. *Nanoscale Research Letters, 12*, 306.
15. Cullity, E. D., & Stock, S. R. (1958). *Elements of x-ray diffraction.* Addison-Wesley Publishing Company. ISBN-13: 978-0201610918
16. Synchrotron radiation research in materials science. *MRS Bulletin*, June 2016.
17. Jenkins, R., & Snyder, R. L. (1996). Introduction to x-ray powder diffractometry. Willey Online Library. ISBN: 9780471513391
18. Jürgen Buschow, K. H., Cahn, R. W., & Veyssière, P., et al. (2001). Encyclopedia of materials: Science and technology. Elsevier. ISBN 978-0-08-043152-9

19. Radamson, H. H., & Thylen, L. (2014). Monolithic nanoscale photonics electronics integration in silicon and other group IV elements. Academic Press. ISBN: 978-012-419-975-0

20. Ayers, J. E. (1994). Measurement of threading dislocation densities in semiconductor crystals by XRD. *Journal of Crystal Growth, 135*, 71–77

21. Ayers, J. E., Ghandhi, K., & Schowalter, L. J. (1992). *Journal of Crystal Growth, 125*, 329.

22. Hordon, M. J., & Averbach, B. L. (1961). Acta Met9, 237.

23. Nabarro, F. R. N. (1987). *Theory of crystal dislocations* (pp. 51–56). Dover.

24. Stehle, H., & Seeger, A. (1956). *Z. Physik, 146*, 217.

25. Scherrer, P. (1918). Nachr. Göttinger Ges, 98.

26. Halliwell, M. A. G., Lyons, M. H., & Hill, M. J. (1984). The interpretation of X-ray rocking curves from III–V semiconductor device structures. *Journal of Crystal Growth, 68*, 523.

27. Vandenberg, J. M., Macrander, A. T., Hamm, R. A., & Panish, M. B. (1991). *Physical Review B, 44*, 3991.

28. Liu, E., & Wu, X. X-ray diffraction analysis of III-V superlattices: Characterization, simulation and fitting. Project work, Corpus ID: 131768216.

29. Radamson, H. H., & Thylen, L. (2014). Monolithic nanoscale photonics-electronics integration in silicon and other group 1V elements. Elsevier Academic Press. ISBN: 978-012-419-975-0.

30. Radamson, H. H., & Hållstedt, J. (2005). Application of high-resolution x-ray diffraction for detecting defects in SiGe(C) materials. *Journal of Physics: Condensed Matter, 17*, S231517.

31. Hansson, G. V., Radamsson, H. H., & Ni, W.-X. (1995). Strain and relaxation in Si-MBE structures studied by reciprocal space mapping using high resolution X-ray diffraction. *Journal of Materials Science: Materials in Electronics, 6*, 292.

32. Fewster, P. F. (2000). X-ray scattering from semiconductors. Imperial College Press. ISBN: 1-86094-360-8.

33. Wortman, J. J., & Evans, R. A. (1965). Young's modulus, shear modulus and Poisson's ratio in silicon and germanium. *Journal of Applied Physics, 36*, 153–156.

34. Nikanorov, S. P., & Kardashev, B. K. (1985). Elasticity and dislocation in elasticity of crystals. "Nauka" Publishing House.

35. Moontragoon, P., Ikonić, Z., & Harrison, P. (2007). Band structure calculations of Si–Ge–Sn alloys: Achieving direct band gap materials. *Semiconductor Science and Technology, 22*(7), 742–748. https://doi.org/10.1088/0268-1242/22/7/012

36. Nikanorov, S. P., Burenkov, Y. A., & Stepanov, A. V. (1971). Elastic properties of silicon. *Sov. Phys. Solid State, 13*, 2516–2519.

37. McSkimin, H. J., & Andreatch, P. (1972). Elastic moduli of diamond as a function of pressure and temperature. *Journal of Applied Physics, 43*, 2944–2948.

38. Herzog, H.-J. (1993). X-ray analysis of strained layer configurations. *Solid State Phenomena, 32*, 523–534.

39. Aella, P., Cook, C., Tolle, J., Zollner, S., Chizmeshya, A. V. G., & Kouvetakis, J. (2004). Structural and optical properties of $Sn_xSi_yGe_{1-x-y}$ alloys. *Applied Physics Letters, 84*, 888.

40. Radamson, H. H., Zhu, H., Wu, Z., He, X., Lin, H., Liu, J., Xiang, J., Kong, Z., Xiong, W., Li, J., & Cui, H. (2020). State of the art and future perspectives in advanced CMOS technology. *Nanomaterials, 10*(8), 1555.

41. Wang, G., Kolahdouz, M., Luo, J., Qin, C., Gu, S., Kong, Z., Yin, X., Xiong, W., Zhao, X., Liu, J., Yang, T. (2020). Growth of SiGe layers in source and drain regions for 10 nm node complementary metal-oxide semiconductor (CMOS). *Journal of Materials Science: Materials in Electronics, 31*(1), 26–33, 5.

42. Hållstedt, J., Blomqvist, M., Persson, P. O. A., Hultman, L., & Radamson, H. H. (2004). The effect of carbon and germanium on phase transformation of nickel on $Si_{1-x-y}Ge_xC_y$ epitaxial layers. *Journal of Applied Physics, 95*, 2397.

43. Wang, G., Moen, M., Abedin, A., Xu, Y., Luo, J., Guo, Y., Qin, C., Tang, Z., Yin, H., et al. (2015). Impact of pattern dependency of SiGe layers grown selectively in source/drain on the performance of 22 nm node pMOSFETs. *Solid-State Electronics, 114*, 43–48.

44. Wang, G., Abedin, A., Moen, M., Kolahdouz, M., Luo, J., Chen, T., & Radamson, H. H. (2015). Integration of highly-strained SiGe materials in 14 nm and beyond nodes FinFET technology. *Solid-State Electronics, 103*, 222–228.

45. Kojima, I., & Li, B. (1999). Structural characterization of thin films by x-ray reflectivity. *The Rigaku Journal, 16*(2), 31–42.
46. Levine, J., Cohen, J. B., Chung, Y. W., & Georgopoulos, P. (1989). Grazing-incidence small-angle x-ray scattering: New tool for studying thin film growth. *Journal of Applied Crystallography, 22*, 528.
47. Levine Parrill, J. R., Georgopoulos, P., Chung, Y.-W., & Cohen, J. B. (1993). GISAXS—Glancing incidence small angle X-ray scattering. *Journal of Physics IV France 3-C8*, 411–417.
48. Jun, Y. S., Lee, B., & Waychunas, G. A. (2010). In situ observations of nanoparticle early development kinetics at mineral-water interfaces. *Environmental Science and Technology, 44*, 8182–8189.
49. Smilgies, D.-M., Busch, P., Posselt, D., & Papadakis, C. M. (2002). Characterization of polymer thin films with small-angle x-ray scattering under grazing incidence (GISAXS). *Synchrotron Radiation News, 15*(5), 35–42.
50. Als-Nielsen, J., & McMorrow, D. (2001). *Elements of modern X-ray physics*. John Wiley & Sons.
51. Müller-Buschbaum, P. (2003). Grazing incidence small-angle X-ray scattering: An advanced scattering technique for the investigation of nanostructured polymer films. *Analytical and Bioanalytical Chemistry, 376*(1), 3–10.
52. Bunge, J. (1982). *Texture analysis in materials science*. Butterworths.
53. Roe, R. -J. (1965). Description of crystallite orientation in polycrystalline materials. III. General solution to pole figure inversion. *Journal of Applied Physics, 36*(1965), 2024. https://doi.org/10.1063/1.1714396
54. Hahn, Th., Klapper, H., Muller, U., & Aroyo, M. I. (2016). Point groups and crystal classes. In *International tables for crystallography* (vol. A, Section 3.2.1, pp. 720–737).
55. Schwartz, A. J., Kumar, M., Adams, B. L., & Field, D. P. (2009). *Electron backscatter diffraction in materials science* (2nd ed.). Springer.
56. Attallah, M. M., Zabeen, S., Cernik, R. J., & Preuss, M. (2009). Comparative determination of the α/β phase fraction in α+β-titanium alloys using X-ray diffraction and electron microscopy. *Materials Characterization, 60*, 1248–1256.
57. Bish, D. L., & Howard, S. A. (1998). Quantitative phase analysis using the Rietveld method. *Journal of Applied Crystallography, 21*, 86–91.
58. Topas. (2003). Bruker AXS GmbH, Karlsruhe.

Chapter 2
Micro-photoluminescence (μ-PL)

2.1 Introduction

Definition

Micro-photoluminescence probes light emission properties of a material with a relatively high spatial resolution. In fact, the spatial resolution this method provides is often limited by the diffraction of light. This is about one *micron* for the visible wavelength range, hence the name of this technique. The more general technique *photo*-luminescence implies that the excitation energy is provided to the sample by light, while *luminescence* stands for the radiation emitted by a solid.

It is important, however, to distinguish luminescence from the black-body radiation or from simply scattered light (Fig. 2.1). In the former case Planck's law describes equilibrium radiation, being a function of temperature (at 300 K the wavelength is ~10 μm, i.e. infrared). Luminescence, on the other hand, is a non-equilibrium radiation: extra energy needs to be provided to the material. Scattered light, on the other hand, refers to the light re-emitted instantly, on the time scale of the period of electric field oscillations (~fs for visible light). In contrast to the scattered light, which is almost elastic (nearly no energy losses) luminescence is normally an inelastic process and its characteristic time constants are much longer (typically from ~ns to ~ms). This is because luminescence is defined by material characteristics both in energy and time domains, rather than by the properties of the excitation light [1]. Therefore, resolving such emission spectrally or temporally yields information about *material-specific* properties. The application of *micro*-PL also adds spatial information and makes the technique highly useful for micro- and nanomaterials.

As opposite to scanning near field optical microscopy (SNOM), described in Chap. 8, no special tip fabrication and light guiding is required for the μ-PL. Standard optical components, such as objective lenses, are normally used, making micro-Pl much less complicated than the SNOM-PL. However, the achievable spatial resolution is roughly an order of magnitude larger than using a near-field microscope tip
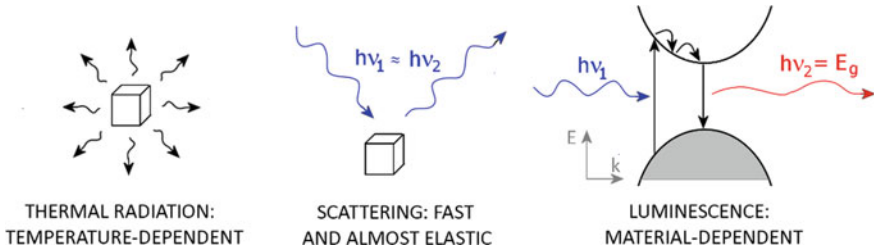
Fig. 2.1 Difference between luminescence and thermal radiation or scattering: higher energy photons are absorbed and lower energy emitted after carrier thermalization to the band edge

(~0.1 μm). To underline the difference micro-PL methods are sometimes described as a *far-field* photoluminescence microscopy.

Therefore, micro-photoluminescence occupies an important niche between ordinary PL, where the whole signal from a large excitation spot is integrated, and SNOM-PL, where a very high spatial resolution can be realized with a specialized equipment. In practice, using μ-PL one can analyze semiconductor bulk materials, heterostructures, quantum wells, or nanowires. Even individual nanoparticles/nanocrystals, such as quantum dots, can be accessed, provided they are dispersed far enough from each other to be resolved in the far-field, corresponding to the surface concentration of less than one particle per square micron.

Implementation

In general, there are different ways to provide extra energy to obtain luminescence from a sample. Electro-luminescence, for example, refers to the carrier injection, while cathodo-luminescence implies hitting a sample with an electron beam. Photo-luminescence is the most practical method among these, since a common laser, or spectrally-filtered light from a lamp can be conveniently used for excitation. It does not require electrical contact fabrication, or a vacuum environment, which is needed for the electron microscope. The characteristics of the emitted light, however, typically do not depend on the means of excitation. This is because, again, they stem from the material properties, as indicated in Fig. 2.1.

Excitation

By varying excitation beam parameters, such as photon energy, intensity, polarization, etc. one can probe different properties of the sample. It is important to remember that intensity refers to the average number of photons per time unit emitted or incident, while the photon energy corresponds to the energy, or wavelength, of the elementary quanta of light in the beam.

The excitation photon energy ($h\nu_1$) can be tuned from the emission energy (resonant excitation) and higher. For lower energy photons, typically below the material bandgap E_g, a semiconductor sample is essentially transparent. It can also be seen from Fig. 2.1, where energy bands of a semiconductor material are schematically shown. Scanning the excitation photon energy is the basis of photoluminescence

excitation (PLE) technique, which allows to extract absorption energy spectra even for individual nanostructures. Increasing energy of incoming photons may also result in different carrier recombination pathways since the electrons are excited high up in the conduction band. Hot carriers are generated here, and their energy dissipation mechanism may vary from the low energy excitation scenario.

Similarly, excitation intensity can be varied, which means increasing the number of impinging photons, although their energy remains the same. A practical measure of this intensity is the excitation power density, which shows incoming energy per area unit of a sample. Intensity modification is used for studying lasing characteristics and for the investigation of non-radiative processes pertinent to high carrier concentrations, such as Auger process.

One can also control and vary polarization of the incoming light. In other words, the orientation of the electric field vector oscillations. For anisotropic samples, polarization can generate a direction-dependent absorption and, hence, light output. Examples include nanorods, or nanowires, which show strong depends on their alignment relative to the electric field vector of the incident photons. Crystallographic anisotropy in the sample can also be probed in this way even for objects without a clear geometrical anisotropy, for example in materials with a *wurtzite* crystal structure.

Finally, non-linear optical effects are sometimes utilized for the excitation, such as two-photon absorption. This is used to filter out excitation and emission beams, using a large difference in the corresponding photon energy.

Emission

Emitted light, in turn, can also be analyzed along these points, revealing various sample properties. The light emission normally occurs via the lowest energy state, after thermalization of the excited carriers (Fig. 2.1, small black arrows). Difference between the excitation and emission energies is converted to heat (lattice vibrations, or phonons) during the thermalization process. In this way the characteristics of the emitted light normally reflect properties of the lowest occupied energy state.

The most common parameter of emission, bearing important information, is the PL spectrum. Typically, the position of the PL peak corresponds to the bandgap, while possible sidebands may reflect specific phonon modes. Spectrum linewidth is defined by the exciton-phonon interaction. It may also manifest possible dispersion of the emission energies in the sample, for example in non-homogeneous ensembles.

Often the light output intensity is monitored as a function of the excitation power. Typically, the dependence is linear until a saturation point, which is defined by the emission lifetime. At that point multicarrier effects, such as Auger process, are becoming relevant.

Temporal characteristics, such as emission lifetime, time-dependent intensity trace, and photon statistics can be also directly recorded. These parameters reveal strength of the optical transitions and possible influence of the local environment in carrier kinetics, which, in turn, reflects the quality of the sample.

Polarization-dependent characterization yields information on the emission state anisotropy, whether geometrical or crystallographic. Non-linear optical response, such as second harmonic generation in nanostructures, is also measurable.

Environment

Micro-photoluminescence studies are often carried out as a function of external parameters, such as in situ magnetic or electric fields, or ex situ chemical or radiation treatment, as well as sample temperature.

Magneto-optical experiments allow to probe fine structure of the lowest state, using the Zeeman effect. Spin degeneracy is lifted by the magnetic field, where the separation between the states is linearly proportional to the field strength. It is detectable in μ-PL as changing peak energy positions as a function of magnetic field.

Electrical field can also shift energy positions of the emission peak in nanostructures, the so-called quantum Stark effect. It happens naturally in some samples, where roaming random charges in the vicinity of emitting centers exert varying electric fields.

Micro-PL often serves as a quick feedback tool for any external modification of samples, whether it is done chemically, by X-rays, or via electron/ion beams. This allows studying effects of surface passivation on the optical output, as well as the role of defects/traps or dopants/impurities on the sample performance.

Another commonly used tunable parameter is sample temperature. Exciton-phonon interaction leads to broadened spectra at room temperature. By depopulating phonon modes at a reduced temperature, one gets a clear access to the energy level structure and can directly detect involved phonon modes as separate spectral sidebands. Temperature scans also give information on possible crystal phase transitions and may provide answers on the physical origin of complex carrier dynamics.

Finally, the lowest energy state may consist of several sublevels with different recombination strengths. This is defined by spin or momentum selection rules, where allowed, forbidden or partially forbidden transitions may exist. Therefore temperature-dependent population of these states, as stipulated by Boltzmann statistics, will be reflected in measured temperature-dependent μ-PL characteristics, such as the lifetime.

Summary

A combination of different excitation/emission configurations of the μ-PL under varying external conditions is a powerful tool to study energy structure, composition, carrier dynamics and general optical properties of semiconductor materials and nanostructures with a rather high spatial resolution (~1 μm).

2.2 Description of the Technique

General Arrangement

Similarly to SNOM-PL the μ-PL can be assembled in a number of different configurations. A central component for light collection and, sometimes, excitation is an objective lens. High magnification (50×, 63×, or 100×) objective lenses are typically used to collect emitted light and to transfer it to the detector via auxiliary optical components, such as filters or other lenses.

A simple μ-PL system can be constructed on top of an optical table, providing that a sample micro-positioning and a stray light shielding are in place. In most cases, however, one can rely on a commercial optical microscope, where most needed functionalities are already included.

Microscope

One should distinguish between a wide-field and a confocal microscope configuration. While the in-plane resolution is diffraction-limited in both cases, the spatial resolution in the out-of-plane (depth) direction is higher for the confocal arrangement. This is due to a small aperture, placed in the optical path of a confocal microscope, discriminating signal by its point of origin (Fig. 2.2). Depending on the system a spatial resolution down to ~50–100 nm in the out-of-plane axis can be achieved, far better than ~1 μm in a wide-field microscope. A disadvantage of this configuration is that the signal from only one point is taken at a time. Therefore, to build a PL image an in-plane scanning of the sample/objective is necessary. In contrast, a wide-field microscope configuration yields instantly a full PL map of the sample. For example, a typical field of view for a 100× objective lens, projecting sample image onto a CCD camera, is ~100 × 100 μm^2. If the laser excitation spot is large enough to cover this whole area a PL map of it can be readily acquired in one shot.

To realize temperature-dependent measurements an isolation of the sample from the ambient is achieved by vacuum. Unless the objective lens is placed inside a vacuum chamber together with a sample, the cooling is done with a cryostat. Then the light excitation and collection take place through the cryostat window, which should have good optical transmittance for the wavelengths of interest. The window inevitably distorts imaging capabilities of standard optics of a microscope. Therefore, special window-correction objective lenses need to be introduced.

Excitation Geometry

The μ-PL technique can be classified by the way the excitation is arranged. In general, bright field, dark field or total internal reflection configurations are used. In Fig. 2.2 the excitation light path for these is shown with blue arrows.

In a bright field geometry, the light is guided via the same path as the light is collected. Excitation and detection beams are then decoupled with a help of a dichroic mirror. This component selectively reflects radiation below certain wavelength, while transmitting light with a wavelength longer than the cut-off. This means that the shorter wavelength (higher photon energy) excitation beam is directed towards the
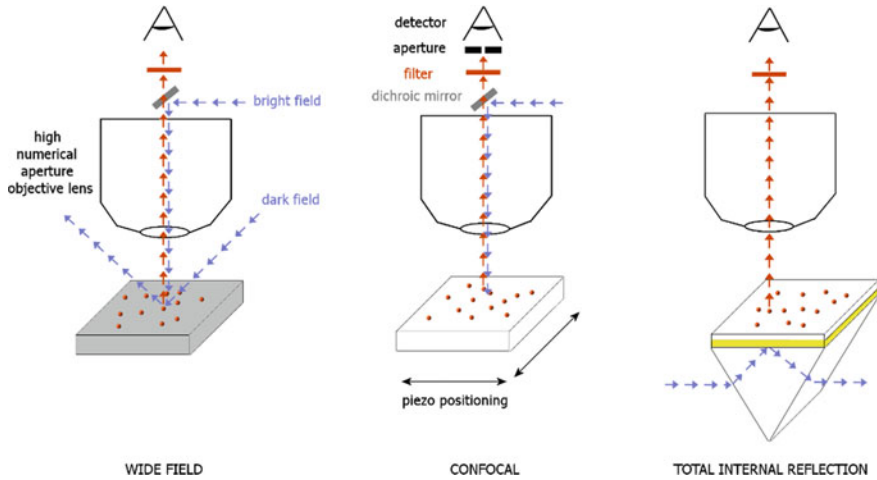
Fig. 2.2 Different configurations of the μ-PL setup: wide-field, confocal and total internal reflection [2]

sample, but not to the detector. At the same time the longer wavelength (lower photon energy) emission from the sample is transmitted.

Because the excitation beam intensity can be high, some light may be back-reflected from the sample and still pass through the dichroic mirror. Therefore, to completely eliminate the excitation light an additional long-pass filter is often utilized. This excitation/collection arrangement is sometimes referred to as *epifluorescence* microscopy. The excitation beam in this case can be focused to a small spot and, therefore, the excitation power density can be high. It allows tuning the excitation power density in a very broad range with the help of neutral density filters placed on the excitation beam path.

On the other hand, to realize the main advantage of a wide-field microscope the excitation spot should cover a large area. It is implemented using so-called dark field excitation geometry. In contrast to the bright field excitation, in dark field excitation the incoming beam is spatially decoupled from the collection beam path, as schematically shown in Fig. 2.2. Implementation of the dark field excitation puts a limit on the type of objective lenses that can be used. Namely, short working distance objective lenses do not allow enough space to guide the light from the side and, therefore, special long working distance pieces are utilized for this purpose.

As a way to avoid this complication an evanescent wave excitation can be introduced. Similarly to SNOM-PL, a near-field excitation is used, but it is arranged not through a sharp tip, but instead via a propagating wave (Fig. 2.2). A prism in usually employed, which makes it difficult to combine this configuration with a cryostat cooling, limiting this arrangement to room temperature measurements.

Detectors

Depending on the characterization modality, such as PL imaging, spectroscopy, or carrier dynamics studies from PL transients, different detectors are utilized. Here, *single* pixel (picture element) and *multi*-pixel type of detectors should be distinguished. In the former, the total incoming signal is integrated into a single channel, while in the latter there is a large array of separate channels, each carrying out individual information.

When an objective lens collects a sample image in a wide-field microscope and transfers it to the detector imaging plane all the spatial information remains preserved (Fig. 2.3). Some distortions may occur due to aberrations of the image transfer optics, yet those are typically limited to the image periphery and are usually minor. If the detector is of a single-pixel type the detected signal resolution will be defined by its area. For very large area detectors the µ-PL advantage becomes void, because all the spatially resolved information is convoluted to a single point. If the detector area is very small, for example similar to the size of a single pixel of a multi-pixel detector, position-specific signal can still be retrieved and analyzed by sample scanning.

Multi-pixel detectors, on the other hand, by design contain a set of separate channels and therefore are directly suitable for PL mapping. Their complex structure, however, does not allow the same performance and flexibility as for single pixel counterparts, especially in time-resolved applications. Therefore, a combination of different detectors, each optimized for a specific task, typically comprise a µ-PL system.
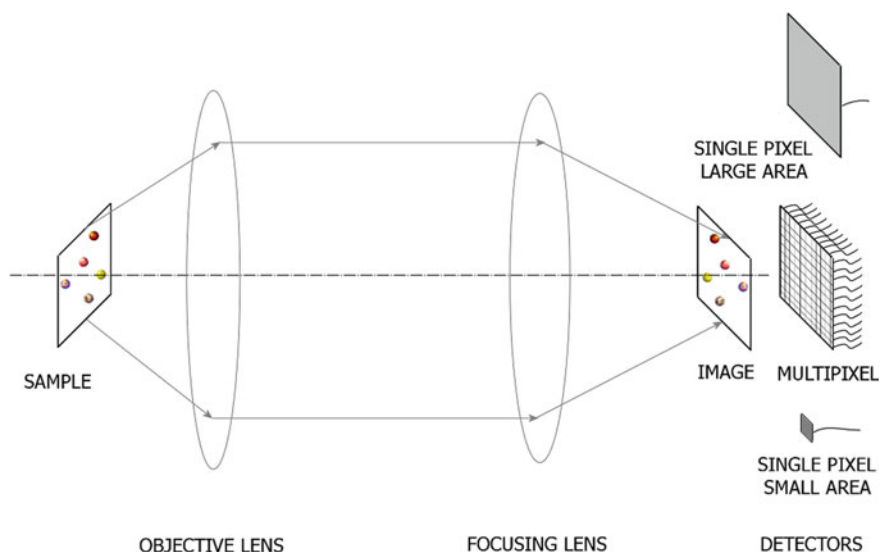
Fig. 2.3 Different types of detectors and their applicability for the micro-photoluminescence

In a confocal microscope configuration, on the other hand, the signal initially contains only a single point (pixel). The choice of the detector type is not as important as in a wide-field microscope in terms of spatial resolution.

Single Pixel

The simplest type of a single pixel detector is a *photodiode*. It is typically a p–n junction designed to detect excited carriers generated by absorption of the incoming light. The principles of p–n junction devices are further outlined in Chap. 7. This effect is sometimes called an *internal* photoelectric effect. Internal means that the excited carriers remain within the detector material. In the ordinary photoelectric effect carriers escape from the material surface, overcoming the work function (Albert Einstein was awarded Nobel Prize in Physics in 1921 for its discovery). In practice, more advanced versions of such p–n junction detectors are used in micro-PL.

A common modification, which is widely employed, is an *avalanche* photodiode (APD). The operation principle consists of the same steps as for an ordinary photodiode with some important modifications. First, absorbed photons generate an electron–hole pair in the intrinsic layer of a p-i-n junction. A semiconductor is used for the detector material, such as silicon, where different doping forms p-, i- and n-layers (i refers here to an intrinsic, or un-doped, semiconductor region). Therefore, only photons with the energy larger than the bandgap can be detected. Then, under a reverse bias the electron–hole pair is separated and the carriers are accelerated to the electrodes (shown for electrons in Fig. 2.4, left). If the electric field is sufficiently high, new carriers can be generated by impact ionization within the detector volume. The new carriers are then accelerated in the field and further electron–hole pairs are generated, hence the name *avalanche*. Thus, a photovoltage is created at the detector output. So, in APDs there is an internal amplification of the signal, which is quantified by the *photodiode gain* parameter. In other words, for a single incoming photon, tens or even hundreds of electrons are generated and detected electrically in this solid-state detector, which drastically enhances the sensitivity of the detector.

Another type of a commonly used single pixel detectors is a *photomultiplier tube* (PMT). Its operation principle is based on the ordinary photoelectric effect. First, photo-induced electron emission takes place from the cathode exposed to the incoming light (Fig. 2.4, right). For this to happen, the work function of the cathode material needs to be smaller than the photon energy. Then emitted electrons are
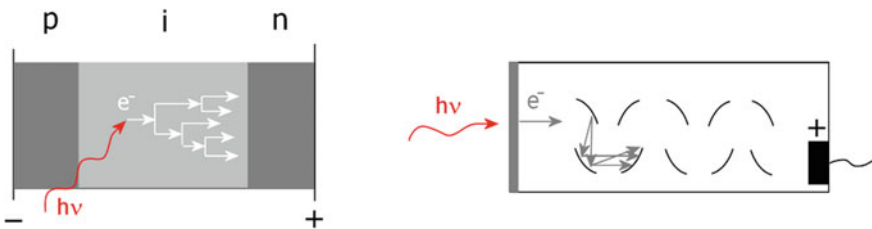
Fig. 2.4 Schematics of single pixel detector operation for APD (left) and PMT (right)

accelerated inside the tube to the first electrode, called a *dynode*. To avoid electrons being stopped by air molecules, the whole device is placed in vacuum tube. At the dynode secondary electron emission takes place, whith a coefficient $\delta > 1$ (typically from 3 to 10). That is, for every incoming electron several more are generated. The process continues through a series of n dynodes, yielding the total amplification gain of δ^n. In practice the gain value can reach one million at the output anode, justifying the name of the device.

When selecting a suitable detector for a particular experiment several parameters need to be considered. First is the detector's *spectral range*. The long wavelength cutoff is defined by the bandgap of the active absorbing layer for APDs and by the work function of the photocathode material for PMTs. Photons with energies below the bandgap, or lower than the work function, will transmit through the device undetected. On the short wavelength side (larger photon energies) the practical limit is set by the absorption in the protective window, covering the detector. A pure quartz window is often used when the detection range needs to be extended to the UV spectral region (> 3 eV). Ordinary glass, albeit less expensive, has many impurities, which absorb photons with energies above the visible range.

Another important parameter is detector's *quantum efficiency*. This quantity shows how likely the process of conversion of the incoming photons to detected charge carriers is. Ideally a 100% conversion probability is desired, but in practice carriers can be lost before being counted. A typical value for the photon-to-electron conversion is in the range of 0.1–30% for these single pixel detectors, and this value may strongly depend on the photon wavelength.

Nowadays the main application area of such detectors is in time-dependent measurements. From this perspective their *response time* is a central parameter. Because every charge readout process takes a finite time, the detector turns "unavailable" for a moment. This time is called the detector's dead time and reflects its resolution in the time domain. Semiconductor detectors, such as APDs, usually feature a faster response time (~ ps), making them suitable even for single photon counting.

Recently a new type of a single pixel detector, combining best features of APDs and PMTs, has become routinely available. It is called a *hybrid* APD-PMT photodetector, borrowing the photocathode part from the PMT and the electron detection part from the APD concept. Essentially an array of dynodes (Fig. 2.4, right), which is often responsible for noise and long response times, is replaced with an APD inside the vacuum tube. In this way a faster detection is achieved than in PMTs, while the charge trapping/release process, pertinent to pure APDs, is suppressed.

For very fast detection applications, down to ~fs scale, a special design is employed, called a *streak camera*. It is based on a photocathode for the conversion of photons to electrons, as in PMTs. To facilitate high temporal resolution capacitor plates are placed right after the photocathode. They bend the trajectories of the accelerated photoelectrons on their way to the detector (a phosphor screen). Switching electronics can quickly modulate bias on the capacitor plates. Therefore, spatial positions of the detected electrons can be correlated with the photon arrival time to the photocathode, resulting in a very high temporal resolution.
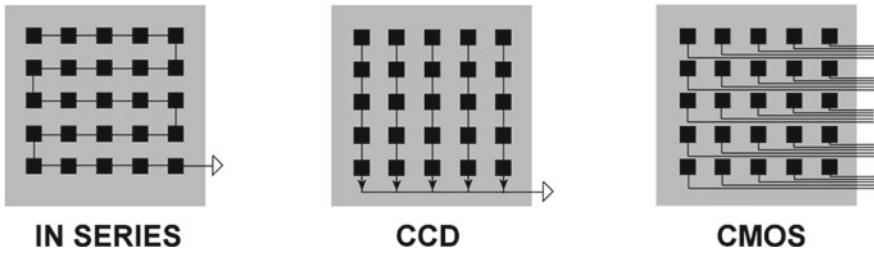
Fig. 2.5 Different layouts of the carrier readout in multi-pixel detectors

It is worth mentioning an example of a single pixel detector design, which is based on a different physical principle than the photoelectric effect. The use of superconducting wires, whose superconductivity state becomes disturbed upon photon absorption, was proposed in 1991 [3], and now such device are commercially available. The detected change in the resistance (from zero to a finite value) corresponds then to a single photon arrival event. A particular point of this design is a uniform spectral dependence of the detector response for a very broad range of wavelengths.

An interested reader may find more information on different types of single-pixel detectors and a detailed comparison between their characteristics elsewhere [4].

Multi-pixel

The advantage of a multi-pixel detector is obvious, when a full image needs to be taken (see Fig. 2.3). But having multiple separate channels immediately leads to a practical question: How to readout charges fast and with a low noise from such an image sensor? One can think of several general technological arrangements, as summarized in Fig. 2.5.

A common read-out thread can go *in series* through all the pixels (Fig. 2.5, left). That would lead to a long readout time and the noise will be amplified on the way of the generated charge movement towards registering. A parallel readout row-by-row would improve time and reduce noise ((Fig. 2.5, middle). This is an arrangement, which is used in *charge-couple device* (CCD) cameras. A fully parallel readout may be the optimum solution (Fig. 2.5, right), but that would require a complex fabrication routine, considering all the leads and amplifiers needed for every pixel. The last design represents *complementary metal–oxide–semiconductor* (CMOS) cameras. Historically, CCD configuration was first to be successfully realized, considering constraints of the available technology in the 1960s. Nowadays, with the advancement of microfabrication technology, CMOS type of multi-pixel image sensors dominates.

One of the first digital image sensors that found a widespread use was CCD. Its impact on technology appeared to be so profound that Nobel Prize in physics in 2009 was awarded for its development [5]. Inventors for the first time succeeded in gathering and reading out signals from a large number of pixels in a short time (Fig. 2.5, middle). The CCD consists of an array of metal electrodes on top of an insulating layer, separating them from the active absorbing material (semiconductor).

Consider an image projected onto such an array, which has brighter and darker parts. Electron–hole pairs will be generated in the semiconductor material proportionally to the incident light intensity. Electrons are then stored in a potential well below the biased electrodes. Collected charge is transferred row-by-row to the readout register by gradual variation of individual biases. Amplification of the collected signal and its visualization by time-to-position conversion is the last step.

Parallel acquisition from n channels reduces signal-to-noise ratio by $\sqrt{n}$, compared to the readout in series (Fig. 2.5, left). It also facilitates a faster readout by a factor of n. A set of individual channels in CCD can be arranged as a full camera for images, such as 1024×1024 pixels. Alternatively, an array of pixels, used primarily in spectroscopy, is designed, for example 1024×256 pixels. Pixel size can vary from several μm^2 to $20 \times 20\ \mu m^2$. Spectral selectivity is an important characteristic, as for single pixel detectors. It is defined by the bandgap of the semiconductor material. Commonly Si or InGaAs cameras are used. The former has a detection range from 400 to 1000 nm, while the latter is used for near-infrared region (from 1000 nm up to 2000 nm).

Because top electrodes block some of the incoming light, recently a different design was introduced, the so called *back-illuminated CCD*. The semiconductor layer is reduced to a thin film and exposed to the incoming light having an array of controlling metal electrodes from the back side. Such detectors boast high quantum efficiencies, approaching 90% in some spectral ranges. Electron multiplication (EM) is often integrated into the device. This is an additional amplification layer before the readout amplifier to allow both higher sensitivity and speed. Nowadays CCD cameras are typically used for low-light applications, where long exposures are necessary to collect the signal.

With the advancement of microfabrication technology CMOS cameras has gradually became the mainstream image sensors (Fig. 2.5, right). They provide fast readout and generally are less expensive. Scientific grade CMOS cameras today have their quantum efficiencies approaching that of CCDs. The only area where CMOS cannot yet compete with the older design is low-light applications. Because many amplifiers exist on the chip they generate parasitic light emission, which may quickly surpass the signal. In CCDs, on the other hand, supplied voltage to the single amplifier can be easily reduced for a long exposure regime.

It is worth mentioning other, less common, image sensor types, which are relevant for micro-photoluminescence. One example is an image intensifier tube with microchannel plates (MCP). Essentially it is an array of small photomultiplier tubes. Instead of dynodes used to amplify electrons (Fig. 2.4, right), the amplification takes place inside an MCP. Here an array of channels carries site-specific information on the intensity level of incoming electrons. At the end electrons can be back-converted to photons by a phosphor screen or read-out electrically to reconstruct an image.

Color cameras can be considered as another example. Those are designed to simulate a human eye with its three color receptors of red, green and blue in every pixel. Main application is to render colors emitted in the visible range as a human eye would detect it. While such images may look appealing, their scientific value is typically limited. Proper spectra acquisition for every point (PL spectrum mapping)

is much more informative. Indeed, as with human eyes, convoluting full spectra into three color numbers loses a great deal of information. In addition, different spectra may produce a very similar set of color coordinates. Therefore, such representation of the emission properties may even be misleading.

Noise

A common characteristic for all detectors is their *dark count* rate, that is counts, which are generated in the complete absence of incoming light. Physical origin of this false signal is a thermal excitation of electrons. Electrons in the material occupy a distribution of velocities according to Fermi–Dirac statistics. So, even at room temperature some particles may statistically acquire enough energy to become excited to the conduction band (as in internal photoelectric effect) or above the work function (as in the ordinary photoelectric effect). Those will generate counts, which cannot be distinguished from the genuine ones.

The problem with this false signal is not its presence as such. If dark counts contributed uniformly over time, they could have been subtracted as a constant pedestal. However, their statistical nature implies fluctuations and, hence, noise. A random process with an average number λ of events per time interval t is described by Poisson distribution with a probability density of having exactly k events per interval:

$$p_k = (\lambda t)^k \frac{\exp(-\lambda t)}{k!} \tag{2.1}$$

So, when the detector dark count rate is given as a fixed number in the specifications, it is the average value λ, which is mentioned. The level of noise can be then estimated from Poisson statistics as proportional to the square root of the average intensity.

It is clear, that a signal below the noise level becomes buried in it and cannot be detected. A quantitative parameter for description of this fact is a *noise equivalent power*. This measure of detector sensitivity shows the incoming light power, which is equal to the noise level (for a half second integration time). The value of this detectability threshold is also often provided in device specifications.

Because dark counts are of thermal origin the most straightforward way to suppress them is to cool the detector. PMTs and CCD cameras are often realized with this option. It can be a Dewar bottle attached to the device, which needs to be refilled with liquid nitrogen (~70 K) from time to time. Alternatively, thermoelectric cooling can be used, which greatly simplifies maintenance of the detector. A multi-stage Peltier element is integrated in this case to reach active area temperatures of ~200 K. This is sufficient to suppress the generation of the thermally excited carriers to a negligible level (few counts per pixel per hour). As the side of the Peltier element attached to the detector refrigerates, the other side heats up, so a running water or a ventilation fan for the heat removal is typically included. Naturally, the cost of the detector becomes higher with these additions, but it is vital, especially for low light applications.

Another way to reduce the level of dark counts is to use a detector with a smaller sensor area. The number of thermally excited electrons in a single pixel is simply proportional to the volume of the material. In practice, APDs with down to $\sim 20\,\mu m^2$ chip area exist and they feature a low dark count rate of a few counts per second. Clearly, alignment of such small area detectors is a price to pay for the improved performance, where signal needs to be focused to such a small spot.

Electrical readout process, including analog to digital conversion and amplification, is an additional source of noise. Every image/pixel has it and the noise level is defined by the device architecture. In CCDs, for example, it is the same for all pixels due to a common readout register. In CMOS cameras, on the other hand, it varies from pixel-to-pixel because an amplifier is built-in to every pixel position. Their average is used as a characteristics value given in specifications.

A natural limit of the noise is a statistical nature of the light itself. If photons arrive to the detector with an average rate I (intensity) their distribution again follows Poisson statistics (Eq. 2.1) for most sources. This is called *shot noise* and it grows as the square root of the total number of arriving photons N. The total signal, on the other hand, is directly proportional to N, so the *signal-to-noise* ratio will grow as a square root of the integration time Δt:

$$SNR \sim \sqrt{N} = \sqrt{I \cdot \Delta t} \tag{2.2}$$

This simple formula implies that the longer exposure is, the better the signal-to-noise ratio becomes. Now the double significance of the detector cooling is obvious. It allows long exposures by suppressing dark count noise in order to collect enough photons to overcome the shot noise. Several hour long exposures are possible, and those are sometimes used in detecting very weak signals in micro-photoluminescence.

Excitation Sources

A general requirement for the excitation source is that the photon energy should *not* be lower than the material bandgap. In practice, green and blue lasers are often used. Both continuous wave (cw) and pulsed regimes can be applied. Monochromatic laser encountered are He–Cd lasers (442 and 305 nm), Kr- or Ar-ion lasers (several lines in blue/green region), dye lasers, and semiconductor laser diodes. A large excitation power is not needed for a micro-PL experiment. A few mW source is typically sufficient, because the light is often focused to a small spot either through objective lenses, or by an external lens in a dark field excitation geometry. Excitation power density is a more relevant parameter in this case. Neutral density filters can then be used to reduce excitation power density in the power scan experiment.

To realize the pulsed pumping regime, acousto-optic modulators can be placed in the beam path. Sometimes they are already built-in to laser diodes and the whole unit can be conveniently operated. If the characteristic times are long (~ms) a mechanical chopper (rotating disc) can be used, although rise and fall profiles are not very steep. The pulsed pumping is especially useful for time-dependent measurements. When luminescence decay is monitored the signal after every pump cycle is summed up
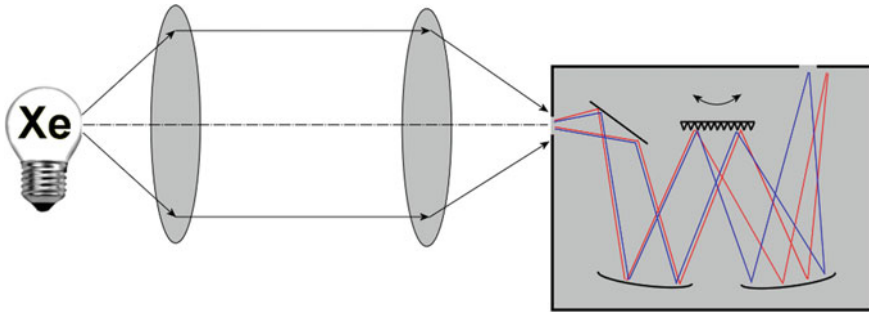
Fig. 2.6 Light from a xenon lamp filtered by a monochromator

to construct the decay curve. It is also practical for pumping of materials with poor heat dissipation properties. In those continuous wave pumping may cause permanent damage and should be avoided. Pulse length and interval between pulses then become important experimental parameters.

In many micro-PL experiments the excitation wavelength needs to be scanned, such as for PLE. Tunable lasers are suitable, e.g. higher harmonics of Nd:YAG or Ti:sapphire ones together with an optical parametric oscillator. The spectral range covered may not be continuous though, if scanning range from different harmonics does not overlap. A combination of different lasers needs to be used.

An alternative solution is a Xe lamp, which emits broadband radiation. A monochromator with a grating is then attached to it to filter out a particular wavelength (Fig. 2.6). Different colors deflected at different angles by the grating, allowing extraction of the wavelength of interest through the output slit. Excitation energy scan is then realized by a simple rotation of the motorized grating. The total output power of the monochromatic light may be low. However, when the beam is focused through the objective lens it is sufficient for the excitation in most cases.

Because the electrical discharge inside the lamp is not stable in time and space the output power of the lamp may fluctuate. A half-mirror is often placed on the beam path to monitor the time-dependent output for subsequent intensity normalization. Recently, laser-driven discharge lamps are being used. They provide much improved stability by defining a small, stable spot inside the bulb, where light-emitting plasma is generated.

When the excitation photon energy equals to the lowest energy state the excitation is called resonant. It requires a tunable light source with sufficient power to precisely match the emission level. An important feature of this excitation regime is the lack of extra energy deposited in the system. Carrier thermalization generally leads to a local temperature increase, and shortens the coherence of the excited state. Therefore, resonant excitation allows much better control of the emission state preparation. A practical issue then is how to decouple stray excitation light with the emission, when their wavelengths are identical. In this case, a polarized excitation light from a tunable laser can be used and, providing the emitted light has a different polarization, it can be filtered out. Alternatively, two-photon excitation, which utilizes non-linear

absorption property, can be used. A very high power laser is needed to create enough photon density for this process. In this configuration a rare case of excitation with photon energy below the bandgap is realized.

Spectral Analysis

Decomposition of the signal by wavelength (photon energy) is one of the most common micro-PL modalities. It can be realized using following tools. A *monochromator* focuses selected signal of a specific wavelength to the output slit for a single channel detector (Fig. 2.6). To build a spectrum one has to gradually turn the grating to collect different wavelengths one by one. A motorized grating is used for this purpose. A *spectrograph* produces spectrum focused to the output port for a multi-channel detector. Essentially, this is a monochromator with a wide output slit, where a multichannel detector, such as a one-dimensional CCD array, can fit. The optics for image transfer from the input to the output port should also be good enough to provide an aberration-free spectral image. Finally, an *imaging spectrograph* can form both a full image, such as for a CCD camera, and a spectrum at the output port plane. Requirements for the in-built optics quality is even higher for this device to avoid aberration in both axes. Imaging spectrographs of Czerny-Turner type (as shown in Fig. 2.6) is a common device type used in micro-PL.

A grating is a key element providing wavelength-dependent dispersion of light (Fig. 2.7). This device just consists of a periodic pattern, covered with metal, such as aluminium. It is made either by mechanical cutting of the grooves (ruled grating) or by etching through the mask, created by a laser interference (holographic grating). Constructive and destructive interference of reflected rays from different points of the pattern produces an angle-dependent diffraction pattern. Within the same order diffraction peak the reflection angle depends on the wavelength of light. This is how a broadband light is decomposed in real space as a function of the wavelength.

Every grating has a zero-order diffraction, which is just a reflection without dispersion from the grating substrate. So, it can also be used as a mirror at the right detection angle. This property is used in micro-PL, when the sample needs to be imaged
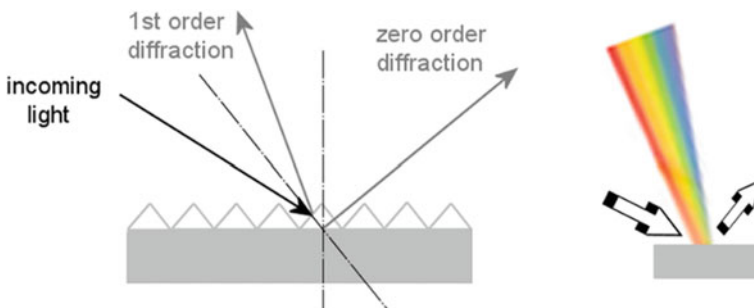
Fig. 2.7 Diffraction grating operation principle

first to find an area of interest. Then, with the same grating, but detecting its first-order diffraction peak, the spectrum from this area can be acquired. The number of grooves per millimeter is a standard grating parameter, defining its *dispersion* power (the more groves the stronger the dispersion). The first order diffraction peak is the strongest in intensity. Sometimes second-order diffraction peak is also utilized, because it provides a higher dispersion, although with a lower intensity. The grating is the most common light dispersing element, but a prism can alternatively serve in a monochromator/spectrograph.

The total performance of a spectrograph is assessed by several parameters. Dispersion of the spectrograph (not only the grating) shows how far different wavelengths end up from each other at the output plane. Linear dispersion is proportional to the angular dispersion of the grating with a proportionality coefficient being the focal length of the spectrograph. So, the larger the device the further apart different wavelengths will be after travelling inside it. While most spectrograph are shorter than half a meter, one or even two meters long devices exist. The purpose is to achieve a high *spectral resolution.*

Resolving power is an ability to distinguish between two closely spaced lines. In other words, two lines may appear as a single peak if the resolving power is insufficient. Gratings with a high groove density, such as 1200 gr/mm, and long spectrographs should be used for high resolving power. When such a high resolving power configuration is applied the total spectral range falling onto the detector becomes very limited. Therefore, in practice a combination of low and high dispersion power gratings is used. Those are typically mounted on a motorized holder. One can make an analogy with an optical microscope, where low magnification lenses can be used for a wide survey imaging (low resolution), and high magnification lenses are suitable for detailed observations of a limited area. Possibility to easily switch between those significantly simplifies whole device operation.

Corrections

It is important to remember that even the best spectrograph/detector system distorts the signal. Wavelength dependence of the detector response was discussed above. In addition, the spectrograph itself may introduce losses along the signal path. For example, grating reflection intensity is usually a strong function of the wavelength. With various mirrors, lenses and filters on the beam path the total system response becomes a combination of spectral dependencies of the individual components. The measured signal $S(\lambda)$ is then a product of the original spectrum $I(\lambda)$ and the system response function $F(\lambda)$:

$$S(\lambda) = I(\lambda) \cdot F(\lambda) \tag{2.3}$$

Because experimental goal is to find $I(\lambda)$, the function $F(\lambda)$ should be identified for each detection system. It can be measured by a source with a known spectrum and a black body radiation lamp with a calibrated output is typically used. In this experiment $F(\lambda)$ is found from the known $I(\lambda)$ and the measured $S(\lambda)$. In many research papers a statement such as "all spectra were corrected by the spectral response" is included.

This correction is especially important for the edges of the detection sensitivity, where the system response curve becomes quite steep. Original spectra can be heavily distorted in that case. On the other hand, when the measured spectra are very narrow, system response variations within the emission line may be neglected.

While the measured intensity is a function of the wavelength in most applications, it is the photon energy, which is important and more relevant for the study of physical processes in materials. The simplest way to convert wavelength to the photon energy is by the definition of frequency $\upsilon = c/\lambda$ (c is the speed of light). After multiplying both parts by the Planck constant it can be numerically written as:

$$\lambda[nm] = \frac{hc}{h\upsilon} \approx \frac{1239}{h\upsilon[\text{eV}]} \tag{2.4}$$

It is important to remember that the measured signal $S(\lambda)$ has the units of [counts/nm]. When integrating it over the whole spectral range we should obtain the total number of counts N. In this case, when changing variables from wavelength to the photon energy, a proper conversion from Eq. 2.3 should be performed:

$$S(\lambda) = \frac{\mathrm{d}N}{\mathrm{d}\lambda} = \frac{(h\upsilon)^2}{hc} \cdot \frac{\mathrm{d}N}{\mathrm{d}(h\upsilon)} = \frac{(h\upsilon)^2}{hc} \cdot S(h\upsilon) \tag{2.5}$$

So, changing only the x-axis according to Eq. 2.4 is not sufficient; for a full and correct change of variables Eq. 2.5 should be used. For narrow lines one can again assume a small variation within the spectrum and Eq. 2.4 is reasonable.

Instrument response also affects spectral resolution of the system. It can be measured using a known light source in a similar way, as a black body source applied for the wavelength-dependent system response. Here, instead, a source of ultra-narrow emission lines is used, typically an argon or a mercury calibration lamp. Atomic transitions are very narrow, reflecting discrete energy states in an atom. The measured spectral linewidth has, however, a finite linewidth due to the broadening, induced by the instrument. The main contribution is the finite size of the source image. It is defined by the input slit of a spectrograph or by the size of the object itself.

Mathematically the effect of broadening can be described by a convolution integral. Consider the original signal of a Gaussian shape $S_0(\lambda)$, every single point of which experiences broadening by the instrument function, which is also of a Gaussian shape $i(\lambda)$ (Fig. 2.8). For the resulting broadened curve one should sum up all the elementary contributions:

$$S(\lambda) = \int\limits_{-\infty}^{\infty} S_0(x) i(\lambda - x) \mathrm{d}x \tag{2.6}$$

It can be shown that a convolution of two Gaussian shapes is a Gaussian curve with a linewidth being square root of the sum of the contributing linewidth squares:
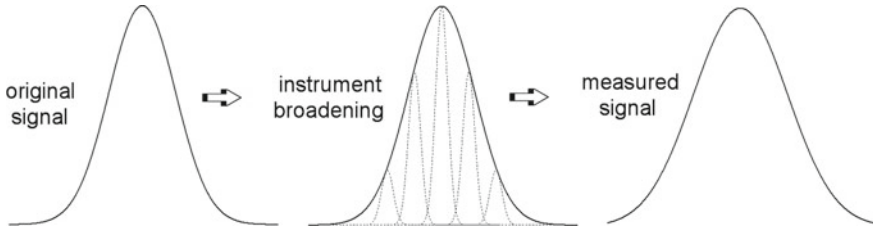
Fig. 2.8 Signal broadening due to the instrument response

$$\Delta S = \sqrt{(\Delta S_0)^2 + (\Delta i)^2} \tag{2.7}$$

This is a convenient formula to find the original linewidth from the known instrument response and the measured signal. This process is called *deconvolution* and it is often used for the evaluation of narrow spectra. Clearly, when the original signal is narrower than the instrument response, the whole signal becomes resolution-limited and only the upper limit of the signal can be extracted. On the other hand, for a signal much broader than Δi the instrument contribution to the broadening can be simply neglected.

Interferometry

In case the spectral resolution provided by spectrographs is not sufficient, the interferometry should be applied. Typically, best spectrometers with CCD arrays can achieve spectral resolution on the order of tens of μeV. Some materials, in particular nanostructures with discrete energy states, such as quantum dots, possess even sharper PL linewidth at low temperatures.

The most common types of interferometers for this application are the Fabry–Perot cavity and the Michelson interferometer. The former consists of two semi-transparent mirrors, forming a cavity. Only the modes supported by the cavity are transmitted, and the scanning of one mirror with the help of piezo elements realizes spectral measurements. Characteristic parameters of this type of an interferometer is a finesse and a free spectral range. The free spectral range is a range, within which a spectrum can be detected; the finesse shows the quality factor (sharpness) of the cavity response. In the Michelson interferometer the signal is split into two arms. One arm acts as a tunable delay line, resulting in the interference in the output signal. Application examples of such devices in micro-PL will be provided below.

Temporal Analysis

The emitted signal can be alternatively dispersed for analysis in the time domain. Pulsed excitation is then typically used, where the pumping is intermittent. During the intervals between the pulses the relaxation of the excitation is monitored as a change of the PL intensity. This is the basis of the most common type of analysis: PL decay measurements. One can monitor dynamics both of the rise and the fall of the PL signal.

Under continuous excitation, on the other hand, other time-dependent characteristics can be extracted. Stability of the emission can be monitored as a PL intensity trace, where effects such as blinking can be revealed. Those expose variations in the carrier recombination scenario on the long time scale ("long" compared to the lifetime). In addition, distribution the photon emission events in time, the so-called photon statistics, provides information about the nature of the emitting state.

Consider first the simplest case of the PL decay measured from a system, where only a single decay channel exists. Then it can be considered as a random process with the average rate $\Gamma[1/s]$. This value shows how many recombination events take place per time unit. Then the number of emitters in the system is decreasing with this constant rate. In other words, the same fraction of emitters becomes relaxed every time unit (first-order kinetics):

$$\frac{dN(t)}{dt} = -\Gamma \cdot N(t) \tag{2.8}$$

Solving this elementary differential equation yields:

$$N(t) = N_0 \cdot e^{-\Gamma t} \tag{2.9}$$

that is, a mono-exponential decay of the number of emitters with time. The product of the average decay rate per time unit and the number of the emitters is the measured PL intensity:

$$I(t) \sim \Gamma \cdot N(t) = \Gamma \cdot N_0 \cdot e^{-\Gamma t} \tag{2.10}$$

A proportionality coefficient includes the detection system collection efficiency, the number of excitation cycles, etc. So, the PL decay curve in this case should be a mono-exponential function of time. Very often the experimental $I(t)$ is plotted in a log-linear scale. Indeed, if we take the natural logarithm of the Eq. 2.10:

$$\log(I(t)) = -\Gamma \cdot t + const, \tag{2.11}$$

the dependence on time now becomes a straight line with a slope being the decay rate.

In real systems, however, carrier dynamics may be complex. Carrier thermalization process may take place at the same time as the recombination, trapping and de-trapping of carriers may occur, separate recombination via emitter centers with different rates may contribute to the signal, etc. If experimental results are shown in a log-linear scale, such a deviation from the mono-exponential character will be obvious, indicating complex dynamics in the system. The PL decay can then be described by a stretched exponential function [6]:

$$I(t) \sim \exp(-\Gamma \cdot t)^{\beta} \tag{2.12}$$

where β is a dispersion factor. When $\beta = 1$ the decay turns into a mono-exponential function, but the more β deviates from unity the stronger the variation in the decay rates in the system.

As with a spectrally dispersed signal, the detection instrument may have limits in terms of temporal resolution. Detector's dead time is one source of the limitations for the recording of fast decays. Excitation laser rise and fall times are also finite. This means that the decays, which are shorter than the excitation fall time, cannot be resolved. Experimental decay curves will be system-response limited in that case.

Photon statistics shows if the emission process is purely random, or an emission event pre-disposes further events in some way. A mathematical measure of this is a correlation function and a convenient way to define it for an intensity trace $I(t)$ is:

$$R(\tau) = \frac{< I(t) \cdot I(t + \tau) > - < I(t) >^2}{< I(t)^2 > - < I(t) >^2} \tag{2.13}$$

where for the lack of correlation $R(\tau) = 0$ and for the full correlation $R(\tau) = 1$. In practice a "second-order autocorrelation function" is also often used:

$$g^{(2)}(\tau) = \frac{< I(t) \cdot I(t + \tau) >}{< I(t)^2 >} \tag{2.14}$$

This definition stems from the photon field correlations, where a "first-order" refers to the electric field, and the "second-order" to the squares of it, i.e. to the intensity. The expression in the denominator is a constant for a given trace. A term "normalized" is sometimes applied to highlight this fact.

A random source of light (no memory in the emission process) possesses Poisson statistics. Laser light is one example of such an emitter. Deviations exist, such as a thermal source, where photons tend to follow one another, the so called "bunching" effect (super-Poisson statistics). Sub-Poisson statistics takes place for a single photon emitter, such as defects in semiconductors or individual quantum dots. It can be understood considering their finite lifetime: the emitter cannot produce another photon right after the previous one. This feature is not only an emitter characteristic, but also a source of single photons for quantum communications and computations.

2.3   Application to Micro- and Nano-scale Materials

In this part we will provide examples how different regimes and modality of micro-photoluminescence are used in studies of micro- and nano-scale materials. It will be shown how analysis of the PL signal in spatial, spectral and time domains gives information on material properties, and how it can be useful for many different applications.

The samples are, in most cases, semiconductor or semiconductor/metal micro- and nanostructures, or combinations of different semiconductor layers, referred to
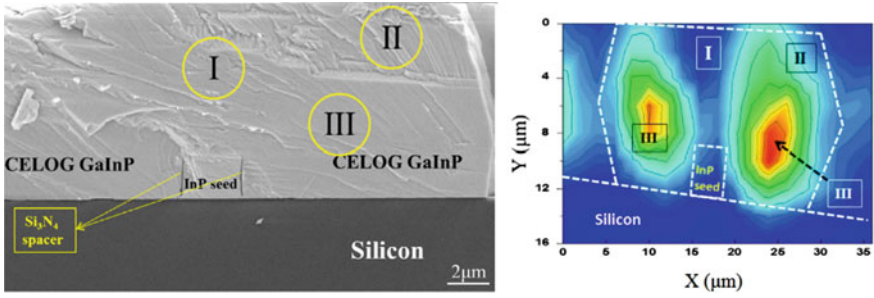
Fig. 2.9 (left) Structural (scanning electron microscope) and (right) optical (micro-photoluminescence intensity map) characterization of heterostructures [7]

as heterostructures. When their thickness approaches Bohr exciton radius the term quantum well (QW) is used (two-dimensional structure with a carrier confinement in one direction). One-dimensional structures are called nanowires (NW) and their inner structure can be resolved as far as the diffraction limit permits. Zero-dimensional objects, the nanocrystals, with sizes similar or below Bohr exciton radius are called quantum dots (QD). These are below the light diffraction limit, because Bohr exciton radius values are typically about several nanometers in semiconductors.

In preparation of such samples for micro-PL it is important to avoid parasitic signal which may quickly surpass the detected signal from the substrate, i.e. to choose a non-luminescent substrate material. In order to probe *individual* nanostructures a sufficient spatial spread (about one object per square micron) is needed to avoid contribution from different objects into the signal from a diffraction limited spot.

Intensity Mapping

This is one of the most straightforward regimes of the micro-PL. A wide field microscope can provide a PL image of an area of about $100 \times 100 \, \mu m$ with a single shot, using a high magnification objective lens. In Fig. 2.9 an example of this modality is shown for a heterostructure of group III–V semiconductors grown on silicon [7].

Integration of direct bandgap semiconductors with the most ubiquitous, yet of indirect bandgap, semiconductor in electronics, is an important research topic. The goal is to allow a combination of photon and electron-based communication across the device. The problem, however, is a large lattice mismatch between silicon and III-V materials, which makes a seamless crystal growth difficult. Many defects are formed at the interface. To avoid those special integration methods can be used, such as a seed-based growth.

In Fig. 2.9, left, a cross-section scanning electron microscopy (SEM) image of an epitaxial lateral overgrowth heterostructure of GaInP, grown from an InP seed on top of a silicon substrate is shown. The bandgap of GaInP (~765 nm wavelength) is different from the pure InP (~980 nm) and from silicon (~1100 nm). In Fig. 2.9, right, a cross-section μ-PL map of the heterostructure detected at the GaInP emission wavelength of 765 nm is provided. Regions I, II and III correspond to the regions
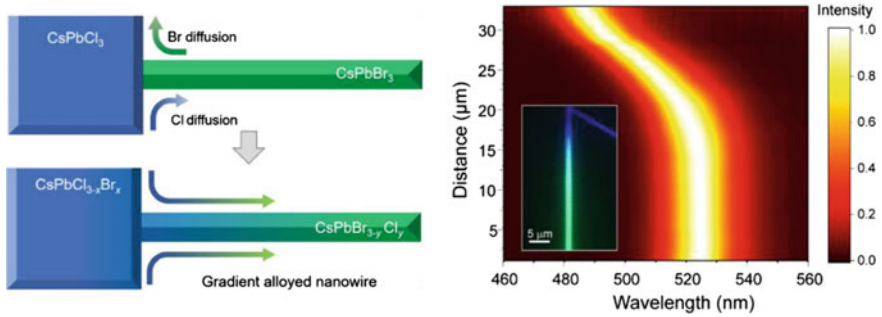
Fig. 2.10 Observation of material inter-diffusion in a sub-μm diameter perovskite nanowire by the μ-PL [8]

seen in the SEM image. A high crystalline quality of the region III corresponds to the maximum PL intensity. Region I, right above the seed, contains many defects, which act as non-radiative recombination centers quenching the GaInP luminescence. InP and Si appear dark at this emission wavelength. The luminescence was excited by a green laser diode.

Spectral Mapping

Similarly to the spatial distribution of the total PL signal, the emission spectra at different points can be analyzed. For example, one can record a series of PL spectra along a real space axis to monitor spatial variations in the sample.

In Fig. 2.10, left, a schematics of a perovskite heterostructure with different anions is shown. Perovskites is a family of materials with a certain crystal structure. They can be relatively easy synthesized by chemical methods. The structurally unstable constituents, however, may undergo inter-diffusion within such heterostructures even at room temperature.

With the help of micro-PL emission spectra, resolved along the nanowire length, this process can be monitored (Fig. 2.10, right). Blueish PL (~480 nm) corresponds to Cl^- anion, while the greenish peak (~520 nm) is from the Br^- containing perovskite compound. By PL spectrum mapping as a function of time and temperature the inter-diffusion constants can be retrieved. Laser line excitation at 400 nm was applied.

PL Spectroscopy

High-resolution micro-PL spectroscopy can provide information about the structure of the lowest energy state. To reduce contributions from thermal vibrations, phonons, such experiments are often performed at low temperature. Phonon modes can then be clearly resolved as separate peaks.

An example of individual CdSe quantum dots is shown in Fig. 2.11. They were synthesized in a colloidal form and dispersed on a glass substrate to access single particles. Embedding these particles (average size ~4 nm) in a polymer layer helped to improve their photostability. A broad spectrum is a signal measured from an ensemble of particles. Different size quantum dots in the ensemble possess different

Fig. 2.11 PL spectra of
individual CdSe QDs and the
enveloping ensemble
spectrum [9]. Copyright
(1996) by the American
Physical Society
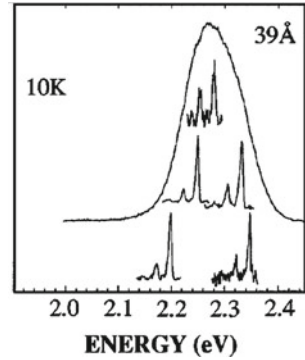
emission wavelengths due to the quantum size effect, leading to a broad integrated
signal. When individual nanocrystals are probed, on the other hand, the emission
peak position varies from dot-to-dot and the spectra become much narrower.

One can also notice lower intensity sidebands next to the main line. Those are
longitudinal optical (LO) phonon modes, where recombination takes place with the
emission of a phonon. This process reduces energy of the photon by the exact amount
of energy, corresponding to the material-specific lattice vibration. A typical signa-
ture of such modes is the repetitive pattern of the sideband with reducing intensity.
Acoustic phonon modes and transverse-optical (TO) phonon modes can also be
detected, depending on the material system. Their presence confirms that the emis-
sion takes place in the core of the nanocrystal. It also reveals the exciton-phonon
coupling mechanism, responsible to the temperature-dependent emission linewidth
broadening.

PLE Spectroscopy

Photoluminescence excitation spectroscopy (PLE) allows to probe higher energy
states, responsible for the light absorption. A standard measurement of sample
absorption is a transmission geometry. However, for small amount of material direct
transmission measurement is not feasible. The absolute absorption is so low that it is
buried in the noise, e.g. shot noise. When the sample is located on an opaque substrate
the standard measurement is also not possible. On the other hand, even individual
objects can be located and probed in micro-PL by PLE.

A direct result of such measurements is an absorption spectrum. For pumping
below saturation, the emission intensity $I(\lambda)$ can be expressed through the excitation
intensity $P(\lambda)$ as

$$I(\lambda) = \sigma(\lambda) \cdot P(\lambda) \cdot QY \cdot \gamma \qquad (2.15)$$

where QY is the quantum yield (efficiency of the excitation light conversion to radi-
ation) and γ is the collection efficiency. From this formula, using recorded emission

intensity and known excitation intensity, one can obtain the absorption cross section wavelength dependence $\sigma(\lambda)$.

In Fig. 2.12 the results of such measurements on individual Si quantum dots is shown [10]. A laser-driven Xe lamp (cf. Figure 2.6) was used, which provided a stable and tunable output for the excitation. Absorption cross section was obtained in a broad energy range. The presence of steps in the absorption curve reveals that energy levels tend to form mini-bands in a quantum dot of this material. Such results would be difficult to obtain by measuring on an ensemble of dots, where curves from many individual objects would sum up resulting in a smoothened shape. Such micro-PL results, obtained with PLE, allow to compare energy structure of nanostructures with theoretical calculations.

The lowest energy state (the emission state) can also be probed by PLE in a resonant excitation regime. The stray laser light needs to be rejected from the detector by other means than spectrally, and polarization is often used. In Fig. 2.13 PLE from scanning of the excitation wavelength by a tunable laser in a resonant excitation regime is shown for InGaAs quantum dots. Very narrow energy levels, corresponding to discrete states can be resolved, as well as their fine structure [11]. In fact, the linewidth is at its natural limit, i.e. radiative lifetime-limited: $\Delta E = h/\Delta\tau$.

Fig. 2.12 PL (left) and PLE (right) spectra of a Si quantum dot. Reprinted with permission from Sychugov et al. [10]. Copyright (2016) by the American Physical Society
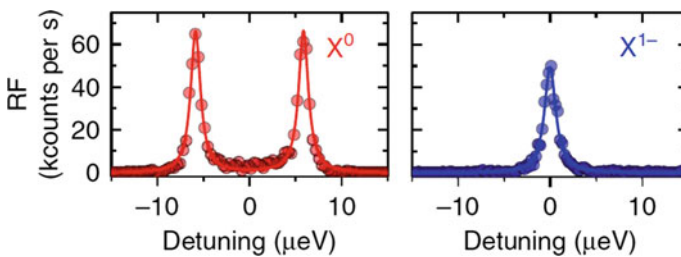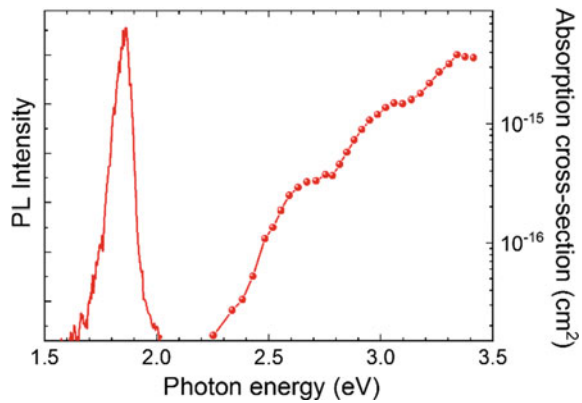
Fig. 2.13 PLE spectra of the lowest emission state (resonance fluorescence) for the neutral (left) and charged (right) exciton in single InGaAs quantum dots [11]. Licensed under a Creative Commons Attribution 4.0 International License
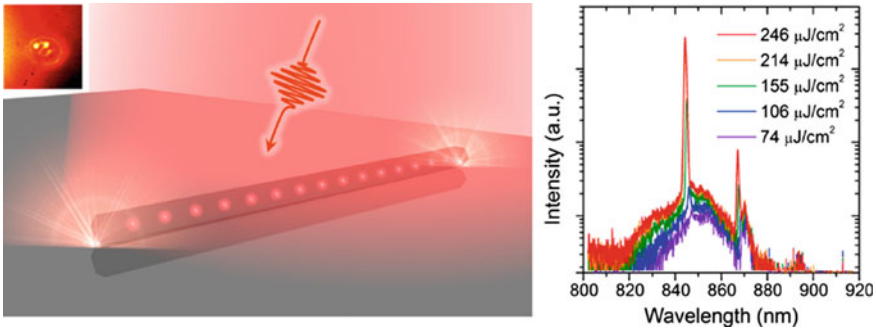
Fig. 2.14   Lasing from a single GaAs nanowire with diameter ~200 μm, acting as a resonator, with InGaAs quantum dots inside. Reprinted with permission from Ho et al. [13]

Excitation Power Scan

Higher energy levels can also be probed without PLE, but using ordinary PL under stronger excitation. When gradually filling the states from the lowest one, the emission from higher states can be spectrally resolved [12]. Another important implication of a power scan is a change in the emission mechanism, for example a transition from spontaneous to stimulated emission.

In Fig. 2.14, inset, a micro-PL image of a ~2 μm long GaAs nano-wire is shown. Two bright spots correspond to the emission from the edges of the nanowire, acting as a resonator for the PL from embedded InGaAs quantum dots. The schematics of such a NW with emitters inside under external excitation is shown in Fig. 2.14, left. When spectrally resolved, the emission has a broad peak at low excitation due to the inhomogeneous broadening (different size quantum dots contribute different wavelengths). It is shown in Fig. 2.14, right, as the weakest spectrum.

With increasing excitation power, the luminescence spectra changes, starting to reveal a sharp line at ~845 nm. This is a signature of a lasing behavior, where ordinary luminescence (spontaneous emission) turns into stimulated emission, thanks to the feedback from the resonator. Another feature of this transition is a radiative lifetime shortening and a shift from the super-Poisson to Poisson statistics, both being measurable quantities in micro-PL, as will be shown below.

Interferometry

To probe very narrow emission lines, such as emission from quantum dots (Fig. 2.13) or from lasing modes (Fig. 2.14), the spectral resolution of a spectrometer may not be sufficient. Interferometers of Fabry–Perot [14] or Michelson type [15] can then be used.

In Fig. 2.15 an example setup for micro-PL interferometry is shown [14]. A green emitting diode laser (~534 nm) is used for non-resonant excitation. To remove thermal fluctuations the samples is cooled by the cryostat to 10 K. Epifluorescence geometry is used to separate excitation and emission beams.
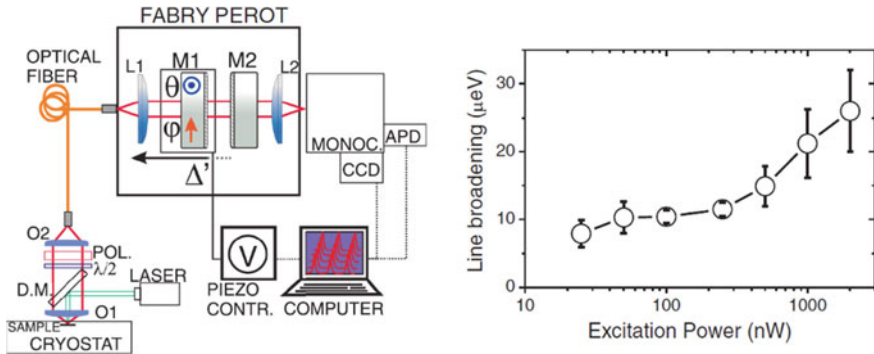
Fig. 2.15 (left) Interferometry setup for single GaAs QD micro-PL. (right) Measured linewidth as a function of excitation power. Reprinted with permission from Abbarchi et al. [14]

Collected PL is coupled to an optical fiber and fed to the Fabry–Perot cavity. It consists of focusing lenses and two mirrors, where one can be moved using a piezo-controller. The theoretical spectral resolution achievable with such an interferometer is a few μeV, which is one to two orders of magnitude better than an ordinary spectrometer can provide.

A result of interferometry application is shown in Fig. 2.15, right. The linewidth from an individual GaAs quantum dot is shown as a function of the excitation power. At higher power the effect of random charges and local heating lead to the broadening of the linewidth. At low excitation power sub-10 μeV linewidth could be resolved by this technique.

Polarization Resolved Spectroscopy

Spectrally dispersed signals can be additionally filtered by its polarization. The plane of electric field vector oscillations shows if the emitting dipole has a preferential orientation, caused by geometrical or crystal structure anisotropy.

In Fig. 2.16 an example is shown for the emission spectrum of a single perovskite quantum dot, taken at low temperature under 570 nm pulsed laser excitation [16]. Red, black and blue curves correspond to 0°, 45°, and 90° polarization angles. From these data it is clear that the emission state has a fine structure with a splitting of ~1 meV. Each emitting state has a distinct polarization, orthogonal to the other. This result not only reveals energy structure of the zero-dimensional nanostructures of this material, but also makes possible applications of such QDs as single or entangled photon sources.

Magneto-Optical Measurements

Another way to get a better insight into the energy structure of the emitting state is magneto-optical measurements. The principle behind this technique is based on the Zeeman Effect. If two levels have very close positions in energy space, yet have a different magnetic moment, they can be separated from each other with the help of a

Fig. 2.16
Polarization-resolved
single-dot spectra of a
CsPbI$_3$ quantum dot.
Reprinted with permission
from Yin et al. [16]
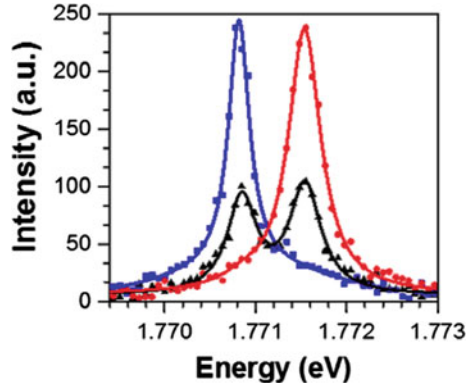
magnetic field. The energy shift is proportional to the strength of the magnetic field B, Bohr magneton μ_B (a constant), and the Landé factor g:

$$\Delta E = g\mu_B B \tag{2.16}$$

In Fig. 2.17 a single particle emission spectrum of a perovskite quantum dot is shown as a function of the external magnetic field in relation to its position in the field absence. Indeed, the separation between two contributions grows linearly with the applied field. From these data values of the Landé factor for electrons and holes can be separately extracted.

Intensity Trace

By monitoring the emission output intensity as a function of time one can detect any long-term fluctuations ("long" compared to the lifetime), or degradation of the investigated material. The quality of the active material and its surroundings can be probed in this modality. It is preferable to target an individual object or a very limited volume to avoid any averaging effects. For example, fluorescence intermittence, so called blinking, is often encountered in quantum dot emission [18]. Such a process would be impossible to identify with ordinary PL, where averaged signal from a large area, containing many emitters, is measured.

The model to explain blinking in QDs typically invokes a charge trap site, which temporarily render the particles charged by trapping holes or electrons [18]. Subsequent non-radiative recombination by the Auger process leads to the observation of "dark" periods, where all external excitation is transformed to heat (phonons) instead of detectable radiation. Emission of the trapped carrier neutralizes the QDs and signifies the recurrence of the "bright" period.

In Fig. 2.18 examples of a PL intensity trace, taken by a micro-PL setup, using 532 nm laser irradiation, is shown [19]. CdSe quantum dots without a shell ("bare")
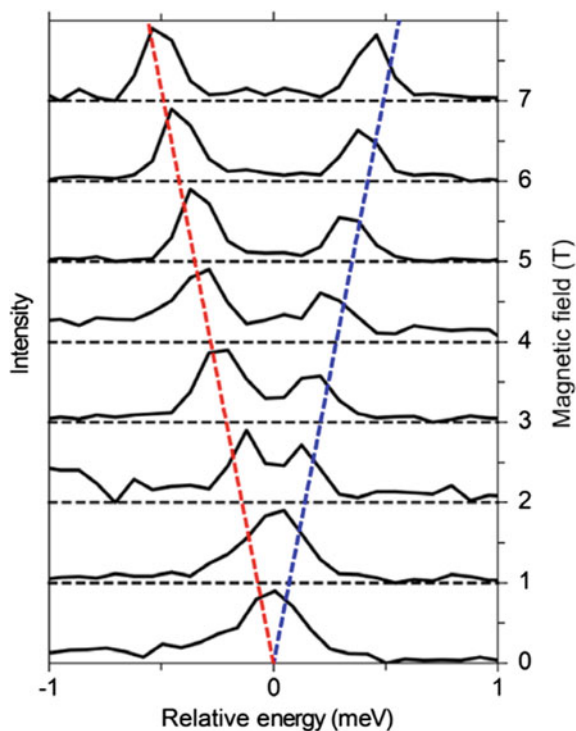
exhibit frequent blinking, while the presence of a shell ("ZnS") makes the PL inten-
sity trace more stable. This illustrates the importance of a high-quality "passiva-
tion" material to avoid charge trapping and subsequent energy loss in QDs for light
emission applications.
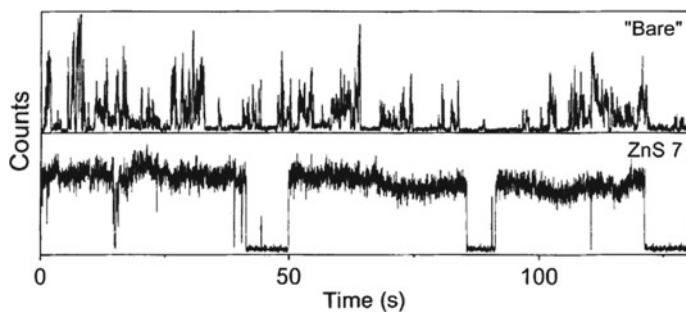
Fig. 2.18 PL intensity trace from a single CdSe QD with and without protective shell. Reprinted
with permission from Nirmal et al. [19]

Lifetime

The lifetime of the excitation can be extracted by PL decay measurements. A pulsed laser excitation is typically used and the emission intensity in between pulses is monitored.

Results are typically shown in a logarithmic-linear (lin-log) scale, as outlined in the previous chapter. Figure 2.19 shows an example of PL decays from individual CdSe quantum dots with a ZnS shell. A pulsed laser of ~458 nm wavelength and a ~5 MHz repetition rate was used for pumping. Decays from three different nanocrystals are shown. In this representation one can clearly see if the decay is of mono-exponential character (for the object III) or of a stretched-exponential appearance (objects IV and V). Characteristic decay times (the lifetime) can be extracted from fitting of the experimental data with a corresponding function (single or multi-exponential).

Photon Statistics

Analysis of the emitted signal in the time domain can be extended to the statistics of photon emission. Are photons generated randomly, or is there any pre-disposition for the subsequent emission after the light generation event?

In Fig. 2.20, left, schematics of the μ-PL system used for a high purity single-photon generation is shown [21]. PL from individual GaAs/AlGaAs quantum dots is detected by a superconducting single-photon detector (SSPD). The sample is placed in a cryostat and solid immersion with aspheric lenses (SIL and AL) are mounted for efficient signal collection. A quarter wavelength plate (QWP) is used for polarization state control, while the beam-splitter (BS) and the polarizer with a notch filter (Pol and NF) serve for the separation of excitation and emission. In addition, both excitation and collection signals are purified spectrally with the help of gratings (reflection,
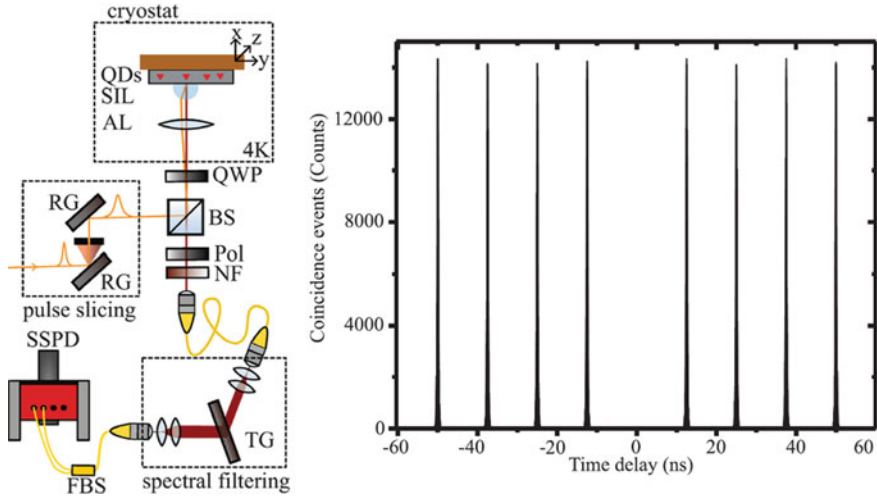
Fig. 2.20 Photon statistics measurements from a quantum dot under two-photon excitation [21]. Distributed under a Creative Commons Attribution (CC BY) license

RG, and transmission, TG). Two-photon ensures resonant excitation with very little extra energy to the system, resulting in reduced local heating and charge trapping.

Results of these measurements in shown in Fig. 2.20, right. The autocorrelation function under a pulsed excitation is shown, with a clear dip at zero delay. It reveals nearly perfect single photon emission characteristics of the source. Such a source of single photons can serve as a convenient tool for quantum communication and computation applications.

Second Harmonic Detection

This is an example of a non-linear processes in a sample, observable in a micro-PL setup. In general, the structure of the micro-fabricated sample, together with the crystal structure of the original material, can support second harmonic generation.

Consider nanopillars of GaP [22]. In an epifluorescence configuration, typical for micro-photoluminescence, the generated signal then becomes of higher energy than the excitation beam. The bandgap of GaP corresponds to the wavelength of ~550 nm and, when excited with a ~840 nm beam, the material is transparent and no photoluminescence can be detected. At the same time, the second harmonic (~420 nm) generation from the nanopillar area can be spatially resolved with the help of epifluorescence micro-PL configuration (Fig. 2.21).

Fig. 2.21 Schematics of
second harmonics generation
from GaP nanopillars.
Reprinted with permission
from Sanatinia et al. [22]
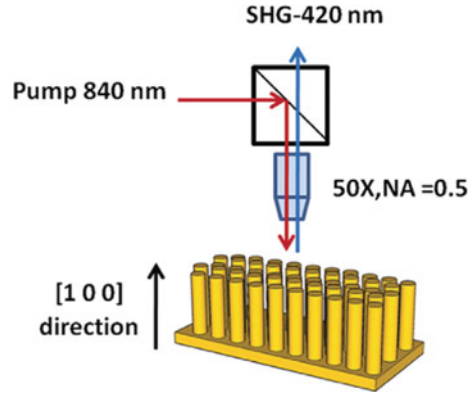
References

1. Valenta, J., & Pelant, I. (2012). *Luminescence spectroscopy of semiconductors*. Oxford University Press.
2. Sychugov, I., Valenta, J., & Linnros, J. (2017). Probing silicon quantum dots by single-dot techniques. *Nanotechnology, 28*, 072002.
3. Goltsman, G. N., Semenov, A. D., Gousev, Y. P., Zorin, M. A., Gogidze, I. G., Gershenzon, E. M., Lang, P. T., Knott, W. J., & Renk, K. F. (1991). Sensitive picosecond NbN detector for radiation from millimeter wavelengths to visible light. *Superconductor Science and Technology, 4*, 453–456.
4. Bulter, A. (2015). Single-photon counting detectors for the visible range between 300 and 1000 nm. In P. Kapusta (Ed.), *Advanced photon counting*. Springer.
5. Boyle, W. S., & Smith, G. E. (1970). Charge coupled semiconductor devices. *The Bell System Technical Journal, 49*, 587–593.
6. Greben, M., Khoroshyy, P., Sychugov, I., & Valenta, J. (2019). Non-exponential decay kinetics: Correct assessment and description illustrated by slow luminescence of Si nanostructures. *Applied Spectroscopy Reviews, 54*, 1–44.
7. Omanakuttan, G., Stergiakis, S., Sahgal, A., Sychugov, I., Lourdudoss, S., & Sun, Y.-T. (2017). Epitaxial lateral overgrowth of $Ga_xIn_{1-x}P$ toward direct $Ga_xIn_{1-x}P$/Si heterojunction. *Physica Status Solidi A, 214*, 1600631.
8. Pan, D. X., Fu, Y. P., Chen, J., Czech, K. J., Wright, J. C., & Jin, S. (2018). Visualization and studies of ion-diffusion kinetics in cesium lead bromide perovskite nanowires. *Nano Letters, 18*, 1807–1813.
9. Empedocles, S. A., Norris, D. J., & Bawendi, M. G. (1996). Photoluminescence spectroscopy of single CdSe nanocrystallite quantum dots. *Physical Review Letters, 77*, 3873–3876.
10. Sychugov, I., Pevere, F., Luo, J.-W., Zunger, A., & Linnros, J. (2016). Single-dot absorption spectroscopy and theory of silicon nanocrystals. *Physical Reviews B, 93,* 161413.
11. Kuhlmann, A. V., Prechtel, J. H., Houel, J., Ludwig, A., Reuter, D., Wieck, A. D., & Warburton, R. J. (2015). Transform-limited single photons from a single quantum dot. *Nature Communication, 6,* 8204.
12. Bayer, M., Stern, O., Hawrylak, P., Fafard, S., & Forchel, A. (2000). Hidden symmetries in the energy levels of excitonic 'artificial atoms.' *Nature, 405*, 923–926.
13. Ho, J. F., Tatebayashi, J., Sergent, S., Fong, C. F., Ota, Y., Iwamoto, S., & Arakawa, Y. (2016). A nanowire-based plasmonic quantum dot laser. *Nano Letters, 16*, 2845–2850.
14. Abbarchi, M., Kuroda, T., Duval, R., Mano, T., & Sakoda, K. (2011). Scanning Fabry-Perot interferometer with largely tuneable free spectral range for high resolution spectroscopy of single quantum dots. *Review of Scientific Instruments, 82*, 073103.

15. Santori, C., Fattal, D., Vuckovic, J., Solomon, G. S., & Yamamoto, Y. (2002). Indistinguishable photons from a single-photon device. *Nature, 419*, 594–597.
16. Yin, C. Y., Chen, L. Y., Song, N., Lv, Y., Hu, F. R., Sun, C., Yu, W. W., Zhang, C. F., Wang, X. Y., Zhang, Y., & Xiao, M. (2017). Bright-exciton fine-structure splittings in single perovskite nanocrystals. *Physics Review Letters, 119,* 026401.
17. Fu, M., Tamarat, P., Huang, H., Even, J., Rogach, A. L., & Lounis, B. (2017). Neutral and charged exciton fine structure in single lead halide perovskite nanocrystals revealed by magneto-optical spectroscopy. *Nano Letters, 17*, 2895–2901.
18. Efros, A. L., & Nesbitt, D. J. (2016). Origin and control of blinking in quantum dots. *Nature Nanotechnology, 11*, 661–671.
19. Nirmal, M., Dabbousi, B. O., Bawendi, M. G., Macklin, J. J., Trautman, J. K., Harris, T. D., & Brus, L. E. (1996). Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature, 383*, 802–804.
20. Schlegel, G., Bohnenberger, J., Potapova, I., & Mews, A. (2002). Fluorescence decay time of single semiconductor nanocrystals. *Physical Review Letters, 88*, 137401.
21. Schweickert, L., Jons, K. D., Zeuner, K. D., da Silva, S. F. C., Huang, H. Y., Lettner, T., Reindl, M., Zichi, J., Trotta, R., Rastelli, A., Zwiller, V. (2018). On-demand generation of background-free single photons from a solid-state source. *Applied Physics Letters, 112*, 093106.
22. Sanatinia, R., Swillo, M., & Anand, S. (2012). Surface second-harmonic generation from vertical GaP nanopillars. *Nano Letters, 12*, 820–826.

Chapter 3
Raman Spectroscopy, Fourier Transform Infrared Spectroscopy (FTIR) and X-Ray Photoelectron Spectroscopy (XPS)

3.1 Principles of Raman Spectroscopy

When the light from a laser with a well-defined wavelength λ_{laser}, shines on, for instance, a molecule it may be scattered elastically or inelastically, i.e., without or with a change of energy in the event. In the collision the oscillating electromagnetic field of the photon creates a polarization of the molecular electron cloud. As a result, the molecule possesses a higher energy state due to the energy of the photon transferred to the molecule. This condition lasts for a very short time and it will lead to an unstable virtual state for the molecule and almost instantaneously the added photon energy is emitted as scattered light. There are now three possibilities for the wavelength of the scattered photon, $\lambda_{scatter}$: $\lambda_{scatter} > \lambda_{laser}$, referred to as Stokes Raman scattering, $\lambda_{scatter} = \lambda_{laser}$, or Rayleigh scattering and $\lambda_{scatter} < \lambda_{laser}$, or anti-Stoke Raman scattering, as shown in Fig. 3.1. Rayleigh scattering is considered elastic since no energy is lost, while Stokes-Raman scattering and anti-Stokes-Raman scattering are inelastic processes, loosing and gaining energy, respectively.

The discovery of Raman spectroscopy extends to 1928 when the Indian physicist C.V. Raman was shining light on different liquids from a mercury lamp and investigated the scattered light. In these experiments, he observed a new spectral line of outgoing light with a frequency lower than the incident light [1]. After his discovery, a series of reports followed, performed in a similar way to his experiments, but on crystals [2]. It was established that when the wavelength of the emitted light is same as the excitation light, elastic scattering, or *Rayleigh scattering*, occurs, while, *Raman scattering occurs if the wavelength has changed*. In Rayleigh scattering, no energy is exchanged but the direction of the incident and emitted light may differ. Raman scattering is an inelastic collision with both direction and energy change. The difference in energy between emitted and incident wavelengths is named after Sir George Gabriel Stokes, and may be positive, Stokes shift, or negative, anti-Stokes shift. For anti-Stokes shift, the outgoing light has larger energy than the incoming, which may appear strange, but will be explained below.
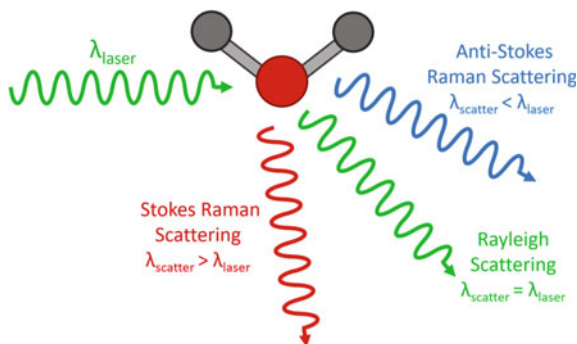
In Raman spectroscopy, the inelastic scattering of light occurs when the incident photons collides with atoms or molecules. A part of the atomic, or molecular vibrational/rotational energy is picked up the photon or, vice versa, energy is absorbed and transferred to the atom/molecule. For Raman scattering by a lattice of atoms or molecules, the incident light interacts with quantized lattice vibrations, phonons. When the interaction of light is with acoustic phonons, having lower energy, it is called *Brillouin scattering*, while higher energy photons are referred to as optical phonons, with which the collision results in a larger shift of the photon energy [3].

The difference between Rayleigh and Raman scattering is described in more detail by Fig. 3.2. Initially, the molecules (or atoms) are in the ground state (E_0) but there are also excited vibrational states with higher energy, such as E_1. When the incident light of frequency v_0 interacts with the molecules, electrons may absorb a photon, resulting in a transition to virtual higher energy states. Electrons occupying a high energy state can then return to their original ground state while emitting a photon. For the Raman process, the de-excitation of the electron occurs to another level than it was excited from. This results in two different energies of the photons emitted in the Raman process. The energy difference between the two photons can be written as

$$\Delta E = h(v_0 - v) = h\Delta v$$

Here, h is Planck's constant, $E_0 + hv_0$ and $E_1 + hv_0$ are the energy of the virtual states and Δv stands for the frequency shift in Raman spectroscopy. Anti-Stokes line appears when the emitted photon frequency is higher than the original frequency in incident photon [3].

The Raman shift is defined as the frequency difference between the incident light and the Stokes, or anti-Stokes lines. The size of the Raman shift is equal to the transition energy difference and has the same magnitude for both Stokes and anti-Stokes lines for the same molecular energy level. Meanwhile, the molecules are often in the ground state during normal conditions and, therefore, the intensity of Stokes line is remarkably stronger than anti-Stokes line and they are more often used and investigated in Raman spectroscopy analysis, as shown in Fig. 3.3.

Molecules or atoms   Molecular vibration

(a) Rayleigh Scattering   (b) Raman Scattering
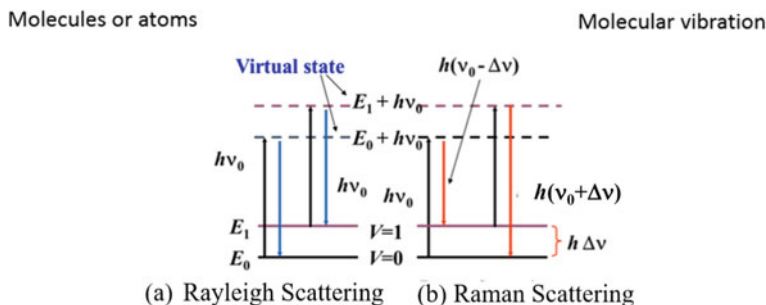
Fig. 3.2 Schematic of light scattering **a** for Rayleigh scattering when the frequency of light is preserved after scattering, and **b** for Raman scattering as the photons of incident light is in exchange of their energy with the molecules or atoms. Black arrows are excitations and blue and red arrows are Rayleigh and Raman de-excitations, respectively
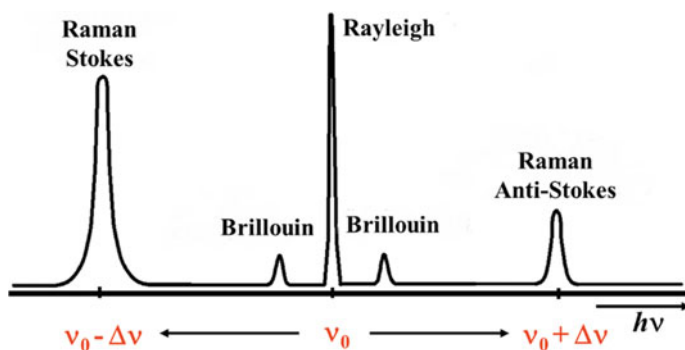
Fig. 3.3 Schematic of the energy distribution of scattered light

Both Rayleigh and Raman scattering are relatively inefficient processes. The Rayleigh scattering has a typically intensity three orders of magnitude lower than the incident light, whereas Raman scattering intensity may even be six orders of magnitude lower the original light [3]. This means that a powerful laser is required for Raman scattering spectroscopy.

Raman spectra from solid materials represent an interaction of light with phonons and results in vibrational modes in the solid. The vibrational motion can occur orthogonal and parallel to the propagation of the disturbance and these modes are referred to as transvers and longitudinal, respectively. In addition, the vibrational modes are further divided in acoustic (low energy) and optical (high energy) branches, where atoms are vibrating coherently in the acoustic mode, but out of phase with their neighbors in the optical mode. Totally, this gives rise to four general branches in a phonon spectrum: transverse acoustic (TA), transverse optical (TO), longitudinal acoustic (LA) and longitudinal optical (LO). Many other modes and peculiarities of phonon spectra exist, depending on lattice types, defects, atomic basis etc. For instance, in

case of strained materials, TO phonons carry the signature of the strain type and they shift to lower and higher frequencies in presence of tensile and compressive strain, respectively [4–7].

3.2 Raman Measurements

3.2.1 SiGe Alloys Analysis

The Raman spectra of strained, or relaxed SiGe layers consists of three peaks where their energies depend on the Ge content and strain. These three peaks are generated from the vibration of Ge–Ge bonds at ~290 cm^{-1}, Si–Ge at ~405 cm^{-1} and Si–Si at ~502 cm^{-1} [8–10]. Figure 3.4a, b illustrates Raman spectra from a relaxed and strained SiGe layer grown on Si. The signature peaks of SiGe have shifted due to Ge content (or strain).

In the analysis of Si$_{1-x}$Ge$_x$ layers the following equations have been proposed to identify the composition, x [11]. The position of Ge–Ge (GG), Si–Si (SS), Si–Ge (SG) peaks can be estimated from the following equations:

$$\omega_{SS}(x) = 520.7 - 66.9x \tag{3.1a}$$

$$\omega_{SG}(x) = 400.1 + 24.5x - 4.5x^2 - 33.5x^3 \tag{3.1b}$$

$$\omega_{GG}(x) = 280.3 + 19.4x \tag{3.1c}$$

A relationship between the strain (ε), mismatch (f) and strain relaxation (r) can be written as:
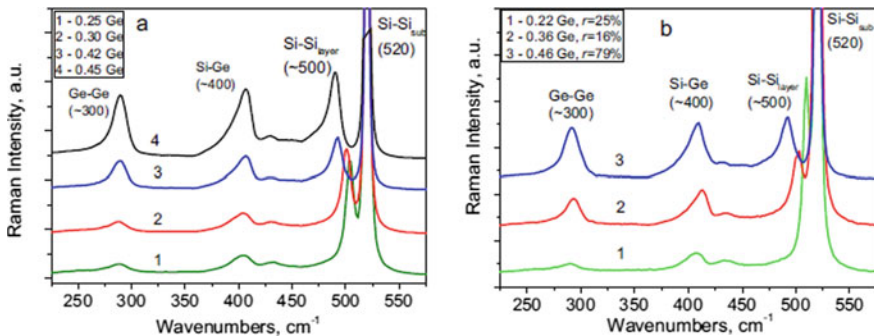
Fig. 3.4 Raman spectra of SiGe layer grown on Si with different Ge contents when **a** strain relaxed SiGe, and **b** strained SiGe. The Si–Si peak of the SiGe layer and the substrate are separated

$$\varepsilon = -f(1-r) \tag{3.2}$$

where f can be calculated from linear interpolation using Vegard's law, or a parabolic function of the Ge content, as discussed in Eq. 1.47 in Chap. 1 [11].

One important point in the Raman measurements is how to determine a precise energy position and intensity of a peak in the spectra. For these measurements, unpolarized light is used at room temperature in backscattering geometry. The measurements are performed by using, for example, an Ar^+ ion laser operating at 514.5, 488 nm, and a HeCd laser for 457 and 325 nm. In order to omit heating the sample and to avoid the artificial shift of peaks, the power of the laser has been kept in a range of just a few mW. The beam is passing microscope objectives to be magnified and deliver a spot size of 1–3 μm.

The scattered beam from a sample is collected and directed to the spectrometer. As mentioned before, signature peaks of SiGe are for Ge–Ge, Si–Ge, and Si–Si vibrational modes. The position of Si–Si peak obtained from the Si substrate is widely used as a reference for analysis of SiGe layers. Suitable software may be used to fit the Raman peaks by using different functions to mimic the peak shapes. The Si–Ge peaks here are fitted with an asymmetrical function (see example in Fig. 3.5), whereas Si–Si peaks from the Si substrate are fitted by a Lorentzian function, or by a mixture of Lorentzian and Gaussian functions [12]. For fitting an asymmetrical peak, the half width at half maximum (HWHM) of the peak, the intensity and the wave number of an asymmetrical peak are the parameters of interest. Note that the width of symmetrical peak is determined by the full width at half maximum (FWHM).

The final intensity as a function of wave number can be obtained by applying the following complicated equation [12]:

$$I(\omega) = \frac{1}{2} \times \frac{[1 - sign(\omega - \omega_0)] \times I_0}{\left(\frac{\omega - \omega_0}{W_1}\right)^2 + 1} + \frac{1}{2} \times \frac{[1 - sign(\omega_0 - \omega)] \times I_0}{\left(\frac{\omega_0 - \omega}{W_2}\right)^2 + 1}$$

Fig. 3.5 Asymmetric and Gaussian fitting methods for experimental peak of a Raman spectra
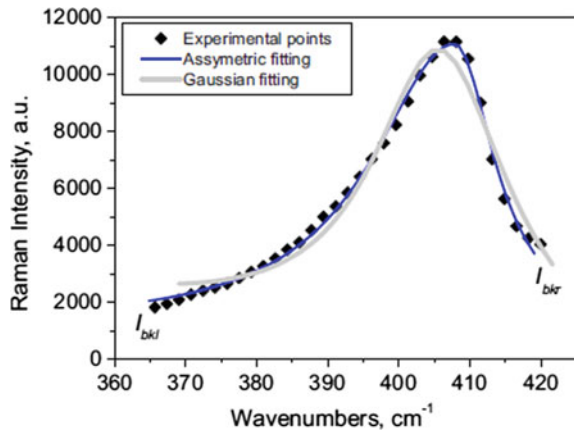
$$+ \frac{1}{2} \times [1 - sign(\omega - \omega_0)] \times I_{bkl} + \frac{1}{2} \times [1 - sign(\omega_0 - \omega)] \times I_{bkr}$$

$$(3.3)$$

where I_0 and ω_0 are the intensity and wavenumber of a peak, W_1 and W_2 are HWHM at lower and higher frequency, and I_{bkl} and I_{bkr} are lower and higher background intensity. The asymmetrical function is the best choice for the fit since the shape of Si–Ge peak becomes more asymmetrical for high Ge contents.

Another method to analyze the peaks in Raman spectra is the so-called intensity method. In contrast to the frequency method, discussed above, the intensity method has the benefit of not needing a fitting function. The method is also applicable for both strained and relaxed layers, but the drawback is that the intensity of Si–Ge and Ge–Ge peaks are strongly dependent on how the measurement are done. This problem becomes more severe when there is uncertainty to determine where the baseline for a particular peak begins and where it finishes. For Raman spectra of a SiGe layer, the relative intensity of peak modes in respect to Ge content can be written as:

$$\frac{I(\text{Si} - \text{Si})}{I(\text{Si} - \text{Ge})} = \frac{A(1 - x)}{2x} \tag{3.4a}$$

and

$$\frac{I(\text{Ge} - \text{Ge})}{I(\text{Si} - \text{Ge})} = \frac{Bx}{2(1 - x)} \tag{3.4b}$$

Due to resonance effects, the intensities of the different modes change with the wavelength of the incident light. Therefore, coefficients A and B in Eq. 3.4a, b, respectively, can be obtained experimentally for certain wavelength. In Eq. 3.4a, b, x determines the composition according to $Si_{1-x}Ge_x$, which can be estimated, for example, by using XRD. The intensities of different phonon modes can be calculated by using commercial software program for peak profile analysis.

3.2.2 Crystal Orientation

As an example, Raman spectroscopy is widely used for characterizing diamond films. In general, diamond films are epitaxially grown on diamond substrates, but the growth on other types of substrates, such as boron nitride (BN), may induce strain in the films (sp^3-bonded carbon), or sp^2-bonded carbon can be formed. The grown diamond material has one peak at ~1332 cm^{-1} as well as an asymmetric broad band centered at ~1530 cm^{-1}, presenting sp^2 carbon, as shown in Fig. 3.6. The sp^3/ sp^2 bonding ratio can be determined from the relative intensity of the above bands. Furthermore, the quantity and type of stress in the diamond films can be estimated

Fig. 3.6 Raman spectra of a diamond film grown on (111) surface of a c-BN microcrystal

c-BN(TO)

1550 cm^{-1} band

c-BN(LO) diamond

Intensity (arbitrary units) →

800 1000 1200 1400 1600 1800

Raman shift/cm^{-1}
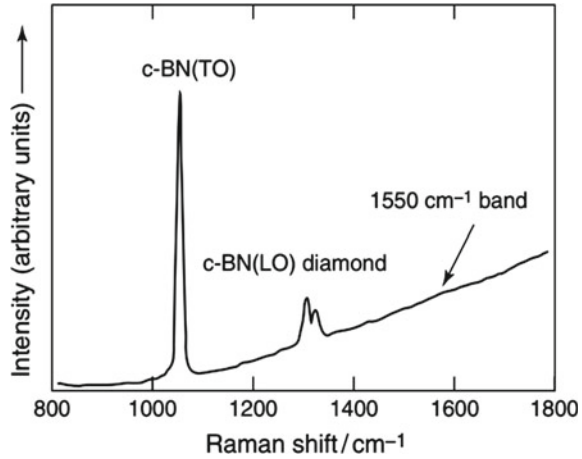
by measuring the amount and direction of frequency shift of the Raman peaks of diamond [13, 14].

3.2.3 Strain Characterization

Figure 3.7 illustrates Raman spectra of the cross section of a 4H–SiC MOSFET measured in the backscattering configuration. In the spectra, the peak intensities are ranked from highest to lowest as: (FTA), 203.5 cm^{-1}; (FLA), 610.5 cm^{-1}; (FTO), 776.0 cm^{-1}; and A_1 (FLO), 964.0 cm^{-1} (F is so called folded modes). The shift of the peaks can give information on compressive and tensile strain. In these measurements, the intensity of mode FTO (2/4) was too strong and the frequency shift of FTO (0) peak at 790 cm^{-1} was measured to determine the stress amount over the cross section the 4H–SiC MOSFET [15].

3.2.4 LO Phonon-Plasmon Coupled Modes

In material physics, plasmons are quasiparticles consisting of an oscillating plasma volume of the free electron gas. Plasmons can be induced by, for instance, a laser pulse, and in III–V semiconductors, the excited plasmons can interact with the LO phonon through the macroscopic electric fields and form a longitudinal optical phonon–plasmon coupled (LOPC) mode. The LOPC modes appear in broad shapes and when the free carrier density increases in the material the associated Raman peaks shift to a higher frequency.
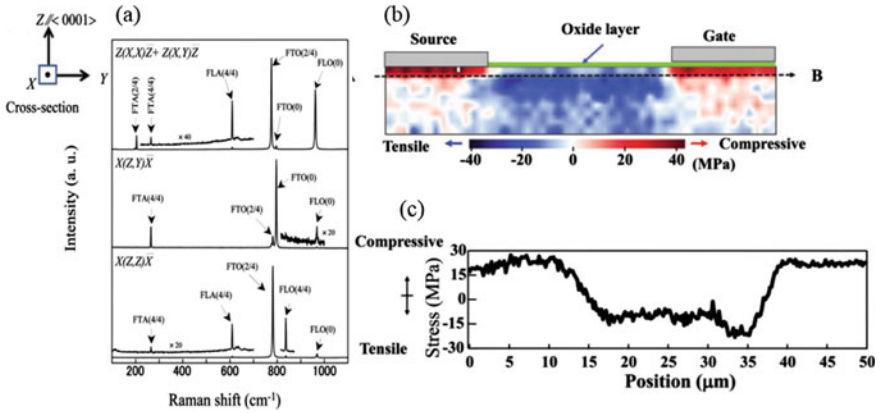
Fig. 3.7 **a** Raman spectra measured at different positions across the 4H–SiC MOSFET, performed in the backscattering configuration, **b** map and **c** line profile of the obtained stress values from the 4H–SiC MOSFET [15]
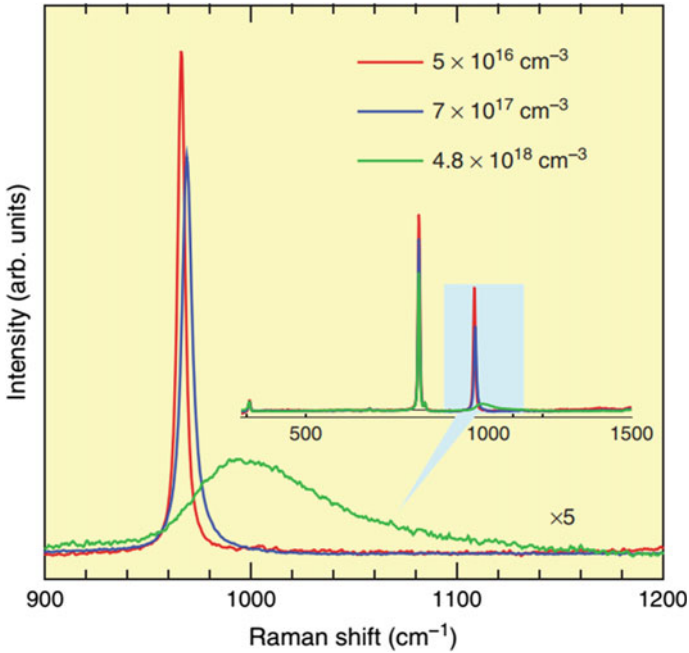
Fig. 3.8 Raman spectra of LOPC modes for an N doped 4H–SiC wafer with different carrier concentration [15]

The sensitivity of LOPC modes to carrier concentration has also been proven for SiC material (Group IV). Figure 3.8 illustrates how Raman line-shape of LOPC modes changes for different nitrogen (N) donor doping in SiC wafers. The LOPC mode becomes broader and the peak is positioned at higher frequency with increasing carrier concentration. In this way, there is a possibility to estimate carrier concentration in 4H–SiC by using the carrier density dependence of LOPC modes.

The peaks of LOPC modes has also been investigated in SiC growth to identify SiC polytypes. Polytypes are different forms of SiC crystals, where the crystal planes are stacked in different order, leading to very different electronic properties between polytypes, although the nearest neighbor Si–C bonds and the mass density are preserved. Commonly found polytypes are 3C, 4H and 6H, where 3C has a cubic structure and 4H and 6H are hexagonal. Raman intensity profiles of different SiC polytypes for the FTA and FTO modes have been estimated theoretically.

Table 3.1 Raman frequencies data for the folded of the FTA and FTO modes with E(E1, E2) symmetry and FLA and FLO modes with A1 symmetry are displayed for typical SiC polytypes

Polytype	$x = q/q_B$	Frequency (cm^{-1})			
		Planar acoustic	Planar optic	Axial acoustic	Axial optic
		FTA	FTO	FLA	FLO
3C	0	–	796	–	972
2H	0	–	799	–	968
	1	264	764	–	–
4H	0	–	796	–	964
	2/4	196, 204	776	–	–
	4/4	266		610	838
6H	0	–	797	–	965
	2/6	145, 150	789	–	–
	4/6	236, 241		504, 514	889
	6/6	266	767	–	–
8H	0	–	796	–	970
	2/8	112, 117	793	–	–
	4/8	203		403, 411	917, 923
	6/8	248, 252		–	–
	8/8	266	768	615	
15R	0	–	797	–	965
	2/5	167, 173	785	331, 337	932, 938
	4/5	255, 256	769	569, 577	860
21R	0	–	797	–	967
	2/7	126, 131	791	241, 250	
	4/7	217, 220	780	450, 458	905, 908
	6/7	261	767	590, 594	

These studies are based on a model where the Raman polarizability consists of contributions from all related bond Raman polarizabilities, which are dependent on the bond length variation. The intensity of the Raman profile is dependent on the atomic displacements, where profiles are sensitive how to input the force constants [14]. In such models there is an effort to consider short-range force constants which extended to the third neighbors in a one-dimensional approach. In this case, the force difference between cubic and hexagonal crystal has to be considered [14]. Since 4H– and 6H–SiC are uniaxial crystals, the LOPC mode appears anisotropically between the mode with the atomic displacement parallel to the c-axis (axial mode) and that with the atomic displacement perpendicular to the c-axis (planar mode) (Table 3.1).

3.2.5 Graphene Analysis

Raman spectroscopy has widely been applied to determine the quality and the number of monolayers for graphene. In such analysis, different Raman frequency bands, e.g. D, 2D and G bands, are analyzed.

The D band is an indicator of the defects in the graphene material, or shows the existence of carbon sp^3 bonds. In Raman spectra, if the graphene has high quality, the intensity of the D band is low. This peak normally appears at the edges of graphene flakes [16].

The 2D band is an indicator of the number of graphene sheets stacked on top of each other. For a monolayer of graphene, a single peak with narrow full-width-half maximum (FWHM) appears in the Raman spectra, while the peak of the 2D band for a multilayer of graphene is broad with a number of small peaks. As an example, a bilayer of graphene has 2D peaks at wavelengths of 514 and 633 nm in the spectra [17]. Figure 3.9 shows different cases for graphene material.

Fig. 3.9 Raman curves of graphene with different number of monolayers with a focus on 2D peak [18]
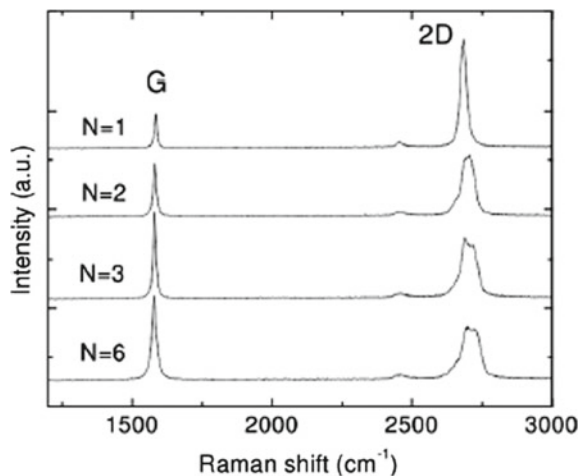
Fig. 3.10 Evolution of G
band with different number
of graphene sheets [19]
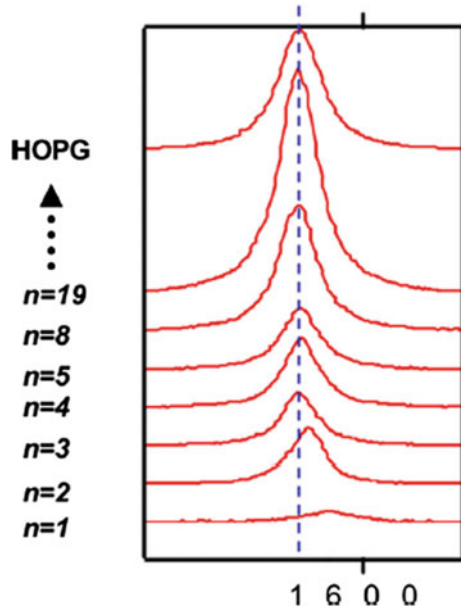
The G band is another signature for the number of graphene sheets. The intensity
and position of this band are two important factors in the analysis. For a single
graphene sheet, the intensity of the G band is low and the peak is located at around
1587 cm^{-1}, but for multilayered graphene, the intensity increases and peak may
slightly move towards lower wave numbers for more graphene layers, as shown in
Fig. 3.10.

3.3 Fourier Transform Infrared Spectroscopy

The Fourier transform infrared spectroscopy (FTIR) is a non-contact and non-
destructive technique with high-resolution ability and large luminous flux for char-
acterizing materials in many fields of science. The technique is rooted in the mathe-
matical Fourier transform, which converts a function into a new function describing
the frequency content of the original function. With the development of modern
computer technology and high-speed, high-resolution data acquisition equipment,
the advantages of FTIR measurement have become more prominent.

3.4 Set-Up and Working Principle of FTIR

The core component of the FTIR spectrometer is a Michelson interferometer, as shown in Fig. 3.11 [20]. The complete FTIR spectrometer includes fixed mirror, movable mirror, beam splitter, infrared light source, detector, corresponding optical collimation system and control system. The Fourier transform spectrometer modulates the incident light by the Michelson interferometer, and the detector converts the interference signal into an electrical signal, which is later changed into a digital interferogram after A/D conversion. There is a Fourier transform relationship between the interferogram and the spectrum of the incident light. By performing inverse Fourier transform on the interferogram, the spectral information of the incident light can be obtained.

The main working principle of the FTIR spectrometer is the following: a well-collimated infrared light beam passes through the interferometer and then it is divided into two beams by a beam splitter, one reflected and one transmitted beam. The transmitted beam proceeds to the fixed mirror (M_2), and is reflected back to the beam splitter. In the reflected path the beam meets the movable mirror (M_1) and is then reflected back, so the reflected and transmitted meet at the beam splitter again (the original transmission is now reflected and the original reflection is transmitted). The beams are now again on the same optical path before they finally reach the detector. By adjusting the movable mirror, the distance travelled by this beam is changed, so when the two beams unite, the path difference will cause interference. The spectrometer requires that the movable mirror must be capable of a high-precision uniform linear movement. The interfered light is transmitted to the IR detector to record the interference intensity versus optical path difference, and then the signal undergoes a Fourier transform to obtain a general infrared spectrum graph of absorption intensity versus frequency, or wavelength.

Assuming that the light source is an ideal monochromatic point light source which is divided into two light beams after passing the first time through the beam splitter.

Fig. 3.11 A schematic view of FTIR set-up based on the invention by Michelson et al. [20]
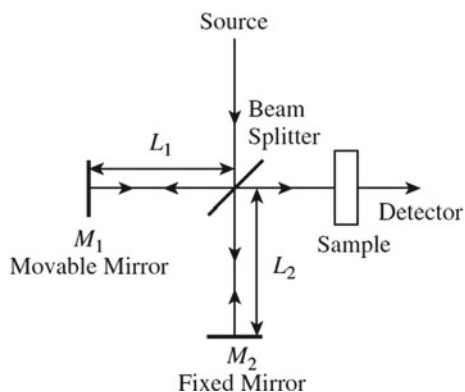
After joining the beams again, the interference of the light, the optical path difference between the two beams is:

$$x = 2|L_1 - L_2| \tag{3.5}$$

When $L_1 = L_2$, the optical path difference is zero ($x = 0$). At this time, the two beams reflected from the fixed mirror and the movable mirror back to the beam splitter have the same phases, and the intensity of the detected light is equal to the sum of the intensity of the two beams.

When the movable mirror moves by $\lambda/4$, the optical path difference will be $x = \lambda/2$ and the two beams will have opposite phases when they meet again. The phase difference is exactly half a wavelength, resulting in destructive interference and the light intensity detected by the detector is equal to zero.

When the movable mirror moves an additional $\lambda/4$ in the same direction, the current optical path difference becomes $x = \lambda$. The two beams of light reflected from the fixed mirror and the movable mirror back to the beam splitter have the same phases again, resulting in constructive interference, which is exactly the same as the interference at zero optical path difference. According to the principle of light wave superposition, if there are two monochromatic light beams, the wave number is v, the propagation direction and the polarization direction are the same, the light intensity is I', and the optical path difference between the two beams is x. The interference light intensity is:

$$I = 2I' + 2I' \cos(2\pi v x) \tag{3.6}$$

In actual measurements with an FTIR spectrometer, the situation is much more complicated. It is impossible for the detector to detect a continuous spectrum, which can only be done by discrete sampling at certain intervals. The optical path sampling interval determines the spectral range of the Fourier transform spectrum. According to the Rayleigh criterion, the condition to avoid any 'spectral line confusion' is that the sampling interval is less than or equal to half of the minimum wavelength. Theoretical analysis shows that the measurement range of the optical path difference determines the spectral resolution of the Fourier transform spectrum. By increasing the measured optical path difference, the spectral resolution of the Fourier transform spectrum can be improved, but the increase of x makes more difficult requirements for computer work. In addition, the beam divergence cannot be zero. Therefore, the mathematical model of Fourier transform of spectral interferogram formation will be influenced by the instrument linearity.

For many cases, an attenuated total reflection (ATR) FTIR can be applied through an internally-reflected infrared beam. This can be obtained when an infrared beam is focused on a crystal sample with high refractive index at a specific angle. As a result of internal reflection, a transient wave is created which can come from the surface of the crystal, and then the beam enters into the sample.

FTIR-SE (SE means spectroscopic ellipsometry) spectrum measurement is a highly efficient detection technology. With the development of Fourier transform
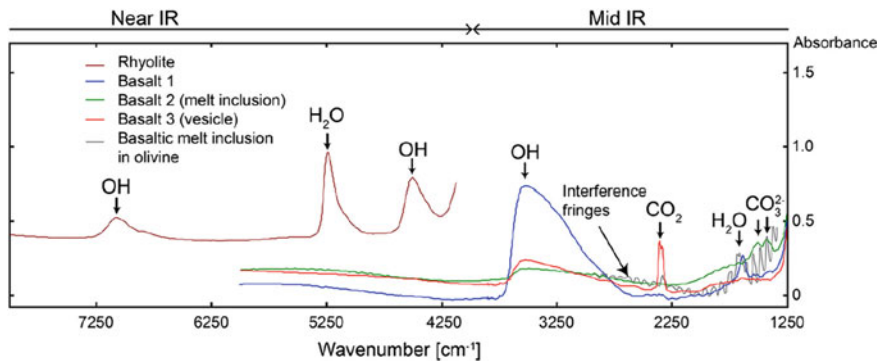
Fig. 3.12 Spectra of volcanic glass in the near- and mid-IR regions [21]

spectroscopy methods and computer technology, many instrument manufacturers have developed various forms of Fourier transform infrared spectroscopy instruments. With the renewed understanding of spectroscopy technology and its applications, NIR (near infrared) spectroscopy analysis technology has also been developed rapidly. By adjusting or changing the light source, beam splitter and detector of the FTIR system, the detection range can be extended from the mid infrared region to the near infrared region. Figure 3.12 shows the spectra of volcanic glass in the near- and mid-IR regions [21]. In recent years, various types of FTIR spectroscopy techniques have been developed and applied in the fields of chemistry [22, 23], environmental monitoring [24], and spectral analysis [25].

In the early days, FTIR spectroscopy was mainly used for molecular structure analysis, quantitative or qualitative analysis of mixed components [26–28]. For quantitative analysis, there are various stoichiometric methods based on Lambert's theorem (theoretical work on the two-body orbit) available. Generally, a certain qualitative analysis is required before quantitative analysis of an FTIR spectrum. FTIR spectroscopy has a good theoretical basis for the qualitative analysis of unknown components, even mixed unknown components. However, the practical use of FTIR spectroscopy for qualitative analysis is subject to various restrictions due to the influence of instrument parameters or environmental parameters [29].

3.5   FTIR's Applications

In semiconductor industry, quality test of the products and batch-to-batch control are constantly being improved. Therefore, nondestructive characterization techniques, e.g. FTIR, spectroscopic ellipsometry (SE), XRD, etc. become more required. Spectroscopic ellipsometry is mostly used in the today's labs. SE operates in the UV/ visible/near IR wavelengths in the range of 250–850 nm, and the technique is used to determine the layer thickness, refractive indices and compositions.

Although, SE provides very important information about material profile, the effects of dopants are difficult to observe due to the wavelengths used. Electrical four-point probe (4PP) measurements (see Chap. 7) can measure non-destructively the sheet resistance and the data can later be converted to doping concentration if the layer thickness is known. There are a lot of efforts to use far-IR measurements of reflectance by adding FTIR spectrometers (or FTIR-SE) [30]. FTIR-SE offers higher sensitivity to thin layers than only reflectometry because both amplitude and phase of the reflected light are determined and not only the intensity.

A typical FTIR tool consists of two sets of arms, one is for SE wavelength region (250–1700 nm) using a Xe lamp light source and the other one for the far-IR region.

IR part consists of a spectrometer with globar (a SiC resistor heater or black-body radiator) as source, an MCT (HgCdTe) IR detector, which is cooled by liquid nitrogen, and grid polarizers. The spot size in FTIR is typically a few mm^2 and angles of incidence are in the range 70–75°. For the ellipsometer part, a rotating-analyzer mode, with four interferograms using analyzer angles of 0, 45, 90, 135°. In these measurements, the two parameters, tan ψ (amplitude ratio) and cos Δ (phase differ-ence) are derived by using Fourier analysis and the spectra are usually collected within a few minutes to obtain better signal/noise ratio.

The tan ψ and cos Δ spectra are analyzed by applying standard SE multilayer tech-niques, where the Fresnel equations are used. The measured data are fitted by using the Drude Law to explain-carrier absorption effects on the real (ε_1) and imaginary (ε_2) parts of the dielectric function as defined [31]:

$$\varepsilon_1 = \varepsilon_\infty - \frac{\omega_p^2}{\omega^2 - \omega_\tau^2} \tag{3.7a}$$

and

$$\varepsilon_2 = \frac{\omega_p^2 \omega_\tau}{\omega(\omega^2 - \omega_\tau^2)} \tag{3.7b}$$

where ε_∞ is permittivity of Si in NIR region, ω_p is plasma frequency and ω_τ is scat-tering frequency. The carrier concentration can be obtained from plasma frequency according to:

$$n \ or \ p = \frac{m^* \varepsilon_0 \omega_p^2}{e^2} \tag{3.8}$$

where ε_0 is permittivity of free space, e is carrier charge and m* is effective mass. The data will be fitted by using the least-square method together with Marquadt's algo-rithm. In this task, the measured spectra are considered and ε_∞, ω_p, ω_τ parameters, as well as the layer thickness are changed to obtain the best fitting curve.

As an example, Fig. 3.13 illustrates FTIR-SE spectra obtained from four Si epilayers deposited on doped substrates. In these measurements interference fringes

rise as a result from the optical contrast between the epilayer/substrate interface. Therefore, the layer thickness can be determined.

Figure 3.14 shows another example dealing with boron doped SiGe on a highly doped substrate and undoped Si buffer. This case shows how the spectra change and the data can be simulated.

The simulation spectra indicate a variation of 20% in dopant concentration of 1×10^{19} cm^3 for 50 nm SiGe layers which is usually a factor of 2 higher than the noise level of the measurement tool. The fitting of the measured data contains an error of up to 30% from the nominal values, but the possibility to estimate doping levels in thin semiconductor alloys with good precision (even with undoped spacer layers) has been reported [31].
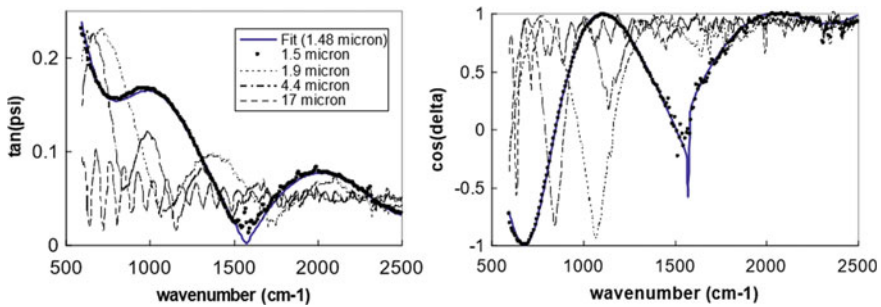
Fig. 3.13 Spectra of four Si epilayers on doped Si substrate where a fit to 1.48 µm thick layer has been performed [31]
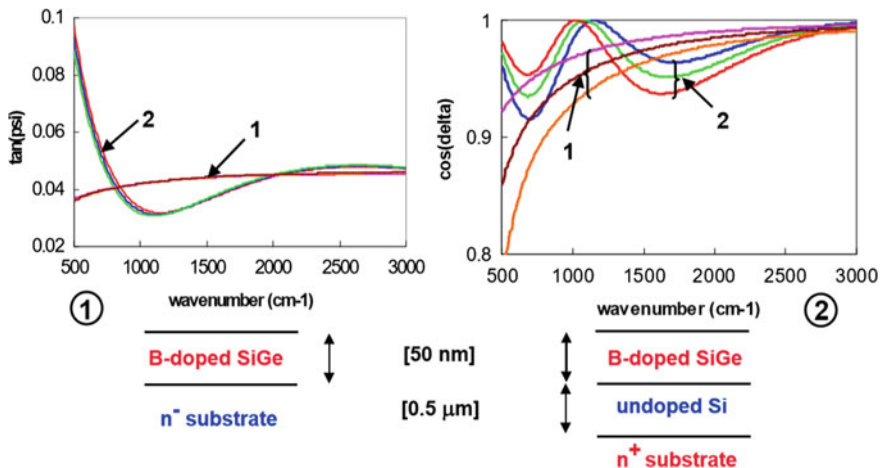
Fig. 3.14 Simulated spectra for B-doped 50 nm SiGe layers. The B concentration varied as 0.8, 1.0, 1.2 × 10^{19} cm^{-3} [31]

Apart from the complexity of the phonon behavior, a major difficulty in Raman spectroscopy is the interference of fluorescence of impurities, or the sample itself. As a final example of the versatility of FTIR we will now discuss how the Raman spectra can be improved by combining with FTIR [5]. While Raman active transitions involve a change in the polarizability of molecules, or lattice phonons, IR active changes in the vibration modes relate to the dipole moment of a module. This feature makes Raman and IR spectroscopy complementary, particularly for symmetric vibrations.

FTIR-Raman instrument consists of high-reflectivity mirrors together with high efficiency detectors to minimize optical loss. A crucial issue for the measurement is the wavelength of the laser has to be in near infrared region. One of the most important point in the FT-Raman measurements is low background intensity to have clear peaks. As an example, Fig. 3.15 demonstrates an improvement in FT-Raman measurement can be obtained by exchanging a Nd:YAG laser operating at 1.064 µm to a laser with the longer wavelength in near-infrared at 1.3 µm [5–7].

It is important to mention here that even though all the phonon modes are not optically active and are not coupled to the electromagnetic radiation field, they can indirectly couple to the light in a Raman measurement (Raman active). Application of Raman spectroscopy for different materials are presented in the following part.

FTIR spectroscopy has the following advantages compared to the infrared spectrum measured by traditional spectrometers:

1. Fast scanning speed: The scanning speed of the Fourier transform spectrometer is faster than that of the dispersive instruments, and information of all frequencies
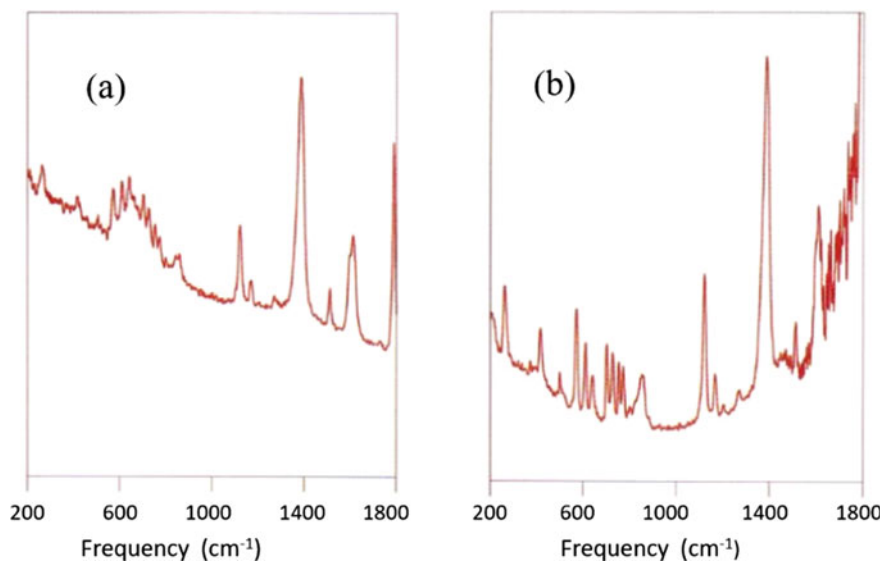
Fig. 3.15 A focus on background intensity: **a** FT-Raman spectrum of a polyimide film excited at 1.064 µm; **b** FT-Raman spectrum of a polyimide film excited at 1.339 µm

of the radiation source can be obtained during a measurement time, so-called "multiplexing".

2. High signal-to-noise ratio: Unlike dispersive, or other types of spectrometers, FTIR spectrometers measure light intensity information formed by interference of all spectra at the same optical path difference each time. The signal-to-noise ratio is proportional to the measurement time. By increasing the measurement time, the signal-to-noise ratio of the spectrum measured by the FTIR spectrometer can be greatly improved.

3. High luminous flux: FTIR spectrometers usually do not have slits, or other light-limiting components to ensure that the interferometer system has a high luminous flux.

4. High measurement accuracy and high resolution.

5. Able to perform wide band measurement.

Inevitably, FTIR analysis technique also has some disadvantages:

1. Fourier spectral analysis technology is an indirect analysis technology, which requires the use of known spectral data to calibrate the spectral information, so it requires more data accumulation. The accuracy of the results of the Fourier spectral analysis is closely related to the quality of the calibration.

2. Because analysis must rely on calibration, the infrared spectroscopy analysis technique is not suitable for detecting dispersive samples that often change in the measured item.

3.6 Development of XPS Technique

X-ray photoelectron spectroscopy (XPS), which is also known as electron spectroscopy for chemical analysis (ESCA), is another surface sensitive material analysis method based on incident photons, but the analyzed species is now emitted electrons. The technique provides important information about the material composition, surface adsorption, surface states, surface valence electron structure, chemical structure of atoms and molecules and chemical bonding of solid samples. The tool is mostly used in condensed matter physics, basic research of electronic structure, thin film analysis, semiconductor research, condensation and surface migration research, molecular adsorption and desorption research, chemical research (chemical state analysis), electronic structure and chemical bond (molecular structure) research.

The XPS instrument, using ultra-high vacuum, was developed in 1970s and a major improvement on the equipment was done by Swedish scientist and Nobel laureate K. Siegbahn. The group worked on the inner layer electron binding energy and the related technical problems about electron energy analysis. Their work focused on measuring the orbital binding energy of each element in the periodic table, and then successfully applied it to many practical chemical related issues.

The early XPS tools had a large area spot-size of 10×10 mm^2 which provided an average analysis across a large area, and the excitation source used was a non-monochromatic x-ray source, which did not provide a very high energy resolution.

More recently, highly sensitive monochromatic XPS with much smaller beam spot has been introduced, referred to as Mono XPS [32], small-area, or small-beam spot XPS, also named selected area XPS (SAXPS) [33] and also imaging XPS (IXPS) [34], have rapidly been developed and received large attention. The monochromatic SAXPS (Mono SAXPS), using for instance synchrotron electron accelerators to generate high intensity and well-defined photons, provides extreme energy resolution, high signal-to-noise ratio, and XPS signals within selected micrometer regions (currently up to ~15 μm). Imaging XPS technique provide informative images of elements and their chemical state distribution within a specified analytical region.

3.7 Basics of XPS Systems

Instrumentation

An XPS energy spectrometer operates under ultra-high vacuum (UHV) and consists of the following parts: x-ray source, energy analysis system, electronic control system, data acquisition and data processing system. A modern electronic spectrometer has three main functions: Mono XPS, SAXPS and iXPS [35].

The UHV system is an important part of modern surface analysis, where the source of excitation e.g. light source of spectrometer, sample chamber, analysis chamber and detector are installed. The demand for such a vacuum system is high pumping speed, generally obtained by multistage combined pump system, providing a pressure lower than 1×10^{-8} mbar.

The x-ray source mainly consists of a filament, anode target and filter window. In XPS tools, Al or Mg double anode x-ray sources, with energies of 1486.6 eV and 1253.6 eV, respectively, are commonly used to excite electrons in each shell of the target element (both inner and outer shells).

The energy analysis system consists of an electronic lens system, an energy analyzer and an electronic detector. The energy analyzer is used to measure the energy distribution of electrons emitted from the surface of the sample within a certain energy range and satisfying the requirements of energy resolution, angular resolution and sensitivity. It is the core component of electronic energy spectrometer, and its resolution, sensitivity and transmission performance are the three main credentials.

Electron spectroscopy analysis involves the acquisition, storage, analysis and processing of a large number of complex data. The data acquisition and processing system is composed of online real-time computer and corresponding software. The on-line computer can directly control the spectrometer and collect and process the experimental data at the same time. The experimental data can, furthermore, be processed by the data analysis system, using mathematical and statistical tools, and combine the results databases to obtain qualitative and quantitative analysis of the tested samples.

Physical Principles

The electrons occupying different energy levels in an atom have different binding energies. When a beam of incident photons with a certain energy interacts with the atoms in the sample, a single photon may deliver its energy to an electron at a certain energy level in the atom. If the energy of the photon is larger than the binding energy (E_b) of the electron, the electron will break away from the original bound energy level of its atom and any remaining energy will be transferred into kinetic energy of the electron. To also break away from the surface and be emitted into the vacuum outside the solid state, an additional energy is needed, which is called the work function of the material. At this point we have one free electron possessing a certain kinetic energy, while the atom has become a positive ion [36].

During photoionization, the energy conversion follows the famous Einstein photoelectric emission equation:

$$h\nu = E_K + E_b + \Phi \tag{3.9}$$

where the $h\nu$ is the incident energy of the photon irradiating the sample, E_K is the kinetic energy of the electron escaping the binding of the atom and surface. Moreover, E_b is the binding energy of the corresponding atomic orbital, where the energy of a certain element has a relatively fixed value. In the actual analysis, the fermi level (E_F) of the system is used as a reference, and the binding energy of the particular electron can be evaluated in relation to E_F. The work function, Φ, of the material, is normally known for most materials. Since the energy of the incident photon, $h\nu$, is known and, by determining the kinetic energy of the free electron with high precision, small changes in the binding energy can be studied and provide information about the electronic states in the near surface regions of the sample, typically a few nm.

3.8 XPS Energy Spectrum

The photoemission process is often conceived as a three-step (three-step model):

(1) absorption and ionization (initial state effect).
(2) atomic response and photoelectron generation (end-state effect).
(3) electron transport to surface and escape into vacuum (external loss).

All these processes contribute to the structure of the XPS spectrum. Each element has a unique set of core levels, and the binding energy typically varies with the level $E_B(1s) > E_B(2s) > E_B(2p) > E_B(3s)$, etc.

For a sample with unknown chemical composition, a full spectrum scan should be made to determine the chemical composition of the surface. When performing XPS analysis, the full spectrum energy scanning range is generally 0–1200 eV, because the strongest peaks of almost all elements are within this range. By a full-spectrum scan of sample, we can identify all elements present on the surface of, for instance, a silicon

Fig. 3.16 A full spectrum energy scanning of XPS survey spectra of the gate stacks of HDO/Si [37]
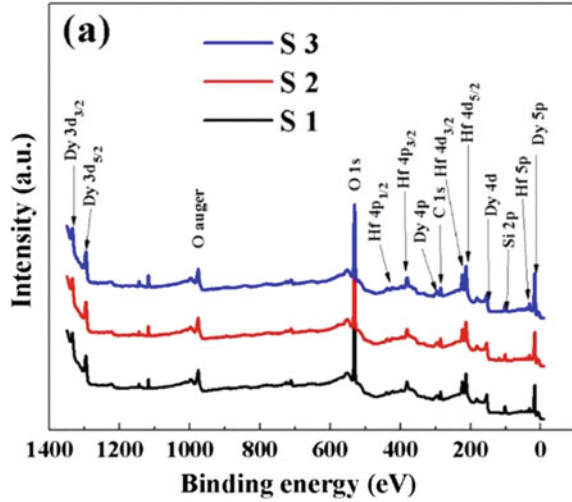
wafer, including natural oxide and surface contamination, the binding energy of each element and the atomic percentage of each element. In terms of general analytic processes, we firstly identify the spectral lines of those elements that always exist on natural surfaces, especially C and O. Secondly, the strong lines of the main elements in the sample and the related lines can be identified. Finally, the remaining weak spectral lines can be analyzed. For the identification of p, d, f lines, it should be noted that they generally have spin-double-line structures, and there is a certain energy interval and intensity ratio between them. As shown in Fig. 3.16 is the evolution of the interfacial chemical states of HDO/Si gate stacks investigated systematically by x-ray photoemission spectroscopy (XPS). Hydrodeoxygenation (HDO) (or hydrotreating) relates to a high-pressure and moderate-temperature process which results in oxygen atoms are rejected by a catalytic reaction with hydrogen.

First, a full spectrum scan is performed. Spectrum S1 represented the non-passivated Si wafer and S2 and S3 represented passivated wafers via the 15 deposition cycles of Al_2O_3(~1.8 nm) and 20 cycles of HfO_2(~1.8 nm) [37].

A series of spectral lines can be seen in the XPS spectrum of the same element, which correspond to different electron orbitals of the elements. The spectral line with the highest intensity is called the main line of the element, which is the main basis for the qualitative characterization of the element. In general, the photoelectrons from the same shell, the larger the total angular quantum number, the greater the intensity of the spectral line will be. Often, in the case of multiple elements, the main line may overlap other spectral lines, in which case the secondary main line can be used.

The peaks of several elements of interest can originate from a narrow region of the spectra, therefore high-resolution fine scanning is necessary in order to obtain more accurate information. In this way, the exact location of the binding energy is obtained to identify the chemical state of the element. If the initial energy state of the atom changes, which can be the case, for instance, when chemically bonding with

other atoms, the electron binding energy in this atom will change. This type of shift in XPS spectra is called a chemical shift. By high-resolution fine scanning of known elements of the material, for example, by sampling the surface of an oxidized silicon wafer, the Si 2p electron layer in XPS spectrum is obtained, which now can have five components, or states, of Si which can be recognized by curve fitting. Silicon of various oxidation states (SiO_x, x = 1–2) appear at high electron binding energies of Si at 103.67 eV. The metal Si showed bimodal peaks at 100.30 eV (Si $2p_{1/2}$) and 99.69 eV (Si $2p_{3/2}$). Moreover, the XPS signal attenuates as the measured elements go deeper, so the thickness and depth of the oxide, or otherwise modified layer at the surface, can also be inferred by the XPS spectrum of the detected metal Si signal.

In order to evaluate the removal of native oxide residuals on a Si substrate, the Si-2p and O-1s were studied by high-resolution XPS spectra. Figure 3.17 demonstrates how the Si substrate (Si–Si) can be composed of the Si 2p3/2(~99.22 eV) and Si 2p1/2 (~99.8 eV), and the Si–O combining peaks are convoluted into four peaks located at binding energies of 100.31, 101.07, 101.65, and 102.87 eV, corresponding to Si^{1+}, Si^+, Si^{3+} and Si^{4+}, respectively. Another peak centered at 102.07 eV can be attributed to the formation of a silicate layer. It can be clearly noted that the intensity of the Si^{1+}, Si^{2+} and Si^{3+} peaks reduces for S2 and S3 samples by using atomic layer deposition (ALD) of Al_2O_3 and HfO_2 IPL (interfacial passivation layers), compared to S1 sample without pretreatment. Furthermore, S3 sample still has a reduced tendency in the intensity of Si–O subpeaks compared to S2 sample. Furthermore, it can be noted that there is no significant difference in the Si^{4+} peak area among the three samples with different surface treatments [37].

Fig. 3.17 Si 2p XPS core-level spectra for S1, S2 (with layer of Al_2O_3), and S3 (with Al_2O_3) samples [37]
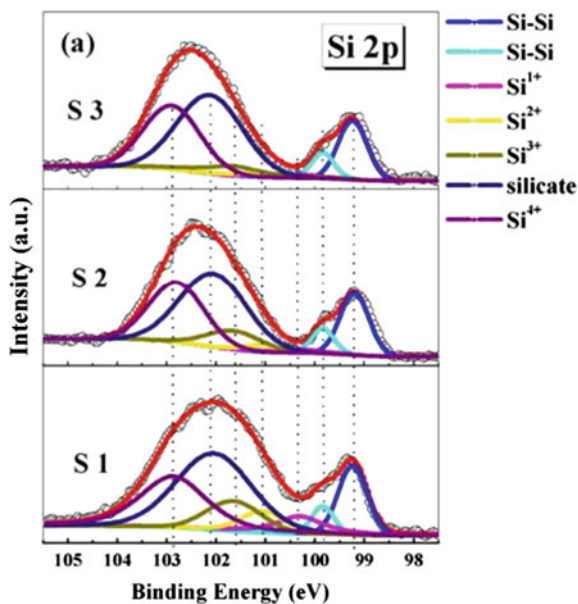
In the process of photoelectric emission, the binding energy of electrons is closely related to the final state of the system, where the effect of different final states on the binding energy of electrons caused by various excitation during ionization is called the final state effect. Complex phenomena, such as relaxation, multiple splitting and multi-electron excitation, are closely related to the electronic structure of the system.

3.9   Characteristics of XPS Systems

X-ray photoelectron spectroscopy is a surface sensitive spectroscopic technique. Although x-ray can penetrate the sample deeply, only a thin layer of photoelectrons emitted from the near surface of the sample can escape into the vacuum. Therefore, XPS can analyze the thickness of the material generally 0–10 nm from the surface. The chemical composition of the material surface can be qualitatively and quantitatively analyzed by XPS, and the information of the chemical states and electronic states of the elements can be obtained. XPS becomes a powerful tool for surface analysis, which is often more convenient compared to analysis of the electron states provided by Auger electron spectroscopy (AES). Furthermore, the XPS technique may cause a little damage to the sample, but it is basically a non-destructive analysis. Since the surface sensitivity of XPS is so high, it is often necessary to prepare a surface that has been exposed to air before measuring. For such samples, ion beam sputtering with high energy particle beams (0.5–5 keV) is used to remove the oxidized surface, although this treatment can also lead to some physical damage.

3.10   Application of XPS Systems

As a typical example of the utilization of XPS, the evolution of the interfacial chemical states of HDO/Si gate stacks have been investigated systematically by x-ray photoemission spectroscopy (XPS). An optimized MOS capacitor based on an Al/HDO/HfO_2/Si gate stack with controlled interface chemistry and excellent electrical properties have been observed [37]. XPS analysis of HfO_2–PMMA (poly methyl methacrylate) hybrid thin film are consistent with the development of an organic–inorganic hybrid network with phases properly linked by the cross-linking agent. It has been demonstrated that there is a strong binding between the inorganic metal oxide and coupling agent molecules, as shown in Fig. 3.18 [38]. The survey XPS spectra of as-deposited and annealed HfO_2 thin films [39]. In this way the possible bindings are detected as marked as shown in the Fig. 3.18.

Another example of XPS applications can be its ability to provide information about the oxidation states of Si nano particles. One issue for the study of nano
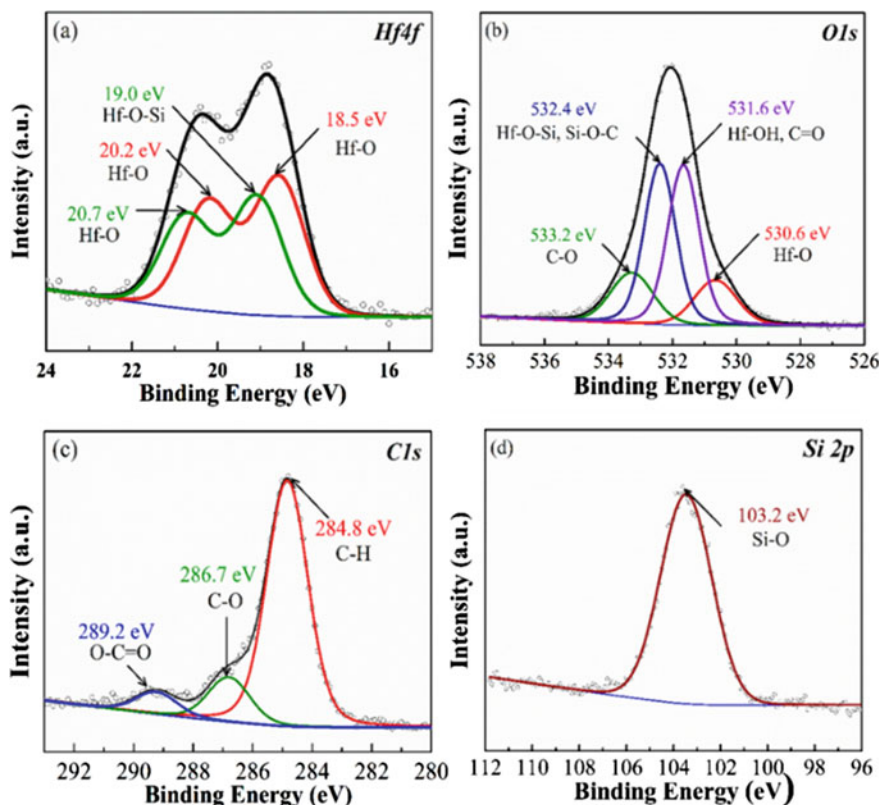
Fig. 3.18 XPS analysis of HfO$_2$–PMMA hybrid thin film in the **a** Hf 4f, **b** O 1s, **c** C 1s, and **d** Si 2p binding energy regions [38]

particles is the interpretation of results when the substrate contribution cannot easily be distinguished. In order to solve the substrate problem, the nanoparticles have to be free. There are reports which show how to form a beam of isolated Si nanocrystals (NCs) and study them by synchrotron radiation XPS (SR-XPS) as shown in the spectrum of Fig. 3.19 [40]. Si-NCs are in powder form where a natural oxide shells exists around the nano crystals. In the target station of the synchrotron beam line, Si-NCs are transported by an Ar carrier gas through aerodynamic lens system (ADLS) before crossing the soft X-ray beam.

When the size of NC is ultra-small (< 4 nm) strain effect appears where the Si 2p core level shifts. The other effect for ultra-small Si-NC is that the electronic property is affected and this causes also an impact on the work function [41]. For the oxidized Si-NCs, a work function for bulk Si is used (4.87 eV) due to the coincidence of the Si $2p_{3/2}$ line position in spectra and the reported one for bulk Si. The relative
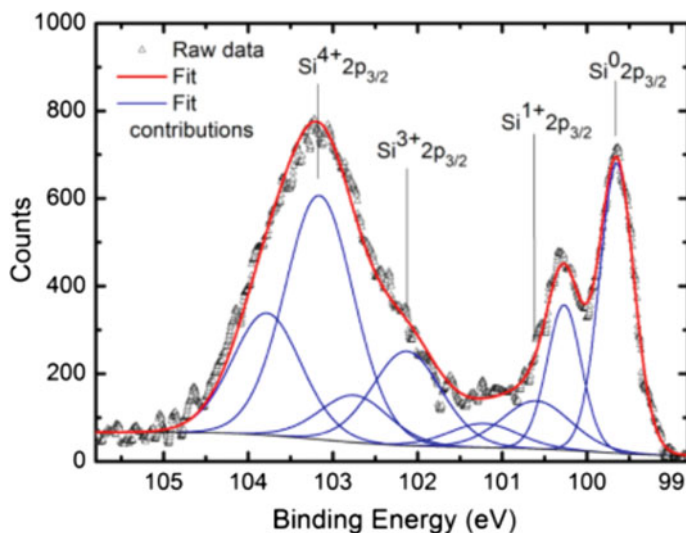
Fig. 3.19 A high-resolution SR-XPS spectrum on isolated Si-NCs. The size of Si-NC is in average 14 nm and the data was collected during 3 h [40]

intensities measured for the oxidation states of Si listed as: Si^0, Si^{1+}, Si^{2+}, Si^{3+}, and Si^{4+}, are respectively 0.67, 0.27, 0, 0.56 and 1 "normalized counts", according to the fitting procedure shown in Fig. 3.19. The binding energies relative to the Si^0 positions for each Si $2p_{3/2}$ contribution are 0.95, 2.48, and 3.51 eV for Si^{1+}, Si^{3+}, and Si^{4+}, respectively.

In conclusion, the XPS technique has gained its position in material science for bulk material as well as nanocrystals (NC) and for understanding the surface chemistry of materials. In special, to form free NC from materials and study them in a synchrotron radiation facility provides a new possibility to study the NCs.

Generally, XPS can provide the following information:

(1) Identification of all elements of atomic concentration $> 0.1\%$, except for H, and He.
(2) Determine surface element composition with an error $< \pm 10\%$, avoiding strong matrix effects.
(3) Sub monolayer sensitivity with a detection depth 1–20 Monolayer (< 10 nm).
(4) Determine chemical shifts and various end-state effects, provide information about the molecular environment of elements, oxidation states, atomic bonding, electronic structures and certain geometric information.
(5) In-depth profile analysis of elements within about 10 nm of the sample.
(6) Analysis under high vacuum is generally non-destructive and the damage of x-ray beam is usually insignificant.

References

1. Raman, C. V., & Krishnan, K. S. (1928). A new type of secondary radiation. *Nature, 121*(3048), 501–502. https://doi.org/10.1038/121501c0
2. Landsberg, G., & Mandelstam, L. (1928). Über die Lichtzerstreuung in Kristallen. *Zeitschrift Für Physik, 50*(11–12), 769–780.
3. Schroder, D. K. (2005). Semiconductor material and device characterization (3rd ed). John Wiley & Sons, Inc. ISBN:9780471739067
4. Smith, E., & Dent, G. (2019). Modern Raman spectroscopy: A practical approach (2nd ed). Wiley & Sons, ISBN 9781119440550.
5. Chase, B. (1987). Fourier transform Raman spectroscopy. *Mikrochimica Acta, 93*(1), 81–91.
6. Schulte, A. (1992). Near-infrared Raman spectroscopy using CCD detection and a semiconductor bandgap filter for Rayleigh line rejection. *Applied Spectroscopy, 46*(6), 891–893.
7. Asselin, K. J., Chase, B. (1994). FT-Raman spectroscopy at 1.339 micrometers. *Applied Spectroscopy, 48*(6), 699–701.
8. Renucci, M. A., Renucci, J. B., & Cardona, M. (1971). Light scattering in solids. In M. Balkanski (Ed.), *Flammarion sciences* (326p).
9. Brya, W. J. (1973). Raman scattering in Ge–Si alloys. *Solid State Communications, 12*, 253.
10. Cerdeira, F., Pinczuk, A., & Bean, J. C. (1985). Observation of confined electronic states in Si strained-Ge_xSi_{1-x} Si layer superlattices. *Physical Review B, 31*, 1202.
11. Pezzoli, F., Bonera, E., Grilli, E., Guzzi, M., Sanguinetti, S., Chrastina, D., Isella, G., Von Känel, H., Wintersberger, E., Stangl, J., & Bauer, G. (2008). Raman spectroscopy determination of composition and strain in $Si_{1-x}Ge_x$/Si heterostructures. *Materials Science in Semiconductor Processing, 11*, 279. https://doi.org/10.1016/j.mssp.2008.09.012
12. Perova, T. S., Wasyluk, J., Lyutovich, K., Kasper, E., Oehme, M., Rode, K., & Waldron, A. (2011). Composition and strain in thin $Si_{1-x}Ge_x$ virtual substrates measured by micro-Raman spectroscopy and x-ray diffraction. *Journal of Applied Physics, 109*, 033502.
13. Yoshikawa, M., Ishida, H., Ishitani, A., et al. (1990). Study of crystallographic orientations in the diamond film on cubic boron nitride using Raman microprobe. *Applied Physics Letters, 57*(5), 428–430.
14. Yoshikawa, M., Nagai, N. (2006). Vibrational spectroscopy of carbon and silicon materials. In Handbook of vibrational spectroscopy. John Wiley & Sons, Inc. https://doi.org/10.1002/047 0027320.s6301
15. Yoshikawa, M. (2006). Infrared spectroscopy and Raman spectroscopy of semiconductor. In Encyclopedia of analytical chemistry. John Wiley & Sons, Inc. https://doi.org/10.1002/978047 0027318.a9652
16. Wall, M. (2012). Raman spectroscopy optimizes graphene characterization. *Advanced Materials and Processes, 170*(4), 35–38.
17. Ferrari, A. C., Meyer, J. C., Scardaci, V., Casiraghi, C., Lazzeri, M., Mauri, F., Piscanec, S., Jiang, D., Novoselov, K. S., Roth, S., & Geim, A. K. (2006). Raman spectrum of graphene and graphene layers. *Physical Review Letters, 97*(18), 187401.
18. Casiraghi, C., Hartschuh, A., Lidorikis, E., Qian, H., Harutyunyan, H., Gokus, T., Novoselov, K. S., Ferrari, A. C., & Harutyuyan, H. (2007). Rayleigh imaging of graphene and graphene layers. *Nano Letters, 7*(9), 2711–2717.
19. Gupta, A., Chen, G., Joshi, P., Tadigadapa, S., & Eklund, P. C. (2006). Raman scattering from high-frequency phonons in supported n-graphene layer films. *Nano Letters, 6*(12), 2667–2673.
20. Schroder, D. K., Rubin, L. G. Semiconductor material and device characterization. Wiley. ISBN: 9780471739067.
21. von Aulock, F. W., Kennedy, B. M., Schipper, C. I., Castro, J. M., Martin, D., Oze, C., Watkins, J. M., Wallace, P. J., Puskar, L., Bégué, F., et al. (2014). Advances in Fourier transform infrared spectroscopy of natural glasses: From sample preparation to data analysis. *Lithos, 206–207*, 52–64. https://doi.org/10.1016/j.lithos.2014.07.017

22. Craig, A. P., Franca, A. S., & Oliveira, L. S. (2012). Evaluation of the potential of FTIR and chemometrics for separation between defective and non-defective coffees. *Food Chemistry, 132*, 1368–1374. https://doi.org/10.1016/j.foodchem.2011.11.121

23. Ulrichs, T., Drotleff, A. M., & Ternes, W. (2015). Determination of heat-induced changes in the protein secondary structure of reconstituted livetins (water-soluble proteins from hen's egg yolk) by FTIR. *Food Chemistry, 172*, 909–920. https://doi.org/10.1016/j.foodchem.2014.09.128

24. Ren, Y., Li, Y., Wang, J., Wang, X., Liu, B., Zhang, L., & Zhang, L. (2005). Reconstruction of air contaminant concentration distribution in a two-dimensional plane by computed tomography and remote sensing FTIR spectroscopy. *Journal of Environmental Science and Health. Part A, Toxic/Hazardous Substances & Environmental Engineering, 40*, 571–580. https://doi.org/10.1081/ese-200046592

25. Basiri, S., Mehdinia, A., & Jabbari, A. (2017). Biologically green synthesized silver nanoparticles as a facile and rapid label-free colorimetric probe for determination of Cu(2+) in water samples. *Spectrochima Acta A Molecular and Biomolecular Spectroscopy, 171*, 297–304. https://doi.org/10.1016/j.saa.2016.08.032

26. Griffith, D. W. T. (1996). Synthetic calibration and quantitative analysis of gas-phase FT-IR spectra. *Applied Spectroscopy, 50*, 59–70. https://doi.org/10.1366/0003702963906627

27. Chen, Y., Furmann, A., Mastalerz, M., & Schimmelmann, A. (2014). Quantitative analysis of shales by KBr-FTIR and micro-FTIR. *Fuel, 116*, 538–549. https://doi.org/10.1016/j.fuel.2013.08.052

28. Delaney, M. F., Warren, F. V., & Hallowell, J. R. (1983). Quantitative-evaluation of library searching performance. *Analytical Chemistry, 55*, 1925–1929. https://doi.org/10.1021/ac00262a022

29. Bacsik, Z., Mink, J., & Keresztury, G. (2004). FTIR spectroscopy of the atmosphere. I. Principles and methods. *Applied Spectroscopy Reviews, 39*, 295–363. https://doi.org/10.1081/asr-200030192

30. Tiwald, T. E., Thompson, D. W., & Woollam, J. A. (1998). Optical determination of shallow carrier profiles using Fourier transform infrared ellipsometry. *Journal of Vacuum Science and Technology B, 16*, 312. https://doi.org/10.1116/1.589802

31. Pickering, C., Leong, W. Y., Glaspe, J., Boher, P., & Piel, J.-P. (2002). Non-destructive characterization of doped Si and SiGe epilayers using FTIR spectroscopic ellipsometry (FTIR-SE). *Materials Science and Engineering B, 89*, 146–150. https://doi.org/10.1016/S0921-5107(01)00821-2

32. Yu, X. R., & Hantsche, H. (1990). Pressure dependence of the charging effect in monochromatized small spot X-ray photoelectron spectroscopy. *Journal of Electron Spectroscopy & Related Phenomena, 50*(1), 19–29.

33. Drummond, I. W., Cooper, T. A., Street, F. J. (1985) Four classes of selected area XPS (SAXPS): An examination of methodology and comparison with other techniques. *Spectrochimica Acta Part B Atomic Spectroscopy, 40*(5), 801–810.

34. Augustin, B., Krishnamurthy, B., Willinger, W. (2009). Internet exchange points (IXPs): Mapped. In *Proceedings of the 9th ACM SIGCOMM conference on internet measurement 2009, (ACM 2009) Chicago, Illinois, USA, November 4–6 (2009).*

35. Hantsche, H. (1989). Comparison of basic principles of the surface-specific analytical methods: AES/SAM, ESCA (XPS), SIMS, and ISS with X-ray microanalysis, and some applications in research and industry. *Scanning, 11*(6), 257–280.

36. Moulder, J. F., Stickle, W. F., Sobol, P. E., & Bomben, K. D. (1992). Handbook of x-ray photoelectron spectroscopy. Perkin-Elmer Corp.

37. Wang, D., He, G., Hao, L., et al. (2019). Comparative passivation effect of ALD-driven HfO_2 and Al_2O_3 buffer layers on the interface chemistry and electrical characteristics of Dy-based gate dielectrics. *Journal of Materials Chemistry C, 7*.

38. Mullapudi, G. S. R., Velazquez-Nevarez, G. A., Avila-Avendano, C., et al. (2019). Low-temperature deposition of inorganic-organic HfO2-PMMA hybrid gate dielectric layers for high mobility ZnO thin-film transistors. *ACS Applied Electronic Materials.*

39. Vinod, A., Rathore, M. S., Rao, N. S. (2018). Effects of annealing on quality and stoichiometry of HfO2 thin films grown by RF magnetron sputtering. *Vacuum, 155*, 339–344.
40. Sublemontier, O., Nicolas, C., Aureau, D., Patanen, M., Kintz, H., Liu, X., Gaveau, M.-A., Le Garrec, J.-L., Robert, E., Barreda, F.-A., Etcheberry, A., Reynaud, C., Mitchell, B., Miron, C. (2014). X-ray photoelectron spectroscopy of isolated nanoparticles. *The Journal of Physical Chemistry Letters, 5*, 3399−3403. https://doi.org/10.1021/jz501532c
41. Kim, S., Kim, M. C., Choi, S. H., Kim, K. J., Hwang, H. N., & Hwang, C. C. (2007). Size dependence of Si 2p core-level shift at Si nanocrystal/SiO$_2$ interfaces. *Applied Physics Letters, 91*, 103113.

Chapter 4
Electron Microscopy

4.1 History of Optical and Electron Microscope

The concept to create a magnified image from an object extends to 1538, when Italian physician Girolamo Francastoro in his publication, Homocentrica, proposed that an enlarged image can be obtained by using two spectacle glasses. Then it took almost sixty years before a compound optical microscope, composed of eyepiece and objective lens, was manufactured by Dutch spectacle makers, Hans and Zacharias Jansen and Hans Lippershey in the Netherlands. After many developments in lens manufacturing and design, the optical microscope was commercialized during the nineteenth century [1].

De Broglie is the first scientist who introduced the theory of wave-like character of electrons and particles with mass, such as electrons can be represented by waves, where the wavelength depends on their energy and that electron wavelengths can be significantly smaller than wavelengths of visible light. This theory indicates that the electrons can be used to create images of two adjacent points which are significantly smaller than what is possible to "see" by visible light due to the diffraction limitation.

The transmission electron microscope (TEM) was invented by Max Knoll and Ernst Ruska in 1931 and the first commercial instrument was manufactured in 1938. The purpose of building the early microscopes was to prove that electron beams could be applied for visible images of materials. The first microscopes showed a resolution of ~10 nm and after almost a decade the resolution could be improved to 2 nm [2].

Indeed, the main goal for development of electron microscopes was to achieve atomic resolution. Therefore, decades of technical developments were focused on making high voltage electron guns, minimizing aberrations (or imperfections) by making better electromagnetic lenses and providing better vacuum systems.

The first scanning TEM (STEM) was demonstrated in 1938 by the German baron Manfred von Ardenne, working at Siemens. The first technically developed STEM equipment was unfortunately destroyed during Second World War [3–5] and no

further development was seen for STEM technique during this time. After the war, the first scanning electron microscope (SEM) was manufactured by Charles Oatley in 1952, and was commercialized in 1965. SEM analysis inaugurated a large possibility to study the topography, and composition variation in the materials.

In these early days of materials science, small particles were mostly studied and later in the 1950s, the analysis of crystal defects, e.g. stacking faults and dislocations in thin films could be performed.

After twenty more years, ultra-high voltage (3 meV) TEM was introduced and a new filament material, lanthanum hexaboride (LAB_6) was used to provide brighter electron source. The development of a new STEM model, which was abandoned in 40s, began again in 70s by Albert Crewe and coworkers at the University of Chicago. They built up a modern STEM using a field emission gun and high-quality electromagnetic lenses. Crewe's team demonstrated for the first time the ability to visualize single heavy atoms on a thin carbon substrate by using annular dark field detector [6].

After more than fifteen years, the resolution of STEM could be improved to 0.2 nm and atomic structure of materials could be imaged for the first time in early 90s. Since that time, high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) has been widely used for material analysis along with chemical analysis instruments e.g. energy dispersive x-ray (EDX) and electron energy loss spectroscopy (EELS) for spectroscopic mapping [7].

The impact of the electron microscopy techniques on our understanding of materials, including biological materials, is difficult to fully appreciate. Since these tools are constantly being improved, it is also likely that we will continue to be amazed by high resolution images of previously "unseen" matter for many years to come.

4.2   Photons and Electrons for Microscopy

There are several types of microscopes using photons, or electrons to create an object image. These microscopes consist of a photon or electron source, glass or electromagnetic lenses and apertures to form the image of the object. Figure 4.1 illustrates the general setup typical for both light and electron microscopes. Note that the position of the sample differs for the SEM.
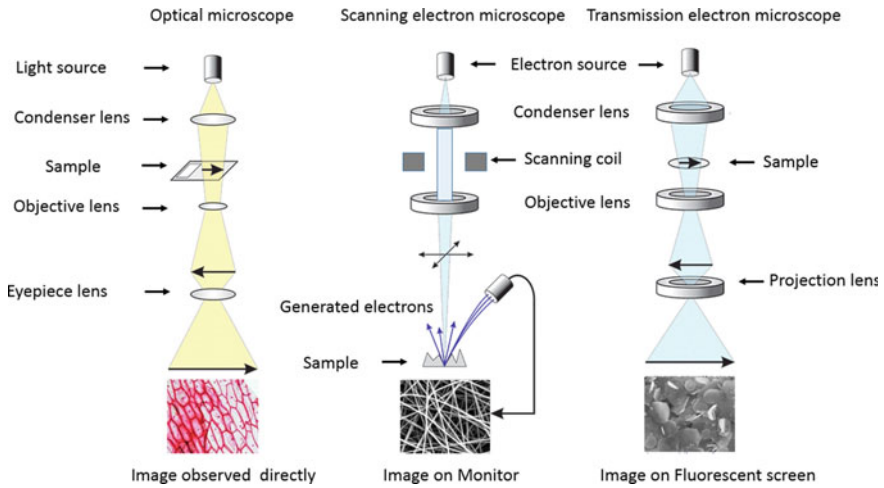
Fig. 4.1 Set-up of photon and electron microscopes

4.3 Transition from Photons to Electrons for Imaging

4.3.1 Basic Definitions

Resolution: Photons, as well as electrons, have a dual character of wave- and particles and may, for instance, diffract when they pass through a circular aperture resulting concentric rings, so called Airey rings, as shown in Fig. 4.2. The resolution is usually discussed in terms of two-point resolution, the smallest distance between two closely spaced point objects where the points still can be clearly resolved. The way to obtain the best resolution is defined by the Rayleigh criterion. This criterion states that the two points are resolved if the center of Airey points (maximum intensity) from one point coincides with the zero intensity (places between the Airey rings) from the second point as shown in Fig. 4.2. The maximum resolution can be written as a resolution term r_1 from Abbe's equation as:

$$r_1 = \frac{d_1}{2} = \frac{0.612\lambda}{n \sin \alpha} \tag{4.1}$$

where n is the refractive index, λ is the wavelength and 'α' is the aperture's angle, that is the half angle of the cone of angles that can be accepted by the lenses. In Abbe's equation, the term '$n\sin\alpha$' is called numerical aperture (NA). In order to increase the resolution, the r_1 term has to be minimized.

Depth of field (DOF): The depth of field is defined as the distance when the top and bottom points from a 3D object still appear sharp with all details. DOF should not be confused with depth of focus which indicates for a fixed object position to which

Fig. 4.2 Diffraction of light
and appearance of Airey
rings: **a** single point, and
b double points, where
Rayleigh criterion is also
marked. The resolution term
is r_1 in the figure and has to
be minimized in order to
obtain better resolution [8]
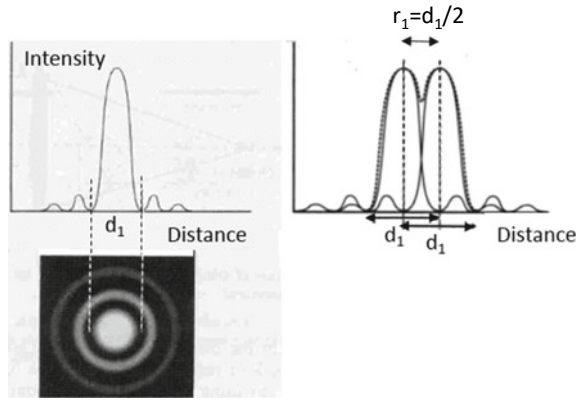
extension the image can be viewed without being out of focus. Figure 4.3 illustrates a simple picture of how the DOF (h), the aperture angle (α), and resolution term (d_1) are connected in a microscope. An equation for depth of field for the microscope is defined by taking into account tanα in the figure and inserting the resolution term, in which case h is written as following:

$$ h = \frac{0.61\lambda}{n \sin \alpha \tan \alpha} \tag{4.2} $$

For a microscope, both high resolution and high DOF are desired. Unfortunately, both these terms cannot be simultaneously increased in an optical microscope. For example, by decreasing aperture angle 'α' in Eqs. 4.1 and 4.2, 'h' is increased but the resolution is decreased.

This situation is different for an electron microscope where the high energy electrons have remarkably smaller wavelengths. The beam of electrons is also focused by strong electromagnetic lenses and this substantially reduces the angle α, which implies that sin$\alpha \approx$ tan$\alpha \approx \alpha$. Furthermore, electrons have to be in high vacuum and the refractive index, n is one. Therefore, the Eqs. 4.1 and 4.2 for electrons are modified as:

$$ r_1 = \frac{0.61\lambda}{\alpha} \tag{4.3} $$

and

Fig. 4.3 An illustration of
DOF (h) and related
resolution term d_1 (where r_1
= d_1/2). In the figure, α is
the half aperture angle
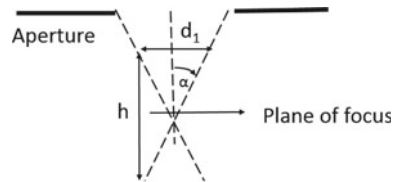
$$h = \frac{0.61\lambda}{\alpha^2} \qquad (4.4)$$

The resolution is dependent on wavelength and electrons accelerated to high energy have very small wavelengths. Hence significantly higher resolution can be achieved in electron microscopes compared to optical microscopes. Theoretically, a resolution in the range of 0.02 nm could be obtained for a 100 keV electron, having $\lambda = 0.0037$ nm, and an $\alpha = 0.1$ rad (about 5.7°). However, this resolution cannot be achieved in practice due to spherical aberrations, described below. A realistic value for the resolution for 100 keV electrons is in the order of 1 nm. For the depth of field, on the other hand, the α^2-term in the denominator significantly increases the DOF for a decreasing aperture angle.

Aberrations: The basic limitations of optical and electron microscopes come from aberrations of the optical elements, such as lenses or mirrors. Aberrations can be of various types, but roots from non-ideal manufacturing of lenses, which results in a small spread of the beam size and blurred edges. There are three main aberrations for optical, or electron microscopes: chromatic, diffraction and spherical aberrations.

Chromatic aberration is caused when the electrons, or photons with different energies are focused to different planes. The minimum probe diameter, d_c, is proportional to the energy variation (E) of the beam, chromatic aberration coefficient (C_c) and the divergence (or convergence) angle of the lens, 'α' according to the following equation:

$$d_c = C_c \alpha \frac{\Delta E}{E} \qquad (4.5)$$

where C_c is approximately equal to the working distance (WD) of the SEM. This distance is the distance between an object and the nearest lens at which the object is still in focus. In SEM, for low acceleration voltages, i.e., electron energies < 1 keV the chromatic aberration has a dominant effect because it varies inversely with the beam energy. The chromatic aberration appears on the beam profile by forming a "skirt" around the beam. This skirt increases the probe diameter and lowers the resolution and the signal contrast in the image.

Aberration due to diffraction occurs as result of the wave-like behavior of electrons (or photons). As the electrons are converged by a lens, they form a diffraction pattern, or Airey rings, at the focal point with a width λ/α. The width of the electron beam, d_{diff}, between the first Airey disc, zero order diffraction ring, is estimated to:

$$d_{diff} = 0.61 \frac{\lambda}{\alpha} \qquad (4.6)$$

Spherical aberration occurs when the peripheral rays of the electron beam are not focused in the same plane as the central rays and it is derived from:

$$d_s = \frac{1}{2} C_s \alpha^3 \qquad (4.7)$$

Fig. 4.4 Different
aberrations for electron
microscope [9]
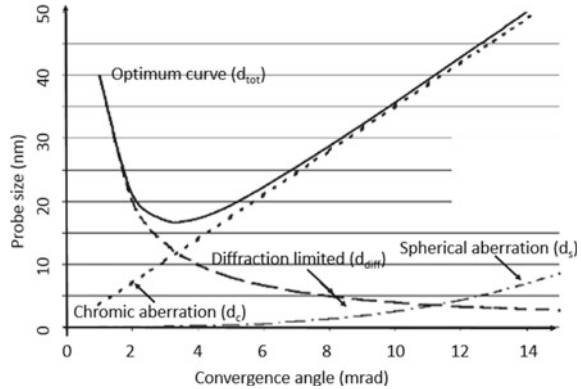
C_s is the spherical aberration coefficient. The final probe size in an electron micro-
scope is determined by adding up the contribution of spherical aberration (d_s), chro-
matic aberrations (d_c), diffraction contribution (d_{diff}) and the original diameter of
the beam emitted from the filament (d_{source}). A simple way to estimate the total
contributions (d_{tot}) is given by:

$$d_{tot} = \sqrt{d_{source}^2 + d_s^2 + d_{diff}^2 + d_c^2} \tag{4.8}$$

Figure 4.4 illustrates the impact of the individual above aberrations separately, as
well as the total contribution, d_{tot}, as a function of the convergence (or divergence)
angle. As the convergence angles are less than about 3 mrad (about 0.17°), d_{tot} is
mostly dominated by the diffraction effect, which decreases as $1/\alpha$. But for α above
3 mrad chromatic aberration is the main limiting factor for the probe size.

The minimum point in the d_{tot} curve is known as the 'diffraction limited' condition
for electron microscopes and it gives the smallest probe size that can be obtained, but
at the same time it is also where the highest current can be reached. In practice, the
best performance is achieved by applying a short working distance and optimized
convergence angle, which is a few mrad. Under such conditions, the beam diameter
is typically governed by diffraction and chromatic aberration to a size of 10–20 nm.
The small values of the numerical aperture (NA) of the objective lens gives a large
DOF, which significantly reduces the beam current. The low beam current results in
poor signal-to-noise ratio for imaging.

Since the aberrations are limiting the resolution in an electron microscope,
different techniques have been proposed to solve the problems. The 'aberration
corrected' SEM (ACSEM) is equipped with devices such as corrected electrooptical
systems (**CEOS**) [10, 11], which consists of a combination of multi-pole magnetic
lenses located around the beam axis. The devices are quadro-, hexa- and octo-pole
element lenses to correct the aberrations. The beam correctors (or stigmators) are
placed on top of the objective lens and the electrons are then forced to focus on the
specimen. Then, as a result, C_c and C_s aberration coefficients are fully controlled

Fig. 4.5 A typical stigmator
for an electron microscope
used for correcting the
aberrations
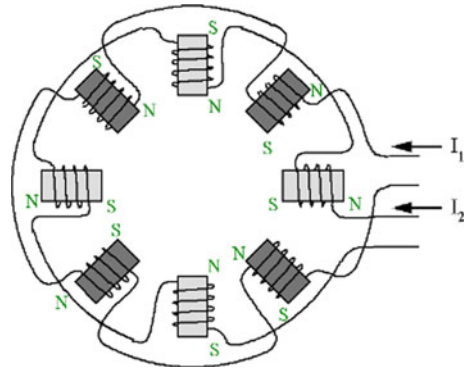
and can be set to zero, positive, or negative by the strength of the various elements
in the multi-pole device (Fig. 4.5).

4.4   Emission Materials for Electron Microscopes

Electron emission from a cathode filament occurs via thermionic, or electric field
emission, or a combination of these, sometimes referred to as Schottky emission. The
emission processes are illustrated in Fig. 4.6. In thermionic emission, the electrons at
the Fermi level, E_F, in the cathode can gain sufficient energy by heating to overcome
the work function barrier, ϕ_w. The emitted current density, J_c, for thermionic emission
is given by the Richardson equation:

$$J_c = AT_c^2 \exp(-\phi_w/k_B T_c) \tag{4.9}$$

where A is a cathode material constant, T_C is the critical temperature of the filament
at which the emission occurs and k_B is Boltzmann's constant.

Fig. 4.6 Different forms of
electron emissions from a
metal filament
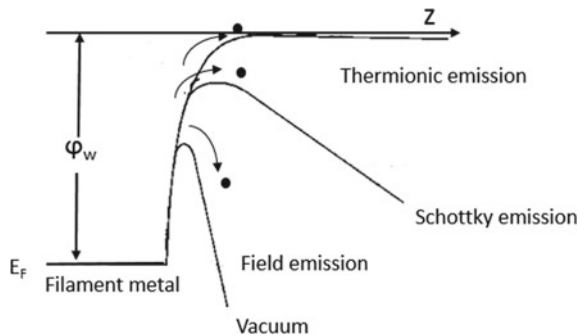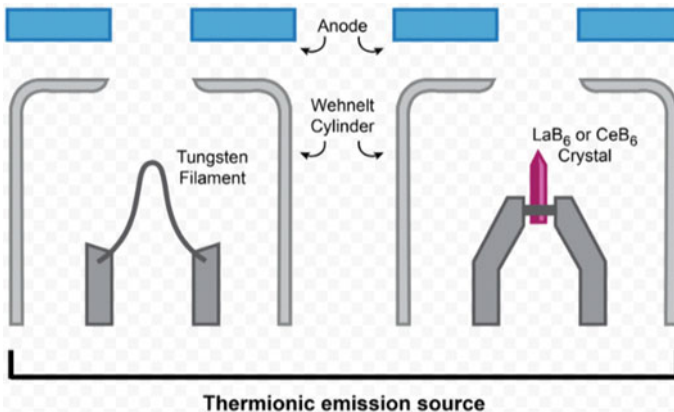
Fig. 4.7 Schematic view of electron-guns for an electron microscope. A V-shaped filament is used for thermionic emission of the tungsten filament and a sharp tip is used as filament for hexaboride filament

When an electric field is applied, the barrier can be substantially lowered and electrons can be emitted and accelerated towards the anode in greater number. The electron emission is then partly controlled by quantum mechanical processes, such as tunneling, and has been described by, for instance, by R. H. Fowler and L. W. Nordheim. Also Walter Schottky made important contributions to the understanding of the field emission process.

Considerable research efforts have been invested in finding suitable materials for electron emission filaments, where the main criteria are high intensity and long lifetime. Materials such as hexaborides (LaB_6 and CeB_6) and tungsten (W), with work functions of 2.5, 2.7 and 4.5 eV, respectively, have demonstrated excellent thermionic emission. The filament is heated resistively by a current and a positive voltage, in respect to the cathode filament ensures that the electrons are extracted from the electron source (see Fig. 4.7).

Tungsten filaments have V-shaped and are commonly heated up to $T_c = 2500$–3000 K, whereas for hexaborides, a temperature of $T_c = 1400$–2000 K is needed to obtain high current density. The key point for performance of these filaments is their emission current which, according to Eq. 4.9, is estimated to only 3 A/cm^2 for W, while it is 20–50 A/cm^2 for LaB_6 and CeB_6, although the latter materials have a drawback since they require high vacuum, at least ten times lower pressure than W [12].

The electron source is often referred to as electron gun, and in Fig. 4.7 two configurations of the electron gun are illustrated. A so-called Wehnelt electrode surrounds the cathode filament. The Wehnelt electrode, having a negative potential in respect to the cathode, acts as a gate which densifies the electron plasma and also determines the size of the emitting area. The positive anode creates a strong electric field, which extracts the electrons and shape them into a beam. The beam is

then further accelerated to the desired energy. In practice, the filament and Wehnelt electrode are connected to negative potential, while the anode is grounded.

4.5 Scanning Electron Microscope (SEM)

Scanning electron microscopes are today widely used in most fields of scientific research, as well as for industrial applications. These microscopes provide information about surface topography, morphology and compositional variation with lateral resolution down to a few nm. They have been commercially available for over 50 years and there exists today a large number of vendors.

4.5.1 Interaction of the Electron Beam and Materials

An electron beam with a specific kinetic energy and direction penetrates into a specimen and interacts with the target particles in a number of different ways. As these primary electrons in the well-focused beam penetrates deeper, the electrons will lose their energy in the various interactions and also change their direction. The electrons will spread out in an onion-shaped volume, typically reaching a depth in the specimen of 1–5 µm depending on the initial energy and the properties of the target material. A series of secondary electrons and photons are generated in this excitation volume, as illustrated in Fig. 4.8.

The produced electrons are mainly categorized into (1) backscattered, (2) secondary and (3) Auger electrons. New electrons, or photons can also be created when for example if a generated x-ray is absorbed and results in another x-ray (fluorescence effect), or electron (photoelectric effect). X-rays can be generated by the
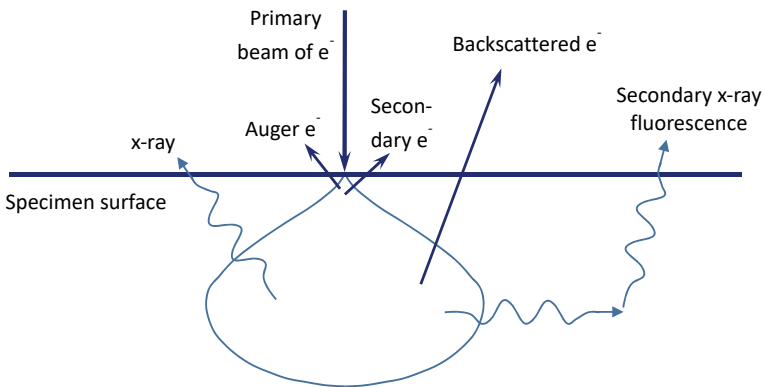
Fig. 4.8 The excitation volume and different electron/photon interactions

bremsstrahlung effect, as energetic electrons change their direction. If the electrons have more than about 30 keV they can also generate x-rays from the K-shell for light elements and L- or M-shell from the heavier elements.

The depth of electron penetration as a function of their energy (E_o) is approximately given by [13]:

$$x(\mu m) = \frac{\sqrt{E_0^3}}{10\rho} \tag{4.10}$$

4.5.2   Electron Signals

As mentioned above, when the electron beam impinges the specimen several types of electrons are generated and give different information. Two main types of electron signals are used to create an image in a SEM, backscattered (BE) electrons and secondary electrons (SE). In the instrument, the electron beam is raster scanned in X–Y directions by magnetic coils and the generated electrons are collected by the detector in coordination with the beam's X–Y position on the sample. The signals from the detector are interpreted into an image on the CRT monitor (see Fig. 4.1). In SEM, magnification, 'M' of an area is defined as M = L/l where l is a raster over the sample and L is a raster in the CRT monitor.

The BEs are a result of elastic collision with atoms. When the electron beam meets the specimen, the electrons enter the excitation volume and interact with atomic nuclei by the Coulomb force in a similar way as a comet will interact by the gravitational force of a planet. The electrons have negative charge and are attracted by the positive field around the nuclei. For sufficient high energies, the electrons will not be captured by the nuclei, but circle around the nucleus in a parabolic trajectory and turn back without losing energy. These electrons are known as backscattered and, from some distance, they seem to move in straight paths in and out of the sample (Fig. 4.9).

One way to detect these electrons and make "backscattered electron image" (BEI) is to place a detector on their return path above the sample. The atomic number of the target material will affect the number of BEs, as show in Fig. 4.10.

Fig. 4.9 Formation of backscattered electrons
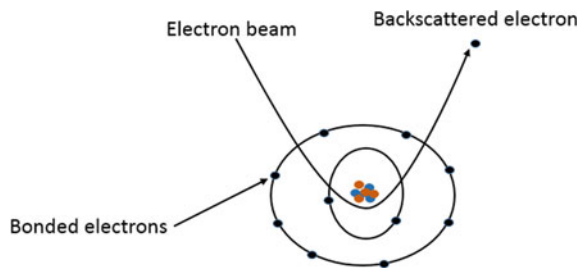
Fig. 4.10 An illustration of backscattered and secondary electrons yield to atomic numbers [14]
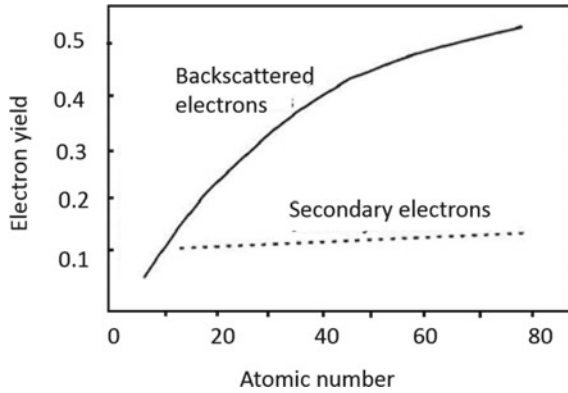
The number of backscattered electrons produced from a material may be quantified by its backscattering coefficient, η. This coefficient depends strongly on a sample's average atomic number, Z [8].

$$\eta = -0.254 + 0.016Z - 1.86 \times 10^{-4} Z^2 + 8.3 \times 10^{-7} Z^3 \tag{4.11}$$

An average Z value for the material is calculated using the weight fractions (w) of each element:

$$Z = w_1 z_1 + w_2 z_2 \cdots + w_n z_n \tag{4.12}$$

Thus, the Z-value for SiO_2, with 0.4674 Si and 0.5326 O by weight is:

$$Z_{SiO_2} = 0.4674 \times 14 + 0.5326 \times 8 = 10.8044$$

A composition variation in the sample will then cause a contrast, so-called Z-contrast, where the heavier atoms appear brighter in BEI compared to the lighter ones as shown in Fig. 4.11a.

For crystalline samples, there is another contrast in BEIs, which is known as channeling contrast. This contrast is much weaker than Z-contrast and it roots from the orientation of the crystal in respect to the incident beam direction. Therefore, channeling contrast mostly appears in polycrystalline materials where the grains have different crystal orientations, leading to the backscattered coefficient variations as shown in Fig. 4.11b.

The secondary electrons are generated through inelastic collisions when the energetic electrons in the beam kicks out electrons from the outer shells of target atoms. These electrons are emitted from the near surface region and will carry important information about the topography of the sample. To create an image from secondary electrons, the detector has to be placed close to the sample which also makes it
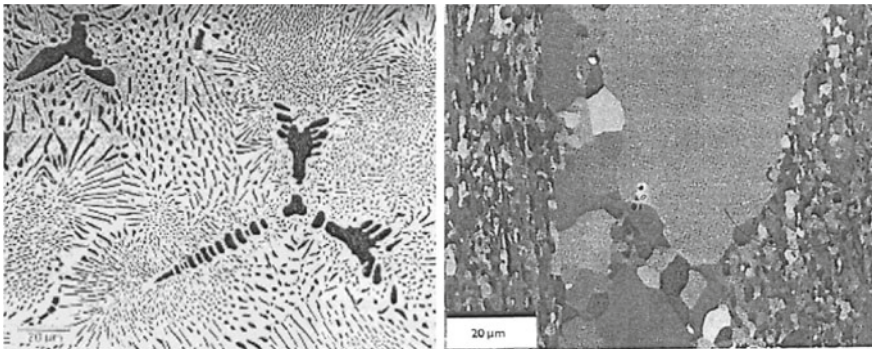
Fig. 4.11 **a** Z-contrast in SEM BEI and **b** electron channeling contrast [15]

possible for the BEs to enter the detector as background to the secondary electron signal. Therefore, the secondary electron image (SEI) is a mixture of both backscattered and secondary electrons.

One innovative design to create BEI and SEI is using an Everthart-Thornley detector as shown in Fig. 4.12. Such a detector consists of a Faraday cage in the front part and an active area inside the cage which is made of scintillator material (e.g. ZnS and CdS). The detector is located beside the sample holder and is angled to effectively meet the secondary electrons. The scintillator material converts the electrons into light prior to arriving at the photomultiplier for amplification. By biasing the cage negatively, the low energy secondary electrons are avoided and the electron signal is purely BEI, while a zero or positive bias would attract both secondary and backscattered electrons providing a SEI.
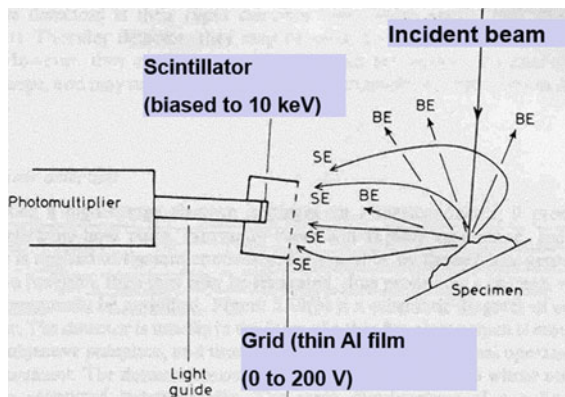
Fig. 4.12 A drawing of a typical detector configuration for SEI & BEI formation in a SEM [14]

4.5.3   Probe Size in SEM

In order to have high resolution SEM image, the diameter of the probe beam has to be as small as possible and the scanning motion of the beam has to be slow enough to generate sufficient statistics from each spot. The distance between consecutive spots also has to be in the same range as the probe diameter. These two requirements imply high precision electromagnetic lenses for focusing the beam at the sample surface, as well as an accurate scanning coil system to move the electron beam in small steps.

Figure 4.13 illustrates a simple picture of how an image is formed in a SEM instrument. The parameters noted in the figure are: the distance between the condenser and the electron gun, u_1, the focal length the condenser, v_1, distance between condenser focal plane and objective lens, u_2, and the focal length of the objective, v_2, which is similar to the working distance (WD) previously mentioned. The diameter of the filament, d_0, is also an important parameter for the lateral resolution of the instrument. The initial beam diameter d_0 has to be as small as possible, therefore, a V-shape filament is commonly used. The parameters v_1 and v_2 can be varied by the electromagnetic lenses, but the WD can also be adjusted "manually". The final probe diameter on the sample is noted d and is derived from the following equation in terms of the first image, d_1 after the condenser lens:

$$d_1 = d_0 \times v_1/u_1 \text{ and } d = d_1 \times v_2/u_2 = (d_0 \times v_1/u_1) \times WD/u_2 \qquad (4.13)$$

In order to minimize d, v_1 and v_2 (WD) has to be minimized. This means that moving the sample closer to the objective lens and applying a higher voltage to the condenser lens are the key points for obtaining a smaller probe diameter.

Fig. 4.13 A drawing of probe beam path in SEM [14]
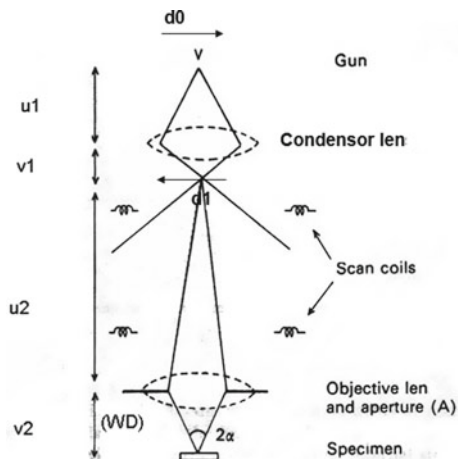
4.5.4   Sample Preparation for SEM

Since the electron beam carries negative charge, the sample will also become nega-
tively charged it is necessary to protect the sample and detectors in the SEM tool for
too high accumulated charge. First, the samples have to be conductive and electri-
cally grounded to avoid the negative charging, which will disturb the SEM imaging.
Therefore, non-conductive samples need to be coated with a conductive layer. Typi-
cally, a thin metal layer, e.g. gold, platinum, or chromium is often sputtered on the
sample to improve the conductivity and make grounding the sample possible. Sput-
tering heavy metals will, furthermore, improve the SEM imaging for light materials
since the emission of backscattered electrons is increased due to the Z-contrast.

   Second, it is also important that the sample is dry and does not contain water or
oil, since it will be under high vacuum in the SEM tool and may contaminate the
vacuum chamber.

   In some cases, conductive metals cannot be used to coat the sample and, for such
samples, low currents and accelerating voltages (e.g. 0.5 to 4 kV) are applied in
order to avoid damaging the sample. It can also be necessary to lower the vacuum
conditions (increase the pressure) in the SEM chamber compared to standard SEM
operation [15]. Since low voltages are applied for SEM, then it is important to use
field emission guns for these measurements in order to obtain high brightness and
small spots size.

4.5.5   SEM Analysis of Nanostructures

Scanning electron microscopy is widely used to distinguish layers in multilayered
structures using imaging of Z-contrast. As an example, SiGe/Si multilayer structure
are used for processing vertical gate all around (VGAA) transistors. In this case, the
SiGe layer is selectively etched while the Si remains unaffected by the etch. This
produces a channel layer with thickness of just a few nanometers [16]. Since the
channel thickness has to be accurately measured then a series of SiGe etching is
performed to determine the etch rate. SEM is used as a characterization tool which
can be applied to determine the etch depth for different etch time. Figure 4.14a, b
show the etch profile of SiGe in SiGe/Si multilayers using different etch methods.

4.6   Principles of TEM

The principle of TEM imaging is to use electrons as the illumination source, having
electromagnetic lenses, compared to transmission optical microscope where photons
and typically glass lenses are used. The resolution of TEM microscope is dependent
on the wavelength of the probing particles. Electrons with very short wavelength
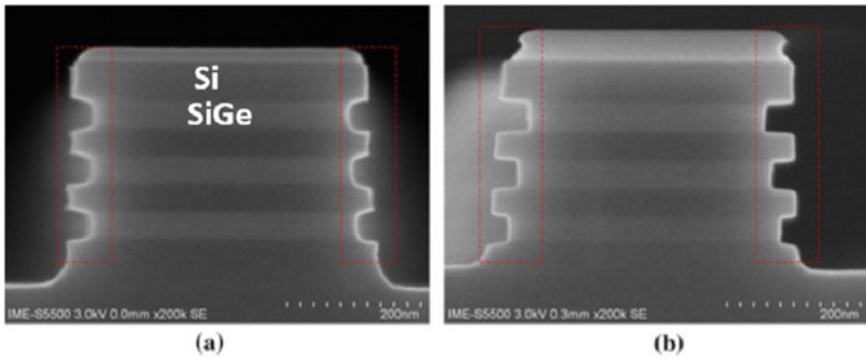
Fig. 4.14 SEM cross-section images of showing the etch-profile of SiGe using: **a** wet etching, 6% HF/30% H_2O_2/99.8% CH_3COOH = 1:2:4 etching 8 min, and **b** ICP dry etching, CF_4/O_2/He = 4:1:5 [17]

offer a high theoretical resolution of 0.1 nm, which is far better than optical microscope resolution limit of 200 nm. Meanwhile, the resolution of TEM like other electron microscopes is limited by spherical aberration and this limits the theoretical resolution to nanometer levels. In contrast to optical transmission microscope, the image contrast in TEM does not mainly depend on absorption, but on scattering and diffraction of electrons in the sample.

In comparison with SEM with accelerating voltage up to 50 kV, TEM employs typically 100–300 kV which allows the electrons to penetrate easily many hundred nanometers inside a sample. The electrons experience collisions with atoms and will be scattered, transmitted or absorbed. If nonrelativistic energies are assumed (?), the wavelength of electrons is determined by the accelerating voltage, V, according to equation:

$$\lambda = \frac{h}{\sqrt{2m_0 E}} \approx \sqrt{\frac{1.5}{V}} \text{ nm} \tag{4.14}$$

Here, h is Planck's constant, m_0 is the electron rest mass, E is the electron energy in J. The equation simplifies to the right-hand expression with the wavelength in nm and gives a rough estimation of the wavelength, at least up to energies of 25 keV? for which the electron velocity is 30% of the speed of light. The setup of a TEM is more complicated than a SEM, as shown in Fig. 4.15. The tool consists of two main parts: the first part contains the e-gun, condenser and objective lenses. This part takes care of focusing the incident electron beam on the sample and to form the primary image. The second part magnifies the primary image using a series of lenses and forms the final image.

The objective lens forms a diffraction pattern at its back focal plane. The diffraction pattern of the primary image is in fact a Fourier transform of the scattered electron wave. This two-step process is the foundation of image formation in high-resolution
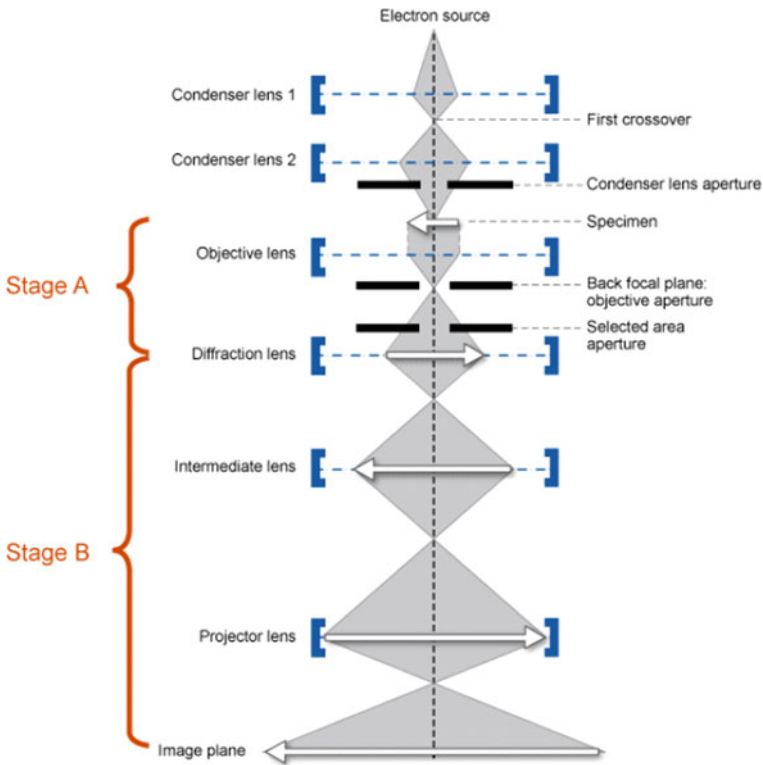
Fig. 4.15 Setup illustration of lenses in a typical TEM instrument [17]

mode. For TEM, high-resolution image is a result of interference of beams which are formed at the back focal plane of the objective lens.

(1)   Illumination system including electron gun and condensers.

Electron gun is the illumination source, consisting of a cathode, a gate, and an anode. Electron-emitting cathode filaments are V- or sharply point-shaped. Usually the electron microscope acceleration voltage is between 60 and 350 kV. The electrons emitted from the cathode tube form a beam of rays through a small hole in the Wehnelt electrode gate, and are extracted by the anode voltage, accelerated and directed to the condenser.

The condenser lens has the function to converge the electron beam onto the sample plane as well as to adjust the beam density and spot size on the sample plane.

(2)   Imaging part, including three lenses, objective lens, intermediate lens and projector lens.

The objective lens determines the resolution and contrast, while the other two lenses relate to magnification of the formed image by the objective lens. Objective lenses form a magnified image (usually 100–200 times) as well as being responsible for

a diffraction pattern. The level of the objective lens aberration has to be as low as possible. For example, the spherical aberration coefficient is required to be in range of 0.7 nm.

The intermediate lens is a relatively weak elecromagnetic lens with low magnification ability (0–20 times). High-resolution electron microscopes may have two intermediate and projector lenses to ensure high-quality images. The intermediate lens projects the formed image, or diffraction spectrum, by the objective lens for the next magnification by the projector lens.

The projector lens works on the final magnified image and diffraction spectrum formed by the intermediate lens on the fluorescent screen. The demands for the projector lens are not as severe as for objective lens. The aberration is regulated by an astigmator.

In all electron microscopes (specially in TEM), high vacuum is demanded to ensure a good transport of the highly accelerated electrons for high resolution imaging. In order to obtain high vacuum, a two-stage vacuum pump which consists of a rotating mechanical pump for the low vacuum and an oil diffusion or turbomolecular pump is used for the high vacuum.

The electron microscope needs two types of power supplies; a 'small current high voltage' power supply for accelerating the electrons, and a 'high current low voltage' one for electromagnetic lens. Therefore, there are large demands for both voltage and current stability to remove aberrations and changes in the electron wavelength.

4.6.1 Diffraction Pattern

Similar to x-ray diffraction, electron diffraction also follows Bragg's law. The incident angle, θ, for electrons in TEM is small and this results in interference pattern from many crystal planes. This can also be explained by the Ewald sphere theory when the radius of the sphere is large, resulting in many intersect with lattice points in the reciprocal space. The appearance of the diffraction pattern reveals the phase of the sample. The pattern of electron diffraction for single crystalline materials is well-ordered arrays of sharp dots, while for polycrystalline material, concentric rings appear. For amorphous materials, only diffused rings are seen, as shown in Fig. 4.16.

To further interpret the experimental data, simulation of electron diffraction pattern is today an important help. The simulated electron diffraction pattern can be applied for both single crystal and polycrystalline materials. The most common used simulation programs are ProJECT/PCED3 and ProJECT/SAED4 [18], used for polycrystalline and single crystalline materials, respectively. These programs are, however, not enough for advanced analysis e.g. coexisting multiple phases with fixed orientation and for comparing of two similar diffraction patterns from different phases. As the instruments are improving, there is a continuous demand for improved software.

For indexing spots in a diffraction pattern, two parameters are required, as indicated in Fig. 4.17:

Fig. 4.16 Diffraction
pattern from different phases
of materials
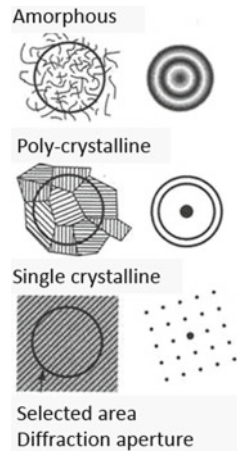
Fig. 4.17 The diffraction
pattern from a single cubic
crystal along zone axis $z =$
011 of Al [19]
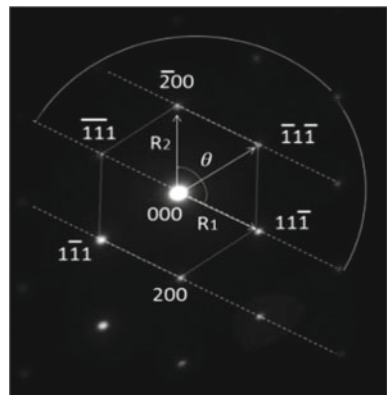
1. The distance (R) between diffraction spots to the center spot (due to the transmitted beam).
2. The angle (θ) between two vectors drawn from the center to two adjacent points.

The zone axis vector (or $z = uvw$) is parallel to the incident beam. The spots of the diffraction pattern are symmetrically distributed around the center spot from the transmitted beam. By applying the vector rules forming parallelograms, R and θ values of the micrograph of the diffraction pattern can be measured and compared with the International Standards [19]. The zone axis $z = uvw$ is determined from the following relations:

$$u = k_1 l_2 - k_2 l_1 \qquad (4.15a)$$

$$v = l_1 h_2 - l_2 h_1 \qquad (4.15b)$$

$$w = h_1 k_2 - h_2 k_1 \qquad (4.15c)$$

where $h_1 k_1 l_1$ and $h_2 k_2 l_2$ are coordinates of each point in the diffraction [19–21].

4.6.2 Bright and Dark Field Image

Bright field image is mainly used in TEM analysis. In such image mode, areas when the electrons are absorbed, or scattered, appears darker, but areas where the electrons are transmitted unscattered, or electrons from the main beam, appear bright. In the bright field image, the main electron beam is chosen by positioning an aperture and the scattered rays are excluded.

The dark-field image appears as a negative of the bright-field image, where those parts which are less reflective, e.g. crystal defects such as pits or dislocations, are visible. TEM dark field and bright field image can be used to characterize the size and morphology of the sample as well as their crystal lattice (e.g. precipitates and dislocations). The main differences between the bright field and dark field mode are which electron populations are used to construct the TEM image as shown in Fig. 4.18.

In brief, dark field image is established by tilting the incident illumination until a diffracted beam from a crystal plane not the incident beam enters the objective aperture in the objective lens back focal plane. Therefore, the dark-field images are generated by the diffracted intensity from a single diffracting plane in relationship with a projected position on the sample, or sample's tilt. The most common sources of image contrast are atomic mass and the crystallinity of the sample. Heavier atoms
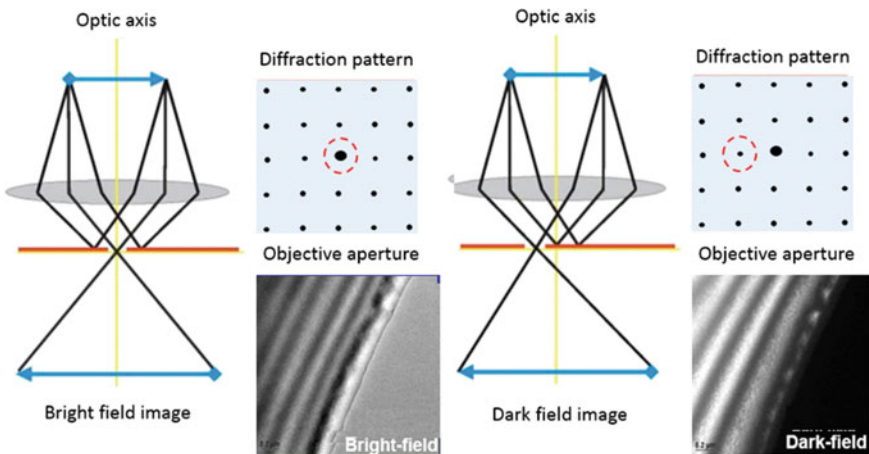
Fig. 4.18 Bright and dark field image principles [22]

scatter electrons more intensely than lighter atoms. Hence, in bright field mode, the regions with lighter atoms are darker, while in dark field mode these regions are brighter. Samples which are crystalline are strongly diffracted and will appear darker in bright field mode, and brighter in dark field mode. The most common sources of error associated with TEM images are associated with aligning and positioning a sample. It is also important to note that the smaller the aperture size, the higher the contrast will be obtained.

In polycrystalline samples, the dark-field images can illustrate the crystals which Bragg-reflection in a certain orientation.

One benefit with dark-field microscopy is that it provides an effective technique free of artifacts for detailed analysis of the material. One drawback of dark-field microscopy is the weak light levels in the image. Therefore, the sample has to be strongly illuminated and extra care is needed to avoid damaging of the sample.

4.6.3   Kikuchi Patterns

The Kikuchi pattern is generated through Bragg reflection of inelastically scattered electrons and they locate on the Bragg reflection spots. These scattered electrons are distributed over large angles and this leads to Kikuchi lines (KLs) appearing as pairs of parallel lines (hkl and –h–k–l) for a crystal plane.

It is practically useful to construct the Ewald Sphere and refer it to the Kikuchi sphere [22–24]. In this case, a sphere of radius λ^{-1} is established which is centered at the origin of the reciprocal space, as shown in Fig. 4.19. A second sphere with same radius λ^{-1} is centered on the reciprocal lattice point of interest, is drawn and the small circle obtained at the intersection of these two spheres illustrates the Kikuchi line (KL). In a similar way, the Kikuchi line can also be obtained by finding the intersection of the Kikuchi sphere and a plane perpendicular to and bisecting the wave vector **K**.

The map of the Kikuchi lines and how they intersect in diffraction provides information about the arrangement of crystal planes. The Kikuchi lines can help to tilt a crystal sample and find the crystal planes.

Kikuchi lines for high quality crystal appears as sharp lines. The Kikuchi lines with low intensity relate to defected material and they are visible on the near side of the incident beam direction, but the high-intensity (excess) lines appear far side of the direction as shown in Fig. 4.20.

4.6.4   Specimen Preparation

In contrast to SEM, TEM analysis requires advanced sample preparation. For instance, a typical sample thickness of 100 nm is necessary for Si-based materials to obtain sufficient brightness and contrast in the transmission analysis. In fact, more
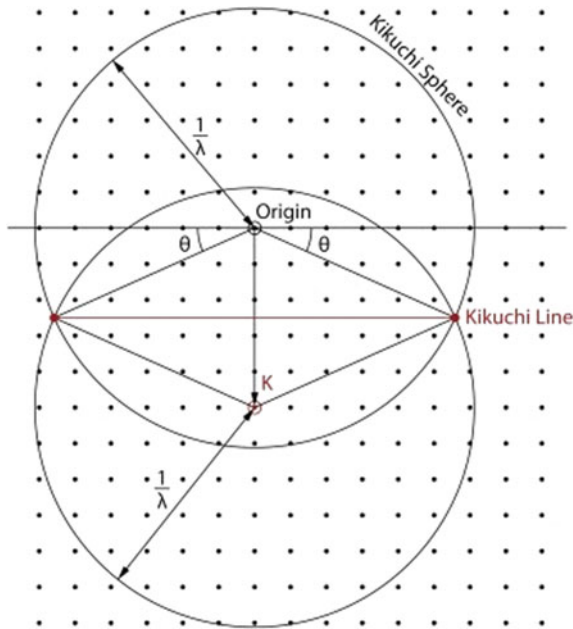
Fig. 4.19 Formation of Kikuchi lines through Ewald's theory

time can be spent on preparing the samples than on the actual analysis in the TEM instrument.

There are several methods for sample preparation. Traditionally, for cross sectional analysis, the samples are cut in small pieces and glued together in a stack, face-to-face. The whole sample is then cut, or sawn into disks, which are polished in several steps to be thinner. In the final step, these disks are mounted on a sample holder for Ar ion milling, where a beam of Ar ions are used to sputter away material and thin down the samples to the desired thickness [25].

As an alternative to this tedious thinning procedure, a focused ion beam (FIB) tool can be used for precise tailoring of samples for TEM analysis [26, 27]. The focused ion beam instrument uses a beam of ions which are scanned over an area of sample to remove material and shape the sample. Gallium is commonly used as ion source in standard FIB instruments to produce a high ion beam current with large sputtering yield. There are also other material alternatives, e.g. metal alloys such as Au with Si or Ge for certain applications. A FIB system is equipped with manipulators and may even have a built-in SEM to monitor the erosion process in situ. It is possible to fabricate 2D, or 3D features with sizes less than 10 nm from almost any solid material.

Figure 4.21 shows a schematic setup of FIB. The liquid metal ion source (LMIS) is commonly used, where gallium is sprayed from the tip of an electrically charged tungsten needle, resulting in ionization. The formed ions are accelerated to an energy of 5–50 keV and are then focused on the sample by using electrostatic lenses. FIB has
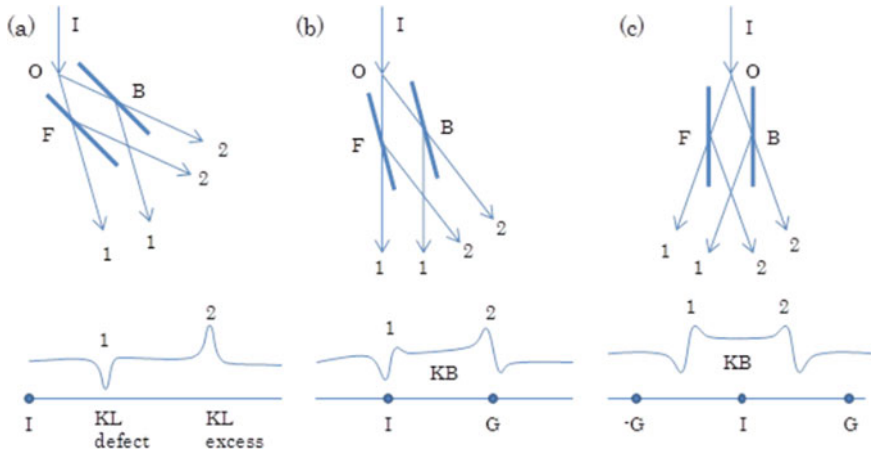
Fig. 4.20 Schematic of KLs formation when inelastically scattered electrons impinges on point O of a crystal plane with the front and back face F and B. The low intensity (defects) and excess intensity lines at Bragg reflections 1 and 2 are formed: **a** The incident beam I has a large angle to the crystal plane, **b** KLs are transformed to Kikuchi band (KBs) when the crystal plane is inclined at small angles to the incident electrons I. This results in the bell-shaped intensity of KLs becomes weaker and the symmetrical feature is lost and asymmetric intensity are visible **c** Kikuchi bands (KBs): When the incident electron impinges symmetrically on the crystal planes, the amplitude scattering become equal for the Bragg reflections 1 and 2 and the bell-shaped intensity disappears. Then a high intensity band is formed due to a strong dynamical diffraction effect [23]. A series of research efforts have been performed to simulate KLs. By computing electron diffraction intensities to reciprocal lattice points the kinematic Kikuchi patterns is simulated [24]

many other applications ranging from micro-electro-mechanical systems (MEMS), microfluidic chips, to biological sciences.

Another option for FIB is using gas field ion sources (GFIS). The gas field source can generate narrower beams compared to LMIS, creating a better resolution, but
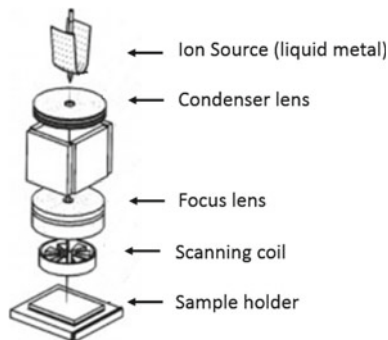
Fig. 4.21 A schematic view of a FIB setup. The lenses focus the ion beam where the scanning coil deviate the beam in a certain direction [27]

have relatively short lifetime. The ion beams can be formed by H_2 gas as well as noble gases e.g. He. The gases are condensed and are ionized by the electrically charged tungsten needle. As an example, Xe ion beam provides better milling compared to gallium beam since it offers higher sputter yields.

4.6.5 Application of TEM for Nano Devices

Transmission electron microscopy has become perhaps the most powerful analytical technique in the field of material science for nm and sub nm samples and new varieties and applications of the technique are still being invented. In recent years, the analysis of electron diffraction pattern has, for instance, been used to determine strain in nanoscale material. Here, we discuss how the strain in nano-scale transistors has been measured by the nano-beam diffraction (NBD) technique. The idea behind these measurements is that the lattice constant is disturbed in presence of strain and the diffraction pattern is affected by the strain amount. The diffraction pattern can also be theoretically analyzed by applying simulation program, such as True Crystal Strain Analysis. In the analysis of NBD, the interplanar distance of (220) planes is commonly measured and simulated.

Figure 4.22 shows a cross-section and diffraction pattern images from the metal gate, channel and the body parts of a 22 nm transistor. The analysis was performed by using an electron beam of nano scale diameter and it can be seen that the diffraction pattern from the metal gate material has weak intensity with diffused airy rings and only a few spots (top left figure). This indicates that the material in this part is of poor polycrystalline quality mixed with the amorphous phase. Comparing with theoretical estimates, it can be seen that the strain in the channel is ~0.5 GPa [28].

Although NBD is an impressive analysis method, but a drawback with the technique is strain relaxation may occur during sample preparation. More advanced and careful sample preparation is required for a successful and precise NBD analysis.

4.6.6 Advanced TEM Instruments

In conventional TEM, the electrons propagate through a thin sample for analysis. In scanning transmission electron microscope (STEM), the electron beam has a small spot size of 0.05–0.2 nm which is scanned over an area of the sample. In this way, the sample is analyzed in each point by a raster illumination system, similar to the SEM. Therefore, STEM also provides an excellent platform for analytical techniques such as annular dark-field imaging, and energy dispersive x-ray (EDX) and electron energy loss spectroscopy (EELS) for spectroscopic mapping.

Most of today's conventional TEM instruments are equipped with necessary circuitry and coiling system with detectors which allows the instrument to be used as
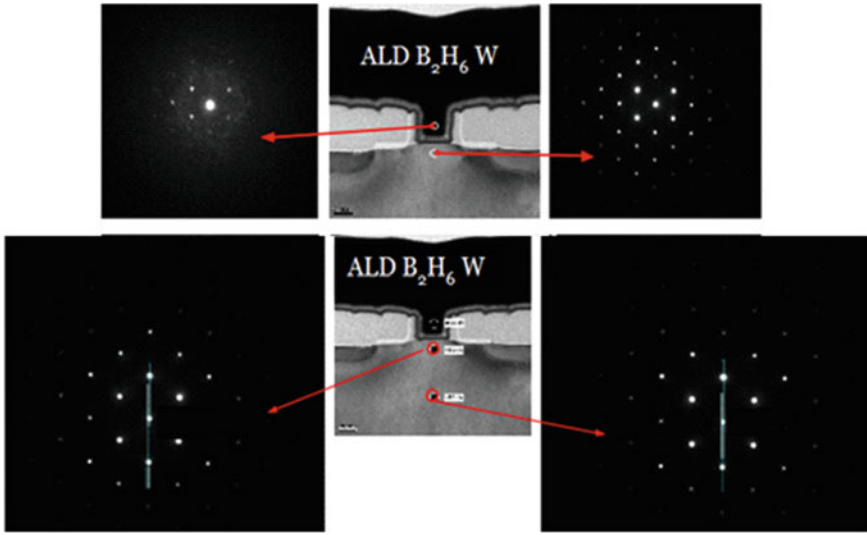
Fig. 4.22 NBD analysis from metal gate, channel and the body regions of a 22 nm CMOS transistor with induced strained in the channel region [28]

STEM. The electron signals can be applied for imaging and spectroscopic application simultaneously. One important point for STEM microscope is the requirement for a vibration-free and temperature-controlled environment, due to their extreme lateral sensitivity with atomic resolution.

4.6.7   Chemical Analysis

There are two methods for chemical analysis, or studying the elemental composition, in a TEM microscope: x-ray energy dispersive spectrometry (EDS) and electron energy loss spectrometry (EELS), as shown in Fig. 4.23.

The electron beam interactions with the sample may excite atoms leading to x-ray emission, as they relax. The energy of the emitted x-ray is a fingerprint and can be used to identify the atoms. In EDS, the detector is shielded by an ultrathin window which is cooled by liquid nitrogen to eliminate thermal noise. The EDS spectrum then consists of a background generated by bremsstrahlung x-rays and the characteristic peaks from the interatomic electron transitions, which are used to identify the chemical elements of the material.

The EDS x-ray detector typically consists of a Si(Li) detector or a silicon-drift detector (SDD), a collimator, an electron trap to avoid contribution from electrons, and a window to isolate the detector under vacuum. The EDS detector operates when the monitored x-ray creates electron–hole pairs in the detector material. By applying a high voltage across Si(Li) crystal the two electrodes collects the electrons and
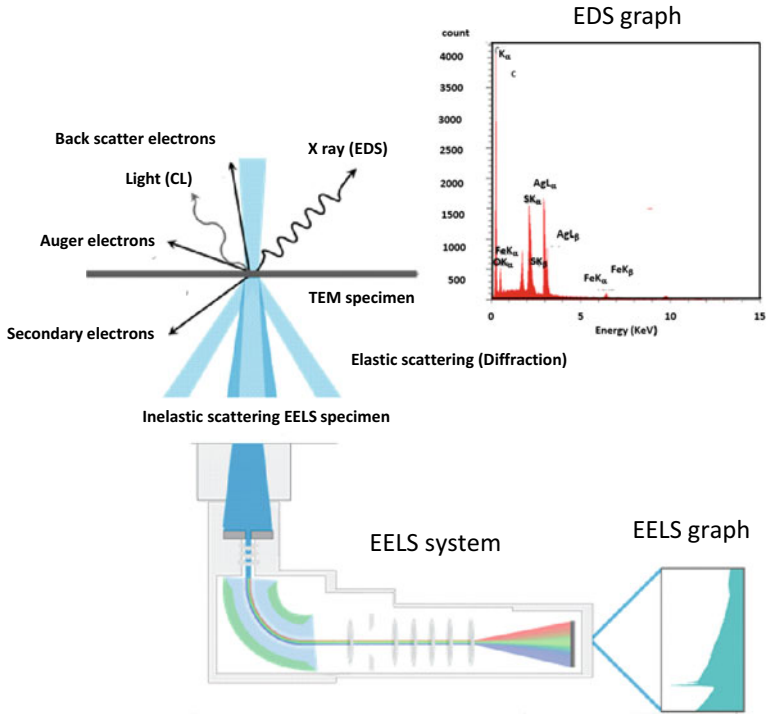
Fig. 4.23 Chemical analysis equipment in TEM instrument

holes, where the signal strength is proportional to the x-ray energy. As an example, to detect Ni, or Al K_α x-rays, with characteristic energies of 7.477 and 1.487 keV, respectively, about 1970 and 390 electron–hole pairs are generated in the detector crystal for the Ni and Al cases.

EDS technique compared to other x-ray technique, e.g. XPS, has significantly deeper penetration length in the materials, but the detected photoelectrons are emitted from only a few nm depth. XPS is excellent for chemical analysis of surfaces or nanosheets, while EDS is a powerful technique for bulk material. Meanwhile, EDS provides significantly smaller volumes for analysis. The Fig. 4.24 compares the analysis of materials using EDS and XPS.

Electron energy loss spectroscopy (EELS) is an analysis mode often integrated in TEM tools to determine the energy distribution of electrons which transmit a thin sample. TEM tools typically use high energy electrons (60–300 kV) and EELS deals with elastic and inelastic electrons to extract information about the sample. This means that the tool contains an electron energy analyzer. Figure 4.25 demonstrates the features of an EELS spectrum.

EELs peaks can be categorized in the following types:

1. Zero loss peak (ZLP): elastic scattering: this provides minor information.
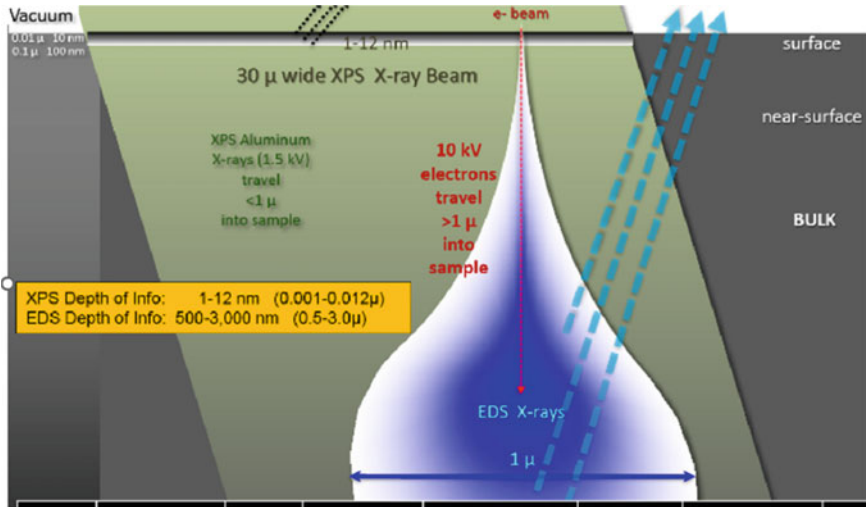
Fig. 4.24 Schematic view of the regions for EDS (blue interaction volume) and XPS (green area, where only the top 12 nm are used) for material analysis in a TEM
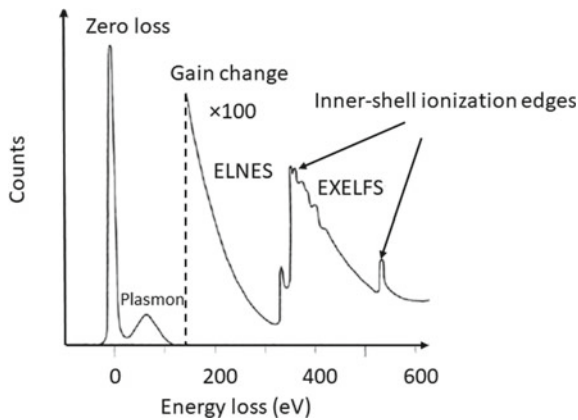
Fig. 4.25 An energy loss spectrum with different distinguishable features in regions: zero-loss region, low-loss distribution and core-loss region

2. Inelastic collision with outer shell electrons (low loss distribution). This region appears with a plasmon peak and the peak provides information about band structure and interband transitions (near zero-loss features).
3. Inelastic collision with inner shell (ionization edges or core loss region). This part of the spectra provides information about elemental composition (or core-loss edge), chemical bonding and oxidation state (density of unoccupied states). This high energy part of the spectrum is called near edge fine structure. Further information about the distribution of near neighboring atoms can be extracted

from the part of spectrum by a technique called extended energy loss fine structure (EXELFS) [29].

EXELFS shows a fine structure which lies with energies of 40–200 eV above the absorption-edge energy of the EELS core-loss spectrum. EXELFS has a close similarity to extended x-ray absorption fine structure (EXAFS) in x-ray spectroscopy and provides similar information about the local atomic structure, but it offers a better sensitivity for low Z elements with nanometer lateral resolution. The EXELFS is generated from scattering of electrons, which are created through excitation by neighboring atoms from the inner-shell state to the states in the conduction band. In order to obtain a proper data acquisition a few issues have to be considered in EXELFS measurement e.g. monitoring thickness and energy resolution variation, radiation damage during measurements, and better dark-current correction.

High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) is a TEM mode which operates at relatively large angles of the primary beam (~50–200 mrad) and uses an annular dark-field (ADF) detector to detect inelastically scattered electrons or thermal diffuse scattering (TDS) [30, 31] (see Fig. 4.26). A STEM image displays the intensities of the transmitted electrons in respect to the incident probe position (see Fig. 4.27). In a HAADF image, the intensity is proportional to about $Z^{1.4}$-Z^2 of the atomic number Z, Therefore, heavy atoms give bright contrast, while light atoms will be dark, or difficult to observe. However, the light atoms are visualized by the annular bright-field (ABF) STEM and using low-angle annular dark-field (LAADF) STEM. There are two advantages with the HAADF technique:

(1) Since the scattering cross-section of TDS is small at high operating angles, very few multiple scattering events occur.
(2) There is no interference effect of electrons and the obtained image is a non-interference image.

The resolution of the HAADF image depends on the incident probe diameter and is usually better than 0.05 nm. When HAADF technique is combined with EELS, element analysis of atomic? column by column can be obtained since the electrons are transmitted through the center hole of the ADF detector.

4.6.8 Scanning Moiré Fringe Analysis

Scanning moiré fringe (SMF) imaging is widely applied to estimate the strain in crystal structures. The SMF technique is a kind of harmonic interference which appears due to an artifact that occurs in the middle range of magnification in HAADF-STEM image. One of the benefits with SMF technique is that it requires no reference crystal. The condition to obtain SMFs in HAADF-STEM image is when the scanning grating pixel size, d_s, is close to the lattice plane spacing, d_l, or is a multiple thereof. Figure 4.28a–c illustrates schematic pictures how SMF is formed when two patterns
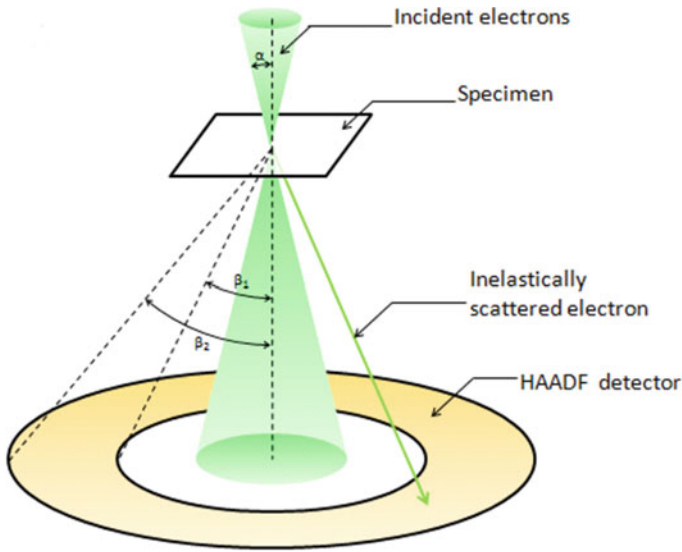
Fig. 4.26 An illustration of HAADF-STEM technique. The inner and outer detector angles are β_1 = ~50 mrad and β_2 = ~200 mrad, respectively. The convergence angle, α is typically 25 mrad for a 200 kV TEM. In such a setup, an ABF detector and a LAADF detector are placed below a HAADF detector [32]
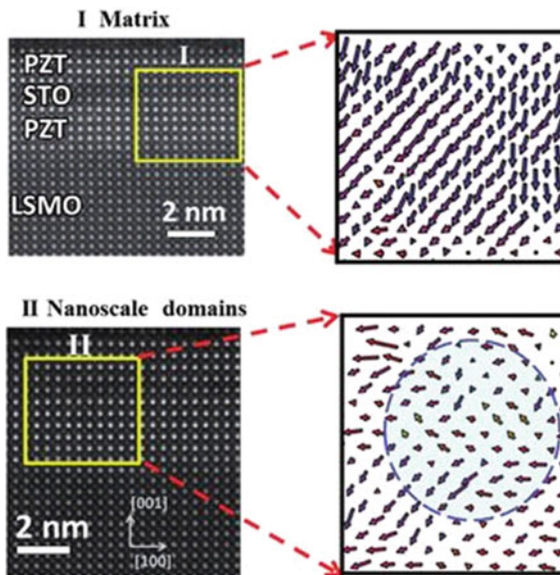
Fig. 4.27 Cross-sectional STEM HAADF images from PZT_2ucSTO ($PbZr_{0.2}Ti_{0.8}O_3$_2uc $SrTiO_3$) material (left) and the polarization displacement maps of the zoomed area (right) with two distinct topologies each illustrating (i) matrix and (ii) a bubble domain [33]

(a)

(b)

(c)

d_{SMF}

d_l = lattice spacing d_s = scanning grating spacing $d_l < d_s$ $d_l = d_s$ $d_l > d_s$
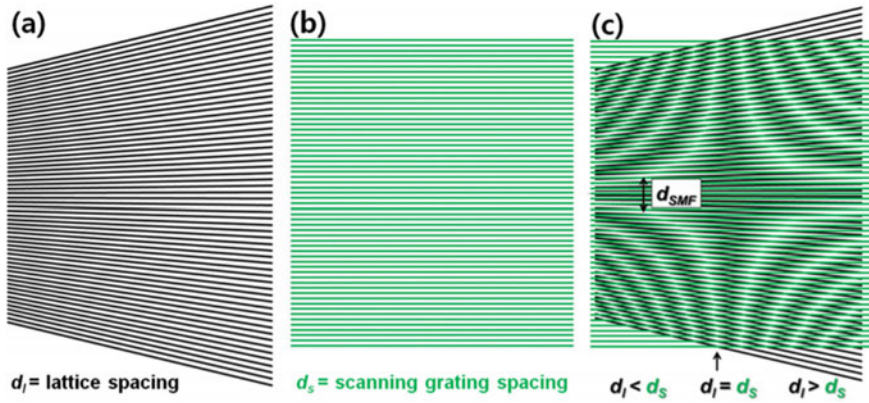
Fig. 4.28 Schematic view of SMFs formation when the harmonic interference occurs between lattice spacing d_l and the scanning grating with a spacing d_s, as following: **a** d_l with a strain field which linearly increases along the horizontal direction, **b** d_s of STEM imaging, and **c** formed SMFs by overlapping of d_l and d_s [34, 35]

interfere with each other. The SMF spacing, d_{SMF}, increases when the lattice spacing increases but with $d_l < d_s$, whereas it decreases when the lattice spacing increases but with $d_l > d_s$ [34].

The lattice spacing can be formulated from the moirés fringes:

$$d_1 = \frac{d_{SMF}d_s}{d_{SMF} + d_s}, \ d_1 < d_s \quad d_1 = \frac{d_{SMF}d_s}{d_{SMF} - d_s}, \ d_1 < d_s \qquad (4.16)$$

In order to provide a better insight about SMF technique an example for SiGe/Si hetereostructure is presented here. Figure 4.29a–d illustrates HAADF-SMF technique measured at a $Si/Si_{1-x}Ge_x$ interface layer. In the SMFs, the image demonstrates the amount of curvature as a result of a strain field at the interface. Figure 4.29c shows how the SMF spacing, d_{SMF}, increases continuously towards the $Si_{1-x}Ge_x$ layer. In order to determine the strain field at the interface, d_{SMF} has to be estimated along the [001] growth direction, as marked by the dashed line in Fig. 4.29a. After determining d_{SMF}, Eq. 4.16 can be used to estimate the (220) d-spacing of the region of interest, d_{ROI}. The strain amount can be determined by applying the following equation:

$$Strain(\%) = \frac{d_{ROI} - d_s}{d_{Si}} \times 100 \qquad (4.17)$$

where d_{Si} stands for (220) d-spacing of the unstrained Si reference ($d_{Si} = 0.1920$ nm). In fact, the strain field of the $Si_{1-x}Ge_x$ in the Si matrix can be determined relative to the unstrained Si substrate [36]. In Fig. 4.29b, positive values of strain stand for tensile stress and the negative values presents compressive stress. By using Eq. 4.17, the tensile strain of $0.4\% \pm 0.1\%$ is estimated at the Si layer in the interface, and this increases toward the $Si_{1-x}Ge_x$ layer to a maximum value of $1.1\% \pm 0.1\%$ at the
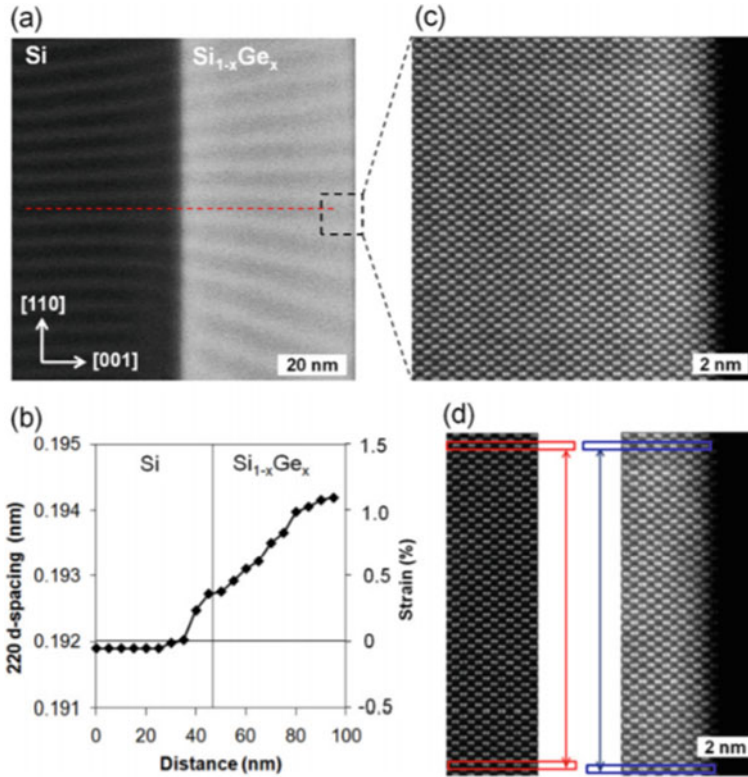
Fig. 4.29 **a** A HAADF-SMF image of the $Si/Si_{1-x}Ge_x$ interface. **b** Strain profile extracted from the marked area in the SMF image, **c** HAADF-STEM image of the $Si_{1-x}Ge_x$ layer obtained at dashed-line box in the SMF image. **d** HAADF-STEM image of the unstrained Si as reference substrate and $Si_{1-x}Ge_x$. The (220) lattice spacing between the two boxes number 60. The dark area with contrast indicates a vacuum [36]

marked area in Fig. 4.29a. The error in this exercise can be derived by the standard deviation after repeating the measurements. In order to confirm the estimated strain using SMF imaging, a high-resolution STEM image of the $Si_{1-x}Ge_x$ layer can be performed to estimate independently the strain amount in the $Si_{1-x}Ge_x$ layer.

Figure 4.29c shows the STEM image of the $Si_{1-x}Ge_x$ layer at the region where the strain has the maximum value, enlarged from Fig. 4.29a as indicated by the dashed lines. By comparing the spacing of sixty (220) fringes between the $Si_{1-x}Ge_x$ and Si crystal as marked by the arrows in Fig. 4.29d, the strain of the $Si_{1-x}Ge_x$ layer was directly estimated to be $1.1\% \pm 0.1\%$. This is consistent with the strain obtained by the SMF image. SMF imaging is also sensitive to defects in the crystal structure where an equivalence in SMF images is disturbed [36].

Nowadays, HAADF-SMF technique is widely used to analyze lattice distortion locally in different nano structures. The technique will contribute to more advanced debugging of the performance of nano devices in future.

References

1. Bradbury, S. (1967). *The evolution of the microscope.* Pergamon Press. ISBN-10:1483131904.
2. Palucka, T. (2002). Overview of electron microscopy. History of recent science & technology.
3. M. von Ardenne. (1938). Das Elektronen-Rastermikroskop. Theoretische Grundlagen. *Zeitschrift für Physik, 109*(9–10), 553–572. https://doi.org/10.1007/BF01341584
4. von Ardenne, M. (1938). Das Elektronen-Rastermikroskop. Praktische Ausführung. *Z. Tech. Phys. 19*, 407–416.
5. McMullan, D. (1995). SEM 1928–1965. *Scanning, 17*, 175–185. https://doi.org/10.1002/sca.4950170309
6. Crewe, A. V., Wall, J., & Langmore, J. (1970). Visibility of a single atom. *Science*, *168*(3937), 1338–1340. https://doi.org/10.1126/science.168.3937.1338
7. Shin, D. H., Kirkland, E. J., & Silcox, J. (1989). Annular dark field electron microscope images with better than 2 Å resolution at 100 kV. *Applied Physics Letters, 55*(23), 2456. https://doi.org/10.1063/1.102297
8. Goodhew, P. J., Humphreys, J., & Beanland, R. (2001). *Electron microscopy and analysis* (3rd ed). CRC Press is an imprint of Taylor & Francis Group. ISBN-13: 978-0-7484-0968-6 (pbk).
9. Joy, D. C. (2005). The aberration corrected SEM. *AIP Conference Proceedings, 788*, 535. https://doi.org/10.1063/1.2063015
10. Zach, J., & Haider, M. (1995). Aberration correction in a low voltage SEM by a multipole corrector. *Nuclear Instruments and Methods in Physics Research (Section A), 363*, 316–325.
11. Krivanek, O. L., Delby, N., & Lupini, A. R. (1999). Towards sub-Å electron beams. *Ultramicroscopy, 78*, 161–180.
12. Ludwig, R. Scanning electron microscopy physics of image formation and microanalysis. Springer. ISSN: 0342-4111.
13. Potts, P. J. (1987). *A handbook of silicate rock analysis.* 336. Springer. ISBN 978-94-015-3988-3.
14. Goodhew, P., Humphrey, J., & Beanland, R. (2001). Electron microscopy and analysis. Taylor & Francis. ISBN-13:978-0748409686
15. Hortolà, P. (2005). SEM examination of human erythrocytes in uncoated bloodstains on stone: Use of conventional as environmental-like SEM in a soft biological tissue (and hard inorganic material). *Journal of Microscopy, 218*(2), 94–103. https://doi.org/10.1111/j.1365-2818.2005.01477.x
16. Radamson, H. H., He, X., Zhang, Q., Liu, J., Cui, H., Xiang, J., Kong, Z., Xiong, W., et al. (2019). Miniaturization of CMOS. *Micromachines, 10*(5), 293, 48. https://doi.org/10.3390/mi10050293
17. Li, J., Wang, W., Li, Y., Zhou, N., Wang, G., Kong, Z., Fu, J., Yin, X., Li, C., Wang, X., et al. (2020). Study of selective isotropic etching $Si_{1-x}Ge_x$ in process of nanowire transistors. *Journal of Materials Science: Materials in Electronics, 31*, 134–143. https://doi.org/10.1007/s10854-019-02269-x
18. Reid, J., Crane, D., Blanton, J., Crowder, C., Kabekkodu, S., & Fawcett, T. (2011). Tools for electron diffraction pattern simulation for the powder diffraction file, 19, 32–37. https://www.cambridge.org/core/journals/microscopy-today/issue/2701B0C54B4BC9E460A3606221EAF0ED. https://doi.org/10.1017/S1551929510001240
19. Janecek, M., Kral, R., Asadi Asadabad, M., & Jafari Eskandari, M. (2016). Modern electron microscopy in physical and life sciences. *Electron Diffraction*, 3–25. OpenTech. https://doi.org/10.5772/61781 (ISBN: 978-953-51-2252-4).
20. Ball, C. J. (1971). *An introduction to the theory of diffraction.* Pergamon Press. ISBN 10:0080157866/ISBN 13:9780080157863.
21. Hammond, C. (1997). *The basic of crystallography and diffraction.* Oxford Science Publications. ISBN 978-0-19-873868-8.
22. Hirsch, P., Howie, A., Nicholson, R., Pashley, D. W., Whelan, M. J. (1965/1977). *Electron microscopy of thin crystals* (Butterworths/Krieger). ISBN 0-88275-376-2.

23. Jeol Glossary: Kikuchi. https://www.jeol.co.jp/Kikuchi
24. Herron, A. D., Coleman, S. P., Dang, K. Q., Spearot, D. E., & Homer, E. R. (2018). Simulation of kinematic Kikuchi diffraction patterns from atomistic structures. *Methods X, 5*, 1187–1203. https://doi.org/10.1016/j.mex.2018.09.001
25. Duan, S. Q., Yu, Q. Q., Chen, L., et al. (2016). Preparation methods of the TEM samples for the nanoscale chip. *Semiconductor Technology, 41*(12), 945–950.
26. Melngailis, J. (1987). {FIB} Focused ion beam technology and applications. *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena, 5*, 469. https://doi.org/10.1116/1.583937
27. André, R. M., Marques, M. J. B., Frazão, O. (2015). FIB fiber microstructure sensors based on focused ion beam technology. *Photoptics*, 3–15.
28. Wang, G., et al. (2017). pMOSFETs featuring ALD W filling metal using SiH4 and B2H6 precursors in 22 nm node CMOS technology. *Nanoscale Research Letters, 12*(1). https://doi.org/10.1186/s11671-017-2080-2
29. Qian, M., Sarikaya, M., & Stern, E. A. (1995). Development of the EXELFS technique for high accuracy structural information. *Ultramicroscopy, 59*(1–4), 137–147. https://doi.org/10.1016/0304-3991(95)00024-U
30. Mundy, J. A., Hikita, Y., Hidaka, T., Yajima, T., Higuchi, T., Hwang, H. Y., Muller, D. A., & Kourkoutis, L. F. (2014). Visualizing the interfacial evolution from charge compensation to metallic screening across the manganite metal–insulator transition. *Nature Communications, 5*, 3464. https://doi.org/10.1038/ncomms4464
31. Friel, J. J., & Lyman, C. E. (2006). Tutorial review: X-ray mapping in electron-beam instruments. *Microscopy and Microanalysis, 12*(1), 25. https://doi.org/10.1017/S1431927606060211
32. Pennycook, S. J., & Jesson, D. E. (1991). High-resolution Z-contrast imaging of crystals. *Ultramicroscopy, 37*(1–4), 14–38. https://doi.org/10.1016/0304-3991(91)90004-P
33. Zhang, Q., Xie, L., Liu, G., Prokhorenko, S., Nahas, Y., Pan, X., Bellaiche, L., Gruverman, A., & Valanoor, N. (2017). Nanoscale bubble domains and topological transitions in ultrathin ferroelectric films. *Advanced Materials, 29*(46), 1702375. https://doi.org/10.1002/adma.201702375
34. Kim, S., Lee, S., Oshima, Y., Kondo, Y., Okunishi, E., Endo, N., Jung, J., Byun, G., Lee, S., & Lee, K. (2013). Scanning moiré fringe imaging for quantitative strain mapping in semiconductor devices. *Applied Physics Letters, 102*, 161604. https://doi.org/10.1063/1.4803087
35. Su, D., & Zhu, Y. (2010). Scanning Moiré Fringe imaging by scanning transmission electron microscopy. *Ultramicroscopy, 110*, 229–233.
36. Denneulin, T., Rouviere, J. L., Beche, A., Py, M., Barnes, J. P., Rochat, N., Hartmann, J. M., & Cooper, D. (2011). The reduction of the substitutional C content in annealed Si/SiGeC superlattices studied by dark-field electron holography. *Semiconductor Science and Technology, 26*, 125010.

Part II
Material Characterization Using Ions

Chapter 5
Rutherford Backscattering Spectrometry

5.1 Introduction

Rutherford backscattering analysis (RBS) is one technique in a family of material characterization techniques were energetic ions are used to probe thin films and solid surfaces. The historic origin of RBS dates back to the first decades of the previous century, when Ernest Rutherford (1871–1937) used positively charged alpha particles to investigate the interaction between these particles and thin foils. In his classical experiment, Rutherford studied the deflection of alpha particles as they passed through gold foils. With the prevailing model of how mass and charge was distributed in the atoms, he expected only small deflections of the alphas from their original path. To his surprise he found that this was not the case; occasionally alpha particles where even scattered backwards, opposite of their original direction. Rutherford described this as "The most astonishing that ever happened to me in my life. It was as incredible as if you fired 15 inch shells on a piece of tissue paper and they bounced back at you" [1]. By careful experiments and clever calculations, Rutherford could demonstrate that the only explanation to his findings was if all the positive charge and most of the mass of atoms was contained in a small nucleus and not smeared out. The negatively charged electrons he placed in orbits around this nucleus much like planets orbiting the sun. Rutherford's result constitutes a foundation for modern physics and our conception of the atom.

Scattering experiments, such as Rutherford's, have been used extensively in physics to probe deeper and deeper into matter. With ever increasing energies of the colliding particles it has been possible not only to induce nuclear reactions and break up the atomic nucleus, but also to unveil the fine structure of the basic building blocks of the nucleus, neutrons and protons. As the fundamental research continues and the understanding of the interactions between energetic particles evolves, it has been realized that the scattering techniques can also be used to help us understand material behavior and composition on a nanometer scale. Since the fundamentals of the collisions and energy loss mechanisms are now well established, ion scattering

experiments with selected incident particles can be used to obtain information about, for instance, the mass of unknown elements in a film, how elements are distributed in the near surface regions and the stoichiometry of compound layers.

Figure 5.1 gives an overview of the most commonly employed techniques where energetic ions are used for material analysis. This chapter will focus on Rutherford backscattering spectroscopy, which still is the most useful and practically accessible method of these. However, first we describe more generally what happens to energetic ions as they are injected into a solid matrix. This overview not only provides insight in RBS, but also gives a background to secondary ion mass spectrometry (SIMS), which is described in the following chapter.
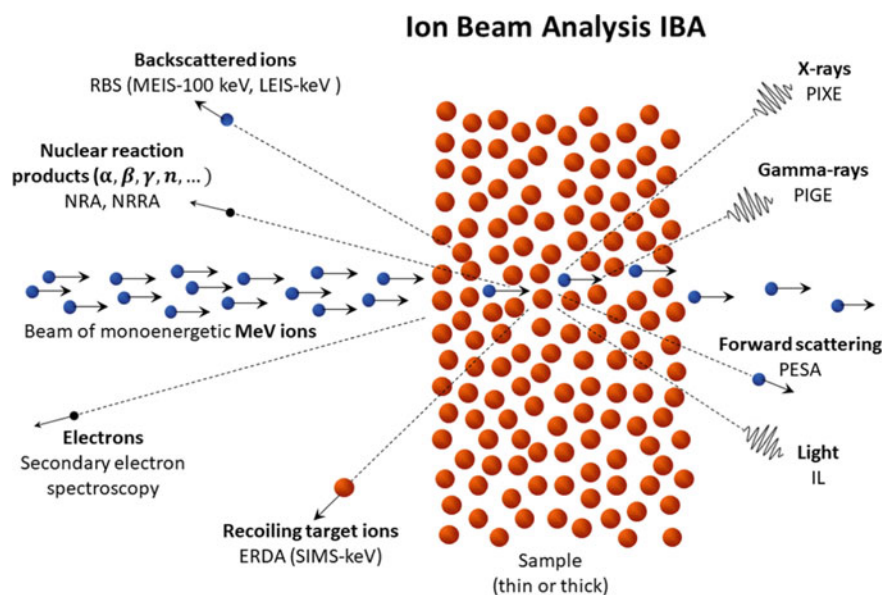
Fig. 5.1 Common ion scattering techniques used for ion beam analysis (IBA) of materials. Backscattering is also measured with about 100 keV incident ions (medium energy ion scattering, MEIS) and incident ions of a few keV (low energy ion scattering, LEIS). Notice also that detection of recoiling target ions, i.e., secondary ions, with primary incident ions of tenths of keV is termed secondary ion mass spectrometry (SIMS) and will be described in the next chapter. Other abbreviations in the figure: NRA-nuclear reaction analysis, NRRA-nuclear resonance reaction analysis, ERDA-elastic recoil detection analysis, IL-ionoluminescence, PESA-particle elastic scattering analysis, PIGE-particle induced gamma emission, and PIXE-Particle induced induced X-ray emission

5.2 Interaction Between Ions and Solids

An atom, or molecule which is stripped of one or more electrons becomes positively charged, while atoms or molecules that attach extra electrons become negatively charged. Such particles, or ions, can easily be accelerated in an electric field and obtain a desired velocity and corresponding momentum and kinetic energy. If the ion encounters a surface of a solid material it will penetrate into this material to a certain depth, giving up this kinetic energy to the solid in numerous collisions with target atoms and electrons. Depending on the velocity of the ion, different interaction processes will dominate the energy loss. This is illustrated in Fig. 5.2, where the energy dependence of two main types of interaction mechanisms are shown, referred to by Niels Bohr as nuclear and electronic stopping [2].

5.2.1 Electronic and Nuclear Stopping

The nuclear interactions are collisions between positively charged ions and positively charged target nuclei, involving repulsive Coulombic forces, where the nuclear charge of the ion, Z_1, and the target atom, Z_2, can be more or less reduced by screening from inner shell electrons. Substantial energy and momentum transfer may occur in these collisions, leading to large angular deflection of the ions and also displacement
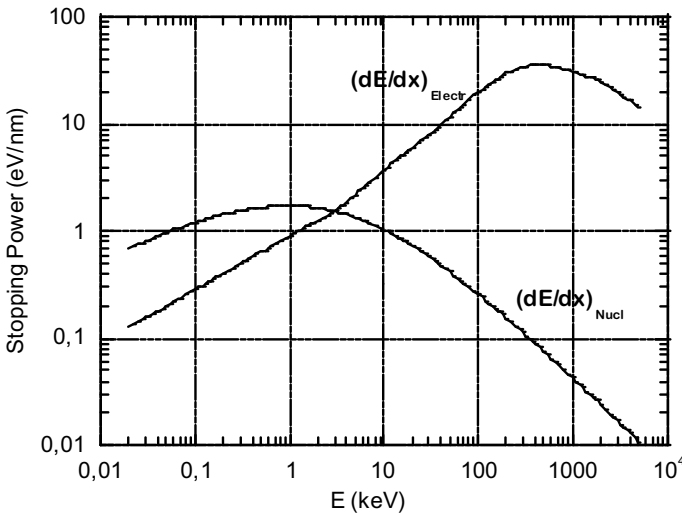
Fig. 5.2 The average energy lost in electronic and nuclear scattering events by incident He ions to a Si target as a function of the ion energy. The unit for the energy loss is shown in eV/nm. The energy loss is separated in electronic and nuclear stopping, where electronic loss totally dominates the interactions at larger energies

of the target atoms. Recoiling target atoms may even gain so much momentum in these collisions that they are able to cause further displacement collisions with other target atoms. Nuclear collisions are considered as elastic collisions and resembles collision between two hard solid spheres, like billiard balls. The electronic stopping involves the much more frequent collisions between (screened) ion and target electrons of attractive Coulombic nature, that leads to excitation and ionization of target atoms. Due to the abundance of these collisions and the relatively small amount of energy transfer in each collision, it is fruitful to look at the electronic stopping as a continuous force, very much like mechanical friction. The electronic interactions are often termed inelastic, since some of the energy is absorbed in internal atomic processes.

As can be seen in Fig. 5.2, both the nuclear and electronic stopping increases with energy for lower energies, passes through a maximum and then decreases as the kinetic energy is increasing. At what energy the maxima occur, and the value of the energy loss per unit path length, dE/dx, depend on the mass and charge of the ion and the types of atoms in the target material. The figure shows the case of He ions dissipating their kinetic energy in a silicon target. To the right in Fig. 5.2, at very high kinetic energy, or large velocities, the penetrating ion will be stripped of its electrons in electronic collisions and move at such a speed that the energy transfer eventually becomes very low. Although relativistic effects may play a role in this energy region, such energetic particles are interacting very little with the target and may penetrate to large depths. Rutherford backscattering analysis is normally performed at an energy of the incident particle which gives a strong electronic energy loss, $(dE/dx)_{electr}$, that can easily be detected, around a few MeV. The average nuclear energy loss, $(dE/dx)_{nucl}$, is relatively low in this region, but nevertheless provides a sufficient number of backscattered particles to make the analysis time efficient. In contrast, the energy used for the primary ions in SIMS (Chap. 6) is chosen to be around the maximum of the nuclear energy loss, since this energy range is most useful for ejecting secondary particles. Also, heavier ions than He are generally used in SIMS to enhance the secondary ion yield from the uppermost surface layers.

The term "nuclear energy loss" does not imply that the ions induce nuclear reactions, but should be seen as elastic collisions between hard spheres. However, at certain energies for the impinging ion, target nuclei may reach a resonant state or, for high enough energies, the ion may penetrate the Coulomb barrier of the target nuclei and trigger nuclear reactions. Such scattering events constitutes a third type of energy loss mechanisms, in addition to the nuclear and electronic stopping, and can also be utilized for material characterization. Nuclear reaction analysis (NRA) and nuclear resonance reaction analysis (NRRA) are valuable complementary techniques to RBS, since these techniques offers possibilities to probe light elements, such as hydrogen, nitrogen and oxygen, which is often difficult, if not impossible, with standard RBS.

5.2.2 Collision Cross Sections

The total scattering cross section, σ, is thought of as an area which is proportional to the probability that an impinging particle will experience a collision with a target particle. It is a very useful concept, since it makes it possible to link microscopically sized particles with measurable macroscopic quantities. Thus, the probability for a certain scattering event between monoenergetic particles in a beam of intensity I and particles in a foil of thickness t and atomic density N will be

$$P = \frac{Number\ of\ events}{Number\ of\ incident\ particles} = \frac{I\sigma Nt}{I} = \sigma Nt \qquad (5.1)$$

The quantity Nt corresponds to the areal density of particles (cm^{-2}) and, by multiplying with the cross section (cm^2), the dimensionless probability for a scattering event will be given. The probability is a measurable quantity that can be determined with high precision due to the many particles normally involved in these types of experiments. Since foil density and thickness can be decided, the cross sections can readily be determined.

The total scattering cross section, σ, describes the collision probability and is specific for each incident and target particle pair. It also varies with the energy of the impinging particle, as illustrated in Fig. 5.2 above, showing the energy loss of an incident He ion in a target matrix consisting of Si. In practical experiments it is often more convenient to study the differential cross section, $d\sigma/d\Omega$, which gives the probability for scattering a particle into a differential solid angle $d\Omega$, centered around the scattering angle, θ, as illustrated in Fig. 5.3.
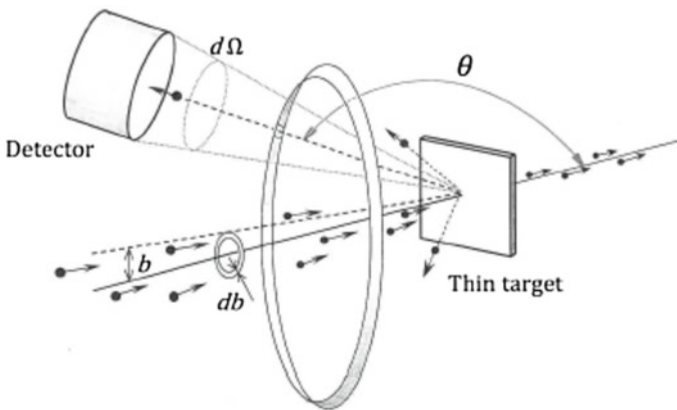
Fig. 5.3 A schematic view of a scattering experiment where an incident particle interacts with a particle in a thin target and is (back-) scattered at an angle θ into a detector of area, A, subtending a solid angle segment $d\Omega$. As indicated in the figure, most particles in the incident beam pass through the foil without being deflected in nuclear interactions. The impact parameter, b, is greatly exaggerated in the figure

In the figure, the particle passes through a slice of a sphere surrounding the scattering center. For isotropic scattering considered here, the scattering probability is equal around this ring. Also, the average differential scattering cross section associated with this scattering angle can be defined as

$$\sigma(\theta) = \frac{1}{\Omega} \int_{\Omega} \frac{d\sigma}{d\Omega} d\Omega \tag{5.2}$$

For a particular detector with an active area A positioned at a distant l from the beam-target interaction point the detector subtends a solid angle $\Omega = A/l^2$ steradians (sr). As a comparison, from the earth, the solid angel subtended by the sun, Ω_{sun}, is about 6.8×10^{-5} sr. For a typical RBS setup, $\Omega_{detector} \sim 100\Omega_{sun}$. The detector, of which there exist a large number of types, counts the number of backscattered particles and also determines their individual energy. The total scattering cross section can be obtained from the differential cross section by integrating in all 4π directions:

$$\sigma_{tot} = \int_{4\pi} \frac{d\sigma}{d\Omega} d\Omega \tag{5.3}$$

5.3   The Principles of Rutherford Backscattering Spectrometry

We will start this description with the fact that energy and momentum are conserved in elastic binary collisions. From this fundamental property the so-called kinematic factor is deduced, which links the energy of the backscattered particle to the mass of the target particles at the surface of the solid. In a second step, we use the well-known values of electronic stopping to find the reduction in energy of the incident particle as it penetrates and collides deeper into the solid. For collisions with the same element as found at the surface, this makes it possible to relate the outgoing particle energy with the depth in the target and obtain depth scales for different target elements. Thirdly, we use the concept of cross sections to estimate the abundance of various target elements present in a sample. Since the cross sections are also well-known, the number of backscattered particles within the detector solid angle, i.e. the yield in relation to the number of incoming particles, gives the number of scattering centers. The aim here is to introduce RBS and the reader is referred several excellent dedicated reviews that have been published about the RBS technique over the years for more in-depth understanding [3–5].

5.3.1 Nuclear Energy Loss—Elastic Binary Collisions

For the nuclear stopping events that lead to backscattering very close to the surface, energy and momentum of the system are carried by the incident particle. This means that we can use conservation laws to obtain a relation between the energy of the ion before and after the scattering event. The system before and after a collision between an incoming ion of energy E_0, mass M_1, and charge Z_1, and a stationary target atom of mass M_2 and charge Z_2 is described in Fig. 5.4.

For nonrelativistic velocities, the conservation of momentum and energy gives the following expressions:

$$E_0 = \frac{1}{2} M_1 v_0^2 = \frac{1}{2} M_1 v_1^2 + \frac{1}{2} M_2 v_2^2 \tag{5.4a}$$

$$M_1 v_0 = M_1 v_1 \cos\theta + M_2 v_2 \cos\varphi \tag{5.4b}$$

$$0 = M_1 v_1 \sin\theta - M_2 v_2 \sin\varphi \tag{5.4c}$$

These equations can, among other things, be used to derive a relation between the incoming particle energy, E_0, which can be controlled in the experiment, and the energy of the backscattered particle, E_1, which is measured:

$$E_1 = E_0 \left(\frac{\sqrt{M_2^2 - M_1^2 \sin^2\theta} + M_1 \cos\theta}{M_1 + M_2} \right)^2 = K E_0 \tag{5.5}$$

For the above expression it is also assumed that $M_1 < M_2$. The proportionality constant $K = K(M_1, M_2, \theta)$ is the kinematic factor, mentioned above. Since the mass of the incident particle, M_1, is known and the scattering angle, θ, is where we place our detector (also known), the mass of the target particles, M_2, can readily be
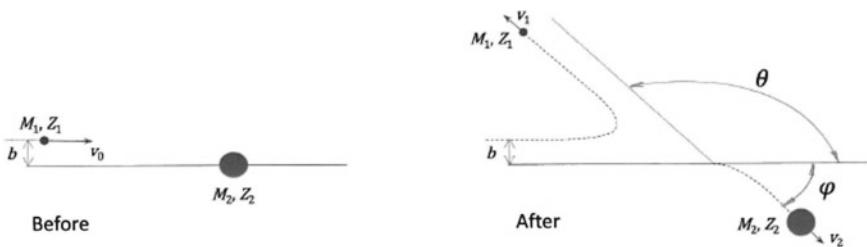
Fig. 5.4 A two-body collision with a central repulsive Coulombic force between a charged particle M_1, moving at an impact parameter b towards a stationary charged massive particle, M_2. The figure shows the situation before and after the scattering event, in which momentum and energy are conserved

determined. Notice also that for $\theta = 180^o$ the energy of the backscattered particle has its lowest value

$$E_1 = E_0 \left(\frac{M_2 - M_1}{M_2 + M_1} \right)^2 \tag{5.6}$$

Indeed, as M_1 approaches the mass of the target, M_2, all the energy will be absorbed by the target nuclei and no particles will scatter backwards.

As an example, consider a very thin foil consisting of two elements of, for instance silver and gold, with masses $M_{Ag} = 108$ u and $M_{Au} = 197$ u which is deposited on a Si substrate. The Au–Ag-film is about 20 nm thick and contains 50% Au and 50% Ag. The RBS spectrum measured at a backscattering angle of 170° from such an experiment, using 2.0 MeV He ions is shown in Fig. 5.5.

We see that the two different masses of Au and Ag can be clearly resolved and He ions backscattered from gold obtain a larger energy than He scattered from silver. The kinematic factor for He backscattered from Au at 170° is 0.9225, resulting in a maximum energy of 1845 keV, as can be seen in the figure. If we did not know the elements in this film, we could estimate their masses from the kinematic factor with known values of M_1, θ, E_0 and E_1. The yield of backscattered particles counted in the detector is shown on the ordinate (y-axis). Later we will see that the yield of an element is related to the stoichiometry of the elements in the film.
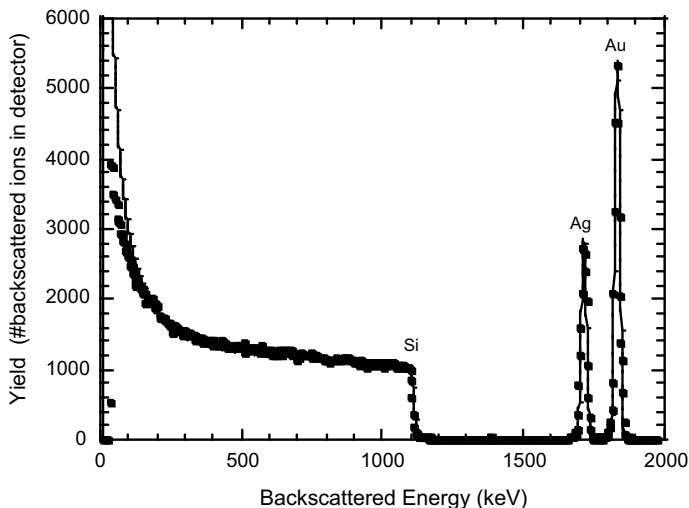
Fig. 5.5 RBS backscattering spectrum from a thin foil containing silver and gold atoms. The energy of the incident He ions ($M_1 = 4$ u) is 2.0 MeV and the scattering angle θ is 170°. The dots are measured data and the line is calculated with the SIMNRA software [6]

5.3.2 Electronic Energy Loss

As stated above, the electronic energy loss, $(dE/dx)_{electr}$, gradually reduces the kinetic energy of an incident particle. For instance, according to Fig. 5.2, a 2.0 MeV He ion will lose about 25 eV per nm in a silicon matrix. In a thick sample, this means that, if the incident particle of energy E_0 experiences an elastic collision with a target particle at some depth t into the target, its energy will be lower than E_0. However, since the values of electronic and nuclear stopping are nowadays well established and incorporated in simulation tools we can used this information to obtain the depth position for the backscattering collision. The energy of the incident particle just before the collision will be

$$E_0 - \int_0^t \left(\frac{dE}{dx}\right)_{electr} dx, \tag{5.7}$$

where the integral gives the electronic energy loss of the incident particle on its way in the material until the collision at depth t. Since the $(dE/dx)_{elect}$ varies linearly within some hundred nm from the surface, a range typically accessible by RBS, the lost energy can be written as:

$$\int_0^t \left(\frac{dE}{dx}\right)_{electr} dx \approx t \left(\frac{dE}{dx}\right)_{electr}^{in} \tag{5.8}$$

Here $(dE/dx)_{electr}^{in}$ is an average energy loss during the distance t. The energy of the incident particle immediately after a collision with a target particle will then, according to Eq. 5.5, be:

$$E_t = K \left[E_0 - t \left(\frac{dE}{dx}\right)_{electr}^{in} \right] \tag{5.9}$$

In the collision, the particle is scattered backwards at an angle θ and will then return towards the surface, giving up additional energy to the target in electronic processes. The final energy of the particle, as it emerges from the surface and can be detected, will be:

$$E_{final} = K \left[E_0 - t \left(\frac{dE}{dx}\right)_{electr}^{in} \right] - \frac{t}{\cos\theta} \left(\frac{dE}{dx}\right)_{electr}^{out} \tag{5.10}$$

Here $(dE/dx)_{electr}^{out}$ is an average energy loss for the backscattered particle during the distance $t/\cos\theta$ on its way out of the material. Since the backscattering events are relatively rare, we can assume that the ion coming back out of the target has only

experienced one collision. In an amorphous target the collisions will occur randomly at any depth, although there is an increased probability for nuclear scattering as the ion slows down deeper in the sample, as can be seen in Fig. 5.2. This means that the backscatter spectra, i.e., the number of backscattered particles as a function of their energy, for a thicker target will have a continuous distribution that increase for lower energy. This is also seen in Fig. 5.5 for the case of 2.0 MeV alpha particles $\left(^{4}_{2}\text{He}\right)$ incident on a thick silicon $\left(^{28}_{14}\text{Si}\right)$ sample with a thin Au/Ag-layer on top. The highest energy of the particles backscattered from Si atoms just under the film is given by Eq. 5.10 above. The average energy loss for the ion on the way through the thin Au/Ag-film both on the way in and on the way out amounts to about 15 keV, so the E_{final} for these He ions penetrating the thin film and hitting the topmost Si atoms is given by (neglecting Si isotopes with other mass than 28 u):

$$E_{final} = 0.5645(2.000 - 0.015) - 0.015 MeV \approx 1.11 MeV \qquad (5.11)$$

This value can be confirmed in Fig. 5.5. The abscissa (x-axis) in Fig. 5.5 can then be converted to a depth scale (x) in the Si sample, starting at the Si surface and moving to the left in the figure, deeper into the sample for lower energies. Depending on the experimental conditions, RBS thus makes it possible to resolve depths on a nm scale and, with the typical energies used, reach over 1 μ m deep in targets.

The situation is somewhat more complicated when the sample consists of several elements, a compound. However, the close encounters between the incident particle and target atoms occur at random and the total electronic energy loss is the sum of all losses in individual collisions. This means that the total loss will be the sum of the losses for each element weighted by the atomic abundance of each element. This is referred to as Bragg's rule and states that, for instance, for a binary compound consisting of the two elements A and B with masses M_A and M_B, present in the compound with the fraction m and n, respectively, the stopping cross section will be:

$$\varepsilon^{A_m B_N} = m\varepsilon^A + n\varepsilon^B \qquad (5.12)$$

where the individual stopping cross sections are given by ε^A and ε^B. Generally, the dominating electronic stopping cross section is defined by the relation $\varepsilon = \frac{1}{N}(dE/dx)_{electr}$, where, N is the volume density of atoms in an elemental compound. Finally, it should be mentioned that Bragg's rule is not accurate in several cases, for instance if there is a large mass difference between the constituents as in many hydrates. Also, for carbides, oxides and nitrides, Bragg's rule may be inaccurate and instead, experimental stopping values should be used for better accuracy.

5.3.3   Rutherford Cross Section

A calculation of the Rutherford cross section as a function of energy and scattering angle is fairly tedious and involves a transformation from a laboratory system to a

centre of mass system and back again. Here we will just state the final answer and refer to [3] for a classical and [7] for a quantum mechanical derivation. The average differential scattering cross section around an angle θ for low mass ions, such as protons $(^1_1 H)$ and alpha particles, with typical RBS energies, 1–3 MeV, is given by:

$$\sigma(E,\theta) = \left(\frac{Z_1 Z_2 e^2}{4E}\right)^2 \frac{4}{\sin^4\theta} \left(\frac{\sqrt{1 - \left(\frac{M_1}{M_2}\sin\theta\right)^2} + \cos\theta}{\sqrt{1 - \left(\frac{M_1}{M_2}\sin\theta\right)^2}}\right)^2 \tag{5.13}$$

A practically useful formula can be extracted from this expression for the case $M_1 \ll M_2$ and large scattering angles. If the energy E of the incident ion is given in MeV, the approximate cross sections, given in the unit barn per steradian (b/sr, 1 b $= 10^{-24}$ cm^2), is given by:

$$\sigma(E,\theta) \approx 0.021 \left(\frac{Z_1 Z_2}{4E}\right)^2 \left[\left(\frac{1}{\sin^4\frac{\theta}{2}}\right) - 2\left(\frac{M_1}{M_2}\right)^2\right] \tag{5.14}$$

Notice particularly that the cross section will increase as $1/E^2$ for decreasing energy and also increase with the square of the nuclear charge of the target particles Z_2.

As was shown in Eq. 5.14, the cross section is defined as a measurable quantity relating the number of scattered particles in a certain angle with the number of incident particles. The possibility to also analytically calculate the cross sections makes Rutherford backscattering analysis a truly quantitative technique.

A final example will demonstrate how to make use of the results demonstrated under sections (a), (b) and (c) to interpret the RBS spectrum shown in Fig. 5.6. This spectrum shows backscattering yield of 2.0 MeV alpha particles from a thin film of nickel silicide (Ni$_2$Si) on top of a silicon substrate.

The silicide film consists of two thirds Ni and one third of Si atoms and has a certain thickness, as can be seen from the width of the Ni peak, where the yield reaches a saturated value and the top is flat. Since Ni is heavier than Si, He ions backscattered from Ni will have a larger energy than those scattered from Si. Since the nickel is only present in the film, the spectrum will show a Ni-peak well separated from the silicon part, where the ions with lowest energy comes from He ions colliding with Ni atoms at the interface to the Si substrate. If the mass density of the film is known, it is easy to convert the energy to a depth scale and determine the film thickness. The remaining part of the spectrum comes from He ions backscattered from Si atoms and there is a step-like increase of the yield (y-scale) visible around an energy of 1.04 MeV. This step marks where the film ends and the substrate begins and here the density of silicon atoms changes from 33 to 100 atomic percent and therefore the yield suddenly increases with a factor of three for larger depths (lower energies). By comparing the relative yield of the Ni peak and the Si in the film, and also considering
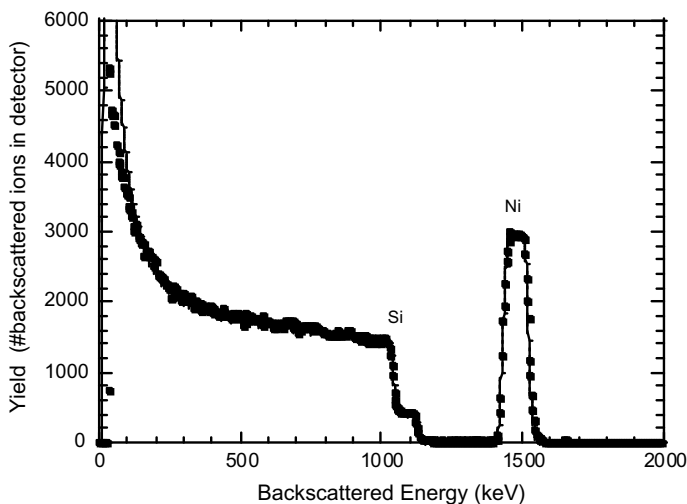
Fig. 5.6 RBS backscattering spectrum from a thin nickel silicide film formed on top of a silicon sample. The probing ions are 2 MeV He and a scattering angle θ of 170° is used. The spectrum is also calculated with the SIMNRA software [6]

that the yield scales as $(Z_2)^2$, we can confirm that the stoichiometry of the film is correct.

5.4 Running RBS

In the final part of this chapter the tools required and some practical considerations for RBS are discussed. It is also described how samples with ordered arrays of atoms, such as single crystalline materials where the collision events will no longer occur at random, can be used in RBS analysis. This makes it possible to explore a variety of RBS called channeling RBS, or C-RBS, which increase the sensitivity and makes it possible to also obtain structural information about the sample.

5.4.1 Equipment

To perform RBS analysis, one needs an accelerator to produce a beam of monoenergetic low mass ions. Although smaller dedicated systems are commercially available, these facilities are typically found at research institutes, or universities. Apart from the accelerator, the equipment needed for RBS is relatively simple. The sample is loaded in a vacuum chamber in front of the beam and a detector is mounted in a backscattering position, as shown in Fig. 5.7.
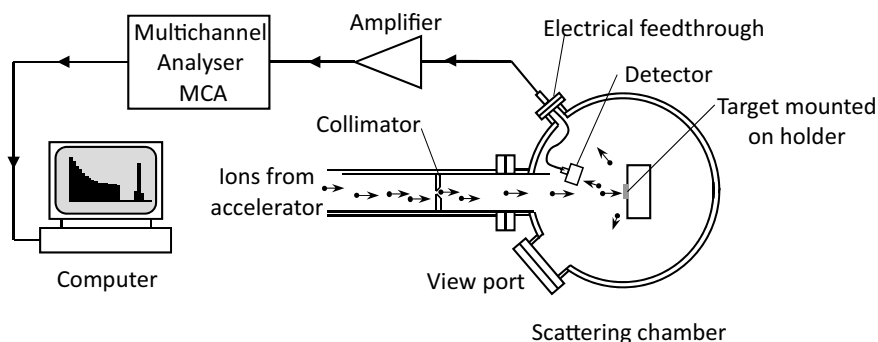
Fig. 5.7 Schematic view of the experimental set-up needed for RBS analysis. The beam of low mass ions is typically provided by a linear electrostatic accelerator and is transported in a vacuum tube to the target chamber. Along the beamline ion optical elements are introduced to steer, focus and collimate the beam so it has a typical spot size of 1 mm at the target. The target is placed in the middle of the chamber and the detector is mounted in a backscatter position. The pulses from ions are amplified and fed into a multi-channel analyser (MCA) and then displayed in the computer

Typical detectors are so called surface barrier detectors, that consist of a reverse biased semiconductor junction. Backscattered ions are stopped in this device and produce an amount of ionization through electronic stopping events proportional to the particle energy. Since an electric field is present in the reverse biased detector, the produced charge from individual ions can be collected and converted to an electrical pulse where the pulse height will be proportional to the incident particle energy.

Surface barrier detectors provide an energy resolution of 10–15 keV, which is sufficient for most RBS investigations. If better energy resolution is required, many other types of detector systems can be used, but the simplicity of the surface barrier detector makes it very popular. The pulses of each backscattered He ion that is stopped in the detector is pre-amplified and then fed into the electronic data acquisition system consisting of a multi-channel analyzer (MCA) to sort the pules according to their energy. The rate of the pulses can be tuned by the current of the incoming beam so individual pulses can be resolved in time, but still be fast enough to make the time for reaching a desired statistical certainty reasonable. As an example, a rate of incident He ions of about 10^{10} s^{-1}, corresponding to a He beam current of 1–2 nA, typically produces a count rate of a few thousand pulses per second. The rate will, of course, increase with larger detector area and also increases proportional to $Z_2{}^2$ of the target atoms, as seen by Eq. 5.14, but a spectrum with sufficient statistics can typically be reached within minutes.

The information given by the pulses, where each pulse gives the energy for one backscattering collision in the target, is then structured by the electronics and software to yield the typical RBS spectrum, where the energy axis (abscissa) is divided into a large number of energy intervals, or "channels". The number of pulses with a specific energy is then accumulated in the respective channel and the y-axis (ordinate) gives the yield. With proper control of the experimental conditions, it is possible to

compare the results quantitatively with calculations, however, for practical reasons, it is customary to use well known samples for calibration of the x- and y-scales.

5.4.2 Channeling

The general discussion has so far neglected the structure of the samples and it has been assumed that the samples are amorphous or, in case of crystalline samples, the experimental situations is such that the incident beam directions is not aligned with any major crystal direction, so called random directions. This means that the probability for close encounters resulting in backscattering events depends on the areal density of target atoms, which is isotropic. However, for crystalline samples it is possible that the incident beam direction aligns with one of the major crystalline axes or planes. Figure 5.8 shows a silicon model tilted in such a way that the [100] direction is pointing out of the page. Obviously, the silicon atoms, represented by the grey spheres, are not randomly distributed; there are clearly areas that are much less densely packed. Incident ions entering in these "channels" are furthermore steered along the axis in gentle, low angle collisions with the target atoms, drastically lowering the probability for large angle scattering events [8, 9]. This has the effect that the probing ions to a large extend will continue down the channels to large depths and very few will be backscattered. Typically, just a few percent are backscattered for a well aligned sample, as compared to a randomly oriented one.

The channeling effect is a beautiful illustration of the elastic collisions experienced by the incident MeV ions and the ordered target atoms, and can also be utilized to obtain structural information about crystalline samples that is not accessible by standard, random, RBS. Since the backscattering yield for probing ions aligned with a major crystal direction is drastically reduced, the sensitivity of RBS can be greatly enhanced. Channeling RBS also allows for investigations of interstitially placed atoms, impurities or misplaced lattice atoms, that will block the channels.

Fig. 5.8 A silicon crystal viewed in the $[\bar{1}00]$ direction. The axial $[\bar{1}00]$ channel is seen in the middle and various planar channels crossing at this point can also be distinguished
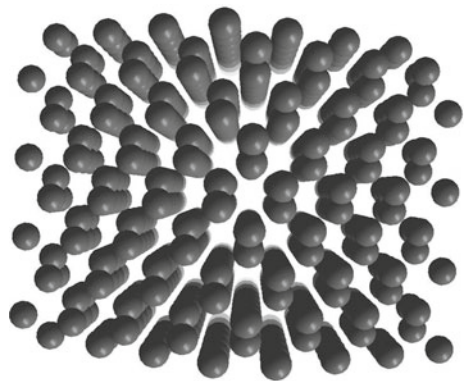
Figure 5.9 shows random and channeled backscattering spectra from 2.0 MeV He on 4H-SiC (hexagonal silicon carbide of the 4H polytype), which is a so-called wide bandgap semiconductor material. For the channeled spectrum the beam was aligned with the principle $[000\overline{1}]$ direction of the hexagonal structure which results in a drastic decrease in yield. Only around 2% of the incident He ions are backscattered. The rest of the incident ions are steered down into the crystal too larger depths from which they will not return. One can also notice that the first few atomic layers are not as well aligned with the rest of the crystal since the yield from the outermost layers in the channeled spectrum give considerably higher yield, seen as two bumps just below the steps at the carbon and silicon edges. This tells us that the topmost 1–2 nm of the material is disordered and the carbon and silicon atoms in this layer are more randomly oriented than the bulk crystal. This surface feature is not observable in the non-aligned spectrum and by using the channeling effect, one can enhance the sensitivity of the RBS technique (Fig. 5.10).
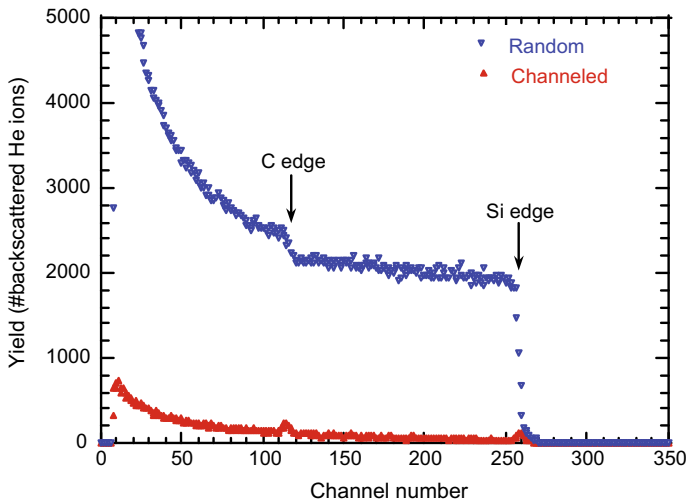
Fig. 5.9 Non-aligned (random) and aligned (channeled) backscattering spectra of 2 MeV He ions on single crystal silicon carbide (SiC). The x-axis is in units of channel number, which is directly proportional to the backscattered energy. In the random spectrum one can see the steps from backscattering at the heavier silicon atoms at the surface around channel 260 and the additional contribution from backscattering at surface carbon atoms at channel 115
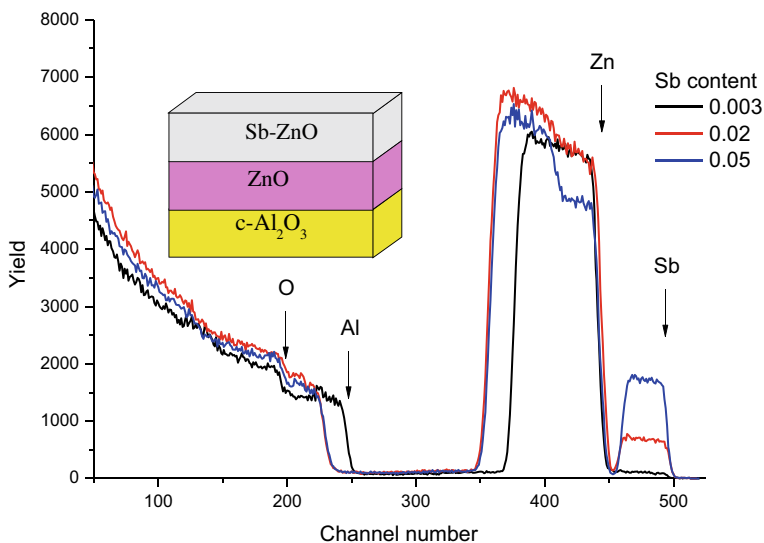
Fig. 5.10 RBS data from 2.0 MeV ^{4}He ions backscattered at 165° from three samples of MBE grown ZnO films. The top 150 nm of the ZnO is doped with Sb at three different concentration that could be accurately determined down to a concentration 0.3 atomic %. Zn and Sb are clearly resolved and the thickness of these layers could also be confirmed. (Unpublished data, courtesy of A. Azarov.)

5.5   Application Example

5.5.1   Composition and Depth Distribution

A typical example of the usefulness of RBS is given in Fig. 5.10. The figure shows RBS results from an epitaxial ZnO layer grown by molecular beam epitaxy (MBE) on a crystalline Al_2O_3-substrate (sapphire), where the top part of the ZnO-film is doped by antimony (Sb), an atom considerably heavier than Zn (and O). The thickness of the Sb-ZnO layer is 150 nm, while the undoped ZnO layer is 200 nm, except the sample with only 0.3% Sb in the top layer, which has about 100 nm undoped ZnO. The structure is shown schematically in the inset.

He-ions backscattered from heavy Sb atoms at the surface of the structure receives the highest energy and shows up to the far right in the spectra. Since the thickness of this Sb-doped layer is only 150 nm, the He-ions backscattered from Sb atoms close to the interface to the undoped ZnO layer will still have higher energy than He-ions backscattered from surface Zn and the two distributions, Sb and Zn, will be clearly separated. The Sb concentration in the 3 different samples can be deduced from the yield of Sb in relation to that of Zn.

5.6 Summary

Ion scattering techniques are extensively used for material characterisation, down to nano-meter depth resolution. The most often used form of this family of techniques is Rutherford backscattering spectroscopy, RBS. Here we have tried to give a brief overview of the method, also discussing the physical background and the close coupling of the technique to fundamental physics. The technique typically requires access to large accelerator facilities to produce the beam of probing ions of MeV energies, but apart from that, the technique is relatively inexpensive, fast and straight forward, and enables measurements with depth resolution down to nm, which is very useful in modern thin film technology.

References

1. Rutherford, E. (1911). The scattering of alpha and beta particles by matter and the structure of the atom. *Philosophical Magazine, 21*, 669–688.
2. Bohr, N. (1915). On the decrease of velocity of swiftly moving electrified particles on passing through matter. *Philosophical Magazine, 30*, 581–612.
3. Chu, W.-K., Mayer, J.W., & Nicolet, M.-A. (1978). *Backscattering spectrometry*. Florida: Academic Press Orlando. ISBN 0-12-173850-7.
4. Alford, T.L., Feldman, L.C., Mayer, J.M. (2007) *Fundamentals of nanoscale film analysis*. USA, Springer Science. ISBN 10: 0-387-29260-8.
5. Wang, Y., & Nastasi, M. (Eds.) (2009). *Handbook of modern ion beam material analysis*. Warrendale: Material Research Society. ISBN 978-1-60511-216-9.
6. Mayer, M. (1997). SIMNRA User's Guide, Report IPP 9/113, Max-Planck-Institut für Plasmaphysik, Garching, Germany.
7. Johnson, R. E. (1982). Introduction to atomic and molecular collisions. New York: Plenum Press. ISBN 0-306-40787-6.
8. Gemmell, D. S. (1974). Channeling and related effects in the motion of charged particles through crystals. *Reviews of Modern Physics, 46*, 129.
9. Morgan, D.V. (Ed.). (1973). *Channeling: theory, observations and applications*. London: Wiley. ISBN 0-471-61510-2.

Chapter 6
Secondary Ion Mass Spectrometry (SIMS)

6.1 Introduction

Even very small amounts of impurities can dramatically affect almost all physical properties (including optical, magnetic and electrical properties) of semiconductor materials. Therefore, a precise determination of impurity concentrations as well as their depth distributions is vitally important for design and fabrication of modern electronic devices, especially taking into account continuous demands for scaling and integration of device components. Secondary ion mass spectrometry (SIMS) is a one of the most sensitive techniques for determination of concentration of dopants and impurities with a detection limit even better than 10^{12} atoms/cm^3 for some cases.

Historically, the first mention of observation of secondary charged particles emitted from the surface under ion bombardment was done already in 1910 by Thompson [1]. The next important step was related to the development of ion sources for mass spectrometry in the mid of the last century [2] and one of the first SIMS instruments was then developed only in 1960s under a NASA contract for investigation of moon rocks [3]. The results obtained were much better than expected and the great potential of SIMS became evident for the rapidly expanding semiconductor industry at that time. Very soon, in the late 1960s, CAMECA (France) delivered one of the first commercial SIMS instruments and after that, SIMS becomes one of the most frequently used techniques in chemical, metallurgical, and semiconductor industry, as well as analytical technique for basic research in physics, chemistry and geology. Today, SIMS instruments have even been developed also for biochemistry and medical applications. Some of the main application areas of SIMS usage today are impurity analysis, dopant profiling, diffusion studies and analysis of interface/ surface contaminations. At present there are several different types of SIMS instruments varying from the primary ion source to the type of mass spectrometer used for the analysis of secondary ions.

6.2   Basics

Figure 6.1 illustrates schematically the ion-surface interactions which form the basis
of SIMS measurements. A focused primary beam is rastered over a small area (typi-
cally a few tens, or hundreds of μm) and interacts with surface atoms, leading to
the erosion of the surface through sputtering of the target atoms. A small fraction of
the sputtered particles (monoatomic particles, or molecules) are ionized (positively,
or negatively) and, therefore, these secondary ions can be accelerated by applying
appropriate electrical potential and further be directed to a mass spectrometer and
detector systems. Therefore, ion sputtering and ionization yield are crucial parame-
ters determining detection efficiency of impurities. The two fundamental processes
sputtering and ionization will be discussed below in detail. It should be noted that
SIMS is considered as a destructive technique since erosion of the surface occurs
during the measurements.
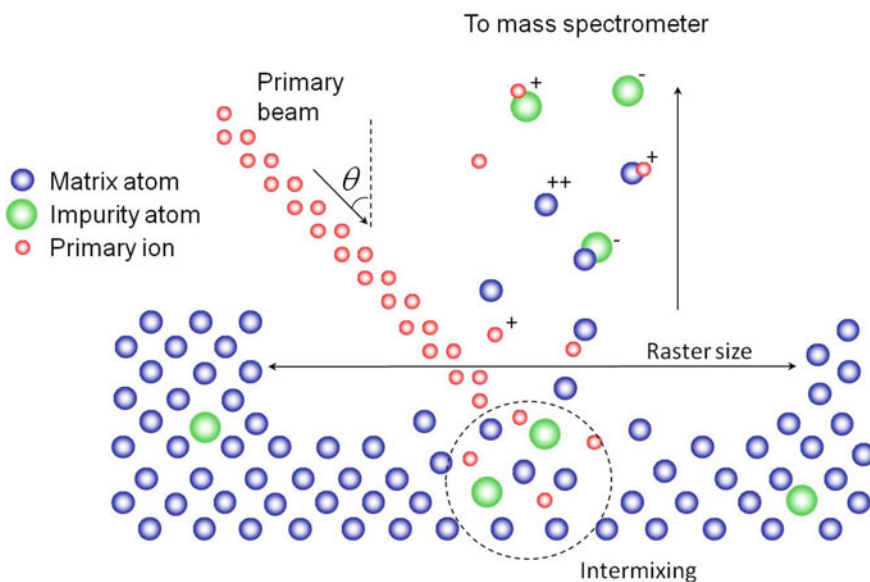
Fig. 6.1 Schematics of interactions between a primary beam and a sample containing impurity
atoms. The elastic collisions between the primary beam particles and target atoms leads to sputtering
of surface atoms and a crater is formed, as the primary beam is scanned across a certain area of the
sample. Charged sputtered particles leaving the surface can be accelerated and further detected by
a mass spectrometer. The dashed circle shows the region where ion beam induced mixing of atoms
occurs (2–10 nm)

6.2.1 Sputtering

The basics of sputtering theory was developed already in the middle of the past century (see, for example, the work by Sigmund [4]). In general, a energetic projectile, for instance an ion, impinging on a target loses its energy due to interaction with target electrons resulting in ionization, excitation and charge transfer (electronic energy loss) and elastic collisions with target atoms (nuclear energy loss). The latter one is the dominant process for the energy range typically used for SIMS measurements (see also Chap. 5). Three qualitatively different cases can be distinguished, depending on the energy and momentum transferred in the collisions:

(i) Single knock-on regime. In this regime recoil atoms from ion-target collisions receive sufficiently high energy to be sputtered, but not enough to generate successive recoils, involving many target atoms. All secondary ions ejected from the surface originate from the uppermost atomic layers. This regime occurs for incident light elements having relatively low energy (< 1 keV)

(ii) Linear cascade regime, where recoil atoms from ion-target collisions receive sufficiently high energy to generate recoil cascades. The density of recoiling atoms is sufficiently low so that knock-on collisions dominate and collisions between moving atoms are infrequent. In this regime sputtering yield is proportional to beam energy. This regime occurs in keV—MeV energy range for almost all ions, except the heaviest ones, which can generate spikes (see below)

(iii) Spike regime characterized by a very high density of displacements. The density of recoil atoms is so high that the majority of atoms within a certain volume (the spike volume) are in motion.

Figure 6.2 illustrates sputtering of the atoms labeled as 2 and 4 due to series of collisions in a linear collision cascade (regime ii). Sputtering yield (Y) is defined as the number of sputtered atoms per incident ion. For the linear collision cascade regime, Y is a function of the energy of the incident ion (E_0) and the angle between the incident beam direction and the surface normal (θ) (see Fig. 6.1). Within a good accuracy, Y can be approximated as:

$$Y(E_0, \theta) = \frac{K_{it}}{U_0} S_n(\xi) f(\theta) \tag{6.1}$$

where $S_n(\xi)$ is the nuclear stopping cross section, which is a universal function of the reduced energy ($\xi = E_0/E_{it}$), while $f(\theta)$ is an angular dependence. $S_n(\xi)$ can be written as:

$$S_n(\xi) = \frac{0.5 \ln(1 + \xi)}{\xi + (\xi/383)^{3/8}} \tag{6.2}$$

and E_{it} and K_{it} depend on the mass ratio of incident ion and target atom (M_1/M_2) as well as the ratio of their atomic numbers (Z_1/Z_2):

Fig. 6.2 Schematics illustrating sputtering of the atoms 2 and 4 due to series of collisions in linear collision cascade
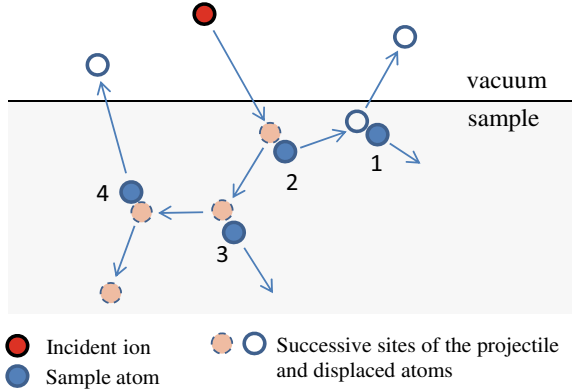
$$E_{\text{it}} = \frac{1}{32.5}\left(1 + \frac{M_1}{M_2}\right) Z_1 Z_2 \left(Z_1^{2/3} + Z_2^{2/3}\right)^{1/2} \quad [\text{keV}] \tag{6.3}$$

$$K_{\text{it}} \approx \frac{1}{3}(Z_1 Z_2)^{5/6} \text{ for } 0.05 \leq Z_2/Z_1 \leq 5 \tag{6.4}$$

The angular dependence is not very easily predicted, but can be approximated as $f(\theta) = \cos^n(\theta)$, where $n \approx 1$–2, and, therefore, sputtering yield reaches a maximum for angles between 60 and 80 degrees and drops down to zero at 90° (glancing incidence). Finally, U_0 in Eq. 6.1 is a surface escape barrier in eV, usually taken equal to the sublimation energy.

Figure 6.3 illustrates that Eq. 6.1 adequately predicts sputtering yield of elemental target, such as Si, for different ions having various energies. It is also seen that that the typical sputtering yield is 1–10 for normal incidence, depending on the primary beam.

Fig. 6.3 Energy dependence of Si sputtering yield for different primary ions at normal incident direction ($\theta = 0$), where solid lines represent Eq. 6.1 [5]
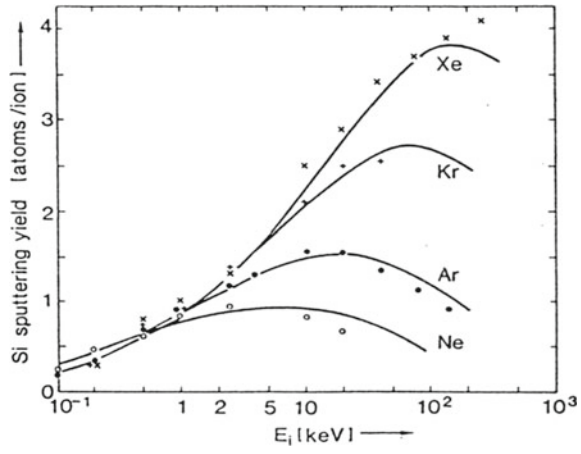
Fig. 6.4 Normalized erosion rate as a function of Ge content in $Si_{1-x}Ge_x$ alloy for different O_2^+ primary beam energies [6]
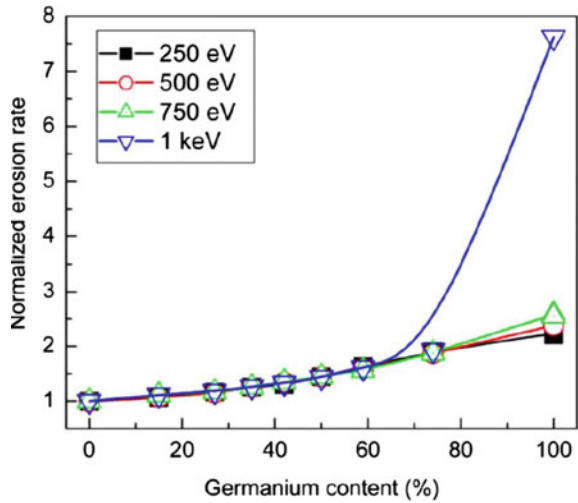
It should be noted that prediction of sputtering yield for multicomponent materials becomes challenging. Indeed, sputtering yield strongly depends on target composition as clearly illustrated by Fig. 6.4, showing normalized erosion rate in $Si_{1-x}Ge_x$ alloys as a function of Ge content. The incident primary ion was O_2^+ and it is seen that erosion rate of Ge is higher than that for Si by a factor of 2, or even higher, depending on the primary beam energy and it is not a linear function of x for $Si_{1-x}Ge_x$ alloys. Furthermore, sputtering of compound material can result in the non-stoichiometric removal of surface atoms leading to changes in surface composition depending on ion type, energy, angle of incidence and even target temperature.

Additional complications arise for polycrystalline samples. Indeed, it is well known that sputtering yield is different for materials in amorphous and crystalline states. Furthermore, in the latter case it depends also on crystal orientation and sputtering yield can vary by a factor of 2 for different grain orientations [7]. This can lead to the pronounced surface roughening during the measurements which, in its turn, may affect accuracy and depth resolution of the SIMS analysis. As an example, Fig. 6.5 shows a crater after SIMS measurements on a polycrystalline iron sample and a dramatic roughness of the crater bottom is clearly visible.

6.2.2 Ionization

In contrast to the sputtering process, there is no complete theory adequately describing ionization of sputtered atoms. In general, ion yield, or ionization efficiency (γ_i) is defined as the fraction of sputtered atoms that become ionized. Typically, this fraction is very low (~10^{-6} ions/primary ion), but can vary by several orders

Fig. 6.5 SIMS crater on a polycrystalline Fe sample measured with a stylus profilometer
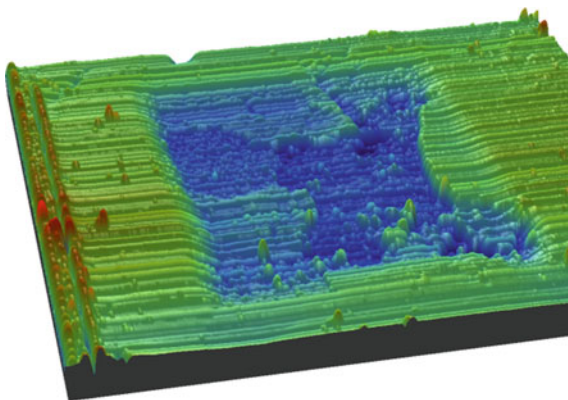
of magnitude depending on the element, the matrix and the chemistry of a sputtered surface.

The low ionization efficiency in many cases can often be remedied by choosing an appropriate primary beam. For example, positive ionization can be enhanced by using an electronegative ion, such as oxygen, for the primary ion beam, while an electropositive primary ion, e.g. cesium, enhances the negative ionization. Indeed, implantation of Cs^+ primary ions at the sample surface reduces the sample work function (due to the low Cs electronegativity) and emission of secondary electrons increases. These electrons can be captured by sputtered atoms (or clusters) with high electron affinity, leading to the generation of negative ions. On the other hand, O_2^+ ions may lead to the formation of metal–oxygen bonds at the surface of metal samples. During sputtering, such bonds can be broken, thus favoring the generation of positively charged metal ions. This is the reason why most SIMS instruments are equipped with ion sources for both oxygen and cesium primary beams.

Two important parameters affecting ionization efficiency are ionization potential (E_i) for positive ions and electron affinity (A) for negative ones. In general, positive ion yield is proportional to $\exp\{C^+(E_i - \varphi)/v\}$, while negative ion yield $\sim \exp\{C^-(\varphi - A)/v\}$, where $C^{+/-}$ are constants, φ is the surface work function and v is the normal component of the ion's emission velocity [8]. Taking into account the above mentioned dependencies, preferential analysis conditions for different elements are shown in Fig. 6.6. It is seen that most of the metals can be efficiently detected using oxygen primary beam, while hydrogen and the right part of the periodic table typically requires Cs. Noble gases can be potentially detected, however, their ionization yield and, therefore, detection efficiency is very low.

In general, for a given element i, the measured secondary intensity (I_i) can be expressed as:

$$I_i = J_P A Y [C_i] \gamma_i T \tag{6.5}$$
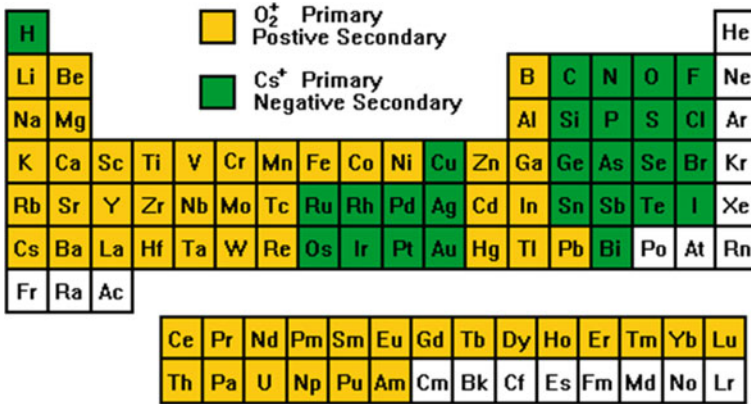
Fig. 6.6 Periodic table of elements illustrating the choice of the primary beam (Oxygen vs. Cesium) for the best detection efficiency [9]

where J_p is a primary ion current density, Y is a sputtering yield, A is an analyzed area, $[C_i]$ is a concentration of species i, γ_i is an ionization efficiency and T is the instrument transmission function, which expresses the amount of the secondary ions that goes through the mass spectrometer all the way to the detector. However, in contrast to all the other parameters, the value of γ_i is generally not known and it strongly depends on ion species and local electronic environment at the origin of ejection (matrix effect). Therefore, in practice Eq. 6.5 cannot be used to convert measured secondary intensities to atomic concentrations. Instead, typically a standard sample with known content of the impurity is measured in addition to the sample with unknown concentration. Issues related to the calibration procedure in SIMS will be discussed more detailed in Sect. 6.4.

As mentioned above in Sect. 6.1, ionized particles are accelerated electrostatically before they reach the mass spectrometer and detector systems. Here, it is important to note that if the measured sample is not fully conductive, or even insulating, a charge build-up occurring due to primary beam exposure can dramatically affect ejection of secondary ions from the sample surface. Furthermore, for thin insulating films a strong electric field can be developed leading to the redistribution of mobile atoms. A charged surface can also lead to changes in the accelerating field for secondary ions, resulting in wrong kinetic energy. There are several approaches which can be used to reduce, or eliminate charging effects. The most commonly used methods involve the use of a conductive coating or a metal grid on the sample, using negatively charged primary ion bombardment (for example, O-) or using an electron gun for charge compensation.

6.3   Instrumentation

The basic components of a SIMS instrument include an ion source, a mass analyzer and a detector. Traditionally, all SIMS instruments can be divided by the type of mass spectrometer of which there are three basic types: quadrupole, magnetic sector and time-of-flight. Magnetic sector and quadrupole SIMS are often operated in a mode called "dynamic SIMS". In this mode, the sample surface is continuously eroded by the rastered primary ion beam. In contrast, time-of-flight SIMS typically operates in a "static" mode, where sample surface is sputtered very lightly, making possible a chemical analysis of the top surface monolayer of the sample. This mode is often preferred for organic samples. In addition to the mass spectrometers, this overview will also describe ion sources and detector systems used in SIMS instruments.

6.3.1   Mass Spectrometers

Magnetic sector analyzer

A schematic of the SIMS instrument based on the magnetic sector analyzer (MSA) is shown in Fig. 6.7. The extracted secondary ions leaving the sample surface have a (non-relativistic) kinetic energy given by Eq. 6.6:

$$\frac{mv^2}{2} = qV \qquad (6.6)$$

where m is the mass of the ion, v is its velocity, q is the charge of the ion and V is the secondary ion acceleration voltage. However, the primary ions leaving the ion
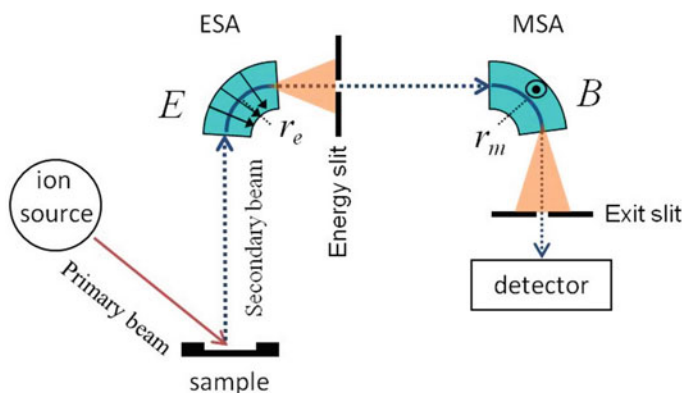
Fig. 6.7 Schematics of SIMS instrument based on the magnetic sector mass spectrometer. Primary and the secondary beam paths are shown by the solid and dashed arrows, respectively. The roles of the energy and exit slits are described in the Sect. 6.4.1

source do not all have exactly the same energy and, therefore, the same velocity and this energy spread will limit the final mass resolution of the secondary ions. This is analogous to the "chromatic aberration" in optical spectroscopy. The energy distribution of the secondary ions extracted from the biased sample can be up to hundreds of electron volts (eV). Therefore, to achieve better resolution, it is necessary to limit the energy spread of the secondary ions extracted from the sample surface. That can be done by adding an electrostatic sector analyzer (ESA).

In the ESA, the ions travel between two curved parallel plates with a radius of curvature r_e. A potential between the two plates permits only those ions with the proper energy to be transmitted. The Lorentz force in that case is equal to the centripetal force, yielding:

$$qE = \frac{mv^2}{r_e} \tag{6.7}$$

where E is an electric field. Thus, only ions having a kinetic energy matching r_e will be curved to pass through the ESA-exit slit and proceed towards the MSA. Using the same formalism for ions travelling through the MSA we obtain the expression:

$$qvB = \frac{mv^2}{r_m} \tag{6.8}$$

where r_m is the radius of curvature of the MSA and B is a magnetic field strength perpendicular to the ions path. Assuming that no energy is lost when the ion travels towards the MSA, i.e. v is the same in the Eqs. 6.7 and 6.8, a combination of these two equations results in:

$$\frac{m}{q} = \frac{(r_m B)^2}{r_e E} \tag{6.9}$$

Thus, only ions with a certain m/q ratio will satisfy this equation and pass through the MSA-exit slit and finally hit the detector. The common operational mode of a magnetic sector mass spectrometer keeps the electrostatic sector at a constant potential and varies the magnetic field strength to yield a desired mass-to-charge ratio (m/q).

The main advantages of magnetic sector SIMS are the very good detection limit, high mass resolution and high dynamic range. However, simultaneous detection of many elements is not possible since the mass switching speed is limited. This makes the time to acquire a spectra quite long. In addition, this type of SIMS instruments is usually larger and cost more than other mass analyzers.

Time-of flight analyzer

Another widely used mass spectrometer is a time-of-flight system and Fig. 6.8 illustrates a schematic of ToF–SIMS instrument. For ToF–SIMS analysis, a solid sample surface is bombarded with a pulsed primary ion beam. Secondary ions, emitted from
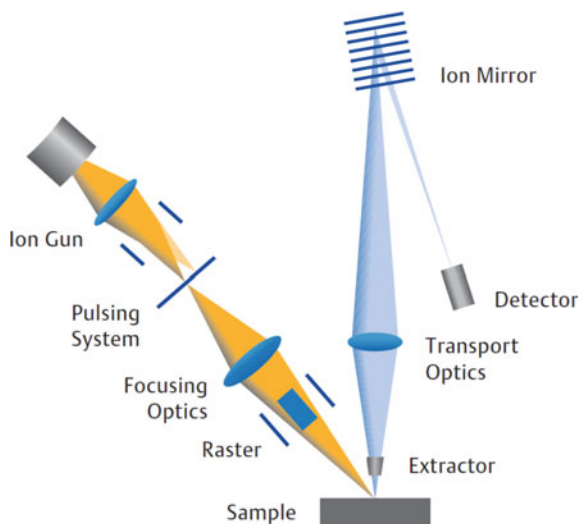
the outer layers of the surface, are extracted and analyzed. This analysis cycle is repeated at high frequency to generate the complete mass spectrum for every pulse.

The mass separation in a ToF–SIMS instrument is based on the principle that the velocity of an ion with a given kinetic energy depends on its mass, so light masses arrive at the detector before heavier masses. This means that the ToF–SIMS provides parallel detection of all masses including isotopic and molecular species with no fixed upper limit. Molecular masses can then reach several thousand amu, where organic secondary ions consisting of several hundred atoms appear.

Since a ToF mass spectrometer measures the ion travelling time from the ion source to the detector, it requires that the starting time (the time at which ions leave the ion source) is well-defined. This is typically achieved by using pulsed primary ion beam, which is a different operation from quadrupole and magnetic sector SIMS instruments where the primary beams typically operate in continuous mode.

Using the Eq. 6.6 and taking into account that the ion velocity is the length of the flight path (L) divided by the flight time (t), the latter one can be expressed as:

$$t = L\sqrt{\frac{m}{q}\frac{1}{2V}} \tag{6.10}$$

However, again the ions leaving the ion source have neither exactly the same starting times, nor exactly the same kinetic energies. That is, the particle energies suffer from what is called "chromatic aberrations" in optical systems. One way to achieve a more "monochromatic" beam is to use an ESA, as discussed above for magnetic sector SIMS. Another way is to use an ion optical device called "reflectron" (or ion mirror) as shown in Fig. 6.8, where schematics of a ToF–SIMS instrument designed by IONTOF GmbH [10] is shown. Typical reflectrons consist of a stack of

metal plates with varying voltages to create an electric field reflecting the ions back to the flight tube. The reflectron decreases the spread in the ion flight times in the pulse of secondary ions of a given mass-to-charge ratio, and therefore improve the mass resolution of the time-of-flight mass spectrometer.

The three main advantages of ToF–SIMS are simultaneous detection of many elements, possibility of analysis of the topmost surface layer and the highest practical mass range among all SIMS analyzers. However, detection limits and dynamic range are not as good as for magnetic sector SIMS instruments. In addition there are limitations when making a depth profile, since the pulsed primary ion source gives very low sputtering.

Quadrupole mass analyzer

Finally, the third type of mass spectrometers typically used in SIMS instruments is a so called quadrupole analyzer and it is schematically shown in Fig. 6.9. A quadrupole mass analyzer consists of two pairs of conducting parallel rods. Two opposite rods have an applied potential of $(U + V\cos(\omega t))$ and the other two rods have a potential of $-(U + V\cos(\omega t))$, where U is a DC voltage and $V\cos(\omega t)$ is a radio frequency (RF) voltage. The applied voltages affect the trajectory of ions traveling between the four rods in a complex way, making them oscillate. By combining DC and RF potentials on the quadrupole rods it is possible to allow a trajectory for a selected mass-to-charge ratio to be transmitted (resonant ion in Fig. 6.9), while all other ions (non-resonant ions in Fig. 6.9) will not have a stable trajectory through the quadrupole and will collide with the quadrupole rods, or the walls of the analyzer, never reaching the detector. A mass spectrum is obtained by monitoring the ions passing through the quadrupole filter as the voltages on the rods are varied. The voltages can be varied in two ways: varying ω and holding U and V constant, or varying U and V with (U/V) fixed for a constant angular frequency, ω.
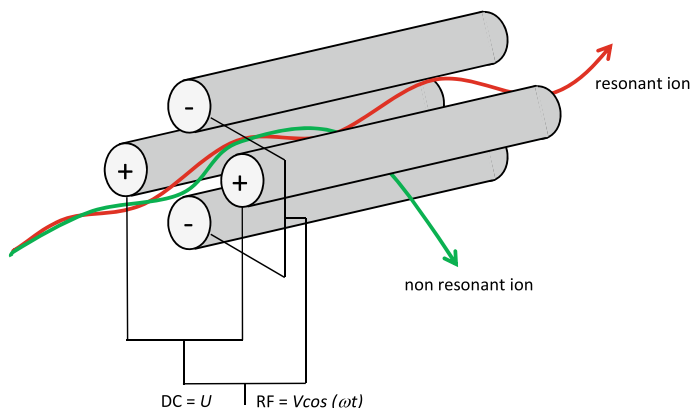
Fig. 6.9 Schematics of quadrupole mass spectrometer

Note that a quadrupole mass analyzer is a relatively small and low-cost system and this makes it suitable for fast operating benchtop SIMS systems. However, poor detection limits as well mass resolution limit the applications of quadrupole SIMS instruments.

6.3.2 Ion Sources

There are three basic types of ion sources which are typically used for SIMS instruments:

  (i)   Ion sources with electron impact ionization (duoplasmatron)
 (ii)   Ion sources with surface ionization (Cs^+ ion source)
(iii)   Ion sources with field emission (liquid metal ion source)

Duoplasmatron

The duoplasmatron was invented by Manfred von Ardenne in 1956 [11]. This ion source can operate with virtually any gas, but oxygen is the most common in SIMS applications because oxygen implantation into the sample surface enhances ionization efficiency for electropositive elements (see the Sect. 2.2). Before this oxygen enhancement effect was discovered, argon was commonly used and is still available for cases where a high sputter yield is needed. A schematic of the duoplasmatron source is shown in Fig. 6.10, where 1 is an intermediate electrode, 2 is a cathode, 3 is an anode, 4 is a plasma and 5 is an extraction electrode.

A cathode filament emits electrons into a vacuum chamber. A working gas (e.g. O_2) is introduced in very small quantities into the chamber, where it becomes charged or ionized through interactions with the free electrons from the cathode, forming a plasma. The plasma is then accelerated through a series of at least two highly charged grids, and becomes an ion beam, moving at fairly high speed from the aperture of the device. This type of ion gun is easy to operate and generates less focused, but high current ion beams.

The duoplasmatron may be used to extract either O^-, O_2^- or O_2^+ depending upon the electrical polarity selected by the operator. In negative mode O^- is the most

Fig. 6.10 Schematics of a Duoplasmatron ion source. 1-intermediate electrode, 2-cathode, 3-anode, 4-plasma, and 5-extraction electrode
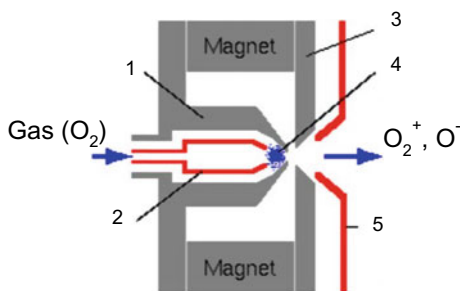
Fig. 6.11 Schematics of Cs ion source
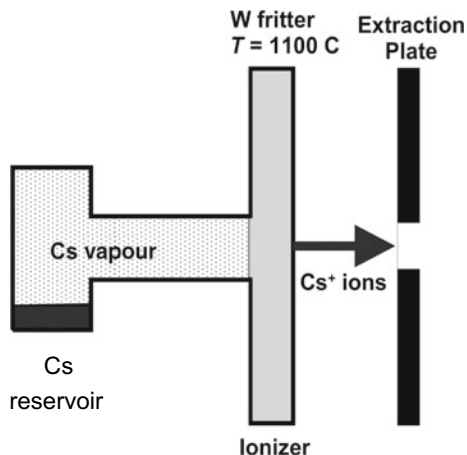
abundant species, while in positive mode O_2^+ is most abundant. When insulating samples are analyzed, O^- has the advantage of preventing charge build-up on the sample surface.

Cesium ion source

Cesium ion source is used to optimize detection sensitivity of electronegative elements (see Sect. 2.2). Probably the first suggestion to use a Cs^+ ion source to deliberately enhance the production of negative ions came from Victor Krohn (Space Technology Laboratories, Inc., California, [12]). Schematic of the Cs ion source is shown in Fig. 6.11 and it might be seen that it consists of two main elements: a reservoir, providing Cs material by evaporation and an ionizer, which is heated to ionize Cs atoms on contact. The ionizer is heated to sufficiently high temperatures to ensure that a desorption rate exceeds the arrival rate.

In general Cs beams are smaller than those generated by the duoplasmatron and sputter material more effectively due to their greater mass. However, operation of Cs ion source is more expensive as compared to that of the douplasmatron because of the high price of a solid Cs compound (Cs-chromate or Cs-carbonate) used in the reservoir.

Liquid metal ion source

Liquid metal ion source operate with metals, or metallic alloys, which are liquid at room temperature or slightly above. The most popular and reliable configuration of liquid metal ion source is shown in Fig. 6.12, where a solid needle is used as a source substrate. The needle is fabricated within the cylindrical capillary and the intermediate region holds the reservoir of the source metal. A high potential is placed on the needle to provide intense electric field at the apex having Taylor cone shape [13]. The source material is indirectly heated and ions are emitted from extremely narrow emitting area.

Fig. 6.12 Schematics of a liquid metal source
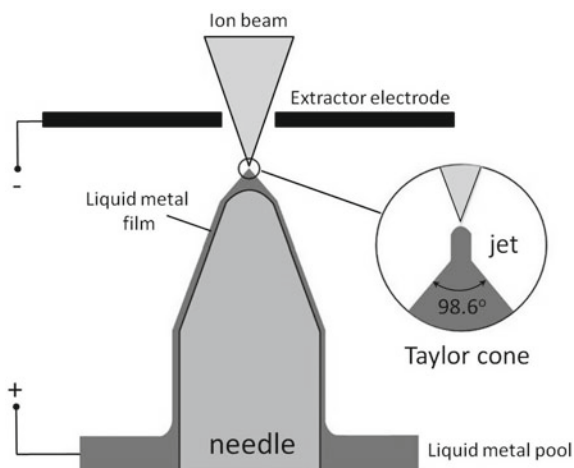
Ga is the most popular source material since it has a low melting point (28 °C), a vapor pressure below 10^{-9} mbar and it is available in a high purity form. This ion source provides a tightly focused ion beam (< 50 nm) with moderate intensity, i.e., high spatial resolution, which is important for mapping chemical elements over the specimen surface. Furthermore, this ion source can work in pulsed mode that makes it ideal for ToF–SIMS instruments.

6.3.3 Detection Systems

Most modern SIMS instruments have more than one detector. The most common secondary ion detectors are the electron multiplier, the Faraday cup, and the micro channel (image) plate. In some instruments only one detector can be used at a time, while advanced systems allow using several detectors simultaneously to give higher precision and rapid data acquisition.

Electron multiplier

The electron multiplier is the most sensitive detector and it must be protected from intense ion beams. An electron multiplier consists of a series of electrodes called dynodes, each connected along a string of resistors. When an ion strikes the first dynode it produces a few (1, 2, or 3) secondary electrons. These secondary electrons are accelerated into the next dynode where each electron produces more secondary electrons. Thus, a cascade or avalanche of secondary electrons occurs. The dynode acceleration potential controls the electron gain and most SIMS measurements use electron multipliers operating with sufficiently high gain to produce a detectable pulse for every single ion that arrives at the detector. However, the detector electronics require a recovery time (dead time) after an ion arrival before a second ion can be

detected. The detector dead time limits the measureable ion arrival rate to around 10^6 counts per second. Thus, the electron multiplier dynamic range extends from below 1 to about 10^6 ion counts per second.

Faraday cup

For detection of secondary ion beams of high intensity exceeding the electron multiplier limit, another detector called Faraday cup is usually used. A Faraday cup detector can detect count rates from 5×10^4 counts per second and upwards. A Faraday cup is just a cup-shaped electrode from which electrical current is measured as a beam of charged particles (electrons or ions) impinges on it. In order to prevent the escape of secondary electrons ejected from the Faraday cup walls by incoming ions or electrons, a negatively biased plate is placed in front of the Faraday cup. It is simple and cheap detector, although its response time is slow.

Microchannel plate

Finally, the micro channel plate consists of an array of miniature electron multipliers. Secondary ion mass spectrometry instruments typically use round arrays with about 2000 channels across a diameter of 25 mm. An ion passing down a channel hits the inner channel wall and produces secondary electrons. The channels are designed so that these secondary electrons initiate an electron cascade down the channel. The pulse of electrons from the back of the detector may either be passed to a second micro channel plate for further gain, or accelerated towards a phosphor screen, where their impact may be viewed directly. Under ideal operating conditions, a count rate of ~ 5000 counts per second should be visible on the phosphor screen.

6.4   Operation Modes and Complicating Factors

There are three main operation modes of SIMS instruments: mass spectrum recording, impurity depth profiling and imaging mode, where two or three dimensional maps of impurities present in a sample can be obtained. In this section all these operation modes will be described along with a discussion about the complicating factors.

6.4.1   Mass Spectra and Impurity Identification

Figure 6.13 shows a mass spectrum obtained from a Mg doped ZnO film grown by metal organic chemical vapor deposition (MOCVD). It shows all the singly charged zinc isotopes (masses 64, 66, 67, 68, and 70 amu) and the intensity of the Zn peaks reflects the abundance of various isotopes. It is important to remember that the intensity of the peaks between different ions cannot be compared in this spectrum,

Fig. 6.13 Mass spectrum of
a Mg doped ZnO film as
measured using an O_2^+
primary beam and a
magnetic sector mass
analyzer. Note the
logarithmic y-scale
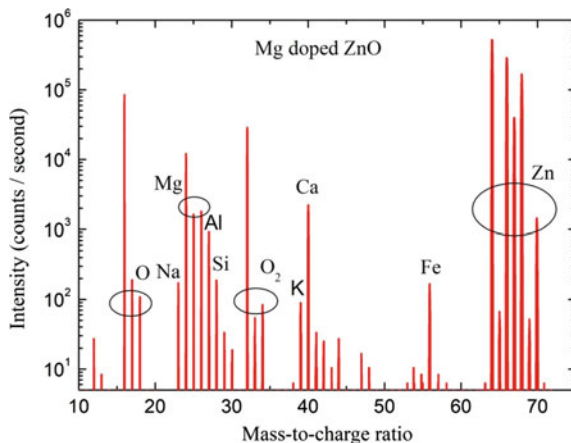
since the yield of the secondary ion species may vary several orders of magnitude.
How to convert the intensity to a true concentration is described below.

The measurements were performed with a primary beam of O_2^+, so the intense
oxygen peak at 16 amu is partly due to the sample content of O, but also the oxygen
from the primary beam. The two stable oxygen isotopes ^{17}O and ^{18}O are also clearly
resolved. Magnesium is represented by its three stable isotopes (24, 25 and 26 amu),
but many other impurities (Al, Si, K, Ca and Fe) can also be detected in the sample.

However, correct identification of impurities is often complicated by a spectral
interference from molecular, or multiply-charged ions having ion mass which is
very close to that of the analytical species to be studied. This has two important
consequences; firstly, it may lead to erroneous qualitative assignments of the mass
numbers and the molecular ion signal is attributed to the analytical species. Secondly,
although care is taken with respect to the identification the molecular interference
can seriously affect the detection limit. There are several ways to avoid, or at least
reduce mass interference issues.

First of all, a correct choice of measured isotope may be a good solution. Indeed,
many elements have more than one stable isotope and some of them can have a strong
mass interference for the given matrix. For example, the detection of $^{10}B^+$ in silicon
matrix is complicated by the mass interference with $^{30}Si^{3+}$ ions, giving the same
mass to charge ratio, even though ^{30}Si has only 3% natural abundance and multiply-
charged ion survival is rare. The evident solution is to use $^{11}B^+$ isotope instead of $^{10}B^+$
and this could give even better signal taking into account that the natural abundance
of ^{11}B is higher than that of ^{10}B by a factor of four. However, more often we have
to detect a less abundant isotope in order to avoid mass interference. For instance,
detection of such important impurity as Fe, which is a well known life-time killer in
Si, is complicated by a strong mass interference of the most abundant ^{56}Fe isotope
with a huge signal coming from $^{28}Si_2^+$ molecular ions. Therefore, to avoid mass
interference we have to measure ^{54}Fe isotope which has a natural abundance of only
6% that strongly affects the signal intensity.

Fig. 6.14 Energy distribution of $^{28}Si^+$ and $^{28}Si_4^+$ ions as measured using O_2^+ primary beam. The inset shows the ESA part of the SIMS instrument, shown in Fig. 6.7, where the direction of the energy slit shift is shown by the arrow
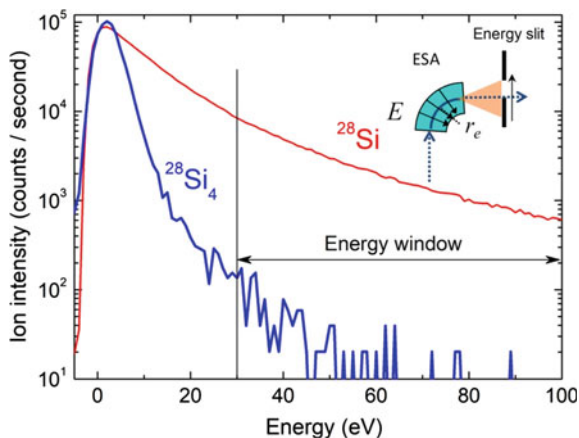
In contrast to the examples given above, another important impurity in silicon, arsenic, has only one stable isotope (^{75}As). Oxygen bombardment of silicon as well as interaction of residual oxygen in the sample chamber with the silicon surface gives rise to a molecular ion at mass 75 ($^{29}Si^{30}Si^{16}O$) and, therefore, detection of arsenic is affected by the mass interference with these molecular ions. In this case a large improvement can be obtained by applying an energy offset utilizing the fact that molecular ions have a much narrower energy distribution than atomic ions, as clearly illustrated by Fig. 6.14 showing energy distributions of $^{28}Si^+$ and $^{28}Si_4^+$ ions as measured by an oxygen primary beam. It is seen that using the shifted energy window a contribution from the large molecules can be dramatically decreased. Note that a change of the energy window is typically realized by moving the energy slits as illustrated in the inset in Fig. 6.14. Here the thin black arrow indicates the direction of the energy slit shift in order to block the low energy part of the ion beam.

However, energy discrimination method requires a substantial difference between the atomic and molecular energy distributions which is typically realized for relatively large molecules consisting of several atoms. In contrast, energy distribution of "simple" molecular ions is not so different from that of atomic ones and, for such a case, high mass resolution mode is needed. This is typically accomplished by closing the energy slits as illustrated in the inset in Fig. 6.15. For example, detection of phosphorous in silicon is complicated by the mass interference with silicon-hydrogen molecules ($^{30}Si^1H$, $^{29}Si^1H_2$, and $^{28}Si^1H_3$). In order to efficiently discriminate between ^{31}P and these molecular species the mass spectrometer must be able to differentiate between 30.98158 and 30.97376 amu ($M/\Delta M = 4000$). Figure 6.15 shows a comparison of mass spectra taken around 31 amu in low and high mass resolution mode in P doped Si sample. It might be seen that ^{31}P, $^{30}Si^1H$ and $^{29}Si^1H_2$ signals are clearly resolved in a high resolution mode; however, the intensity of the signal is much lower as compared to the low resolution one.

Finally, it should be mentioned that in some cases a mass interference can be avoided if an impurity of interest is measured as a complex with a matrix element, or
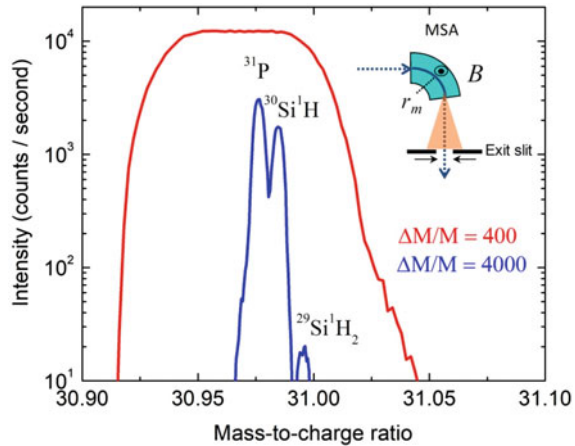
ions from the primary beam. Moreover, for some elements this "trick" can be used to increase its ionization efficiency and improve the dynamic range. For example, nitrogen has a very high ionization potential (14.53 eV) and, at the same time, zero electron affinity (0 eV). That means that N cannot be efficiently ionized neither using oxygen nor cesium primary beam. As a result, both positive and negative secondary ion yields of nitrogen atomic ions are poor, or even absent. However, N can be detected as a molecular ion with matrix a element, or atoms of the primary beam. For example, in common semiconductor materials such as Si, nitrogen is normally profiled with a Cs$^+$ primary beam and monitored using negative molecular ions of ^{28}Si^{14}N. In contract, in compound semiconductors, such as ZnO or SiC, nitrogen can be efficiently measured as ^{14}N^{16}O or ^{14}N^{12}C negatively charged molecular ions, using a Cs$^+$ primary beam [14, 15].

6.4.2   Depth Profiling and Calibration

Depth profile of impurities can be obtained by recording the secondary ion intensity as a function of sputter time and Fig. 6.16 shows concentration versus depth profile of O in Si-SiGe multilayes, as measured by a Cs$^+$ primary beam. The Si matrix signal (not calibrated), measured simultaneously as the O-signal is also shown for layer identification. The modulation of the O concentration in different Si-SiGe layers can be clearly seen.

It should be noted that both depth and concentration calibration is required to convert original ion intensity versus sputtering time spectra to real depth profiles similar to one shown in Fig. 6.16. Since a constant erosion rate can usually be assumed through the analyzed layer, the sputtering time is translated into depth by subsequent determination of the sputtered crater depth by using, for instance, some type of profilometer. It is more complicated to translate peak intensities in a

Fig. 6.16 Concentration versus depth profile of O in Si-SiGe multilayers as measured by Cs$^+$ primary beam. The Si depth profile is shown for layer identification
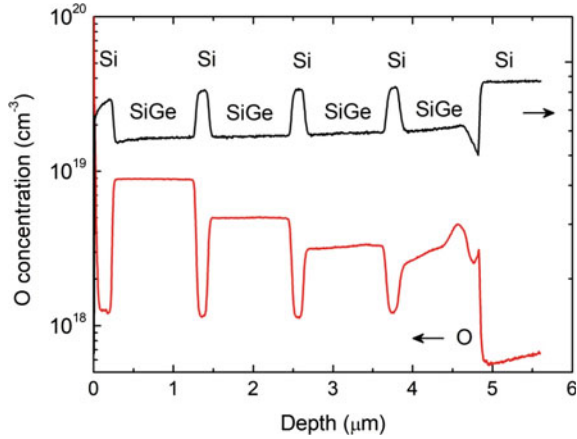
spectrum to concentrations, as it has been mentioned already in Sect. 2.2, that the theoretical prediction of sputtering and ionization yield is challenging. Therefore, the most straightforward intensity calibration method is to use calibration standards. This empirical approach provides the best accuracy in quantitative SIMS analysis. There are two common used empirical calibration methods: by using a relative sensitivity factor (RSF), or by using a calibration curve. Both methods require external standard samples, which must contain the specific ion in a specific matrix. Standards for impurity elements are typically made by ion implantation of the matrix of interest with a known dose of impurity.

The RSF technique is the most widely used calibration method if a concentration of the impurity to be quantified is less than 1% of the total matrix concentration. In this concentration range, a linear relationship exists between the secondary ion intensity and its concentration. Therefore, for relatively low impurity concentration of type i, its concentration value (C_i), typically given in units of cm^{-3}, can be expressed as:

$$[C_i] = RSF_i \frac{I_i}{I_m} \tag{6.11}$$

where I_i and I_m are the ion intensity of impurity of interest and matrix element, respectively, while RSF_i is a relative sensitivity factor of the element i for a given matrix. Using a standard sample implanted with the known dose of the same impurity i, RSF_i can be calculated as:

$$RSF_i = \frac{\Phi I_m T}{d \int I_i(t)dt} \tag{6.12}$$

where Φ is the implanted dose, d is the crater depth, T is the total sputtering time and $\int I_i(t)dt$ is the integrated number of detected secondary ions over the depth profile. There are strict requirements for the ion implanted standard samples. Firstly, the

sample of interest and the standard should match in matrix composition. Secondly, the ion implant dose and energy should be appropriately chosen. The dose should be high enough to obtain good counting statistics and, at the same time, should not exceed 1% of the atomic concentration to avoid ionization changes due to matrix effects. Furthermore, the ion energy should be high enough to place the implant peak relatively deep, about 100 nm, to avoid surface effects. The accuracy of quantification with ion implanted standards is often better than 10%.

For high impurity concentrations (in excess of ~ 1%), matrix effects must be taken into account, since the presence of a high concentration of an impurity may have an impact on both impurity and matrix secondary ion yields. As a result, a relationship between the secondary ion intensity of impurities and their concentration may exhibit strong nonlinear behavior. In such cases, calibration curves can also be used, but several calibration samples containing the element of interest in various concentrations above 1 atomic %. Thus, such a calibration curve provides information on the variation of ion yields with matrix composition also for alloys and compounds.

There are many other complicating factors affecting the detection limit and depth resolution of SIMS profiles, for instance, the so-called crater edge effect. During depth profiling the analyzed area should be localized within the uniform region of the sputtered zone and clearly separated from the edges (see, for example, Fig. 6.17 showing a crater where an information zone is depicted by the hatched area and different sources of the measuring artefacts are indicated schematically). Edge effects can occur because of misalignment, i.e., the gated area is not positioned within the center of the sputtered crater, and of drift or fluctuation of the primary beam position during analysis. Another problem is related to re-deposition of sputtered material and a phenomenon that is usually referred as "memory effect". Re-deposition of the sputtered material from the crater edges on the crater bottom can substantially degrade the detection limit of the measured impurity. Obviously, similar to that of direct crater-edge rejection discussed above, this effect is more serious when the concentration of the element of interest is higher at the crater walls than in the analyzed center area. Furthermore, memory effect is also related to re-sputtering of material from the extraction lenses of the secondary ion optics. In particular, the memory effect is important for instruments with small working distance between sample and the entrance slit of the mass spectrometer.

Another limiting factor is related to the residual gases in the sample chamber. In fact, bad or insufficient vacuum in the sample chamber can dramatically increase the background intensity, specifically for such elements as H, C and O, as well as for elements which may have a mass interference with molecules containing these species. Obviously, these effects can be reduced by improving the vacuum in the sample chamber, or by increasing the sputtering rate. For example, the best detection limit for O in silicon was achieved using ultrahigh vacuum sample chamber equipped with an ion pump and employing primary ion beam of high intensity. The detection limit of O was reported to be as good as 2×10^{15} cm^{-3} for a vacuum in the sample chamber of less than 2×10^{-10} Torr and a sputtering rate of ~ 40 nm/s for 14.5 keV Cs$^+$ primary beam [16].

Fig. 6.17 Schematics of the artefacts affecting the depth resolution of a SIMS profile
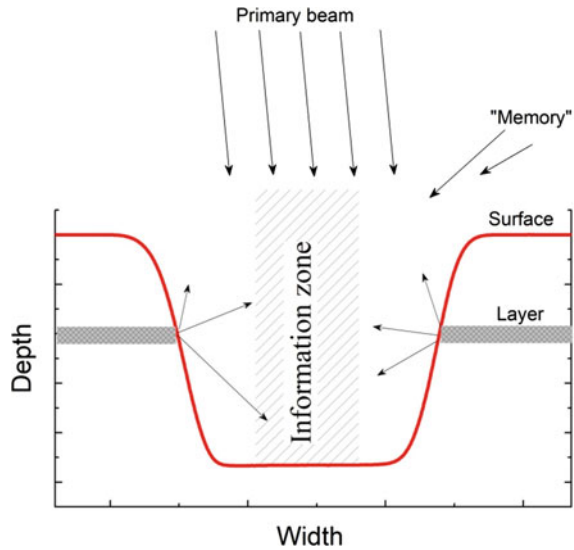
It should also be mentioned that ion beam mixing may lead to a broadening of the measured concentration profiles in respect to the true profiles. Ion beam mixing results from to the atomic displacements within a collision cascades generated by the primary beam in the target. Ion mixing becomes more pronounced when increasing the density of the collision cascades, which depends on the projectile mass and its energy. Therefore, for the correct measurements of narrow impurity profiles, a primary beam with reduced energy is preferable. Thus, it was demonstrated that detection of compositional changes occurring within a very thin (1–2 nm in width) layer is possible with Cs^+ primary beam having the energy as low as 150 eV [17].

A final issue that may reduce the depth resolution of SIMS profiling, is surface roughness, whether such roughness exists prior to the analysis, or emerges during the sputtering process. Obviously, effects of surface roughness on depth resolution become more crucial for very shallow depths, for instance, when measuring nanostructures. It is essential to ensure a very flat sample surface for correct interpretation of the obtained results, even though a very low energy of the primary beam is used.

6.4.3 Imaging Mode (2D) and 3D SIMS

The last operation mode of SIMS instrument discussed in the present chapter is imaging. In this mode the spectrometer is fixed at a particular mass and an element analysis across a given area can be obtained by using either the scanning ion microprobe or the direct imaging approach. Here, the lateral resolution directly depends on the primary beam size. Therefore, a better focused primary beam gives a better resolution and a lateral resolution of less than 0.5 μm can be readily achieved for
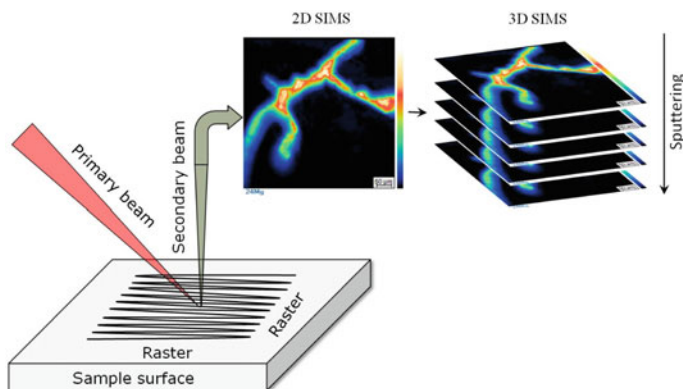
Fig. 6.18 Schematics of 2D (3D) SIMS operation mode

SIMS instruments equipped with conventional Cs, or O ion sources. However, a lateral resolution was reported to be as low as ~ 25 nm for the last generation of SIMS instruments equipped with a co-axial ion optics and a Cs source of a high brightness (CAMECA nanoSIMS 50L) [18].

Figure 6.18 illustrates schematically a SIMS instrument operating in the imaging mode. In order to obtain an image of a particular impurity distribution, the secondary ion intensity is recorded as a function of the primary beam position. An example of such image is shown in Fig. 6.18 where the measurements were performed using O_2^+ primary beam rastered over the area of 500×500 μm^2. In this figure, a 2D distribution of ^{24}Mg in a SiMg alloy is plotted and high Mg content areas are clearly visible.

Several images can be recorded subsequently during sputtering of the sample surface, as depicted in Fig. 6.18. After stitching the images together, a 3D impurity distribution in the analyzed sample can be obtained. However, quantification of the results in imaging mode is challenging and the influence of various effects has to be taken into account. For example, image contrast obtained by 2D SIMS can be affected not only by local concentration variations, but also due to orientation effects, different phases, surface roughness, etc. Indeed, an orientation contrast may appear in polycrystalline material because of a dependence of the ionization probability on the orientation of the crystallographic planes with respect to the primary ion beam. In addition, a phase contrast occurs because the ionization probability is matrix dependent, and for rough surfaces, topographic effects can distort the concentration contrast. However, there is continuing development of new state-of-the-art SIMS instruments capable to overcome existing limitations. For example, it was recently demonstrated that using SIMS instrument with integrated scanning probe microscopy module along with advanced 3D reconstruction software can minimize such surface topographic effects [19].

Fig. 6.19 Depth profiles of
the main Zn isotopes in the
isotopically modulated ZnO
heterostructures as measured
by SIMS
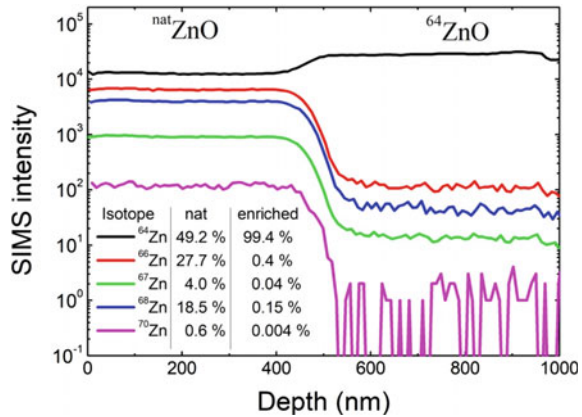
6.5   Examples of Application

6.5.1   Isotopically Modulated Heterostructures

Isotope analysis and trace element abundance is a very important task, in particular,
in geological science. However, isotopically enriched/modulated materials are of a
great interest in material science since they offer a possibility for the investigation
of mass transport in solids. Furthermore, these materials can have properties which
are different from those in the materials with natural abundance. For example, the
thermal conductivity of isotopically enriched silicon is larger compared to Si with
natural abundance [20].

   Figure 6.19 shows depth profiles of the main Zn isotopes in an isotopically modu-
lated ZnO heterostructure, as measured by SIMS using O_2^+ primary beam. The
thickness of the sandwich-type structure is $500 + 500$ nm and it is possible to calcu-
late the isotope ratios for both layers. As indicated in the inset, the second layer is
enriched with ^{64}Zn isotope up to 99.4% and the all other isotopes are significantly
reduced as compared to the ZnO with natural abundance.

6.5.2   Zn Self-diffusion in ZnO

As discussed above, isotopically modulated heterostructures can be used for the
investigation of the mass transport in solids. Indeed, self-diffusion in solids is often
mediated by point defects such as vacancies, or interstitial atoms and formation/
migration energies of these defects determine the kinetics of the process. Further-
more, in a contrast to the migration energy, the formation energy of point defects is a

Fig. 6.20 Concentration versus depth profiles of the ^{70}Zn in the isotopically modulated ZnO heterostructures as measured by SIMS. Arrhenius plots of Zn diffusivity along with obtained activation energies are shown in the inset (adapted from [21])
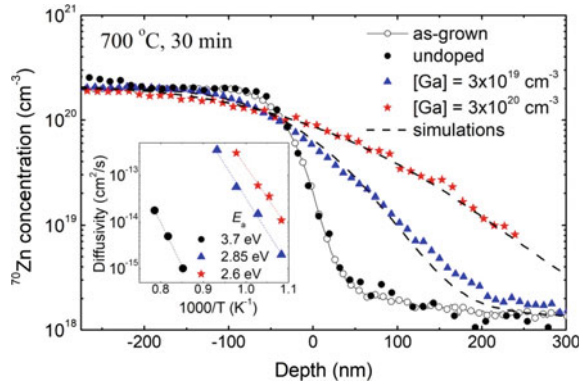
function of Fermi level position and chemical potential (stoichiometry). So, Arrhenius analysis of self-diffusion is a direct method allowing to determine energetics of point defects mediating self-diffusion.

As an example, Fig. 6.20 illustrates an acceleration of Zn self-diffusion in ZnO with increasing Ga concentration, which has a donor behavior and, therefore, shifts the Fermi level towards the conduction band. Indeed, Fig. 6.20 shows concentration versus depth profiles of ^{70}Zn in the isotopically modulated heterostructure after annealing at 700 °C for 30 min. in air ambient. It is seen that the Ga doping dramatically enhances Zn diffusivity. Arrhenius plots of Zn self-diffusion extracted from this data are shown in the inset, where the obtained activation energies are also depicted.

It is seen from this figure that activation energy of Zn self-diffusion decreases with increasing Ga concentration. It is important to note that the obtained results support the Zn vacancy assisted mechanism of Zn self-diffusion in ZnO and allow to draw conclusions regarding Zn vacancy energetics [21].

6.5.3   Indium Segregation in In$_x$Ga$_{1-x}$N Quantum Wells

The last example is related to the characterization of high quality In$_x$Ga$_{1-x}$N quantum wells (QW) for optoelectronic devices able to operate in the blue and UV spectral range. Currently InGaN-based blue lasers and light emitting diodes are widely used in quantum technologies, telecommunications, white lightning, etc. One of the main difficulties in the fabrication of high quality In$_x$Ga$_{1-x}$N QWs is redistribution and segregation of In atoms as well as variations of QWs width. These issues are attributed to the large lattice mismatch between InN and GaN (~10%) and low growth temperature necessary for the InGaN deposition. The uniformity and the absence of In precipitates are key factors determining the properties of the QWs.

A good sensitivity of SIMS for both In and Ga ions makes this technique very suitable for the investigation of In/Ga distribution as well as precipitate formation. In this example, 3D imaging SIMS mode was utilized for the study of a sequence
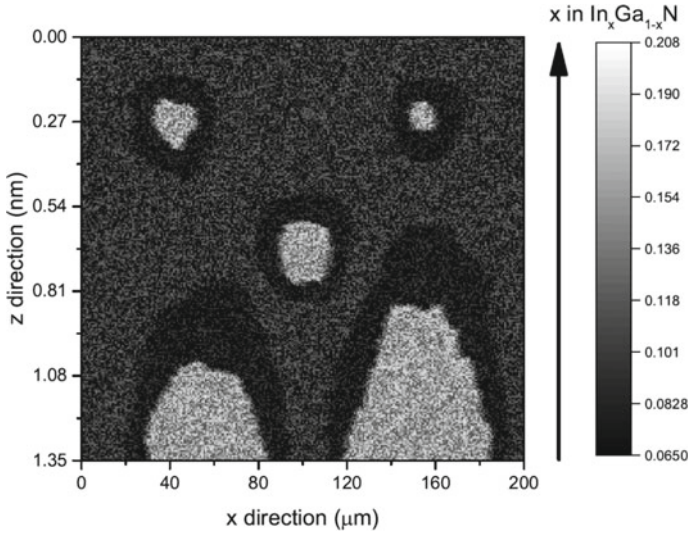
Fig. 6.21 Cross-sectional view of In distribution in $In_xGa_{1-x}N$ QW where z direction corresponds to the depth of analysis and the arrow indicates QW growth direction [22]

of 5 $In_xGa_{1-x}N$ QWs, where each QW has a thickness of about 2 nm [22]. As a result, Fig. 6.21 shows the cross-sectional image of a single QW illustrating a non-uniformity of the In atoms and how it is varying through the QW layer. The measurements were performed using ultra low energy (100 eV) Cs^+ primary beam rastered over an area of $250 \times 250 \, \mu m^2$. Note that the lateral resolution is determined by the size of the primary beam used for the analysis (~ 1 μm), while the depth depth resolution was estimated to be as low as 0.2 nm. Such a high depth resolution was realized by using a primary beam of ultra low energy. Imaging mode SIMS analysis shown in Fig. 6.21 revealed regions within the QW where In atoms seems to getter at certain points (brighter regions) forming an Ín-depleted region (dark contrast) around them. The In content ranges from 6 to 20%, while the average In content was estimated to be about 9%, a value that corresponds well to what was measured with other techniques, such as X-ray diffraction analysis. Thus, this example clearly demonstrates a large potential and advantages of 2D (and also 3D) SIMS for spatial composition analysis of ultra thin complex functional films.

6.6 Summary

Secondary ion mass spectrometry (SIMS) is one of the most sensitive analytical techniques for impurity analysis in solids. Depending on the material and the element of interest, a detection limit in the range of 10^{12}–$10^{16} \, cm^{-3}$ (corresponding to 0.1 ppb-ppm level) can be reached. Potentially, all the elements of the periodic table including

isotopes can be detected in three main SIMS operation modes (mass spectra, depth profiling, and ion imaging). A large dynamic range (> 5 orders of magnitude), a high depth resolution and a good lateral resolution makes this technique one of the most attractive tools for the characterization of modern materials for electronic applications and also many other areas.

However, there are several factors limiting the SIMS applications. For instance, accurate quantification is typically limited to impurity concentrations less than 5 atomic percent due to matrix effects and there are also special requirements for the samples to be measured with SIMS instruments. For example, the samples must be vacuum compatible and, at least partly, conductive. Furthermore, quantification of the results requires special standard samples and can be complicated by mass interference, sample roughness, etc. For a correct interpretation of the obtained results a deep understanding of the involved physical and chemical processes is required. In addition, SIMS instruments are complicated and rather expensive, hence, efficient usage of this technique for material characterization requires experienced personnel.

References

1. Thompson, J. J. (1910). Rays of positive electricity. *Philosophical Magazine, 20*, 252.
2. Herzog, R. F. K., & Viehbock, F. P. (1949). Ion source for mass spectrography. *Physical Review, 76*, 855L.
3. Herzog, R.F.K., Poschenreider, W.P., & Satkiewicz, F.G. (1967). NASA, Contract No NAS5-9254, final report GCA-TR-67-3N.
4. Sigmund, P. (1969). Theory of sputtering. *Physical Review, 184*, 383.
5. Zalm, P. C. (1994). Secondary ion mass spectrometry. *Vacuum, 45*, 753.
6. Morris, R. J. H., & Dowsett, M. G. (2009). Ion yields and erosion rates for $Si_{1-x}Ge_x$ ($0 \leq x \leq 1$) ultralow energy O_2^+ secondary ion mass spectrometry in the energy range of 0.25–1 keV. *Journal of Applied Physics, 105*, 114316.
7. Balden, M., Bardamid, A. F., Belyaeva, A. I., Slatin, K. A., Davis, J. W., Haasz, A. A., Poon, M., Konovalov, V. G., Ryzhkov, I. V., Shapoval, A. N., & Voitsenya, V. S. (2004). Surface roughening and grain orientation dependence of the erosion of polycrystalline stainless steel by hydrogen irradiation. *Journal of Nuclear Materials, 329–333*, 1515.
8. Nørskov, J. K., & Lundqvist, B. I. (1979). Secondary-ion emission probability in sputtering. *Physical Review B, 19*, 5661.
9. Evans Analytical Group (www.eag.com)
10. IONTOF GmbH (www.iontof.com)
11. Wolf, B. (Ed.). (2017). *Handbook of ion sources* (p. 560). CRC Press.
12. Krohn, V. E. (1962). Emission of negative ions from metal surfaces bombarded by positive cesium ions. *Journal of Applied Physics, 33*, 3523.
13. Taylor, G. (1964). Disintegration of water droplets in an electric field. *Proceedings of the Royal Society of London. Series A, 280*, 383.
14. Li, Y., Wang, S., & Smith, S. P. (2006). SIMS analysis of nitrogen in various metals and ZnO. *Applied Surface Science, 252*, 7066.
15. Ber, B.Ya., Kazantsev, D.Yu., Kalinina, E.V., Kovarskii, A.P., Kossov, V.G., Hallen, A., Yafaev, R.R. (2004). Determination of nitrogen in silicon carbide by secondary ion mass spectrometry. Journal of Analytical Chemistry **59**, 250 (2004).
16. Jakiela, R., Barcz, A., Sarnecki, J., & Celler, G. K. (2018). Ultrahigh sensitivity SIMS analysis of oxygen in silicon. *Surface and Interface Analysis, 50*, 729.

17. Michałowski, P. P., Gaca, J., Wójcik, M., & Turos, A. (2018). Oxygen out-diffusion and compositional changes in zinc oxide during ytterbium ions bombardment. *Nanotechnology, 29*, 425710.

18. Saka, S. K., Vogts, A., Kröhnert, K., Hillion, F., Rizzoli, S. O., & Wessels, J. T. (2014). Correlated optical and isotopic nanoscopy. *Nature Communications, 5*, 8.

19. Wirtz, T., Fleming, Y., Gerard, M., Gysin, U., Glatzel, T., Meyer, E., Wegmann, U., Maier, U., Odriozola, A. H., & Uehli, D. (2012). Design and performance of a combined secondary ion mass spectrometry-scanning probe microscopy instrument for high sensitivity and high-resolution elemental three-dimensional analysis. *Review of Scientific Instruments, 83*, 063702.

20. Ruf, T., Henn, R.W., Asen-Palmer, M., Gmelin, E., Cardona, M., Pohl, H.-J., Devyatych, G.G., Sennikov, P.G. (2000). Thermal conductivity of isotopically enriched silicon. Solid State Communications **115**, 243 (2000)

21. Azarov, A., Venkatachalapathy, V., Mei, Z., Liu, L., Du, X., Galeckas, A., Monakhov, E., Svensson, B. G., & Kuznetsov, A. (2016). Self-diffusion measurements in isotopic heterostructures of undoped and in situ doped ZnO: zinc vacancy energetics. *Physical Review B, 94*, 195208.

22. Michałowski, P. P., Grzanka, E., Grzanka, S., Lachowski, A., Staszczak, G., Plesiewicz, J., Leszczyński, M., & Turos, A. (2019). Indium concentration fluctuations in InGaN/GaN quantum wells. *Journal of Analytical Atomic Spectrometry, 34*, 1718.

Part III
Electrical Measurement Techniques

Chapter 7
Electrical Characterization of Semiconductors: *I–V*, *C–V* and Hall Measurements

7.1 Introduction

Fabrication of electronic devices as well electrical measurements of electronic materials typically requires junctions, or interfaces between materials with different properties, such as a metal–semiconductor contact, or an insulator-semiconductor junction. Functionalization of this junction depends not only on the basic properties of the materials, but also on the properties of the interface. In contrast to an ideal junction, a real interface always has defects of various types which not only can affect the contact quality, but often dominate the junction properties. Therefore, understanding the physics of junctions is essential for electronic device fabrication and characterization, and the basic properties of metal–semiconductor contacts are initially considered in this chapter.

Electrical resistivity, measured in Ω cm, is a basic material property and defines how well a material conducts an electric current. An insulating material have a poor electrical conductivity with a resistivity in the range of 10^9–10^{20} Ω cm. In contrast, metals are good electrical conductors with a resistivity of about 10^{-6} Ω cm, or less. Semiconductors conduct current better than insulators, but not as good as metals and they have an intermediate resistivity in the range of 10^{-4}–10^7 Ω cm, strongly depending on the doping. With "doping concentration" in a semiconductor one refers to the chemical concentration of dopant atoms, which consists of both substitutional and interstitial dopants, a concentration which can be obtained by, for instance, SIMS measurements, as described in Chap. 6. This should not be mixed with electrically active dopant concentration, which is obtained from electrical measurements. The active dopant concentration will be equal to the substitutional dopant concentration when an atom contributes a hole (acceptor) or electron (donor) for the electrical transport.

In metallic conductors the current is transported by electrons, while in semiconductors the current can be carried both by both electrons and holes. In a classical view, electrons are negatively charged particles that move in the opposite direction

compared to the direction of an applied electric field, while holes are positively charged particle moving with the field. Conduction electrons are normally transported in the conduction band and holes, equivalent to "lack of an electron", are transported in the valence band of semiconductors.

According to classical Drude theory [1] electrons (and holes) can be compared to molecules in a gas, moving around stochastically in space due to their thermal energy, every now and then colliding with neighbors, or imperfection in their way. When an external field is applied the charged particles respond accordingly and a net motion of particles occurs. This motion should be considered as an average motion of many particles impeded by various scattering events. Therefore, it is possible to assign an average velocity, $\langle v \rangle_{drift}$, to the particles in a volume, depending on the dominating scattering mechanism. The drift velocity is typically substantially lower than the particles thermal velocity, $\langle v \rangle_{th}$, and increases linearly with the applied field.

This particle view of charge transport can also be applied to charge carriers in crystals, although additional scattering events may then occur, for instance scattering against vibrating lattice atoms (phonon or lattice scattering). Scattering may also occur against ionized impurities (impurity scattering). In addition, charge carriers in a crystal will experience a force from the periodic potential formed by the ordered arrays of atoms in such a system. By ascribing an effective mass, m^*, to the carriers confined to a crystal, i.e. a modification of the mass of a free electron, or hole, the carrier transport in the periodic potential can still be described within the classical model.

If carriers in a solid are accelerated by an electric field, *E*, during a certain time *t*, *the equation of motion is written as:*

$$\frac{d}{dt}\left(m^* v_{drift}\right) + \frac{m^* v_{drift}}{\langle \tau \rangle} = qE \tag{7.1}$$

Here we consider an average time, $\langle \tau \rangle$, until a certain dominating scattering event takes place. Meanwhile, during steady state the first term in the above equation becomes zero and rearrangement gives:

$$\langle v \rangle_{drift} = \mu E \,(\text{where} \mu = \frac{q \langle \tau \rangle}{m^*}) \tag{7.2}$$

The proportionality constant, μ, between the average velocity and the field in Eq. 7.2 is called carrier mobility (cm^2/Vs) and is an important parameter in electrical devices. The mobility is a material specific constant and differs normally for electrons and holes, but depends also on lattice imperfections, strain (lattice distortion), dopant level and temperature, since these parameters will affect the average scattering time.

As indicated above, the current density (A/cm^2) in a conductor with a constant cross-sectional area can be described by the average velocity of carriers. For holes in a semiconductor under an electric field in the x-direction the drift current density becomes

$$J_x = qp\langle v\rangle_{drift} = qp\mu_p E_x \tag{7.3}$$

where μ_p is the hole mobility and the concentration of holes in the conductor is given by p (in units of cm^{-3}). Now we can introduce the resistivity, ρ, of a material where the current is carried by holes:

$$\rho = \left(qp\mu_p\right)^{-1} \tag{7.4}$$

A relationship between parameters: resistance, R (Ω) (reciprocal of conductance, G with unit Siemens, S), and resistivity ρ (Ωcm), (reciprocal of conductivity, σ with unit Siemens/meter, S/m), and sheet resistance, R_S ($\Omega/\square$ or ohms per square) is given in Eq. 7.5. It is assumed that the charge carriers move perpendicular to the cross-sectional area $A = Wt$, where W is width and t is thickness, of the conductor, which has a length l.

$$R = \rho\frac{l}{Wt} = R_S\frac{l}{W} \quad \text{where} \quad R = \frac{1}{G} \quad \text{and} \quad \rho = \frac{1}{\sigma} \tag{7.5}$$

In Eq. 7.5, conductivity (σ) could be written in general as a function of both a concentration of charge carriers and their mobility as $\sigma = q(n\mu_n + p\mu_h)$. We also point out the conductance in Eq. 7.5 is the reciprocal of resistance in a DC case, while for an AC case, it is the reciprocal of impedance.

From the discussion above it is obvious that measurements of macroscopic resistivity can lead to an insight in the microscopic behavior of charge carriers in materials governed by scattering processes. Moreover, Eq. 7.5 also provides information of the doping of materials and may reveal the dynamics around defect trapping and de-trapping of carriers, etc. By making the probes and contacts smaller, the measurements move from macroscopic to microscopic and, for instance using scanning probe microscopy (SPM) for electrical measurements, it is possible to assess transport properties for very few carriers at a time, but still maintain the classical macroscopic description. However, for sufficiently small structures, the classical view must eventually be abandoned and replaced by a quantum mechanical description.

The methods based on four-point probe, Hall effect, and capacitance–voltage measurements are perhaps the most widely used techniques for the investigation of the electrical properties of electronic materials. Despite that these characterization techniques were developed already for more than half a century ago, they are still intensively used in semiconductor industry and for research purposes. Meanwhile, the technology roadmap continuously increases the demand for the device scaling and more dense integration of devices on the chips and, therefore, the analysis methods must constantly adapt and improve their sensitivity and geometrical resolution. A spectacular example is the quantum Hall effect, which is drastically different from the ordinary Hall effect and only becomes observable when device sizes scale down to nm sizes, where quantum effects come into play. This quantum Hall effect has unprecedented applications in metrology.

In the present chapter the above-mentioned techniques are reviewed and information which can be obtained from these methods is discussed. We start with the basic properties of metal–semiconductor contacts.

7.2 Metal–Semiconductor Contacts

For any electrical measurement, we typically use metallic conductors, to deliver a current, or applying voltage to a sample. The contact between the metal connector and a semiconductor will then be essential and the properties of the metal–semiconductor interface must be known in order to correctly interpret the measured signals. There are two types of metal–semiconductor contacts—Ohmic and Schottky type. An Ohmic contact shows a linear dependence of electric current on the applied voltage with normally a low resistance which is independent on the applied voltage polarity. In contrast, a Schottky type contact demonstrates rectification behavior, i.e. it allows to pass electric current only in one direction. Realization of one of them depends on the type of semiconductor (*n* and *p*) and the relation between the work functions of a metal and semiconductor. The work function of the metal (Φ_m) is defined as the separation in energy for an electron at the metal Fermi level ($E_F{}^m$) and an electron at the vacuum level (E_{vac}), i.e., a completely free electron. Thus, Φ_m is the energy required to remove an electron from the metal and place it sufficiently far away so it does not sense any force from the metal surface. For a semiconductor, the work function (Φ_S) depends on the doping level and it is defined as $\Phi_S = \chi_S + (E_C - E_F{}^S)$, where χ_S is an electron affinity (the energy difference between the bottom of the conduction band, E_C, and the vacuum level) and $E_F{}^S$ is the semiconductor Fermi level. Table 7.1 shows a type of contacts for different relation between metal/semiconductor work functions and semiconductor type for relatively low doping levels. In the present section, the formation and basic properties of these two types of contacts will be described in detail, specifically for *n*-type semiconductors. However, the similar formalism can be readily applied to describe the contact between a metal and *p*-type semiconductor.

Table 7.1 The type of the metal–semiconductor contact depending on the semiconductor type (relatively low-doped) and relation between the work functions of the metal and semiconductor

	n-type semiconductor	*p*-type semiconductor
$\Phi_m > \Phi_S$	Schottky	Ohmic
$\Phi_m < \Phi_S$	Ohmic	Schottky

7.3 Schottky Contacts

Figure 7.1 shows simplified band diagrams of a metal and an *n*-type semiconductor before and after contact formation of Schottky junction. When the metal and semiconductor in Fig. 7.1a are put in an ultimate contact, the electrons in the conduction band of the semiconductor move to the empty energy states above the Fermi level of the metal, since the Φ_m is larger than χ_S. Transport of electrons leads to the bending of the semiconductor energy bands until the Fermi level becomes constant throughout the structure. When the Fermi level becomes constant throughout the structure, thermal equilibrium is reached, as shown in Fig. 7.1b. As a result, a potential barrier for electrons is formed (E_b) at the interface to impede their movement from metal to semiconductor. This barrier is the so-called Schottky barrier and, due to negligible band bending in the metal, its height follows the Schottky–Mott relation:

$$E_b = \Phi_m - \chi_S \tag{7.6}$$

The fixed positive charges of ionized donors on the semiconductor side then lead to the formation of a depletion region, where its width (W) is determined by the band bending, extending until the bands become flat. For *n*-type semiconductor W can be written as:

$$W = \sqrt{\frac{2\varepsilon_s(V_0 - V)}{qN_d}} \tag{7.7}$$

where ε_s is a semiconductor permittivity, q is a elementary charge, N_d is a net donor concentration, V is the applied voltage and V_0 is zero bias build-in voltage determined
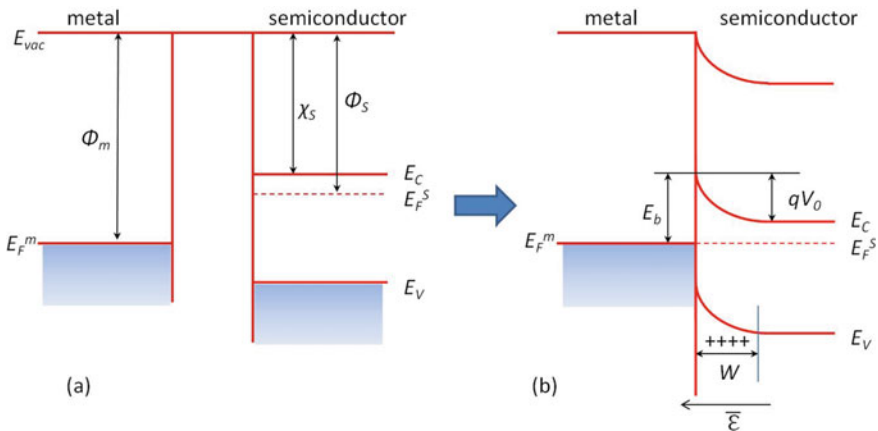
Fig. 7.1 The energy band diagrams (simplified) of a metal and an n-type semiconductor before contact **a** and after a perfect contact **b** illustrating formation of Schottky type contact

from $qV_0 = \Phi_m - \Phi_S$. At equilibrium the motion of electrons across the junction is balanced by the contact potential so that there is no net current. However, behavior of the Schottky junction dramatically depends on the polarity of the applied voltage.

In a forward biased junction (the metal is connected to positive terminal and *n*-type semiconductor connected to negative terminal), the external potential (*V*) is applied in such a way that it opposes the built-in potential (in Fig. 7.2a). The voltage drop is across the depletion region since it has the highest resistivity. As a result, the potential barrier for the electrons on the semiconductor side decreases and becomes equal to $q(V_0 - V)$. For relatively low doping level of the semiconductor when the width of the barrier is relatively large and tunneling effects are negligible, the main mechanism of the current through the contact under the external potential is a thermionic emission. The total current density (*J*) is determined by the carrier fluxes in a forward and reverse directions and it can be expressed as:

$$ J = J_0 \left(\exp\left(\frac{qV}{ck_bT} \right) - 1 \right) \tag{7.8} $$

where k_b is a Boltzmann constant, T is a temperature, c is an ideality factor ($c = 1$ for a thermionic emission) and J_0 is a constant determined as:
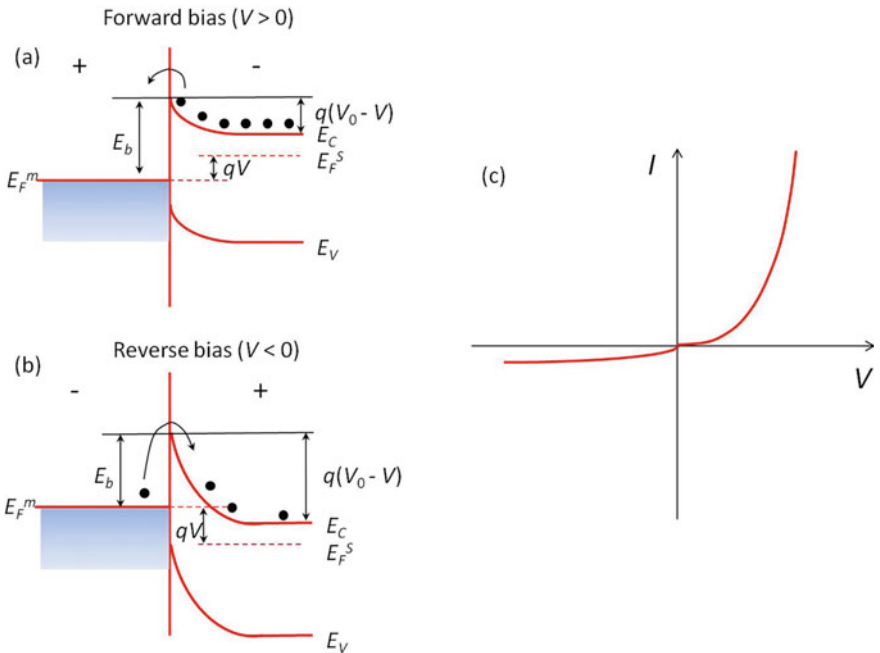
Fig. 7.2   The energy band diagrams of a Schottky junction under forward **a** and reverse bias **b**. Typical I–V characteristics of a Schottky junction showing rectifying properties are plotted in the panel **c**

$$J_0 = RT^2 \exp\left(-\frac{E_b}{k_b T}\right) \tag{7.9}$$

where R is a Richardson constant for thermionic emission and is a material property. According to Eq. 7.8 the current exponentially increases with applied potential in the forward bias. Practically, Schottky diodes never satisfy the ideal situation with $c = 1$, therefore, c is an important parameter characterizing Schottky contact. The ideality factor can be calculated from the slope of the linear region of the forward bias J–V characteristics plotted in a logarithmic scale using the relation:

$$c = \frac{q}{k_b T} \frac{dV}{d\ln(J)} \tag{7.10}$$

In the case of a reverse bias (the metal is connected to negative terminal and n-type semiconductor connected to positive terminal) the external potential is applied in the same direction as the junction potential, as shown in Fig. 7.2b. Therefore, the potential barrier for the electrons on the semiconductor side increases by qV and the current from semiconductor to the metal becomes negligible. Now the total current density is determined by the current from the metal to the semiconductor equal to J_0 and its value does not depend on the applied voltage since it does not change the Schottky barrier height. The reverse bias J–V characteristics can be used for the estimation of the barrier height according to Eq. 7.6. It should be noted that the Schottky barrier height can be also estimated by other methods such as internal photoemission or capacitance–voltage measurements. The typical I–V characteristics of Schottky junction is schematically shown in Fig. 7.2c and it is clearly seen that it acts as a rectifier, i.e. it conducts in forward bias but not in reverse bias.

Finally, it should be noted that there is one more parameter characterizing Schottky contact along with such parameters as the barrier height and the ideality factor. This parameter is called series resistance and it determines behavior of the forward bias I–V characteristics at high current densities. Indeed, in real Schottky contacts the thermionic emission theory describes the forward I–V characteristic for a limited range of current densities. At higher currents the deviations from ideal diode characteristics are observed as an additional voltage drop through the device. This voltage drop is generally attributed to a series resistance (R_s) of the neutral regions of the semiconductor. Therefore, in the Eq. 7.8 the applied or experimentally measured voltage in the Schottky diode circuit follows the relation $V = V_D + IR_s$, where V_D is the voltage across the metal–semiconductor interface.

7.4   Ohmic Contacts

Figure 7.3a illustrates band bending occurring after a contact between n-type semiconductor and a metal having lower work function than the semiconductor. At equilibrium, electrons move from the metal to the empty states in the conduction band
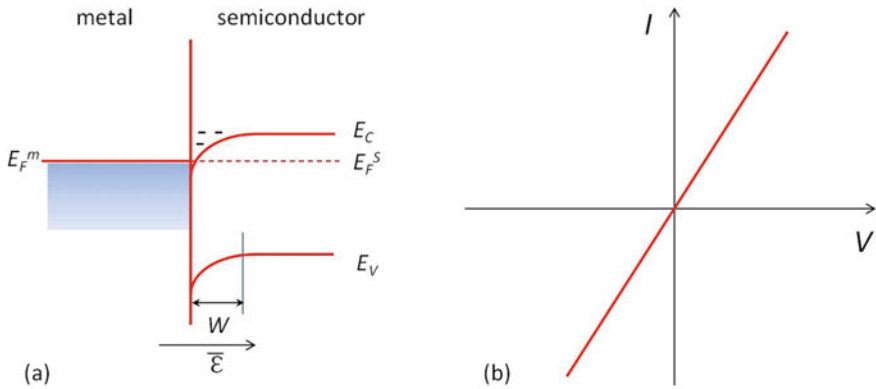
Fig. 7.3 The energy band diagram of a metal and an n-type semiconductor after a perfect contact **a** and typical I–V characteristics of an Ohmic type contact **b**

of the semiconductor, so that there is an accumulation region near the interface (on the semiconductor side). Thus, resulting electric field has an opposite direction as compared to that of Schottky contact. The accumulation region has a higher conductivity than the bulk of the semiconductor, therefore, the resistivity is determined by the bulk resistivity of the semiconductor.

Thus, an Ohmic junction behaves as a resistor conducting in both forward and reverse bias. Since the resistivity of the contact is very low and does not depend on the applied voltage, the *I–V* characteristics exhibit linear (or quasi linear) behavior as clearly illustrated by Fig. 7.3b.

Another type of Ohmic contacts can be realized for highly doped semiconductors even if the metal has a higher work function than that of the *n*-type semiconductor. Indeed, according to the Eq. 7.7, the width of the Schottky barrier decreases with increasing doping level. If this width is very narrow (on the order of 3 nm or less) electrons can readily tunnel through such barrier as illustrated by Fig. 7.4a. In this case the current density is described in the terms of field emission theory:

$$J \sim \exp\left(\frac{qV}{E_{00}}\right) \tag{7.11}$$

where the tunneling parameter E_{00} depends on the semiconductor permittivity (ε_s), a concentration of ionized donors (N_d), and an effective electron mass (m^*):

$$E_{00} = \frac{\hbar}{2}\sqrt{\frac{N_d}{\varepsilon_s m^*}} \tag{7.12}$$

However, the dominant mechanism of the current through the contact depends also on the temperature. Thus, field emission (FE) is dominant for $k_b T \ll E_{00}$,
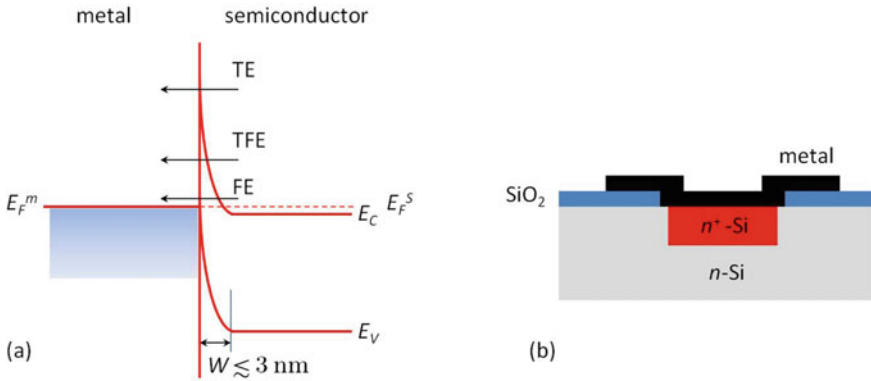
Fig. 7.4 **a** The energy band diagram of a metal and highly doped (degraded) n-type semiconductor after a perfect contact. **b** Possible structure of tunneling contact to n-Si

thermionic-field emission (TFE) dominates for $k_b T \approx E_{00}$, while thermionic emission (TE) becomes dominant for $k_b T \gg E_{00}$ (see Fig. 7.4a).

It should be noted that, normally the doping of the semiconductor is determined by a functionalization of a device and it can be relatively low. To form a tunneling contact to such low doped semiconductor, a thin heavily doped layer of the same conductivity can be formed over this semiconductor as illustrated by Fig. 7.4b showing a possible structure of tunneling contact. This heavily doped area can be formed either by ion implantation or by in-situ doping during epitaxy.

7.5 Non-ideal Contacts

There are several complicating factors distorting the ideal model of metal–semiconductor contacts presented above. The most common factors are related to the:

 (i) Image force
(ii) Surface states

The schematic of the barrier height lowering due to image force is shown in Fig. 7.5a. This force arises due to redistribution of the charge carriers in the adjacent metal. Indeed, when an electron in the semiconductor is close to a Schottky contact, a positive charge will be induced on the metal surface. The force of attraction between the electron and the induced positive charge is equivalent to the force that would exist between the electron and an equal positive charge located at the equal distance in the metal. This positive charge is referred to as the image charge. The electrostatic force between the electron and pseudo particle is described by Coulomb attraction and the electric field of the image force ($\mathcal{E}_{image}$) is opposite to the barrier field, so the barrier height is lowering.
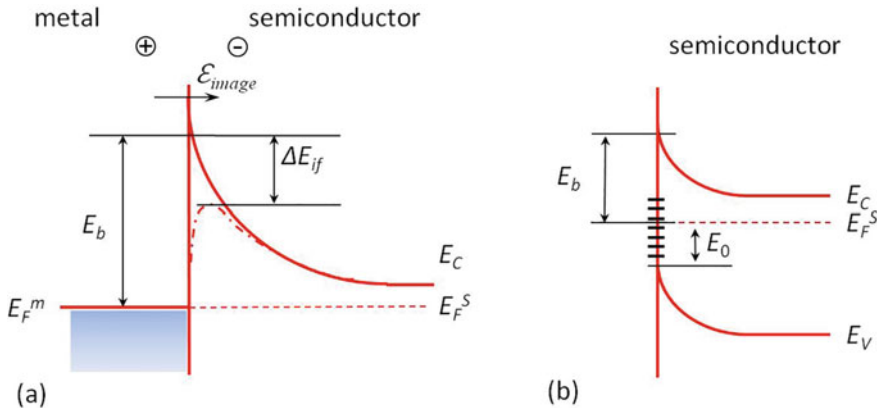
Fig. 7.5 The energy band diagrams schematically illustrating the distortion of Schottky barrier due to **a** image force and **b** surface states

The barrier lowering due to image force (ΔE_{if}) depends on the maximum value of the electric field across the contact which is a function of the doping level and the applied voltage:

$$\Delta E_{if} = \frac{1}{4}\left[\frac{2q^3 N_d}{\pi^2 \varepsilon_s^3}(V_0 - V)\right]^{1/4} \tag{7.13}$$

Note that for forward bias ($V > 0$), the electric field and, therefore, the image force are smaller and the barrier height is slightly larger than that at zero bias. In its turn, for the reverse bias ($V < 0$), the barrier reduction is larger and its dependence on the applied voltage leads to a voltage dependence of the reverse bias current. This barrier lowering can be quite strong at high electric fields and explains why the reverse current of Schottky contact is higher than that of *p–n* junction with a remarkable strong dependence on the reverse bias voltage. However, it should be kept in mind that the barrier lowering due to image force is experienced by free carriers located near the contact; therefore, this effect is not noticeable in a capacitance–voltage measurements.

Another factor affecting the barrier height is related to the existence of surface states. Already in 1947 Bardeen [2] suggested that observed discrepancy between the measured barrier height and the predicted one based on Schottky Moss theory can be attributed to the effect of surface states having energy levels within the bandgap. Density of the surface states in semiconductors can be very high and they can be intrinsic or extrinsic origin. Intrinsic surface states arise due to the presence of dangling bonds at the surface, while extrinsic ones are related to the damage caused to the surface of the semiconductor during the metal deposition process, which are called metal induced gap states (MIGS).

The surface states are characterized by neutrality level E_0 as illustrated by Fig. 7.5b showing the band bending in near the surface region due to surface states. The energy level E_0 causes the surface to be charge neutral if states below this level are filled and states above are empty. If the density of surface states is high, the Fermi level will be pinned at this energy level and the bands of the semiconductor could therefore be bent even without being in contact with the metal. In its turn, when the metal is brought into the contact with the semiconductor, the alignment of the Fermi levels will be accomplished by the transfer of electrons from the filled surface states into the metal, instead of from the semiconductor into the metal. Thus, for high density of the surface states the number of electrons emitted by the surface states will be enough to reach equilibrium. In this case the Fermi level position at equilibrium is pinned and is determined mainly by the surface states, not by the carrier concentration in the semiconductor and work function of the metal. As a result, the barrier height is given by the Bardeen limit:

$$E_b = E_g - E_0 \tag{7.14}$$

Thus, in the Bardeen limit the Schottky barrier height is independent of the work function of the metal, but it depends only on the neutrality level E_0 of the surface states on the semiconductor. The Schottky–Mott limit (Eq. 7.6) and the Bardeen limit are the two extreme cases for the barrier height for metal–semiconductor contacts and, in reality, the actual Schottky barrier height falls somewhere between these two limits.

The Fermi level pinning is strong in many commercially important semiconductors, such as Si, Ge, and GaAs. For example, near all metals form a significant Schottky barrier to n-type Ge and an ohmic contact to p-type Ge, since the Fermi level at metal/Ge interface strongly pinned near the valence band edge of Ge [3]. Furthermore, recombination losses at the contacts due to surface states are now a major factor limiting Si solar cell efficiencies. Therefore, reduction of the surface state density as much as possible is an important issue in the development of semiconductor devices. This requires additional technological steps and one of the solutions is adding an intermediate insulating layer with reduced density of surface (interface) states to unpin the Fermi level. It should be noted that the situation becomes even more severe for nanoscale and low-dimensional structures due to increased surface-to-volume ratio as compared to bulk materials [4].

7.6 Contact Resistance

The voltage drop across the metal–semiconductor contact is determined by the specific contact resistance (ρ_C), which is a figure of merit of the Ohmic contact and it is defined as:

$$\rho_C = \left. \frac{\partial J}{\partial V} \right|_{V=0} \; [\Omega \cdot \mathrm{cm}^2] \tag{7.15}$$

where V is the voltage across the barrier and J is the current density through the barrier. Therefore, for the homogenous contact of a cross sectional area A, the contact resistance is:

$$R_C = \frac{\rho_C}{A} \tag{7.16}$$

As it has been shown above, the current density depends on the dominating transport mechanism in Ohmic contact. Thus, for metal contacts with lowly doped semiconductors where the thermionic emission dominates, the specific contact resistance can be written as:

$$\rho_C = \frac{k_b}{qRT} \exp\left(\frac{qE_b}{k_bT} \right) \tag{7.17}$$

In its turn, for metal contact on heavily doped *n*-type semiconductors where the current due to tunneling effects dominates, ρ_C is:

$$\rho_C \sim \exp\left[\frac{2\sqrt{\varepsilon_S m^*}}{\hbar} \left(\frac{E_b}{\sqrt{N_d}} \right) \right] \tag{7.18}$$

The Eq. 7.18 shows, in particular, that for highly doped material the specific contact resistance depends strongly on the doping concentration and varies exponentially with the factor of $E_b/(N_d)^{1/2}$ for *n*-type contacts.

A well-known and widely used method for contact resistance measurements is the linear transfer length method (TLM) structure. A brief explanation of a measurement between two adjacent metal contacts is illustrated in Fig. 7.6(a) where the total resistance R_T is the sum of the semiconductor bulk resistance (R_{bulk}) and the contact resistances in the metal–semiconductor interface (R_C), so that $R_T = R_{bulk} + 2R_C$.

The bulk resistance of the semiconductor can be written in terms of sheet resistivity (R_{sh}) as $R_{bulk} = R_{sh}(d/Z)$, where d is the spacing between the metal contacts and Z is the contact width. Taking into account that the current flow achieves the lowest resistance which means that not the entire length of the contact will be used (see Fig. 7.6a), it can be written:

$$R_T \approx \frac{R_{sh}}{Z}(d + 2L_T) \tag{7.19}$$

where $L_T = (\rho_C/R_{sh})^{1/2}$ is the transfer length that expresses how much of the contact length (L) is used for the majority of the current flow over the metal–semiconductor interface. Thus, by measuring and plotting the total resistance R_T as a function of the spacing d, the contact resistance R_C can be extracted as illustrated by Fig. 7.6b. The slope of the obtained dependence represents the bulk resistivity, while the intercepts
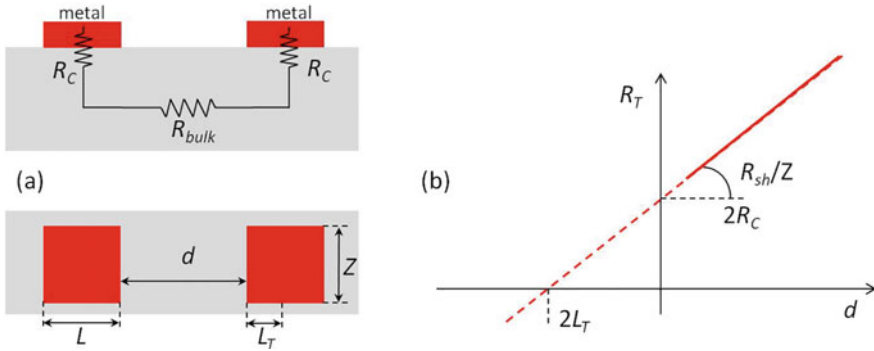
Fig. 7.6 a Metal–semiconductor contact in a simple case (side and top view) and equivalent resistance-based model of the structure. **b** TLM plot of total resistance versus the distance between the contacts

give the values of contact resistance and transfer length. In order to calculate specific contact resistivity, the effective contact area equal to $L_T Z$ should be used, so that $\rho_C = R_C L_T Z$.

The obvious drawbacks of TLM measurements include the indirect measurement of the specific contact resistivity calculated from the R_T measurements, so that at least three measurements are needed with different d for one data point. Furthermore, an isolation of the mesa structures by etching and patterning of the metal contacts is needed. There are some variations of the TLM method. For instance, there is a circular TLM where metal contacts are deposited on the semiconductor with circular gap, or spacing between the metal contacts. Such configuration provides isolation by default and does not require the additional etching process, as required by linear TLM.

7.7 Resistivity Measurements

Electrical resistivity is one of the basic material parameters and to determine the electrical resistivity is often crucial for characterization of semiconductors in research area as well as an industry. The resistivity of a material depends on several factors, including the material doping, processing, and environmental factors such as temperature and humidity. There are many characterization techniques used for the resistivity measurements of bulk materials as well as thin films. The choice of the appropriate technique depends upon the type of material, magnitude of the resistance, shape, and thickness of the material. Among these techniques, Four-Point Probes and van der Pauw are the most effective and widely used methods.

7.7.1 Four-Point Probe Method

Originally, 4-point probe method was invented in 1916 for geological applications (the measurements of the earth's resistivity) [5] and later in 1954 it was adopted for semiconductor wafer resistivity measurement [6]. Today this method is the most common technique used in semiconductor industry to measure the sheet resistance of shallow layers and thin films (as a result of epitaxy, ion implantation, diffusion, evaporation or sputtering) as well as the bulk resistivity of bare wafers.

Figure 7.7 shows schematics of the four-point probe measurements. The setup consists of 4 collinear arranged and evenly spaced metal rods, where a current (*I*) is conducted through the outer rods (1 and 4) and a voltage (*V*) is measured between the 2 inner rods (2 and 3). Such configuration allows to eliminate the contribution from the contact resistance, which is very sensitive to the surface conditions and the contact pressure. In general, when the lateral size of the measured sample is larger than the distance between the probes (*S*) and edge effects can be neglected, the resistivity can be written as:

$$\rho = 2\pi S \frac{V}{I} g \qquad (7.20)$$

Fig. 7.7 Schematics of a common collinear four-point probe geometry for resistivity measurements
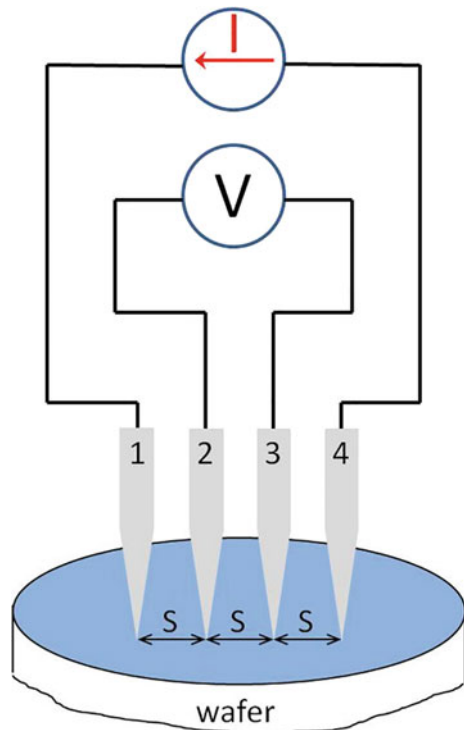
where g is a geometrical factor reflecting the relation between the sample thickness (t) and S. In the case of non-conducting bottom wafer surface boundary, it can be expressed as:

$$g = \frac{t/S}{2 \ln\left[\frac{\sinh(t/S)}{\sinh(t/2S)}\right]} \tag{7.21}$$

Obviously, $g \approx 1$ for a bulk sample when $t/S \gg 1$ and, hence, $\rho = 2\pi S V/I$. In its turn, for very thin samples ($t/S \ll 1$) the geometrical factor reduces to $g \approx t/S/(2\ln(2))$ and the resistivity becomes equal to $\rho = 4.532 tV/I$.

A potential problem in four-point probe measurements is that for correct measurements information about the thickness of the measured layer is required, especially in the case of multi-layered structures. Some variations of the four-point probe setup were proposed to overcome this issue. For example, it was demonstrated that dimensionality of the conductive path can be precisely estimated using modified four-point probe setup where three of the probes are rigidly fixed, while the position of the last probe is changeable [7]. Another serious limitation of four-point probe measurements for shallow junctions is the probe penetration depths from the pressure applied by hard and sharp probe tips and, for high probe loading, the probes can effectively short out a sufficiently shallow junction. Therefore, using "soft" probe tip materials (such as Hg and elastically-deformable metals) as well as the development of probes with significantly lower probe contact force is required in order to reduce this effect. Therefore, micro four-point probes setups [7], where the probe spacing can be as low as a few μm, have a potential because of minimal probe penetration and high special resolution allowing accurate measurements of shallow junctions with local conductivity variations due to, for example, selective ion implantation or non-uniform dopant activation.

7.7.2 Van Der Pauw Configuration

In order to minimize the errors in resistivity measurements due to edge effects and sample size/shape a van der Pauw Resistivity Method can be used. This variation of four-point probe method was proposed by van der Pauw in 1958 [8] with different shapes as shown in Fig. 7.8. In this configuration, the contacts should be sufficiently small and Ohmic.

Two of the probes are used to source current and the other two for measuring the voltage. The probes configuration is switched in order to calculate 2 different values of resistance. Thus, for the sample of arbitrary shape having four-point contacts along the periphery as illustrated by Fig. 7.8, the specific resistivity is given by:

$$\rho = \frac{\pi}{\ln 2} t \frac{R_{12,34} + R_{23,41}}{2} F \tag{7.22}$$
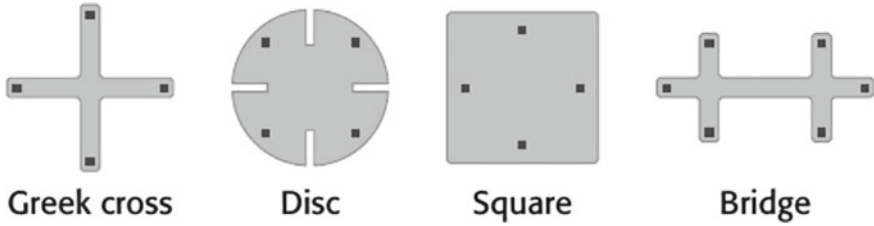
Fig. 7.8 Arbitrarily shaped sample with four contacts

where the resistance $R_{12,34} = V_{12}/I_{34}$ ($R_{23,41} = V_{23}/I_{41}$) and F is a geometrical correction factor which is a function only of the ratio $R_{12,34}/R_{23,41}$ and satisfying the relation [9]:

$$\cosh\left[\frac{(R_{12,34}/R_{23,41} - 1)}{(R_{12,34}/R_{23,41} + 1)}\frac{\ln 2}{F}\right] = \frac{1}{2}\exp\frac{\ln 2}{F} \tag{7.23}$$

The correction factor consists of three factors of (a) thickness (F1), (b) finite lateral dimension (F2), and (c) for probing near the edge of the sample (F3). These correction factors are provided for conductive and conductive substrates.

Most often the samples of symmetrical shape, such as the circle or the square, are used. In this case $F = 1$ and Eq. 7.22 is simplified to:

$$\rho = \frac{\pi}{\ln 2} t R_{12,34} = 4.532 t R_{12,34} \tag{7.24}$$

That is similar to the conventional four-point probe expression and a sheet resistance is $R_{sh} = \rho/t = 4.532 R_{12,34}$.

Practically, a series of 8 measurements is typically performed around the periphery of the sample, as illustrated by Fig. 7.9, to compensate for offsets. After that the results are combined mathematically to compute the resistivity.

Specifically, once all the voltage measurements are taken, two values of resistivity (ρ_A and ρ_B) as well as the average resistivity (ρ) are derived as follows:

$$\begin{cases} \rho_A = \frac{\pi}{\ln 2} t \frac{V_1 - V_2 + V_3 - V_4}{4I} F_A \\ \rho_B = \frac{\pi}{\ln 2} t \frac{V_5 - V_6 + V_7 - V_8}{4I} F_B \\ \rho = \frac{\rho_A + \rho_B}{2} \end{cases} \tag{7.25a–c}$$

where F_A and F_B are geometrical correction factors ($F_A = F_B = 1$ for perfect symmetry). It should be noted that van der Pauw configuration is used not only for the resistivity measurements, but it can be used also for other electrical measurements, such as Hall effect, as demonstrated in the next section.
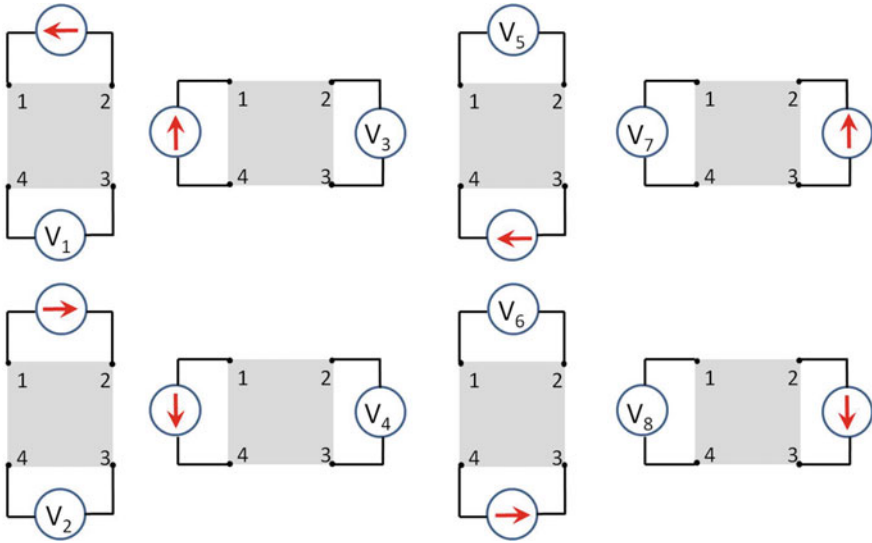
Fig. 7.9   Schematics illustrating a van der Pauw method derived for sample resistivity measurement

7.8   Hall Measurements

Resistivity measurements are often enough to draw conclusions about processes typically used in electronic device manufacturing, for example, to make a fast and reliable testing of doping, efficiency of an electrical isolation in semiconductors, etc. However, resistivity alone cannot provide a complete understanding of electrical transport in materials. In this section a method based on the Hall effect measurements is described which is indispensable characterization technique for determining majority carrier charge type in semiconductors as well as for accurate measurements of such important parameters as carrier density and mobility. Furthermore, performing temperature dependent Hall effects measurements allows, under certain circumstances, to determine the major donor (acceptor) energy levels.

7.8.1   Ordinary Hall Effect

The Hall effect was discovered by Edwin Herbert Hall in 1879 [10], 18 years before discovery of electron by Thompson. In this experiment the sample was placed in a magnetic field (B) directed perpendicular the sample surface and to the current flow (I). It was found that a combination of a magnetic field through a sample and a current along the length of the sample creates a potential proportional to the current and to the magnetic field and developed across the material in a direction perpendicular to both the current and to the magnetic field. Using the measurements, Hall showed for

the first time that it is possible to determine the sign of charge carriers in a conductor. Schematic illustration of the Hall effect in a slab of *n*-type semiconductor is shown in Fig. 7.10. Typically, for the Hall measurements a van der Pauw method is applied and four Ohmic contacts are formed, where a current is applied through the opposing pair of contacts, while the voltage is measured across the remaining pair of contacts.

As the result, electrons drift to the one side of the sample, while the anther side becomes positively charged. Charge redistribution creates an electric field and when equilibrium is established, a potential difference between the two sides of the sample can be measured. This potential difference is called Hall voltage (V_H) and it can be expressed as:

$$V_H = \frac{IB}{ntq} \tag{7.26}$$

where n is a carrier concentration, q is a elementary charge and t is a thickness of the sample. The semiconductor type can be determined from the sign of the V_H. In its turn, n can be easily calculated from the Eq. 7.26 since all three quantities the I, B and V_H are measurable.

Hall electric field intensity E_H is proportional to the product of current density J and magnetic field intensity B, namely:

$$E_H = R_H BJ \tag{7.27}$$

where R_H is called Hall coefficient. After defining the direction of the Hall electric field, the Hall coefficient will also be $+$ or $-$. In Fig. 7.1, the current is set to along the $+$ X direction and the magnetic field is along the $+$ Z direction. Then for the p-type sample, the Hall electric field is along the $+$ Y direction in this case, where $E_H > 0$, $R_H > 0$. For the n-type, similarly, $E_H < 0$ and $R_H < 0$. From this perspective, the sign of the Hall coefficient can help to determine the conductivity type of the

sample. With the assumptions of energy-independent scattering mechanism relaxed, the expression for the hole and electron density become [10, 11]:

$$p = r/(q R_{\mathrm{H}}), \quad n = -r/(q R_{\mathrm{H}}) \tag{7.28}$$

where r_H is the Hall factor, defined by $r_H = \langle \tau^2 \rangle / \langle \tau \rangle^2$, where τ is the relaxation time for a type of scattering in a semiconductor (as defined in Eqs. 7.1 and 7.2). The Hall factor varies due to the different scattering mechanisms in a semiconductor and its value lies in range of 1–2. A better estimation for r could be given for lattice scattering, $r = 3\pi/8 = 1.18$, for impurity scattering $r = 315\pi/512 = 1.93$, and for neutral impurity scattering $r = 1$[11, 12]. The r can be also affected by applied magnetic field and temperature for measuring R_{H} and it is considered as $r_H = R_{\mathrm{H}}$ $(B)/ R_{\mathrm{H}} (B = \infty)$ where for high magnetic field r_H approaches to 1. As an example, r_H in n-type GaAs varies from 1.17 at $B = 0.01$ T (lattice scattering), to 1.006 at $B = 83$ kg [13]. In practice, typical magnetic field values for Hall measurements are between 0.05 and 1 T. The Hall mobility μ_{H}, defined by $\mu_{\mathrm{H}} = |R_{\mathrm{H}}|/\rho = |R_{\mathrm{H}}|\sigma$ differs from the conductivity mobility. In most of the Hall measurement cases, r is treated as the unity but this needs to be simulated [14]. In many heterostructures, there is a lattice mismatch which leads inducing strain and modifying the bandgap. These systems are multi-valleyed, and r_H may significantly differ from unity. As an example, for the case of heavy and light holes valence (hh and lh, respectively) band system, the r_H could be derived from r_{hh} and r_{lh} from the following relation [15]:

$$r_H = \mathrm{f_{hh}r_{hh}}\mu_{\mathrm{hh}}^2 + \mathrm{f_{lh}r_{lh}}\mu_{\mathrm{lh}}^2 / (\mathrm{f_{hh}}\mu_{\mathrm{hh}} + \mathrm{f_{lh}}\mu_{\mathrm{lh}})^2 \tag{7.29}$$

where $\mathrm{f_{hh}}$ and $\mathrm{f_{lh}}$ are the carrier fractions in hh and lh bands, with mobility values of μ_{hh} and μ_{lh}. Mathematically, r, f quantities are written as:

$$r_{\mathrm{hh}} = \langle \tau_{\mathrm{hh}}^2 \rangle / \langle \tau_{\mathrm{hh}} \rangle^2 \tag{7.30}$$

$$\mu_{hh} = e \langle \tau_{hh} \rangle / m_{hh}^* \tag{7.31}$$

$$f_{\mathrm{hh}} = n_{\mathrm{hh}}/(n_{\mathrm{hh}} + n_{\mathrm{lh}}) \tag{7.32}$$

A mean value of relaxation time, τ is given from: [16]:

$$\langle \tau_{hh}^n \rangle = \int_0^\infty \tau^n(E) D_{hh}(E) E_{\exp}(-E/kT) \mathrm{d}E / \int_0^\infty D_{\mathrm{hh}}(E) E_{\exp}(-E/kT) \mathrm{d}E \tag{7.33}$$

where D_{hh} (E) is the hh density of states. A similar expression can be also written for the lh relaxation time.

Fig. 7.11 Dependence of
Hall factor on Ge
composition in $Si_{1-x}Ge_x$
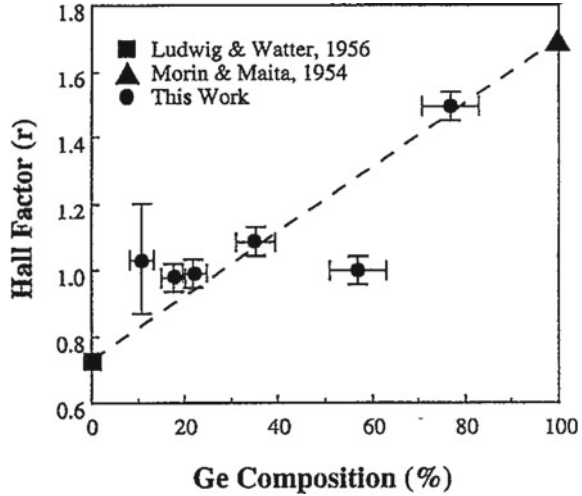
Another example which is discussed here is for compressive strained $Si_{1-x}Ge_x$
alloys. As a result of the strain, the hh and lh energy band are lifted and the curvature
of the bands is changed. Therefore, the carrier scattering is affected. Thus, the r_H-
values in $Si_{1-x}Ge_x$ system are significantly affected in presence of strain as shown in
Fig. 7.11 [17].

The behavior of r_H for holes for strain and relaxed $Si_{1-x}Ge_x$ have been also
calculated by using the relaxation time approximation in Monte Carlo simulation
[18]. The outcome demonstrated that the r_H has a strong relationship with strain
which results in a change in anisotropy of the valence band. It is also demonstrated
that the direction of strain and magnetic field have also impact on the r_H. There are
also reports which demonstrate the effect of doping concentration in $Si_{1-x}Ge_x$ on r_H
[19] when acoustic-phonon, nonpolar optical-phonon, alloy, and ionized impurity
scattering have taken into account.

As we discussed earlier, for device application the carrier mobility (or drift
mobility) is the very important factor.

After deduction, we have $\mu_H = r_H\,\mu_p$, and $\mu_H = r_H\,\mu_n$ for extrinsic *p*- and *n*-type
semiconductors, respectively.

Taking into account that resistivity can be determining by use of the van der Pauw
resistivity measurement technique, mobility (μ_H) can be calculated as:

$$\mu_H = \frac{1}{nq\rho} = \frac{R_H}{\rho} \tag{7.34}$$

Both Hall coefficient and μ_H are determined from Hall measurements. It should
be noted that despite a consideration above was performed for electrons in the doped
n-type semiconductor, a similar formalism can be applied also for the holes. Thus,
for a semiconductor with significant concentrations of both types of carriers, a Hall
coefficient is:

$$R_H = \frac{p\mu_h^2 - n\mu_e^2}{q(p\mu_h + n\mu_e)^2} \tag{7.35}$$

where p and n are the hole and electron concentrations, while μ_h and μ_e are the hole and electron mobilities, respectively. The magnitude and sign of R_H, in that case, depends on both the drift mobility ratio and the concentrations of holes and electrons. Therefore, if holes dominate ($p\mu_h > n\mu_e$), $R_H > 0$ and the material is said to be p-type, while if $R_H < 0$ (as for simple metals), the material is n-type.

Note that there are factors affecting accuracy of the Hall measurements. Some of them are similar to those discussed previously for resistivity measurements such as quality and size of the ohmic contacts, sample uniformity, etc. In addition, photo-conductive effects should be minimized that can be done by measuring in the dark environment. Special attention should be paid to the temperature variations during the measurements because of the temperature affects both the mobility and the concentration of charge carriers in semiconductors. Effects of the temperature on the Hall measurements and how it can be exploited are discussed in the next section.

7.8.2 Temperature Dependent Hall Effect

As it has been mentioned in the previous section, the temperature affects both mobility and concentration of charge carriers in semiconductors. Two major mechanisms dominate the temperature dependence of μ i.e. phonon scattering (scattering due to lattice vibrations) and ionized impurity scattering. In general, ionized impurity scattering dominates at low temperatures, while phonon scattering prevails at higher temperatures.

Impurity and phonon scattering have an opposite behavior with the temperature, so that the first one increases with temperature as $T^{3/2}$, while the latter one decreases as $T^{-3/2}$ giving the temperature dependence of a total mobility as one shown schematically in Fig. 7.11a. The total mobility follows Matthiessen's rule and is determined by the process with the lowest mobility:

$$\frac{1}{\mu} = \frac{1}{\mu_{impurity}} + \frac{1}{\mu_{lattice}} \tag{7.36}$$

Typically lattice scattering dominates over the entire temperature range for semiconductors with low dopant concentration. It should be noted that, at constant temperature the mobility decreases with increasing dopant concentration because of greater concentration of impurities.

The temperature dependence of free charge carrier concentration (n) in doped semiconductor is schematically shown in Fig. 7.12b. For low temperatures n is determined by ionization of dopants and then n exhibits a saturation for the temperature range where all the dopants are ionized. For high temperatures intrinsic regime

dominates and the charge carrier concentration is determined by the intrinsic carrier concentration (n_i):

$$n_i = \sqrt{N_C N_V} \exp\left(-\frac{E_g}{2k_b T}\right) \tag{7.37}$$

where N_C and N_V are effective density of states for the conduction and valence bands, respectively. In this temperature range Arrhenius plot of the carrier concentration data can be used to determine energy band-gap (E_g) according to the Eq. 7.37 and the slope of the Arrhenius plot for that temperature range is $-E_g/(2\,k_b)$.

The saturation regime corresponds to the case when all the dopants are ionized i.e. $k_b T \gg E_d$, where E_d is a dopant energy level. At the same time, the temperature is not high enough, so that the contribution from the intrinsic carrier concentration can be neglected (see Fig. 7.12b). This regime is often called "extrinsic" and n is determined by a net carrier concentration equal to the difference between donor and acceptor concentrations i.e. $n = |N_d - N_a|$.

Behavior of carrier concentration in the doped semiconductor at low temperatures is more complicated and it depends whether or not the material is compensated. For example, for *n*-type semiconductor having $N_d > N_a$ the temperature dependence of n exhibits two distinct regions in the ionization (freeze-out) part of the curve, as shown in Fig. 7.12b. Specifically, for the temperature range where electron concentration is below the donor concentration but still larger than acceptor concentration, i.e. $N_d \gg n \gg N_a$, the activation energy is equal half of the ionization energy of the donors (*half slope* $= -E_d/(2\,k_b)$). As temperature decreases further, n becomes small such that $n \ll N_d, N_a$ and under such conditions the Arrhenius plot of the electron concentration demonstrate the "*full slope*" equal to $-E_d/k_b$.

Thus, temperature variable Hall measurements can provide information about band-gaps as well as major dopant energy levels in semiconductors. However, it should be noted that doping level plays a substantial role on the temperature dependence of the free carrier concentration. For example, lightly doped semiconductors
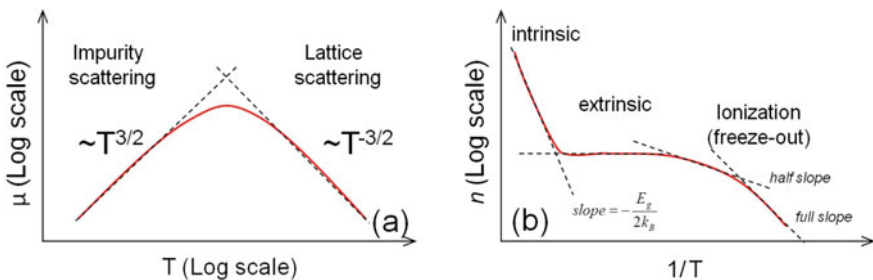

Fig. 7.12 **a** Approximate temperature dependence of mobility with both lattice and impurity scattering in a double logarithmic scale. **b** Charge carrier concentration as a function of reciprocal temperature in a doped semiconductor showing intrinsic, extrinsic and ionization (freeze-out) regions

exhibit freeze-out behavior already at relatively high temperatures and this region moves towards the lower temperatures with increasing doping. Moreover, in heavily doped (degenerated) semiconductors there are free carriers even at zero Kelvin so, they do not show freeze-out effects.

7.8.3 Quantum Hall Effect (QHE)

The ordinary Hall effect theory predicts a linear dependence of Hall voltage on the magnetic field strength, while longitudinal resistance is independent of the magnetic field and it approaches zero as scattering processes are not important. However, it behaves drastically different as structure size scales down and quantum effects come into play. Historically, quantum Hall effect was discovered by Klaus von Klitzing in 1980 [20]—about a hundred years after the discovery of the classic Hall effect. The Nobel Prize was awarded for this work just five years later. The quantized Hall effect is observed in a two-dimensional electron gas (2DEG) for the strong magnetic fields (B) and low temperatures, so that $\hbar q B/m^* \gg k_b T$ (m^* is the effective mass of the electron, $\hbar$ is the Plank constant and q is the elemental charge). The 2DEG needed to observe the QHE can be realized in various types of semiconducting heterostructure devices where the electrons can be confined in a plane. For, example, one of the common methods to form 2DEG is using GaAs/n-AlGaAs heterostructure as illustrated by Fig. 7.13a showing the band bending in this system at equilibrium. AlGaAs has a larger bandgap compared to GaAs, and that the Fermi level in the AlGaAs lies just below the conduction band edge for the high n-doping. Therefore, due to electron rearrangement the conduction band on the GaAs side of the interface has a very narrow triangular potential well, whose minimum lies below the Fermi level. In this system the movement of the electrons is restricted in one spatial direction, while they are free to move only within a plane, so 2DEG forms. A spatial confinement in 2DEG leads to the formation of discrete Landau levels with the gap between each level proportional to the magnetic field:

$$E_N = \hbar \frac{qB}{m^*} \left(N + \frac{1}{2} \right), \quad N = 0, 1, 2, \ldots \tag{7.38}$$

where N is the quantum number of the Landau level. Further, the unavoidable disorder caused by the impurities, which always present in the system, broadens the Landau levels in Landau sub-bands. As a result, even at extremely low temperatures the Landau levels are not delta functions and two different kinds of electronic states (localized and extended states) formed.

Quantization of the energy levels in 2DEG leads to the deviation of the Hall and longitudinal resistances from those in the ordinary Hall effect. The typical dependences of the Hall coefficient (R_H) and longitudinal (R_{xx}) resistances on the magnetic field in 2DEG are shown in Fig. 7.13b. Note that in a two-dimensional system, the Hall resistance is equal to the Hall coefficient resistivity ($\rho_H = R_H$) which is in contrast to
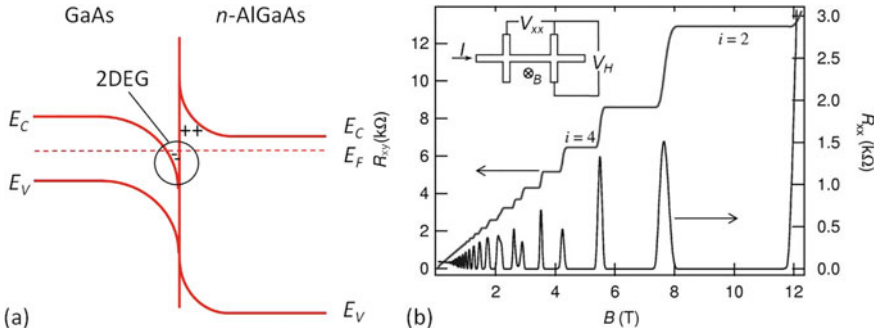
Fig. 7.13 a The energy band diagram illustrating band bending in GaAs/n-AlGaAs heterostructure and formation of two-dimensional electron gas (2DEG) at the interface. **b** The Hall (R_H, left-hand scale) and longitudinal (R_{xx}, right-hand scale) resistances of 2DEG as a function of the magnetic field illustrating the integer quantum Hall effect. A schematic of the Hall geometry is also shown in the inset of the panel **b** [21]

the three-dimensional case, where geometrical corrections are required. It might be seen from the figure that Hall resistance shows an abrupt stepwise dependence with increasing magnetic field, i.e. it is constant over a range of magnetic fields. In its turn, the longitudinal resistance is zero but becomes finite during transitions between the plateaus of the ρ_H. The plateaus of the Hall resistance take on values of:

$$\rho_H = \frac{2\pi\hbar}{q^2\nu}, \quad \nu = 1, 2, 3, \ldots \tag{7.39}$$

where ν is an integer. This is so-called integer quantum Hall effect. Remarkable, that the value of ρ_H is measured to be an integer to an extraordinary accuracy and one of the most impressive applications of the quantum Hall effect is metrology. Indeed, the value of the quantum Hall resistance depends only on the relation between two fundamental constants and it is $2\pi\hbar/q^2 = 25{,}812.807\ \Omega$ on the first plateau. This value is also often called as the von Klitzing constant and it has been used as the primary resistance standard by the most National Metrology Institutes since 1990 [22]. Thus, the resistance standard was redefined in terms of the two fundamental constants and it replaced the artifact standards represented by wire resistors.

It should be mentioned that under specific conditions, the value of ν in the Eq. 7.39 can take on rational fractional values, such as $\nu = 1/3, 2/5, \ldots$ This is so-called fractional quantum Hall effect which was discovered in 1982 for certain devices with reduced disorder [23] and the Nobel Prize was awarded for this discovery in 1998.

7.9 Capacitance Voltage Measurements

Another type of electrical characterization of semiconductors is based on capacitance–voltage measurements. This technique was developed by Hilibrand and Gold in 1960 [24] and after that capacitance–voltage (*CV*) measurements became one of the fundamental characterization techniques for *pn*-homo- and heterojunctions, Schottky diodes, and also metal–insulator-semiconductor interfaces, for instance MOS structures. Nowadays, *CV* technique is intensively used in both industry and research for the fast and reliable characterization of semiconductor structures. It can be used for determining several important parameters such as doping concentration, built-in voltages and density of interface states. In contract to the electrical methods described above (Hall and four-point probe measurements), a material prepared for *CV* measurements requires rectifying junction (such as Schottky contact), which can be fabricated, for example, by depositing a metal with appropriate work function.

7.9.1 Theory of CV

The basic principle of *CV* measurements is that the width of the depletion region of a semiconductor device depends on the applied voltage. Assuming the abrupt edges of the depletion region in conjunction with the complete dopant ionization (depletion region approximation), the depletion region behaves as a parallel plate capacitor. Figure 7.14a shows a schematic of a reverse-biased Schottky junction, formed between a metal and a *p*-type semiconductor, illustrating the formation of the space charge region in the depletion area with a width of *W*.

The junction capacitance (*C*) is defined by the variations of fixed charges in the depletion region (ΔQ) caused by the small amplitude high frequency oscillations of the applied voltage (ΔV):
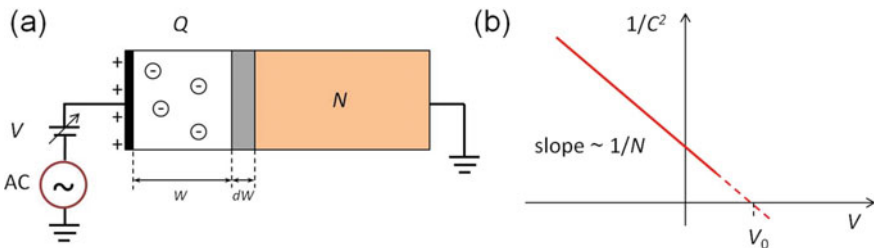

Fig. 7.14 **a** Illustration of a reverse-biased Schottky junction between the metal and p-type semiconductor. **b** A schematic illustration of $1/C^2 = f(V)$ plot which can yield the built-in potential (V_0) and doping density (N)

$$C = \frac{\Delta Q}{\Delta V} = A\frac{\varepsilon_S}{W} = A\sqrt{\frac{q\varepsilon_S N}{2(V_0 - V)}} \qquad (7.40)$$

where A is a diode cross section area, N is the dopant concentration and V is the applied voltage (positive/negative for forward/reverse biases, respectively). The Eq. 7.40 can be rearranged to the function of reciprocal capacitance squared versus the applied voltage change, i.e. $1/C^2 = f(V)$ (in Eq. 7.41), which is typically used for plotting and interpretation of *CV* results. Evidently that $1/C^2 = f(V)$ should be a straight line for a uniformly doped material, where the slope is reversely proportional to the doping concentration:

$$\frac{d(1/C^2)}{dV} = -\frac{2}{A^2 q\varepsilon_S N} \qquad (7.41)$$

In its turn, the built-in voltage (V_0) can be extracted from the extrapolated intersection of the straight line of $1/C^2 = f(V)$ with voltage axis ($1/C^2 = 0$) as illustrated in Fig. 7.14b, where the typical $1/C^2(V)$ function is shown for a uniformly doped semiconductor. It should also be noted that for non-uniform doping, the $1/C^2(V)$ function is not linear and Eq. 7.41 then gives a possibility to determine the depth profile of the dopants. It is important to note that, in contrast to the Hall measurements, where a free carrier concentration is determined, the dopant concentration derived from *CV* measurements is attributed to the fixed charge of ionized dopants.

There are some additional factors complicating the interpretation of the results of *CV* measurements. For example, it should be kept in mind that in contrast to the reverse bias conditions, where only fixed charges contribute to the junction depletion capacitance, for forward bias a diffusion capacitance needs to be taken into account due to the large concentration of mobile carriers. Another limitation is related to the resolving the doping changes at the distances less than the Debye length (L_D). L_D is a characteristic length for semiconductors defined as:

$$L_D = \sqrt{\frac{\varepsilon_S k_b T}{q^2 N}} \qquad (7.42)$$

and it sets the spatial limit (resolution) of measured profile. In its turn, a maximal depth of analysis is limited by the width of the depletion region that can be achieved without inducing a breakdown due to high forward/reverse currents and this can be very restrictive in highly doped materials, where the depletion depths are small.

In order to overcome the fundamental depth limitations and measure relatively deep dopant profiles, an electrochemical *CV* method is often used [25]. In this technique the metal is replaced by an electrolyte and the electrolyte/semiconductor interface behaves as a Schottky junction if the electrolyte is concentrated enough. Thus, the capacitance of the Schottky barrier formed at the electrolyte/semiconductor interface can be measured, while the electrochemical anodic dissolution reaction is used

to remove the material at a controlled rate. Hence, the carrier concentration as a function of depth can be obtained by a repetition of this etch/measure cycle.

In theory, profiles of unlimited depth can be obtained by this method. However, the obvious disadvantage of this technique is that the method is destructive. Furthermore, it is difficult to measure and control the diode area with high enough precision, and the depth resolution is limited by the etch non-uniformity. Despite these limitations, electrochemical *CV* method is widely used for depth profiling of dopants in various semiconductors including two-dimensional electron gas heterostructures and delta-doped layers [26].

A technologically very important usage of *CV* measurements is when insulating structures are measured, for instance a metal–insulator-semiconductor (MIS) structure. By far the most fabricated semiconductor device today is the metal–oxide–semiconductor field effect transistor (MOSFET), where the electrically insulating gate stack (MOS) is a key building block. The down scaling of the MOSFETs is driving the whole electronic revolution and the possibility to assess the gate properties through *CV* measurements is a key ingredients in this development. By studying how the capacitance changes with applied voltage for a MIS structure it is possible to extract invaluable information about gate leakage currents, dielectric properties of the insulator and various electron and hole traps and other charges at the interfaces and in the insulator bulk. The behavior of the MIS junction under reverse and forward bias is more complicated than the metal–semiconductor junction and, because of its importance for micro- and nanoelectronics several text books dedicated to this subject exist, for instance [27]. Here, we will give a condensed delineation of CV measurements of MIS structures.

7.9.2 Process of CV Structure and Setup

For CV measurement of MIS structures, typically, an oxide layer is deposited on a high-quality silicon substrate, and then this structure is usually metallized with a thin film of aluminum. This thin metal layer is patterned to form MOS capacitors with specified areas. The electrically insulated top contact of aluminum (or, alternatively, of highly doped poly-silicon) forms one electrode of a parallel capacitor and can be contacted from the top by a probe, for instance made of tungsten. The backside of the semiconductor has an Ohmic contact and is electrically connected by a conductive "chuck". Typically, the backside contact is held at ground potential and the frontside contact is biased negatively, or positively. In CV measurements, information of both static, or quasi static, voltages and transient response of high frequency voltage variations are utilized.

Figure 7.15 shows a schematic of a CV measurement setup. An AC impedance meter, or LCR meter (for inductance [L], capacitance [C], resistance [R]), measures complex impedances typically with a frequency range of 1 kHz to 10 MHz. AC impedance is measured by applying an AC voltage out of the high current terminal (HCUR). The low current terminal (LCUR) measures the current through the device,

Fig. 7.15 A basic setup for CV measurements
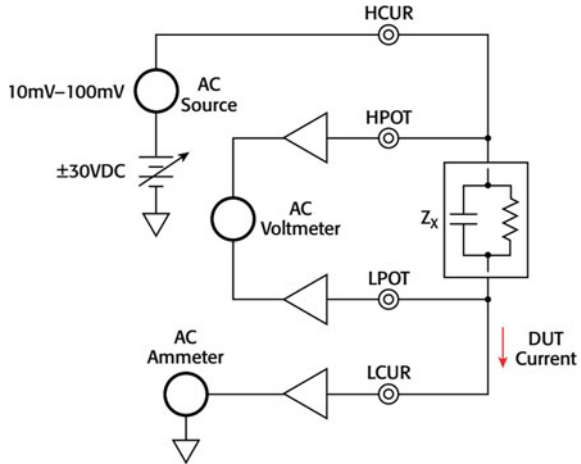
while the high and low potential terminals (HPOT and LPOT) measure the voltage across the device. Both the voltage and current are measured in a phase-locked manner which accurately distinguishes the phase angle between them. Finally, AC impedance parameter can be derived from the obtained amplitude and phase angle.

7.9.3 Theory of CV Measurements

Figure 7.16 illustrates the band diagram and the charge distribution in a MOS structure. Due to the established potential close to the interface between the oxide and the semiconductor (ψ_s), the gate voltage (V_G) is written as:

$$V_G = V_{ox} + \psi_s = \frac{-Q_s}{C_{ox}} + \psi_s \quad \text{where} \quad C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \quad \text{and} \quad C_s = \frac{d(-Q_s)}{d\psi_s} \quad (7.43)$$

It is possible to analytically calculate the charge and capacitance from CV measurements. The potential ψ can be written from Poisson equation as following:

$$\frac{d^2\psi}{dx} = -\frac{q}{\varepsilon_{Si}}[p(x) - n(x) + N_D - N_A] \quad (7.44)$$

where the electron and hole concentrations can be rewritten from intrinsic carrier concentration as following:

$$\frac{d^2\psi}{dx} = -\frac{q}{\varepsilon_{Si}}[N_A\left(e^{-\frac{q\psi}{kT}} - 1\right) - \frac{n_i^2}{N_A}\left(e^{\frac{q\psi}{kT}} - 1\right)] \quad (7.45)$$

The electric field can then be written as:

Fig. 7.16 Band-diagram and charge distribution of a MOS
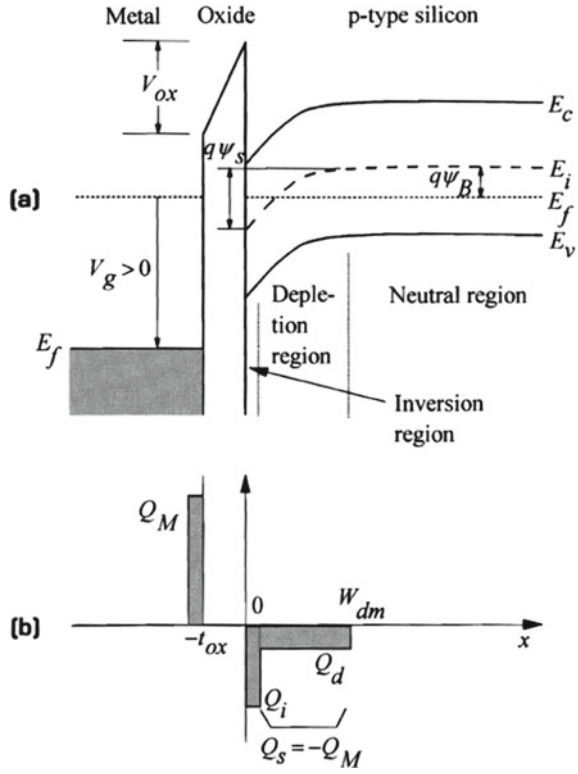
$$E^2(x) = \left(\frac{d\psi}{dx}\right)^2 = -\frac{2kTN_A}{\varepsilon_{Si}}\left[\left(e^{-\frac{q\psi}{kT}} + \frac{q\psi}{kT} - 1\right) + \frac{n_i^2}{N_A^2}\left(e^{\frac{q\psi}{kT}} - \frac{q\psi}{kT} - 1\right)\right]$$

(7.46)

Finally, the charge and capacitance in the semiconductor are obtained and the charge is plotted as a function of the electrostatic potential in Fig. 7.17.

$$Q_s = -\varepsilon_{Si}E_s$$
$$= \pm\sqrt{2\varepsilon_{Si}kTN_A}\left[\left(e^{-\frac{q\psi_s}{kT}} + \frac{q\psi_s}{kT} - 1\right) + \frac{n_i^2}{N_A^2}\left(e^{\frac{q\psi_s}{kT}} - \frac{q\psi_s}{kT} - 1\right)\right]^{\frac{1}{2}}$$

(7.47)

$$C_s = \frac{d(-Q_s)}{d\psi_s} = \pm\sqrt{\frac{2\varepsilon_{Si}q^2N_A}{kT}}\frac{\left(1 - e^{-\frac{q\psi_s}{kT}}\right) + \frac{n_i^2}{N_A^2}\left(e^{\frac{q\psi_s}{kT}} - 1\right)}{2\left[\left(e^{-\frac{q\psi_s}{kT}} + \frac{q\psi_s}{kT} - 1\right) + \frac{n_i^2}{N_A^2}\left(e^{\frac{q\psi_s}{kT}} - \frac{q\psi_s}{kT} - 1\right)\right]^{\frac{1}{2}}}$$

(7.48)

Fig. 7.17 Schematic of total charge density containing fixed and mobile charges in Si versus surface potential, ψ_s in a p-type MOS structure
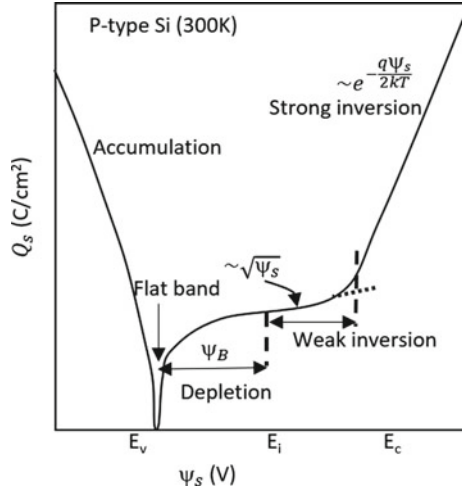
7.9.4 Details About CV Measurements

Electronically, it is relatively easy to measure very small values of the capacitance of a structure, such as a MOS junction. This makes it possible to apply the CV technique to study how small amounts charges can affect the static and transient junction properties. The CV technique is applied to determine mainly capacitance, thickness or the dielectric constant of the dielectric layer, threshold voltage, flatband voltage, surface potential, oxide charge, interface state density, and work function of the gate material.

As seen in Fig. 7.18, there are different types of charges, as well as interfacial states, which have influence on the CV measurements of Si MOSFET, as listed below:

- Mobile ions: These types of ions are mainly acceptors and can move under the influence of electric field and temperature. (Mobile impurities are highly unwanted in devices and may lead to reliability issues.)

Fig. 7.18 Schematic of different charges in a MOS structure
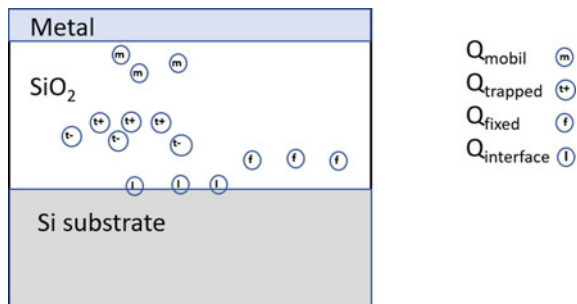
- Trapped charges: these types of charges are created during charge injection for example by radiation. They consist of electrons and holes trapped at imperfections in the oxide, or at the interface. Trapped charges can be emitted to the valence or conduction bands, depending on the temperature and time.
- Fixes charges: these types of charges are commonly positive and located at, or in the vicinity of the Si/SiO_2 interface. They are generated during oxidation and their concentration depends also strongly on the post annealing treatment.
- Interface states: these are atomic/ionic defects at the Si/SiO_2 interface and the concentration depends also on post annealing treatment. At low frequency CV measurements, the interface states establish a capacitance in parallel with semiconductor capacitance.

Generally, CV measurements of a MOS capacitor structure consist of different bias regions depending on the semiconductor doping and the applied gate voltage, as shown in Fig. 7.17. The regions are referred to (i) accumulation, (ii) depletion, and (iii) inversion and they are explained in the following for a MOS structure built on a p-type semiconductor.

Accumulation Region.

The accumulation region of the CV curve is reached when the gate is negatively biased compared to the backside contact. In this case, majority carriers (holes) will be attracted by the negative potential and move towards the gate. Since they cannot enter the insulating oxide layer, they will be accumulated just under this layer. When the applied voltage is sufficiently negative the capacitance is constant and the slope of CV plot is mainly flat. In this condition, the thickness of oxide layer can be calculated from the oxide capacitance. It is important to mention here that for very thin oxide the CV plot does not have a flat slope and the results becomes inconsistent with the true value.

Depletion Region.

When the gate voltage sweeps to positive values, the holes are repelled from the region underneath the oxide and a depleted zone is established in the semiconductor. In this voltage region, the oxide capacitance and the depletion capacitance can be measured in series. For increasing positive gate voltage, the depletion region extends deeper into the semiconductor (compare to depletion width, W, in Fig. 7.19). As the voltage increases, the depletion capacitance, as well as the total capacitance, becomes smaller. Thus, in the depletion region, the slope of CV plot is negative.

Inversion Region.

When the applied gate voltage is increased further, the minority carriers (electrons), present at small concentrations, begins to sense the attractive force from the positive bias. The minority carriers then move toward the semiconductor-oxide interface, forming an inversion of the semiconductor from p- to n-type in a thin layer under the oxide. Further increase in the positive bias does no longer lead to an extension of the
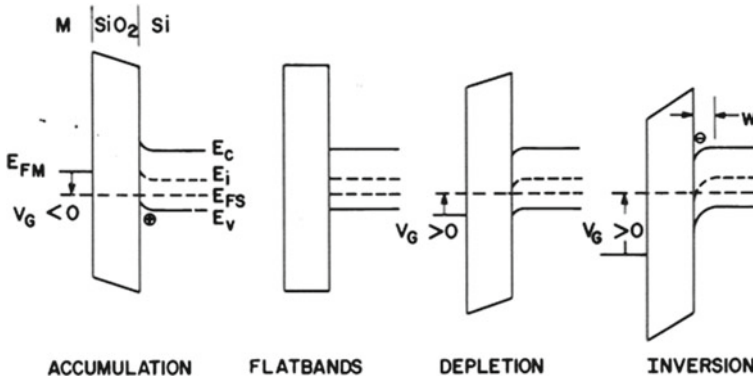
Fig. 7.19   Band-diagram of MOS with p-type semiconductor under different applied gate voltages

depleted layer, but increases instead the concentration of electrons in the *inversion layer*.

The inversion-charge generation, however, is typically slower than a measurement frequency of 100 kHz, used for high frequency CV measurement. The average time for generating an inversion charge is obtained from:

$$\text{Average time} \approx 10\tau_g N_a / n_i \tag{7.49}$$

where τ_g is the generation lifetime, N_a is the acceptor doping level, and n_i the intrinsic carrier concentration cm^{-3}.

For example, for a semiconductor with a doping level of a 10^{15} cm^{-3} and microsecond generation lifetime, electron–hole-pair (ehp) generation cannot respond sufficiently fast to the high frequency signal. As a result, when the depletion region obtains its maximum depth, the measured capacitance by the HF-CV analyzer still relies on the majority carrier position and distribution. The measured total capacitance has contributions of both oxide capacitance and maximum depletion capacitance in series and the slope of the CV plot is flat.

7.9.5   CV Curves and Interpretation

For CV measurements the applied bias is a slowly varying potential in the order of several V, that can be considered static. Added on this slowly varying "DC" voltage, a small signal AC voltage in the order of mV, is superimposed to measure the capacitance. Capacitance voltage measurements can be performed by sweeping the applied voltage either from accumulation to inversion (from − to + voltage for p-type, or from + to − for n-type), or inversion to accumulation (+ to − voltage for p-type; − to + for n-type).

A quasistatic CV measurement is performed by sweeping the bias voltage applied to a MOS capacitor, leading to the change of the semiconductor surface from inversion to depletion and later to accumulation, as shown in Fig. 7.20. In an ideal case, there is no conduction current passes through the oxide layer, which requires an oxide that does not leak. In the case, when the semiconductor surface is either in accumulation or inversion, a layer of charged mobile carriers exits beneath the oxide. Therefore, the measured capacitance (C_{max}) is only the capacitance of the oxide layer (C_{ox}) which is mainly dependent on the area of frontside contact, permittivity of the oxide layer and oxide thickness. In contrast, when the semiconductor is depleted, there are no mobile carriers underneath the oxide layer, i.e., at the Si/SiO$_2$ interface, although mobile carriers are still generated underneath the depletion region. As a result, the measured total capacitance consists of the oxide capacitance and the capacitance of the depletion layer connected in series and this total capacitance will then be less than C_{max}. When the voltage is swept from inversion to accumulation, during depletion the capacitance is decreased from C_{max} to a minimum value, C_{min}, which corresponds to maximum depletion layer width. As the applied bias moves towards the accumulation region the capacitance increases again. The minimum capacitance, C_{min}, can be written in terms of semiconductor capacitance, C_s, and C_{ox} as:

$$\frac{1}{C_{min}} = \frac{1}{C_{ox}} + \frac{1}{C_s} = \frac{t_{ox}}{\varepsilon_{ox}} + \frac{t_{dep}^{max}}{\varepsilon_s} \qquad (7.50)$$

where t_{0x} and ε_{ox} are oxide thickness and the dielectric constant of the oxide, while t_{dep}^{max} and ε_s are the maximum depletion thickness and the dielectric constant of the depletion layer.

When the AC small measurement signal has high frequency, the measured capacitance in inversion is not the same as in the quasistatic one. The reason behind is when an applied AC voltage is in the megahertz range, the response of the inversion layer is too slow and cannot follow the signal. For sufficiently high frequencies, this delay is similar also for ionized dopant impurity atoms and the inversion layer is in

Fig. 7.20 Low- and high-frequency CV plot
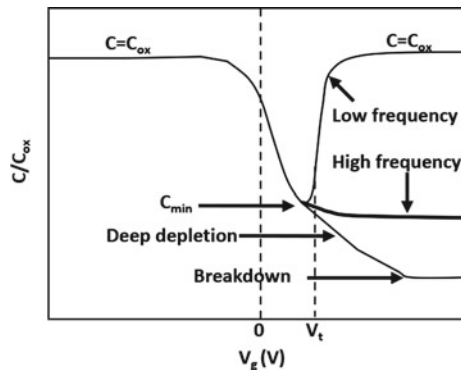
such a case unchanged with respect to the AC component of the bias. Figure 7.18 shows typical CV measurements at high and low measurement frequencies.

One application of using the deep depletion region is the possibility to find out the time constant of carrier generation-recombination processes. In general, the generation-recombination is slow for a high-quality substrate and, as result, the minority carrier lifetime is long, while in the presence of defects, the minority carrier lifetime is remarkably reduced. Furthermore, when minority carrier lifetime becomes shorter, the deep depletion region will be absent and the MOS capacitance increases in inversion, as shown in Fig. 7.18. This allows minority carrier lifetime to be determined by studying the dependence of inversion capacitance on sweep rate. The minority carrier lifetime can be estimated from the derivative of this function extrapolated back to equilibrium conditions.

Many other varieties of CV measurements have been explored in the field of semiconductors. For instance, varying the measurement frequency can in many cases be exchanged by varying the measurement temperature, when studying transient effects related to capturing and emission of mobile charge. Surface illumination can also be explored to provide photo-generated minority carriers which aids in the formation of the inversion layer, which can be useful for wide bandgap semiconductors, where the generation times are typically orders of magnitude longer than for silicon. In the final section of this chapter we will describe an often used CV method to study capture and emission processes in more detail.

7.10 Deep Level Transient Spectroscopy

Deep Level Transient Spectroscopy (DLTS) is a powerful technique to characterize the deep level impurities in semiconductors. DLTS was developed by Lang [28, 29] in 1974, and it is a technique which is based on capacitance transient thermal scanning. This technique applies the capacitance of a Schottky or p-n diode as a probe to display the variations in charge state of a certain deep center. The technique is considered as spectroscopic because it can resolve signals from different traps. DLTS has the benefit over other techniques since the technique is able to distinguish between majority- and minority-carrier traps [30] and it provides e.g. the concentrations, energy and capture rates of those traps. Using the DLTS technique, the deep levels are filled with free carriers induced by electrical or optical methods where the generated thermal emission processes provide a capacitance transient. The transient is characterized by signal processing when the sample temperature is changed at a constant rate. In this way, an entire spectroscopic analysis of the semiconductor band-gap is provided.

7.10.1   Introduction

Electronic and optical properties of semiconductors are to a large extent defined by the shape of the valence band maximum E_V, and the conduction band minimum, E_C. The separation of these electronic energy levels is, in fact, deciding if the material is a conductor, semiconductor, or insulator. Materials where the energy needed to excite an electron from the valence band to the conduction band lies between a 'few tenths' of an electron volt (eV) to 'some' eVs, the material is typically classified as a semiconductor. For instance, silicon has a band gap of 1.12 eV at room temperature. If E_V and E_C, furthermore, appears for the same electron momentum, the semiconductor will allow for direct recombination of electrons from the conduction to the valence band. These materials are called direct semiconductors and energy released in the direct electronic transitions is typically emitted as light. Such materials can be used to manufacture light emitting diodes (LEDs) and semiconductor lasers. Other semiconductors do not have the property of direct transitions from E_V to E_C and cannot be used for light emission. Silicon belongs to these indirect semiconductors.

While the band structure defines most of the fundamental properties of a semiconductor, it is in practical situations very often defects and anomalies that defines the actual usefulness of a material for device applications. Semiconductors are typically very sensitive to small amounts of impurities and disturbances in the crystal lattice, which can greatly change the electronic and optical properties of the material. Consider, for instance, doping atoms that can turn the material into p- or n-type at relatively small concentrations. The low abundance of such impurities and crystal defects makes it difficult to assess them with most of the other techniques described in this book. However, since they influence the electronic properties, it is also possible to study them by electric methods. Defects that largely deviates from the regularity of the lattice, will introduce energy levels in the gap, some distance away from E_V and E_C. Such defects are named deep level defects and they interact with the hosting semiconductor by capturing and emitting electrons from both band edges. These defects should be seen as different from the shallow levels that are introduced by doping atoms that are similar to the host lattice atoms. Deep level defects can impose severe limitations for the functionality of a semiconductor device and, despite that they may occur at even lower concentrations than doping atoms, can make the material useless for device applications. Deep level transient spectroscopy (DLTS) is a very important tool for studying such deep level defects in a material, which is doped by a higher concentration of shallow level donors or acceptors.

With DLTS, one utilizes a semiconductor junction, such as the Schottky and pn-diodes, or a MIS interface, as described earlier in this chapter. Then, by changing the bias of this junction it is possible to impose different boundary conditions on the capture and emission processes of charge carriers to and from deep levels in the semiconductor bulk, or at the interface between semiconductor and dielectric. Typically, the deep levels are present in a semiconductor at a concentration in orders of magnitude lower than the semiconductor doping, but they can still have a strong

Fig. 7.21 A DLTS set-up
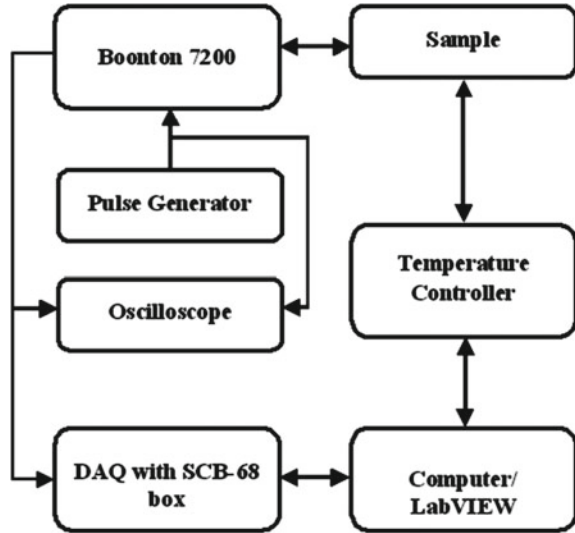
effect on the operation of devices depending on their ability to capture and emit charge carriers.

7.10.2   Set-up

A typical set-up of the DLTS system is shown in Fig. 7.21. In this system, Boonton 7200 capacitance meter provides a fast response and a recovery time of < 50 s after an overload condition [31]. A Hewlett Packard HP 8012B pulse generator supplies the filling pulse with widths in range of 10 ns^{-1} s through the capacitance meter to fill the trap centers in space charge region. The capacitance transient of the sample is measured by the Boonton 7200. Then the signal is fed to the SCB-68 box inputs of the data acquisition DAQ card NI PCI-6251. The oscilloscope shows a view of both the train of pulses and the created capacitance transients. The sample temperature can be controlled by a Lakeshore model 331 temperature controller. The capacitance meter and temperature controller are connected to a National Instruments general purpose interface bus interface [32].

7.10.3   DLTS Measurements

As a p–n junction is biased by a voltage, the number of free charge carriers on both sides of the junction is changed resulting in capacitance change. This effect on the capacitance consists of the change of both junction capacitance and minority carrier

concentration (or diffusion capacitance). In general, junction capacitance has main role under reverse biased condition whereas diffusion capacitance is important under forward biased conditions.

Consider a p–n junction with a deep level energy of E_T and the total density of deep states, N_T and n_T is the density of filled traps. When a reverse bias V_R is pulsed to a p–n junction and then the junction is decreased under a short time to zero. The electrons flow into the depletion region and they can be captured in the levels in this volume. In this way, the emission rates and trap concentrations can be estimated from the variation in the capacitance of a p–n junction for bias pulses. The capacitance transients can be obtained by keeping constant bias and temperature and applying a single filling pulse.

$$C(t) = C_0 \left[1 - \frac{N_T}{2N_D} e^{\left(-\frac{t}{\tau}\right)} \right] \tag{7.51}$$

where C_0 is the capacitance at reverse bias. The resulted transient can then be used to determine the emission rate of the carriers at a certain temperature. Figure 7.22 shows a p^+n junction under three different biasing conditions.

In Fig. 7.22a, the junction has been under reverse bias for a time long enough to leave all the traps in the depleted region, W, empty of carriers. The reverse bias is then removed during the filling pulse, Fig. 7.22b, and carriers are now introduced in the previously depleted region. If the trap bandgap level, E_T, is under the Fermi level, E_F, the traps will capture electrons from the conduction band until they are all filled. The reverse bias is now applied and the depletion region is established, but now the trapped carriers will be emitted to the conduction band and swept away

Fig. 7.22 The band diagram of the three bias stages of a p^+n junction used for DLTS measurements. During stage b, no reverse bias is applied and the Fermi level, E_F, is constant throughout the structure. E_C is the conduction band edge, E_V the valence band edge and E_T shows the position of the deep level. The depletion layer is denoted by W.
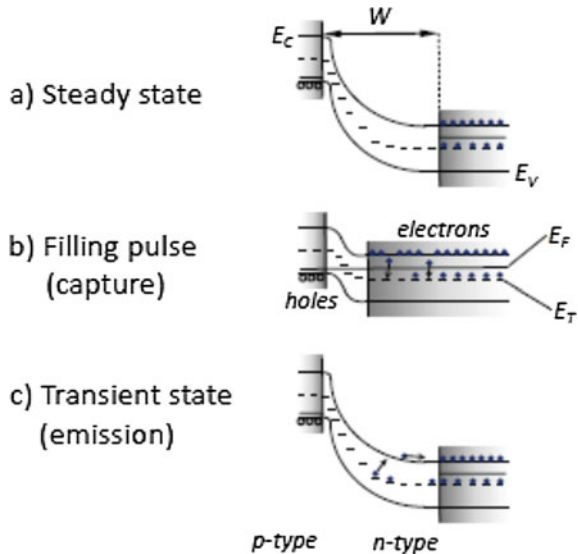
by the electric field. This is the transient stage, Fig. 7.22c, that last until all trapped electrons in the depleted region are emitted, after which the steady state is reached again and the process can be repeated. During the transient stage the capacitance is measured. The capacitance is very sensitive to the trapped charges and will be affected by the change of charges that occur during the transient stage. Figure 7.23 displays the applied bias sequence and the change in capacitance of the space charge region as the electrons are being emitted from a trap.

The thermally stimulated emission of carriers from bandgap states to the band edges increases exponentially with increasing temperature. This means that most often one trap is dominating the charge carrier traffic around a certain temperature, T_a. For lower temperatures the thermal energy is not sufficient to stimulate for instance electrons to jump up to the conduction band during the time window provided by the bias pulse frequency. On the other hand, for temperatures higher than T_a, the thermal emission process occurs so fast that it is not recorded at the set repetition frequency.

Thus, by keeping the measurement frequency constant and ramping the temperature of the sample from low to high values, successively more shallow states will contribute to the capacitance transient. This is shown in Fig. 7.24, which also indicates how the final DLTS signal can be formed during the measurement. Nowadays, the measured transients are often stored in a computer and the final DLTS signal is evaluated by dedicated software to maximize the signal to noise ration and enhance the resolution of closely spaced peaks.
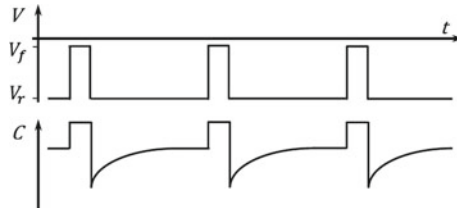
Fig. 7.23 The temporal sequence of voltage pulses that are applied for measuring majority carrier traps (electrons) in an n-type semiconductor. Carriers fill the traps during the filling pulse, V_f. During the longer periods with bias V_r the capacitance is changed due to the emission of electrons. The capacitance is not measured during the filling pulses
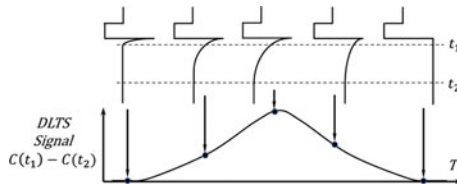
Fig. 7.24 Formation of the DLTS signal according to Lang, using so called box car technique. Here the capacitance transient is sampled at two times after the filling pulse, t_1 and t_2, and the difference between these values is plotted as the temperature is scanned. This produces a DLTS peak when the pulse frequence and the temperature matches the emission rate of the trap

From statistical mechanics, the relation between electron emission, e_n, and capture, c_n, from a deep level is given by the "detailed balance" expression:

$$e_n = N_c c_n e^{-\Delta G / kT} \qquad (7.51)$$

Here N_C is the effective density of electron states in the conduction band and ΔG is the change in Gibbs free energy for the electron while jumping from the conduction band edge to the trap, $\Delta G = G_C - G_T$. For electron transport involving deep levels in semiconductors the correct thermodynamic potential is the Gibbs free energy [33], which is related to the change in entropy, ΔS, and enthalpy, ΔH, by the equation $\Delta G = \Delta H - T \Delta S$. By using the more common electron potential energy, E_T, one often neglects large changes in entropy associated with electron transitions to deep bandgap states. The absolute temperature is given by T and k_b is the familiar Boltzmann constant. This means that the emission can be expressed as:

$$e_n = N_c c_n \chi_n e^{-\Delta H / k_b T} \qquad (7.52)$$

where an entropy factor has been introduced: $\chi_n = \frac{g_0}{g_1} e^{\Delta S / k}$. Here the electronic degeneracy of the state before and after filling with one electron is given by g_0 and g_1, respectively. Since N_C is a known value for the most semiconductors. By measuring the emission as a function of temperature and also using the possibility to vary the measuring frequency, it is possible to deduce the thermal activation enthalpy, often carelessly described as an activation energy, capture coefficient and the entropy factor in a DLTS measurement. Furthermore, since the amplitude of capacitance transient is directly proportional to the number of trapped electrons, the concentration of defects can also be quantified. The reverse bias and the amplitude of the filling pulse can also be adjusted to reach different depths in the semiconductor and in this way concentration versus depth profiles, $N_T(x)$, can be obtained. An analogous treatment can also be used to derive the corresponding equations for hole trapping in a p-type semiconductor.

In DLTS measurement, the emission rates and trap concentrations can be estimated from the variation in the capacitance of a p–n junction for bias pulses. The capacitance transients can be obtained by keeping constant bias and temperature and applying a single filling pulse.

$$C(t) = C_0 \left[1 - \frac{N_T}{2 N_D} \exp\left(-\frac{t}{\tau} \right) \right] \qquad (7.53)$$

where C_0 is the capacitance at reverse bias. The resulted transient can then be used to determine the emission rate of the carriers at a certain temperature.

One important ability of DLTS technique is to establish a *rate window* when the measuring device provides an output only when a transient appears with a rate within a window. Therefore, when the applied voltage to the sample is pulsed and the temperature is changed at a constant rate which results in the emission rate of carriers

from defect centers, and the instrument gives a response peak when the emission rate of defect centers lies within the window.

It is more common sense that instead considering a rate window, the DLTS technique applies a time filter. This means that an output signal is obtained if the transient has a time constant concurring with the center of the time window of the filter.

A very important property of such a filter (time or rate) is that the output is proportional to the amplitude of the transient. Thus, we can excite the diode repeatedly while the temperature is varied and by scanning over a large temperature interval. In this way, the information about which levels are present, what are their concentrations and by using different time/rate windows, the thermal activation energies of the levels can be estimated.

One commonly used method to establish a time filter is introducing of the dual-gate boxcar integrator. This determines the emission rate window and gives signal-averaging capabilities to increase the signal-to-noise ratio, providing the possibility to detect low concentration of defect centers. The capacitance transients are obtained on a fast-response capacitance bridge. These transients are inserted into the double boxcar and the tool measures the capacitance with gates set at t_1 and t_2. The difference $C(t_1) - C(t_2) = \Delta C$ is estimated. Afterwards, the ΔC is filtered and converted into the DLTS output S(T) as following:

$$S(T) = \Delta C_0 \left[\exp\left(-\frac{t_1}{\tau}\right) - \exp\left(-\frac{t_2}{\tau}\right) \right] \tag{7.54}$$

where ΔC_o is the capacitance change of the pulse for $t = 0$.

S(T) has a maximum value for a certain time constant τ_{max}. When a transient is very rapid, it has already finished before the first gate opens and thus $C(t_1) = C(t_2)$, and S(T) = 0. In similar way, if the transient is too slow, it does not change between the two gates and S(T) = 0. Therefore, an output with value S_{max} value will appear only for transients with time constants lie between t1 and t2. The τ_{max} value is obtained by differentiating S, as following:

$$\text{If } x = t_2/t_1 \text{ then } \tau_{max} = (x - 1)t_1/lnx \tag{7.55}$$

$$S_{max} = \Delta C_0 \left[\exp\left(-\frac{\ln(x)}{(x - 1)}\right) - \exp\left(\frac{-x\ln(x)}{(x - 1)}\right) \right] \tag{7.56}$$

Equation 7.56 indicates that the peak height depends both on the ratio of t_1/t_2 and ΔC_o, therefore the defect center concentration, N_T can be estimated.

There are different forms of defects for p- or n-type Si and C-doped Si. Many defects could have complex combinations. For example, vacancy-vacancy could be V_2, V_2^{-2} and V_2^- and with oxygen could be VO, V_2O or V_3O. Furthermore, vacancies can make with hydrogen, VO-H and with substitutional Phosphorous, VP_s. The defects for carbon atoms could be for substitutional (s) and interstitial sites (i) e.g. C_sC_i, and with hydrogen HC.

Defects have distinguished states in the bandgap which are revealed in a DLTS spectra. As an example, when a n-type Si layer is radiated by electrons, in the upper half of the band gap, O–V, E(0.17 eV), two charge states of V–V, E(0.23 eV) and E(0.40 eV), and P–V, E(0.44 eV), are typically appeared in DLTS spectra.

7.10.4   Parameters Characterizing Deep Levels by DLTS Technique

DLTS technique is applied to investigate the defect centers in semiconductors and to determine the following important characteristic parameters:

I.   *Deep level concentration*

This parameter can be determined by the preciously described DLTS capacitance measurements and assuming all the traps are filled.

$$\frac{\Delta C}{C_0} = \frac{N_T}{2N_D} \tag{7.57}$$

II.   *Majority carrier emission rates*

In these measurements, a bias voltage which is large enough to overwhelm the edge effects (but not exceeding the threshold voltage) is employed. The sample is initially cooled and then gradually heated up during the measurements. In this condition, the sample is pulsed with small voltages in order to fill the levels with majority-carriers. In the moment, when the sample returns to the quiescent reverse bias, the levels begin to emit signals and causing a transient. The capacitance is determined during the transient at the pre-set rate window and the DLTS output is estimated versus the temperature. Therefore, a peak is generated for each level with an emission rate within the pre-determined time window in the temperature range of the scan. Since of the emission rate is temperature dependence then this is observed for by determining the peak position for several different time windows.

III.   *Minority carrier emission rates*

In order to obtain of minority carrier emission, the levels have to be filled with minority carriers. This is performed through minority carrier injection when the diode is in forward biasing and then it returns to quiescent reverse bias in a pulse. In this way, the deep level creates minority carriers which generate a transient. Therefore, this measurement is similar to the majority carrier emission but with opposite peak sign.

IV.   *Activation energy*

Under thermodynamic equilibrium, the emission rates for electrons or holes (e_n or e_h) and capture coefficients (σ_n) of a deep level follow an Arrhenius curve. As an example, for electron emission is following:

$$e_n = \sigma_n < v_n >_{th} N_C \exp\left(-\frac{E_C - E_T}{kT}\right) \tag{7.58}$$

where $(E_C - E_T)$ represents activation energy (E_A), $\langle v_n \rangle_{th}$ is the average thermal velocity of the electron, and N_c is the effective density of states in the conduction band. The Eq. 7.58 can be modified for holes where N_V is the effective density of states in valence band, and the average thermal velocity and the activation are replaced to $\langle v_p \rangle_{th}$ and $(E_A = E_T - E_V)$.

In principal the activation energy is obtained from the slope of $e_n(1/T)$ curve where the intersection of the plot is $\sigma_{n,p}$ at $T = \infty$. In Eq. 7.58, both the $\langle v_{th} \rangle$ and effective density of states, N_C (or N_V) are temperature dependence. Thus, the activation energy cannot be determined accurately from $e_n(1/T)$ curves. Therefore, the temperature dependence of $\sigma_{n,p}$ has to be estimated independently and then a correction is necessary to be done to the apparent activation energy.

V.   *Measurement of capture cross-section*

In $e_n(1/T)$ curves, the $\sigma_{n,p}$ value at $T = \infty$ is for an ideal case and the real value of $\sigma_{n,p}$ is far from the capture cross-section. This is due to the temperature dependence of $\sigma_{n,p}$ but also any minor error in extrapolation may result in several orders of magnitude difference for $\sigma_{n,p}$ value.

Measurement of majority carrier capture cross-section can be performed when a fixed emission rate is chosen and the scan is done while the filling pulse-width is changed from scan to scan. It occurs when the pulse width enhances from a small value then the peak height is enhanced to certain maximum value of pulse width. At this moment, all defect centers are entirely filled during a single saturation pulse. A relationship between peak height and the filling pulse width (t_p) is given by:

$$1 - S/S_\infty = exp\left(-\frac{t_p}{\tau}\right) \tag{7.59}$$

where S presents the peak height for a pulse width of t_p and S_∞ is the saturated peak height.

In Eq. 7.59, $1/\tau$ is related to the slope of $1 - S/S_\infty(t_p)$ curve. Meanwhile, capture cross-section, σ_n can be also determined from the following equation.

$$\sigma_n = 1/(n\tau \langle V_{th} \rangle_{th}) \tag{7.60}$$

where n is the dopant concentration. In this case, σ_n can be determined from three input parameters: n, $\langle v_{th} \rangle$ and τ.

It is important to mention here that to determine the capture cross-sections for minority carriers is not a simple task. This is due to the difficulty to accurately estimate the concentration of electrons (or holes), which is dependent on the current in the injection pulse.

7.11 Summary

A comparison of the four-point probe, Hall effect and *CV* measurements is given in the table below, where the main information which can be obtained with these techniques is summarized.

	4-point probe	Hall effect	*C–V*
Main information	Resistivity	Free carrier concentration, mobility	Net carrier concentration
Conduction type	No	Yes	Yes
Depth profiling	No	No	Yes
Destructive	Partly	No	Schottky contact is required

In addition, *IV* characterization can provide useful information about the contact type and such important parameters of Schottky junction as a barrier height, ideality factor and series resistance. Furthermore, temperature dependent Hall measurements can provide information about the major donor (acceptor) energy levels. However, this analysis typically requires an equipment for low temperature measurements. Finally, it should be mentioned that equipment for all these measurements is typically not expensive.

References

1. Drude, P. (1900). Zur Elektronentheorie der Metalle. *Annalen der Physik, 306*, 566–661.
2. Bardeen, J. (1947). Surface states and rectification at a metal semi-conductor contact. *Physical Review, 71*, 717.
3. Nishimura, T., Kita, K., & Toriumi, A. (2007). Evidence for strong Fermi-level pinning due to metal-induced gap states at metal/germanium interface. *Applied Physics Letters, 91*, 123123.
4. Speckbacher, M., Treu, J., Whittles, T. J., Linhart, W. M., Xu, X., Saller, K., Dhanak, V. R., Abstreiter, G., Finley, J. J., Veal, T. D., Koblmüler, G. (2016). Direct measurements of fermi level pinning at the surface of intrinsically *n*-type InGaAs nanowires. Nano Letters 16, 5135.
5. Wenner, F. (1916). A method of measuring earth resistivity. *Bulletin of the Burreau of Standard., 12*, 469.
6. Valdes, L. (1954). Resistivity measurement of germanium for transistors. *Proceedings of the IRE, 42*, 420.
7. Kjeldby, S. B., Evenstad, O. M., Cooil, S. P., Wells, J. W. (2017). Probing dimensionality using a simplified 4-probe method. Journal of Physics: Condensed Matter 29, 394008.
8. Van der Pauw, L. J. (1958). A method of measuring the resistivity and Hall coefficient on lamellae of arbitrary shape. *Philips Technical Review, 20*, 220.
9. Zimney, E. J., Dommett, G. H. B., Ruoff, R. S., & Dikin, D. A. (2007). Correction factors for 4-probe electrical measurements with finite size electrodes and material anisotropy: a finite element study. *Measurement Science and Technology, 18*, 2067. https://doi.org/10.1088/0957-0233/18/7/037
10. Putley, E. H. (1960). *The hall effect and related phenomena*. Butterworths.
11. Smith, R. A. (1959). *Semiconductors* (Chap. 5). Cambridge University Press.
12. Beer, A. C. (1963) *Galvanomagnetic effects in semiconductors* (p. 308). Academic Press.

13. Wolfe, R. D. L. C. M., & Stillman, G. E. (1983). Magnetic-field dependence of the hall factor of gallium arsenide. In: G.E. Stillman (Ed.), *GaAs and related compounds* (Conf. Ser. No. 65, Inst. Phys., Bristol, pp. 569–572)

14. Schroder, D. K. (2005). *Semiconductor material and device characterization* (3rd ed.). Wiley.

15. Fritzsche, H. (1959). *Methods of experimental physics* (vol. 6, Part B, Chap. 8). Academic.

16. Ludwig, G. W., & Watters, R. L. (1956). Drift and conductivity mobility in silicon. *Physical Review, 101*(6), 1699–1701.

17. Chen, Y. C., Li, S. H., Bhattacharya, P. K., et al. (1994). Direct measurement of the Hall factor for holes in relaxed $Si_{1-x}Ge_x$ (0<x<1). *Applied Physics Letters, 64*(23), 3110–3111.

18. Dijkstra, J. E., et al. (1999). Effects of strain and alloying on the Hall scattering factor for holes in Si and $Si_{1-x}Ge_x$. *Journal of Applied Physics, 85*(3), 1587–1590.

19. Fu, Y., Joelsson, K. B., Grahn, K. J., et al. (1996). Hall factor in strained p-type doped $Si_{1-x}Ge_x$ alloy. *Physical Review B Condensed Matter, 54*(16), 11317.

20. Klitzing, K. V., Dorda, G., & Pepper, M. (1980). New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance. *Physical Review Letters, 45*, 494.

21. Jeckelmann, B., & Jeanneret, B. (2001). The quantum Hall effect as an electrical resistance standard. *Reports on Progress in Physics, 64*, 1.

22. Taylor, B. N. (1989). Basic standards and fundamental constants. *IEEE Transactions on Instrumentation and Measurement, 38*, 164.

23. Tsui, D. C., Stormer, H. L., & Gossard, A. C. (1982). Two-dimensional magnetotransport in the extreme quantum limit. *Physical Review Letters, 48*, 1559.

24. Hilibrand, J., & Gold, R. D. (1960). Determination of impurity distribution in junction diodes from capacitance-voltage measurements. *RCA Review, 21*, 245.

25. Blood, P. (1986). Capacitance-voltage profiling and the characterisation of III-V semiconductors using electrolyte barriers. *Semiconductor Science and Technology, 1*, 7.

26. Sell, B., Gatzke, C., & Fernández, J. M. (1998). Electrochemical capacitance-voltage profiling of heterostructures using small contact areas. *Semiconductor Science and Technology, 13*, 423.

27. Streetman, B. G., & Banerjee, S. K. (2015) *Solid state electronic devices* (7th ed.) Prentice Hall. ISBN-13: 978-0133356038.

28. Peaker, A. R., Markievich, V. P., & Coutinho, J. (2018). Tutorial: Junction spectroscopy techniques and deep-level defects in semiconductors. *Journal of Applied Physics, 123*, 161559.

29. Lang, D. V. (1974). Deep level transient spectroscopy: A new method to characterize traps in semiconductors. *Journal of Applied Physics, 45*, 3023.

30. Sah, C. T., & Walker, J. W. (1973). Thermally stimulated capacitance for shallow majority-carrier traps in the edge region of semiconductor junctions. *Applied Physics Letters, 22*, 384. https://doi.org/10.1063/1.1654683

31. Doolittle, W. A., & Rohatgi, A. (1992). A novel computer based pseudo-logarithmic capacitance/conductance DLTS system specifically designed for transient analysis. *Review of Scientific Instruments, 63*, 5733. https://doi.org/10.1063/1.1143356

32. Kumar, S., et al. (2008). Setup for in situ deep level transient spectroscopy of semiconductors during swift heavy ion irradiation. *Review of Scientific Instruments, 79*, 056103.

33. Engström, O., & Alm, A. (2005). *Thermodynamical analysis of optimal recombination centers in thyristors, Solid material and device characterization* (3rd ed.). Wiley.

Part IV
Scanning Probe Techniques

Chapter 8
Scanning Probe Microscopies (SPMs)

Part I: Atom Force microscope (AFM)

8.1 History of AFM

The development of the scanning probe microscopy (SPM) started in the 1970s and the early 1980s when scanning tunneling microscope (STM) was invented by Gerd Binnig and Heinrich Rohrer (Nobel Prize winners in Physics in 1986). After a short time, a technology breakthrough was achieved in nanoscale analysis by introducing the atomic force microscope (AFM) by Gerd Binning, Calvin Quate and Christoph Gerber. They manufactured the first prototype of AFM operating in contact mode with the sample surface [1]. Basically, AFM was developed to overcome the fundamental drawback with STM, which can only provide images from conducting materials, or semiconducting surfaces.

AFM with non-contact mode was introduced in 1987 (See Martin et al., 1987 [2]) for characterizing the soft biological samples. For such measurements, the small cantilever beam, holding the sharp probe tip in one end, could oscillate near to its resonant frequency at a distance of 1–10 nm above the sample surface. During such scans, long-range attractive forces caused changes in the amplitude, frequency and phase of the cantilever. In non-contact mode, the forces acting on the sample are remarkably weaker than in contact mode and no damages result in the soft samples. The first micro-machined tips were fabricated in 1991 by Pratar and the first, AFM with tapping mode was invented in 1993 [3]. In this mode, the cantilever oscillates at its resonant frequency and taps the surface of the sample to decrease the lateral damages. The tapping force is kept constant by a feedback loop, which also gives the output signal to monitor, for instance, height variations.

A new generation of AFM was developed in 1996 when smaller and lighter cantilevers were manufactured for higher resolution and shorter scanning times. The designed cantilevers had a higher resonant frequency leading to a higher scanning speed and a low spring constant [4]. The movement of the cantilever is typically monitored by a laser beam reflected from the cantilever and the newest AFM tools contains small cantilever with a significant smaller laser spot. This model is equipped

with an integrated illumination source, which gives the opportunity for combining an optical microscope with atomic force microscope in the same setup. Today's AFMs cover a wide range of methods where the instrument's probe interacts with a sample in many different ways and AFM is able to characterize electrical properties, such as capacitance, resistance, electrostatic forces, material workfunctions and currents, including the originally detected tunneling currents. In addition, the technique has found widespread applications in the study of magnetic effects, optical spectroscopic features and mechanical properties, such as adhesion, stiffness, friction, dissipation, of basically every type of material with atomic resolution.

Atomic force microscopy has many advantages compared to, for instance, SEM and TEM techniques and can provide high-resolution 3D images of single molecules, or even atoms. Nonconductive materials e.g. polymers can also be studied more easily, since electron beam microscopy require conducting samples. In addition, the tip in AFM tools can be fabricated by, for instance, insulating, highly conductive, or even ferroelectric materials, which opens up numerous varieties for analysis of materials on nano- or sub-nanometer scale [5, 6].

8.2 Set-up

The fundamental constituents of the instrument are the sharp probe on the cantilever, a laser with position sensitive detector and a sample stage with piezoelectric elements in 3 dimensions to feedback the position of the sample, as shown in Fig. 8.1.

The AFM instrument can operate in non-contact mode when the tip, which is mounted at one end of a springy cantilever, is placed at few nanometers distance from the surface of the sample. At such distance, the electrostatic potential between the tip and the atoms of the surface may still be large enough to cause deflection of the tip. As the tip (or sample) is scanned in the horizontal x- and y-plane, the cantilever will be deflected according to the topography of the surface and a topographic image is established. The deflection of the cantilever is monitored by the reflected laser beam and forms a feedback signal to the z-piezo element, which controls the height of the tip over the surface. Typically, the feedback adjusts the z-piezo crystal so the
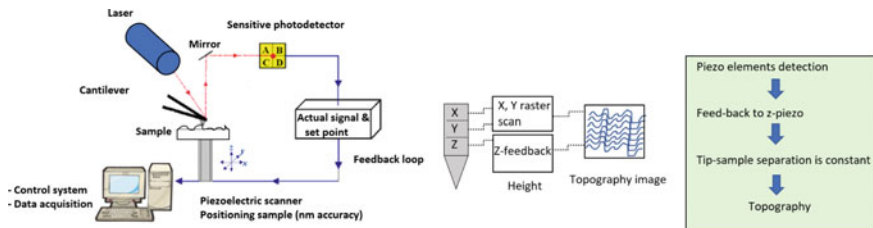

Fig. 8.1 An AFM set-up for nano-scale characterization with piezoelectric elements for a z-direction (vertical) feedback
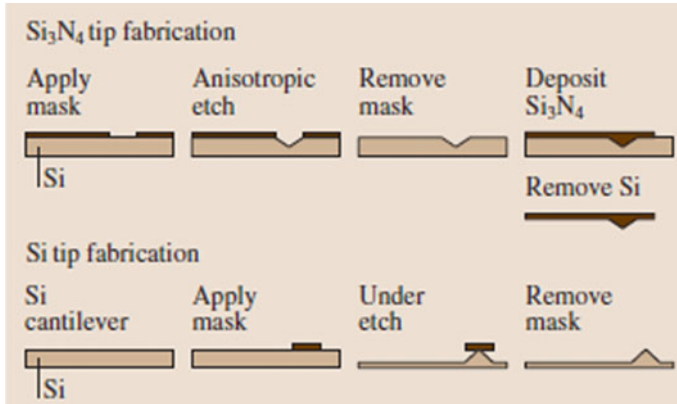
Fig. 8.2 A simplified process flow of Si and Si₃N₄ cantilevers

force between tip and sample is kept constant, i.e. the tip to surface distant remains constant during the scanning. The same feedback signal also provides information to construct the topographic image in the computer. The development of the material for the piezo-electric crystals was an important part of the success of the SPM technique.

8.3 Cantilever Manufacturing Process

Cantilevers with the sharp tip are commonly made of silicon nitride (Si_3N_4) or pure Si through a micro electromechanical system (MEMS) process, according to Fig. 8.4. The Si_3N_4 cantilevers have a typical tip radius (of curvature) of 30 nm and they are rugged and well-suited for scanning of organic/biological samples. Silicon cantilevers/tips can have sharper tips, down to 10 nm radius, and they can also be doped to be more conductive. The MEMS processes for Si and Si_3N_4 manufacturing differ from each, where the Si tip is formed by an under-etch, but the Si_3N_4 tip requires several more steps such as: a V-shape trench etch, Si_3N_4 deposition for the filling the trench and Si back etch as shown in Fig. 8.2 [7]. Many other types of tips are also commercially available, for instance diamond coated tips for very hard materials.

8.4 AFM Operational Modes

In general, AFM has two main operational modes for surface studies: (1) the static, or contact mode and (2) the dynamic mode, which consists of non-contact mode and tapping mode. Figure 8.3 demonstrates the AFM's three operational modes:
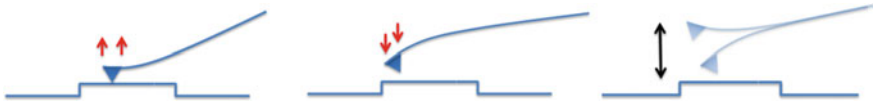
Fig. 8.3 A schematic illustration of three modes of AFM scanning operations. From left: Contact mode, non-contact mode and tapping mode
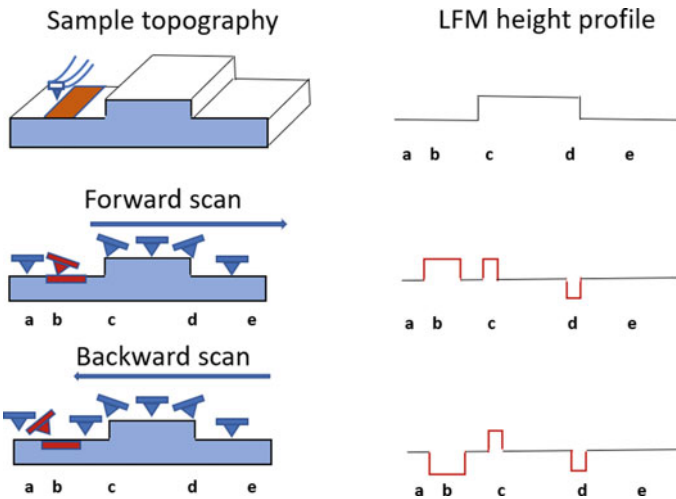

Fig. 8.4 Cantilever displacement and LFM signal resulted as topographic features and material variation. The red-shaded area is material with different composition. The lateral cantilever torsion during the forward and reverse scans with the corresponding LFM signal are also marked

8.4.1   Contact Mode

In this mode, the tip is in constant contact with the surface of the sample, maintaining a static bending of the cantilever where the vertical displacement of the tip in relation to the initial position is z_V. The cantilever experiences a normal force, p, according to: $p = kz_V$, where k is the spring constant. During the area scanning, the z-scanner moves the tip according to the surface topography [5].

There are some drawbacks with the contact mode operation. For instance, there is a high risk of contamination and damaging of the surface, since the exerted forces in vertical and lateral directions cause the probe tip to be dragged along the sample surface which could lead to damage of the tip, or the sample. It should also be mention that imaging in contact mode can be performed in constant-height mode and constant-force mode. In the latter case, the force is kept constant and the repulsive force region bends the cantilever away from the sample, resulting in the initial deflection. In this condition, when the scanner moves the tip on the sample, the contact force makes

the cantilever's initial deflection to change. Meanwhile, in constant height mode, the cantilever's deflection is maintained constant as the z-feedback loop adjusts the cantilever.

8.4.2 Non-contact Mode

This measurement mode is performed when the cantilever oscillates close to its resonance frequency (~ 100 kHz). This resonance frequency can be calculated for a spring-like cantilever and equals $\omega_0 = \sqrt{k/m}$, where k and m are the spring constant and the mass of the cantilever, respectively. The oscillations of cantilevers are created by the AFM's system, but the amplitude of the oscillations is relatively small. In non-contact mode, the cantilever has to be stiffer, i.e. larger k, compared to the contact mode, which enhances the sensitivity to small bends in response to the acting forces. The probe tip has a separation of about 1–10 nm from the surface of sample during the measurements and, therefore, the resonance frequency and amplitude of the vibrating cantilever is altered when the tip is close to the surface and interacts with the sample by attractive or repulsive forces. These interactions are the fundament of the technique. In practice, either the amplitude, or resonance frequency, maintains constant by the feedback signals.

There are also drawbacks with non-contact mode. First, the measured signal is weak and areal xy-scanning must be slow to allow for the interaction between the tip and surface to register also weak forces, e.g. van der Waals forces with a large spring constant cantilever. Most SPMs operate under normal ambient conditions and a thin water layer on the sample surface can be expected due to the moisture in the air. Due to the small dimensions involved in AFM measurement, this water can result in a capillary force between the tip and sample and this force is strongly attractive, in the range of 10^{-8} N, and can "glue" the tip to the surface for some time. These drawbacks typically make the lateral resolution worse, compared to the contact mode.

8.4.3 Tapping Mode

Special care must be taken when using AFM for imaging the surface of soft materials. In tapping mode AFM, the oscillations are of the tip are larger and the tip briefly touches the surface of the sample. In this way, a similar resolution to the contact mode is obtained, but the applied force can be considerably weaker. In tapping mode, the cantilever vibrates close to its resonance frequency with an amplitude in the range of 20–200 nm, which is significantly larger than the non-contact mode. The interaction between the tip and surface changes the resonance frequency and also results in a clear change of the amplitude of the cantilever, leading to a phase angle. Based on

this, there are three types of data output from the tapping mode AFM: (1) height, (2) phase and (3) amplitude data.

(1) Height data: This imaging mode registers the changes in vertical motion of the z-scanning piezo as it tries to keep the distance of the tip from the surface constant. The voltage applied to the piezo cylinder is proportional to the tube's length expansion or contraction. Therefore, the variation in the z-axis motion, in coordination with the xy-position, can be interpreted as a topographical map of the sample surface. The z-piezo voltage is a good indicator of the height of surface features, although the resolution and sensitivity on the rising and falling edges of topographical features is limited.

(2) Phase data: This type of imaging is obtained as a result of the change in phase of the cantilever oscillation drive signal in relation to the phase offset of the measured oscillations of the cantilever. The oscillations of the cantilever are mostly produced by yet another piezo crystal and the phase of this drive signal is compared to the phase of the cantilever response signal given by the photodiode detector. As the probe tip scans the surface of the sample, the phase offset of the oscillating cantilever varies compared to the phase offset of the input drive signal. This occurs, for instance, where areas with different elasticity, or softness, exists on the sample surface and, as a result of the slight frequency change, the phase angle between the two signals shifts. These phase changes can be established quite well by the electronic circuits, and can be plotted in a phase image.

(3) Amplitude data: The amplitude of the cantilever is monitored by the photodiode detector. The value (rms) of the vertical, bending, motion of the laser signal on the detector is mostly recorded to monitor height differences, as described above. The cantilever can also experience twisting forces due to the tip/sample inter-actions and this motion can also be detected by the segmented photo detector. By also monitoring the torque on the cantilever, amplitude images tend to better reproduce the edges of surface features. During operation of the AFM in force mode, the probe tip is first lowered into contact with the sample, then indented into the surface, and finally lifted off from the sample surface.

8.4.4   Lateral Force Microscopy (LFM)

In AFM, there are modes which use the out-of-plane deflection of the cantilever in order to detect the in-plane frictional, or mechanical properties of the sample [8]. This measurement can be done by using the mode lateral force microscopy (LFM). This mode is an important technique to study heterogeneous samples, where the variation in compositions over the sample surface can be determined with nanoscale resolution. The LFM mode operates in contact mode and during the measurement, the focus is on how the cantilever bends in the vertical direction in order to obtain topographical information. At the same time, LFM reveals the torsion of the cantilever in the lateral direction, which can be due to frictional forces. The edges of nanofeatures on the

sample surface, as well as heterostructure materials inclusions, create contrast in the LFM signal, as illustrated in Fig. 8.4.

In a sample, steep edges may twist the cantilever and this motion of the cantilever makes the laser beam deflect sideways, which can be detected by the partitioned optical detector. The cantilever will tilt in opposite directions during the forward and reverse scanning motion of the tip across the sample surface, causing a similar decrease, or increase in the LFM signal, depending on if the tip is making a trace, or retrace. If also the material is varied, so the tip experience different friction, a lateral torsion of the cantilever occurs and a contrast is obtained due to the frictional force variation. By comparing the forward and backward LFM scans, we are able to distinguish between the LFM signals created by topography features and the difference in frictional signal resulting from material heterogeneities.

8.5 Theory of AFM

8.5.1 Basic Forces Acting on the Probe

Atomic force microscopy is strongly dependent on the properties of the tip and cantilever. The technique provides quantitative information about the surface morphology, but also quantitative information concerning the interactions between the tip and specimen. Of utmost importance for the sensitivity of a quantitative measurement is the tip shape together with the cantilever spring constant. Several types of forces are contributing to the final response of the tip, as it approaches the sample surface. In this section we will describe some to the interacting forces and the general potential distribution between tip and sample. Then we will present the basic equations of motions for the tip.

The interactions between the probe tip and sample occurs mainly through attractive van der Waals forces of electrostatic nature, forming a potential distribution depending on the distance, r, between tip and sample. This potential can be approximated by the Lennard–Jones potential, which originally was intended to describe the potential between two electronically neutral molecules or atoms at some distance from each other [6]:

$$U_{LJ}(r) = 4c\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{8.1}$$

Here, c is a value of the minimum depth of the potential well and σ is the distance at which the particle–particle potential is zero, which is sometimes used to describe the size of the particle. The first term in Eq. 8.1 is related to repulsive forces, drastically increasing as the particles comes closer to each other ($1/r^{12}$ dependence) and a second term with negative sign giving an attractive force that vanish at very large particle separation. This is the van der Waals force, which is dependent on the shape of the
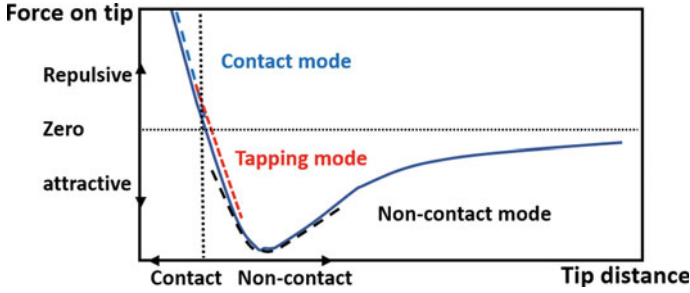
Fig. 8.5 Static forces on the AFM probe versus the distance from the sample surface in different operation modes. The operation region of three common operation modes of the AFM are indicated in blue (contact), red (tapping) and black (non-contact)

tip and is basically of attractive nature, where potential scales with distance as $1/r^6$. Van der Waals forces can be further split in: (a) Keesom force, which roots from permanent dipole–dipole interactions, (b) Debye dipole force, coming from a dipole interaction with an induced dipole and (c) London dispersion force coming from two temporary dipoles interactions [9]. A schematic of the potential between a tip and an arbitrary surface as a function of their separation is shown in Fig. 8.5, which also indicates the typical operational regimes for contact, tapping and non-contact AFM modes.

In real measurements, however, geometrical shapes, additional charges and contamination, such as water on the surface, distorts the Lennard–Jones potential, for instance by the capillary forces discussed above. In addition, the cantilever spring constant is only constant in a limited range, leading to instability problems when the tip is too close to the surface. Using vacuum condition or creating a hydrophobic surface, for example using iso-propanol, may be used to minimize some of these effects.

Other electrostatic forces, of Coulombic nature acting over larger distances, are also affecting the tip and specimen behavior. These forces can be roughly approximated by a parallel plate capacitor model which express the capacitance between the tip and the specimen $C = \frac{q}{\Delta v} = \frac{A\varepsilon_0}{d}$, where q here is the total charge, Δv is the voltage difference between tip and sample, ε_0 is the vacuum permittivity, A represent the active area between the tip and specimen and d is the distance between tip and sample. The Coulomb force between the tip and specimen can be written as

$$F = \frac{Q^2}{2A\varepsilon_0} = \frac{A\epsilon_0}{2d^2}\Delta v^2 = -\frac{1}{2}\frac{\partial C}{\partial d}\Delta v^2 \qquad (8.2)$$

This electrostatic force acting on the tip depends on the geometry of the tip (and the specimen) and the tip-specimen separation. In practice, often the electrostatic force is attractive and is sensitive for a distance in the range of 10 nm [10].

8.5.2 Equations of Motion for the Tip

In a theoretical approach to describe the dynamic AFM operation, the tip vibrations are often treated as a harmonic oscillator with a driving force from the piezo crystal and the spring constant of the cantilever. In addition, various forces of increasing between the tip and sample are introduced to model damping of the motion, resulting in changes in amplitude, frequency and phase shift.

One equation for the cantilever motion can be written by considering a spring type movement. In this case, the spring force is $F = -kz$, where the negative sign in front of the spring constant, k, leads to a force in the opposite directions of the tip motion, z. The driving force for the oscillation, F_ω, induce an amplitude A_0 and oscillation frequency according to $A_0 \cos \omega t$, while the attractive hydrodynamic force, F_h, is given by

$$F_h = \frac{m\omega_0}{Q}\dot{z} \tag{8.3}$$

Here, we use the notation for time derivative $\dot{z} = \frac{d}{dt}z$ and $\ddot{z} = \frac{d^2}{dt^2}z$, Q is the quality factor for a resonant frequency of ω_0 with a deviation frequency $\Delta\omega$ given by $\omega_0/\Delta\omega$ [11]. The equation of the cantilever motion can then be written according to Newton's second law:

$$m\ddot{z} + (\frac{m\omega_0}{Q})\dot{z} + kz = F_{ps} + F_0\cos(\omega t) \tag{8.4}$$

In a first approximation we assume that the probe-sample force is negligible ($F_{ps} = 0$) and, if we also remember that the spring constant for a harmonic oscillator is $\omega_0 = \sqrt{k/m}$, then the Eq. 8.4 is simplified as:

$$\ddot{z} + (\frac{\omega_0}{Q})\dot{z} + \omega_0^2 z = \frac{F_0}{m}\cos(\omega t) \tag{8.5}$$

In general, the solution for the above equation consists of a steady (z_s) and transient part (z_t).

$$z(\omega, t) = z_s(\omega, t) + z_t(\omega, t) \tag{8.6}$$

where each particular solution can be written, using complex numbers, as $Z(\omega, t) = Ae^{i\omega t}$. Therefore, Eq. 8.5 can be written as:

$$-A\omega^2 e^{i\omega t} + A\left(\frac{\omega_0}{Q}\right)i\omega e^{i\omega t} + A\omega_0^2 e^{i\omega t} - \frac{F_0}{m}e^{i\omega t} = 0 \tag{8.7}$$

where the amplitude, A in general form can be written as:

$$A = \frac{\frac{F_0}{m}}{\left(\omega_0^2 - \omega^2\right) + i\omega\frac{\omega_0}{Q}} \quad \text{or} \quad |A(\omega)| = \frac{\frac{F_0}{m}}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\frac{\omega_0^2}{Q^2}}} \qquad (8.8a,b)$$

At the resonance condition the amplitude becomes:

$$|A(\omega)|_{(\omega=\omega_0)} = \frac{Q F_0}{k} \qquad (8.9)$$

and the cantilever angle in response to the driving force will be:

$$\phi = arctan\frac{\frac{\omega\omega_0}{Q}}{\left(\omega_0^2 - \omega^2\right)} \qquad (8.10)$$

This shows that the amplitude of cantilever depends on both the quality factor, Q, and the size of the driving force. The steady (z_s) and transient solutions (z_t) in Eq. 8.6 are following:

$$z_s(\omega, t) = Re\left[|A|e^{i\omega t}e^{-i\phi}\right] = \frac{\frac{F_0}{m}}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\frac{\omega_0^2}{Q^2}}}\cos(\omega t - \phi) \qquad (8.11)$$

$$z_t(\omega, t) = Be^{\frac{\omega_0}{2Q}t}\cos(\omega_d t + \beta) \qquad (8.12)$$

where B and β are different constant values and ω_d is a system resonance frequency and are related to the cantilever resonance frequency through $\omega_d = \omega_0\left(1 - \frac{1}{4Q^2}\right)$.

In the presence of other forces, the spring constant can be modified to an effective spring constant, k_e and the oscillation frequency can be approximated by:

$$\omega = \sqrt{\frac{k_e}{m}} = \sqrt{\frac{k - F'}{m}} = \omega_0\sqrt{1 - \frac{F'}{k}} \qquad (8.13)$$

During the interactions of the tip and specimen, a resonance frequency shift ($\Delta\omega$) is introduced which can be written as:

$$\Delta\omega = \omega - \omega_0 = \omega_0\left(\sqrt{\left(1 - \frac{F'}{k}\right)} - 1\right) \approx -\omega_0\frac{F'}{2k} \qquad (8.14)$$

For attractive forces, F' is positive and $\Delta\omega < 0$ and these conditions are in reverse for repulsive forces as shown in Fig. 8.6.

Due to the small distances involved in the interactions between tip and sample, the equation of motion can also be treated non-classically by quantum mechanical theory. In this case, the motion will be described by the Hamiltonian operator, H:

Fig. 8.6 Dependency of amplitude (A) on the frequency (ω) for different forces
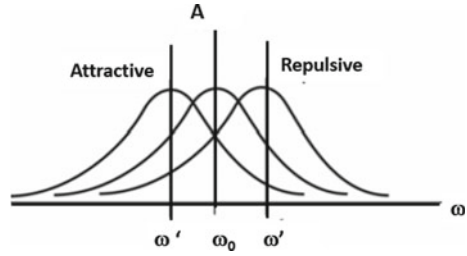
$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2 z^2 \tag{8.15}$$

where p is the momentum of the moving tip. By this approach, it is also possible to model the purely quantum mechanical effect of tunneling of electrons between the tip and sample due to overlapping electron distributions. This was an important part of the original development of the technique and led specifically to the invention of the scanning tunneling microscope, STM, further described below.

Theoretical approaches to model the dynamics of the AFM tip are often useful to understand the general behavior of the system. However, due to the complex system with various active forces that can be extremely sensitive to the geometry, as well as the presence of moisture and other experimental parameters that are difficult to control, it is often difficult to reach quantitative results from these calculations.

8.6   AFM Characterization and Application

Nanotechnology has opened up a new research field. How to observe and characterize the structure and properties of materials and devices on a nano-scale is related to the precision that can be provided by nano-detection techniques. The characterization and application of atomic force microscope, AFM, is today a necessary tool for testing and development of modern devices, particularly in microelectronics. In this section, AFM is reviewed in the three aspects: surface topography analysis, mechanical analysis and nano material processing.

8.6.1   Surface Topography Analysis

The AFM characterizes the three-dimensional topography of the sample surface by detecting the force between the probe and the sample. Today's AFM tools have a resolution of 0.1–0.2 nm in the horizontal (xy) direction and a resolution of about 0.01 nm in the vertical (z) direction, which makes it possible to "see" individual atoms.

Atomic force microscopy has significant advantages compared with other conventional microscopic methods, where the outstanding advantages are not only the ability to detect conductors, semiconductors, insulator samples, but also to perform imaging in vacuum, atmospheric and liquid environments. Using the capacity of modern computers, it is also possible to make 3D simulations of the topography of a sample and provide images suitable for advanced applications.

8.6.2 Mechanical Analysis

Atomic force microscopy utilizes the force between the tip of the probe and the atoms on the surface of the sample and has a sensitivity of the order of pN (10^{-12} N). This makes it possible for AFM to monitor the small force between the probe and the sample, as well as evaluate the mechanical properties of the sample. Early studies demonstrated that the adhesive force between mica, a mineral with very flat surface, and an ordinary silicon nitride AFM tip was about 100 nN. However, simply immersing the sample in water reduces the mechanical force by a factor of 100 to about 1 nN [12]. Later, the effect of van der Waals forces could be estimated and it was found that these forces can also be significantly reduced in a suitable solvent [13]. Other examples of the sensitivity of AFM is a report on the interaction of hydrophilic and hydrophobic AFM tips with argon nanobubbles on HOPG (highly oriented pyrolitic graphite) in water [14]. Another example is the AFM study of mechanical properties of three different nano structures based on suspended carbon nanotubes, CNT: straight CNT, coiled CNT, and a torsional CNT [15].

8.6.3 Nanomaterial Processing

Since the birth of the atomic force microscope, the field of its application has continued to expand. In addition to detecting surface features of samples and analyze its microstructure, today, the tip of the AFM can also be used to induce localized oxidation, local modification, and nano-etching to change the surface topography of the sample. AFM is also used to manipulate single molecules and atoms and perform nanofabrication on a sub-nanometer scale with ordered molecular films, metals and superconductor surfaces.

The earliest nano-etching and nano-machining technologies appeared already in 1987. At the AT&T Bell laboratory [16], a scanning tunneling microscope (STM) tip was used to achieve atomic-scale processing of the surface of single crystal germanium for the first time. The formation of an artificial atomic structure indicates the possibility of nanoscale processing using SPM. After that, there have been many more methods developed for nano-fabrication using SPM, and the materials and conditions required for processing nano materials have also expanded greatly. Scanning probe
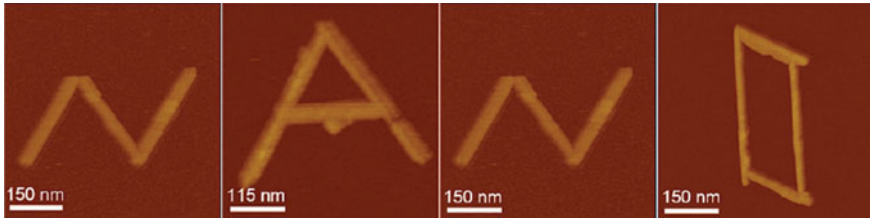
Fig. 8.7 Tapping-mode AFM images showing the word "NANO" composed of individual GAV-9 nanofilaments [20]

nano-machining technology has gradually developed into one core part of nanotechnology due to its applicability, operability and atomic-scale resolution. Many more examples of innovative adaptions of the AFM technique to develop process schemes for nano-fabrication is given below:

- AFM writing technique has been used to peel off resist and produce narrow patterns that can be filled with, for instance Cr to form Cr wires with 40 nm linewidth on a GaAs substrate [17].
- AFM technique has been applied on metallic lines with a 40 nm linewidth as single-electron transistors [18].
- Dip-pen nanolithography (DPN) technique has been used to directly manipulate molecules to form patterns by AFM, making DPN a potentially useful tool for creating and functionalizing nanoscale devices [19].
- An approach has been presented to produce spatially organized individual one-dimensional peptide nanofilaments by AFM and achieved nanofilament extension at designated positions and selected orientations. This method has been applied to obtain a pattern with the word "NANO", as shown in Fig. 8.7 [20].
- A study of electrochemical etching of n-type GaAs substrates investigated the surface topography and the density of grains of the porous GaAs (π-GaAs) samples by AFM, and determining nanocrystalline with a grain size close to 7 nm [21].
- A closed-loop nanoscale precision stage has been integrated with AFM to mechanically fabricate 3D nanostructures according to predetermined designs, e.g., 3D human face nanostructures, nanoline arrays of sine-wave and triangular nanostructures [22].
- AFM technique nano-oxidation has been used to fabricate nickel silicide nanowires on silicon-on-insulator substrate and studied the effects of the width of nanowires on the structural and electrical properties of nanowires [23].
- AFM-based nanolithography can be used to fabricate groove arrays on Si surface with 30–100 nm pitches and 5–32 nm depth [24].
- AFM lithography technique has been applied to manufacture Graphene planar transistors [25].

8.7   Dedicated SPM and AFM Techniques

Scanning probe and atomic force microscopies have become widely used in many advanced applications on the nano-scale. In this last section we exemplify in more detail how these tools can be specialized and tuned for studying a variety of physical, chemical and biological effects.

8.7.1   STM: Scanning Tunneling Microscopy

Scanning tunneling microscopy is applied to image topography, manipulate surface structures, and measure surface properties typically from biological samples to nanoscale devices [1, 26]. Notice that STM is not used as a conventional microscope, since it does not magnify the sample image. The tool consists of a sharp electrically conductive tip, which holds typically just one or two atoms at the apex, contained in a UHV chamber. The tip moves just above an electrically conductive surface, as shown in Fig. 8.8. In operation mode, the tip may be closer than an Ångström (10^{-10} m) from a nano particle on the surface. If a small positive voltage is applied to the tip, electrons may tunnel through the space between the sample and tip, and form a current that can be detected. By scanning the tip along the surface, the tip's position is adjusted to ensure a constant distance of the tip from the surface. If the surface is homogeneous, the current is then kept constant. The tip movement is traced and recorded and can be graphically displayed on a computer screen. Apart from topographic information, the technique also allows for exploration of quantum mechanical effects.

Fig. 8.8 A schematic view of scanning tunneling microscope operation (STM) [27]
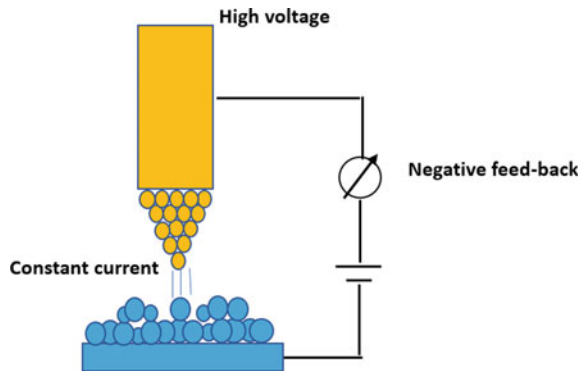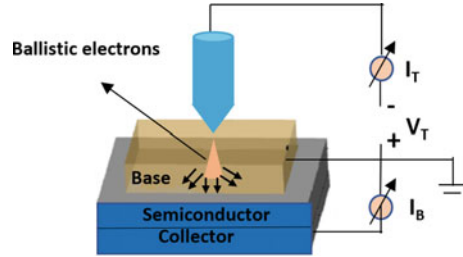
Fig. 8.9 A schematic view of the BEEM operation [28]

8.7.2 BEEM: Ballistic Electron Emission Microscopy

Ballistic electron microscopy (BEEM) technique is based on ballistic electron transport, i.e. uninterrupted motion of energetic electrons, through several layers of materials and interfaces. The technique uses a STM tip to inject electrons into a grounded metal base of a Schottky diode, as shown in Fig. 8.9. The bottom of the structure is positively biased and a few of the injected electrons will move ballistically through the metal to the metal–semiconductor interface where they will experience an electrostatic potential, the Schottky barrier. Ballistic electrons may have enough energy to overcome the Schottky barrier and contribute to the BEEM current collected at the bottom of the structure. Balllistic electron emission microscopy offers nanometer spatial resolution due to the atomic scale positioning capability of the tip. Furthermore, the energy distribution of the tunneled electrons is relatively small, making BEEM a technique with very good energy resolution, about ~ 0.02 eV.

8.7.3 Electrostatic Force Microscopy (EFM) Mode

Electrostatic force microscopy is similar to normal non-contact AFM, but reveals additional information about the electrical properties of a scanned surface by measuring the electrostatic force between the surface and the tip while also biasing the tip [29–32]. In this way, information about surface potential and/or the charge distribution is provided. Furthermore, the scans also reveal the short-ranged van der Waals forces acting between the tip and the sample surface, related to the surface topography just as in ordinary AFM, and these forces can be distinguished from the electrostatic forces. One approach is to sense the distance-dependence of the electrostatic and van der Waals forces, which differ from each other. It can be measured by applying a dual pass lift-mode technique, as shown in Fig. 8.10. There is also a second approach where different frequencies are applied for measuring the electrostatic force, or the topography imaging. The van der Waals force and the electrostatic force are distinguished by employing an extra lock-in amplifier for the EFM signal.

The generated signals of both techniques contain information on the surface topography (called 'Height') and the electrical properties of the surface (called 'EFM
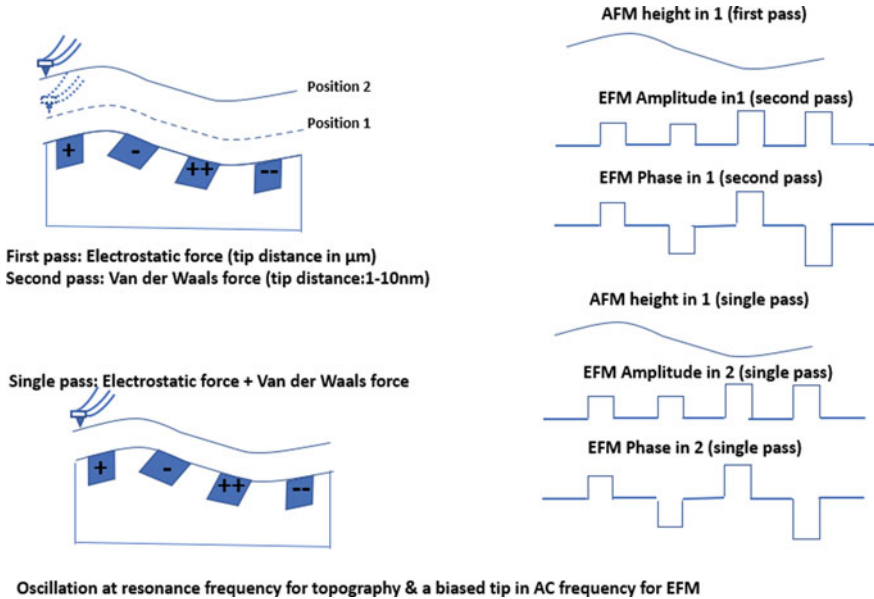
Fig. 8.10 An illustration of dual frequency EFM by using a single pass technique. The topography and electrostatic interactions are simultaneously measured by applying two different frequencies for detection. The generated signals are amplitude and phase EFM, which show the height, as well as the magnitude and the sign of the surface charges

amplitude' and 'EFM phase') generated by the van der Waals and electrostatic forces, respectively. The two different approaches for measurements are introduced separately in Fig. 8.10. To obtain high sensitivity and low signal noise, the scans are done in dynamic mode, applying lock-in detection. First, in non-contact mode the tip scans the surface oscillating at the cantilever's resonance frequency in the regime of the attractive van der Waals force. Second, the scan is performed at constant tip-sample distance, where the tip oscillation at resonance is only sensitive to the electrostatic forces. The amplitude and phase of the oscillation during the second scan provide the magnitude and sign of the surface potential.

8.7.4   Kelvin Probe Force Microscopy

Kelvin probe force microscopy (KPFM) is a technique to perform a nanoscale mapping of the material's workfunction. The technique is also used to determine the contact potential difference (CPD) [6, 11, 33]. A workfunction is defined as the minimum energy needed for an electron to leave the metal surface completely. KPFM takes into account both electric forces and gradients of electric forces between the tip and sample, where the tip and sample are considered as a parallel plate capacitor.
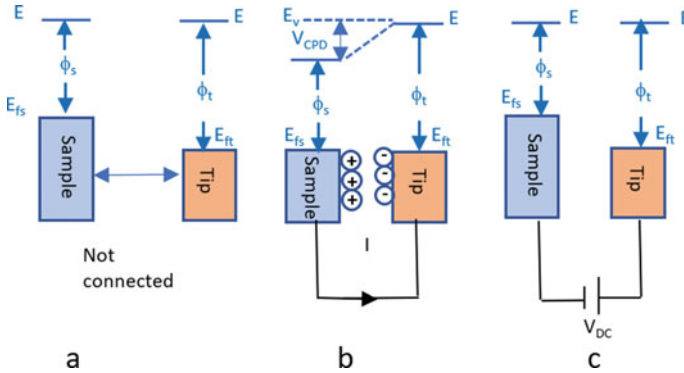
Fig. 8.11 Schematic view of energy band-diagram for different conditions for the sample and tip. The Fermi energy for the tip and the sample are denoted by E_{ft} and E_{fs}, respectively, as well as for the workfunctions are φ_t and φ_s, respectively. The figures a-c refers to when the tip and sample are **a** separated by a distance "d", **b** connected with a conducting wire (shorted) and **c** with an applied voltage between the tip and the sample [11]

Figure 8.11 explains the mechanism of KPFM, and how the workfunction can be obtained.

In the separated case, the Fermi levels are different and the system is not in equilibrium, as shown in Fig. 8.11a. When a conducting wire is connected between the sample and tip, electrons will flow from the sample with a lower workfunction, creating a positively charged tip, which has a higher workfunction. In this situation the Fermi levels becomes equal and an electric field is generated which imposes a contact potential difference, V_{CPD}, as shown in Fig. 8.11b. In the latter case, Fig. 8.11c, a voltage of V_{dc} is applied between the sample and tip. When the condition $V_{CPD} = V_{dc}$ is reached, the electric force between the sample and tip becomes zero. Therefore, the workfunctions can be determined as following:

$$V_{CPD} = \frac{\varnothing_s - \varnothing_t}{e} \tag{8.16}$$

The KPFM operates in two imaging modes: the amplitude modulation (AM) and the frequency modulation (FM) by considering the electric force and the force gradient, respectively [6]. The KPFM can be used in a single-pass or dual-pass mode. In the single-pass mode, both the topography and surface potential of the sample can be measured together in a single scan line. However, for the dual-pass mode, during the initial pass the topography is measured and in the second pass the workfunction is obtained by using the cantilever at a distance of a few nanometers above the sample.

In KPFM, in addition to a DC voltage, V_{dc}, an AC voltage, $V_{ac} = \sin\omega t$, can also be applied between the sample and tip. Now the AFM is operating at the cantilever's resonance frequency and the total voltage difference, ΔV, will be

$$\Delta V = V_{dc} + V_{ac}\sin\omega t - V_{CPD} \tag{8.17}$$

The main force equation for the system can be formulated as:

$$F = -\frac{1}{2}\frac{\partial C}{\partial d}\Delta v^2 = -\frac{1}{2}\frac{\partial C}{\partial d}(V_{dc} + V_{ac}sin\omega t - V_{CPD})^2 \qquad (8.18)$$

Then, the total force consists of three forces F_{dc}, F_ω and $F_{2\omega}$ as follows:

$$F_{dc} = -\frac{1}{2}\frac{\partial C}{\partial d}\left((V_{dc} - V_{CPD})^2 + \frac{1}{2}V_{ac}^2\right)(\text{for topography}) \qquad (8.19)$$

$$F_\omega = -\frac{\partial C}{\partial d}(V_{dc} - V_{CPD})V_{ac}\sin(\omega t)(\text{for contact potential difference}) \qquad (8.20)$$

$$F_{2\omega} = \frac{1}{4}\frac{\partial C}{\partial d}V_{ac}^2\cos(2\omega t)(\text{for capacitance microscopy}) \qquad (8.21)$$

The F_{dc} term is applied to study the topography, while F_ω provides information about the contact potential difference. The CPD is obtained when $F_{dc} = V_{CPD}$ and, if $F_\omega = 0$, then $F_{2\omega}$ term is used for capacitance study [6].

The above equations show that the appropriate electric force to modulate ω is created when an AC voltage is chosen at the resonance frequency. In this condition, the cantilever oscillations are detected by the photodiode in the AFM system [6]. The applied voltage is selected to be at the cantilever resonant frequency. In this condition, the system is sensitive to the detected force F_ω and allows to decrease the ac-voltage [6]. The frequency signal from this position is given into the lock-in amplifier which senses the oscillation amplitude. The signal is at that moment given into the KPFM feedback loop. The feedback regulates the V_{dc} until the oscillation amplitude of F_ω goes to zero. At this moment the force term F_ω in Eq. 8.20 is zero. Hence the output signal is used to map the information about the contact potential difference.

8.7.5 MFM: Magnetic Force Microscopy

Magnetic force microscopy (MFM) is widely used to study magnetic properties of nano materials. The technique utilizes a sharp tip coated with a ferromagnetic material to determine the distribution of magnetic domains on the sample. Magnetic force microscopy can be used in non-contact mode and it creates image by measuring the amplitude and phase of cantilever oscillation. In this way, the signals reveal the magnetic domain distribution of the sample surface, as illustrated in Fig. 8.12.

The MFM mode can also be used to probe magnetic storage devices, hard disk media, and magneto-resistive heads, as well as to image naturally occurring, or deliberately written domain structures in magnetic materials.
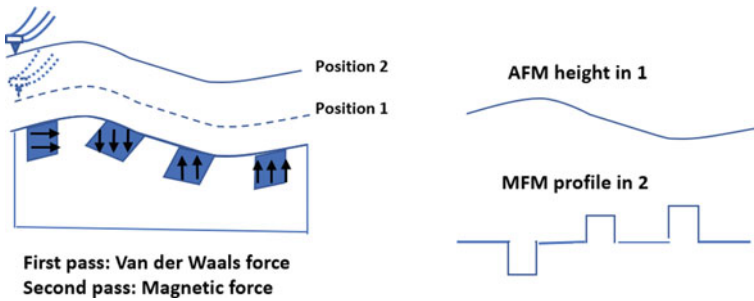
Fig. 8.12 Schematic view of MFM operation and its profile scan

8.7.6 Scanning Spread Resistance Microscopy (SSRM)

The SSRM technique uses a conductive AFM tip to scan a region of a conductive, or semiconducting material, in contact mode. The detected signal will be dominated by the *spreading resistance*, an additional resistance arising from the current spreading around the sharp tip. Scanning spreading resistance microscopy is often used to scan a cross-sectional surface of a semiconductor device with the purpose of determining the dopant distribution in the structure. For SSRM measurements, a good depth resolution then needs careful preparation of beveled, or cleaved surfaces, otherwise the contact resistance value can be significantly affected. Furthermore, different corrections are required to account for junction shifts due to carrier diffusion as a result of the beveled surfaces. The technique offers a high spatial resolution, which is several times less than a tip radius. The technique offers a dynamic range of doping concentrations between 10^{15} and 10^{20} cm^{-3} for both n- and p-type doping in Si.

In general, there are two operating SSRM modes, internal and external, according to how the current amplifier is being used. The internal mode is obtained when the current amplifier is set on a fixed gain and the external SSRM mode is obtained when an external low noise current amplifier with variable gain is used. In this mode, the current range is changed by varying the gain of the amplifier. The SSRM technique is a variety of Spreading Resistance Profiling (SRP) methods which is uniquely applied to analyze nanomaterials. In general, SRP measurements are dual probe techniques and scans the sample surface while measuring the current between the two probes, while in SSRM one conductive AFM tip scans the surface, as illustrated in Fig. 8.13.

In order to have a good depth resolution for SRP analysis, the sample surface needs to be beveled. The quality of the sample preparation strongly affects the quality of contact resistance measurement. The probe tips for SRP is commonly fabricated by tungsten alloy with a large curvature radius of about 10 μm, while SSRM tip is often made of highly doped diamond like carbon (DLC) and the curvature is typically a few tens of nanometers, yielding three orders of magnitude better spatial resolution.
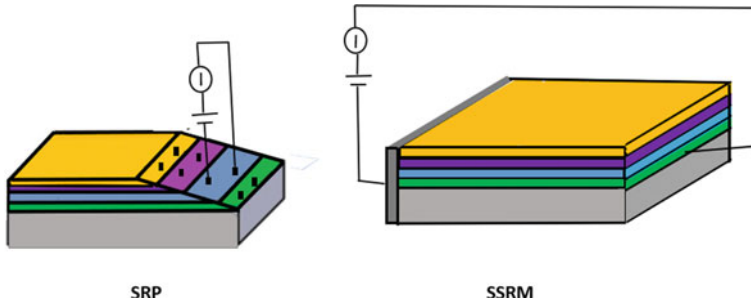
Fig. 8.13   A drawing of how SPR and SSRM measurement is performed

8.7.7   Scanning Capacitance Microscopy (SCM)

Scanning Capacitance Microscopy (SCM) operates in contact mode and provides a topographic image while, at the same time, a corresponding capacitance map is created. In these measurements, the tip is one electrode and the bulk semiconductor is the other electrode, similar to SSRM. However, for SCM the sample surface is not conductive and oxides or other dielectric materials on the surface prevents a current flow. Instead, the capacitance between tip and sample is studied. The capacitance signal is dependent on the contact area of the tip to oxide, the oxide properties, and the charge distribution of the semiconductor. An accurate SCM analysis requires good sample preparation and formation of the surface oxide. Therefore, for a smooth flat sample and a high-quality oxide, the capacitance signal is largely controlled by the doping concentration of the sample, making SCM an alternative to SSRM for measuring the doping distributions in samples. In SCM measurements, both DC and AC bias are used and the output signal is typically the derivative of the capacitance, dC/dV.

A bias applied to the tip will deplete the near surface region of mobile charges and set up a space charge region. The sign of the derivative of the measured capacitance–voltage curve determines if the sample is p- or n-type, while the amplitude of the dC/dV signal is proportional to the doping concentration. As with many of the SPM techniques, the SCM gives mainly qualitative results and modeling work is needed for a quantitative analysis. Alternatively, calibration samples with known doping can be prepared, and measured simultaneously with the same tip (Fig. 8.14).

Fig. 8.14 SCM measurement: the tip is coated by the conductive material as an electrode and the semiconductor sample function as the second electrode. An AC voltage is added on the DC bias voltage. In this figure, the oxide layer of semiconductor and the depletion layer are capacitors in series
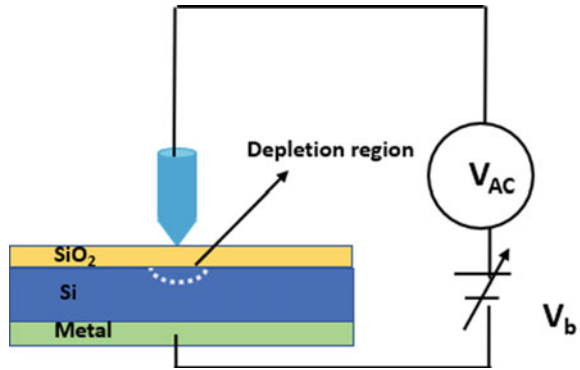
8.8 Part II: Scanning Near-Field Optical Microscope (SNOM)

8.9 Basic Principles

When light is used for imaging of a sample, the resulting spatial resolution is limited by its diffraction properties. In other words, features much smaller than the wavelength λ cannot be distinguished from one another. Abbe diffraction limit stipulates the minimum resolvable distance as:

$$d \approx \frac{\lambda}{2 \cdot NA} \tag{8.22}$$

where NA is a numerical aperture, which is proportional to the refractive index and the sine of the convergence angle. Typical values of the numerical aperture for microscope objective lenses are from 0.2 to 1.5 (the latter is for oil-immersion lenses). Therefore a practical resolution limit is approximately defined as $\sim \lambda/2$. So a characteristic size of the "probe" in this regime would be about half of the wavelength.

This description holds for a *propagating* light wave. However, when facing a subwavelength aperture, the wave normally does not penetrate through the opening. Only in some unique cases of a resonant array of holes in a metal the penetration of the propagating field can occur [34]. At the same time, the wave is not completely back-reflected either, allowing for a small fraction to "leak" through the openings. This part is called an *evanescent* wave. Its intensity quickly decays away from the aperture, and it has the same frequency as the initial propagating wave.

The principle of scanning near-field optical microscopy SNOM is based on using the evanescent wave as a probe, instead of a propagating wave with its inherent resolution limits. A short range nature of the interaction for the evanescent wave leads to the term *near-field*, as opposite to the *far-field*, reserved for the description of more conventional wave optics. Because the signal at any moment is collected

Fig. 8.15 Light intensity
distribution at a
sub-wavelength aperture
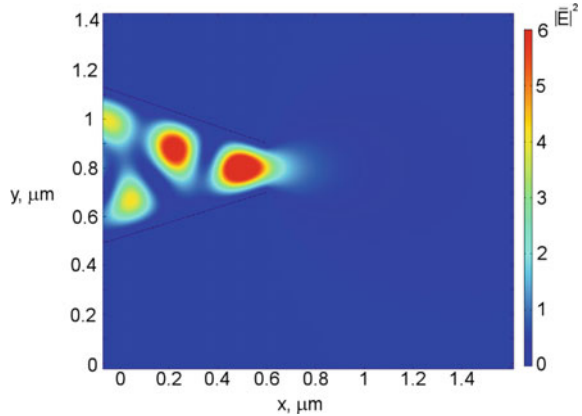
from a very small area, the *scanning* of the probe over the sample surface is necessary
for an image formation.

An example of the light intensity distribution around a sub-wavelength aperture is
shown in Fig. 8.15. It was obtained by the numerical solution of the Maxwell equa-
tions for the electromagnetic wave travelling along x-axis (electric field polarization
is out-of-plane). The spatial distribution of the field intensity (square of the electric
field absolute value) is shown for the area at the apex of the metal-coated silica fiber.
Free-space light wavelength is 500 nm and the aperture diameter is 200 nm.

It is seen that a standing wave is formed in the fiber due to multiple reflections of
the wave from the metallized edge surface. Yet it is also obvious that some of the light
intensity "protrudes" out of the sub-wavelength aperture, where it rapidly decays.
This extended field is relied on to localize the optical probe below the diffraction
limit for a SNOM operation.

The near-field type of interaction, in general, is not limited to subwavelength
apertures. Any oscillating dipoles, such as free electrons in a metal under the external
field, contain a far- and a near-field component. Light-mediated interactions at the
nanoscale through the evanescent wave is central in the area of nano-optics. Interested
reader is referred to specialized literature on this topic, where SNOM is just one of
many possible useful manifestations [35].

8.10 Operation Modes

In practice, SNOM operation includes an interplay between the near-field (evanescent
wave) and the far-field (propagating wave) for the image formation [36]. While the
near-field coupling is used to probe samples, ensuring a high spatial resolution, the

ultimate signal detection normally takes place in the far-field. Depending on the mutual arrangement the following modes can be defined:

- *Illumination.* It is schematically shown in Fig. 8.16, left. Here the illumination is realized via near-field, while the collection is done in the far-field. A probe with an aperture (an optical fiber with a deposited metal coating) focuses an illuminating wave to the sub-wavelength region. By means of a near-field excitation the energy is transferred to the sample, which, in turn, radiates light to the far-field for eventual detection with auxiliary optics. Here the resolution is defined by the aperture, which limits the excitation area.
- *Collection.* In this regime, shown in Fig. 8.16, second from the left, the operation sequence is inverse to the previous case. Far-field radiation illuminates the sample. Near-field coupling of the surface structures to the probe transfers part of the energy locally to the tip. The excited edge of the probe re-emits obtained energy to the far-field, which is collected at the far end of the fiber. The resolution is again defined by the aperture, which here limits the collection area.
- *Illumination-collection.* Next in Fig. 8.16 is a combination of these two regimes. Here both the excitation and collection is done via near-field coupling. First, a conversion from the far- to the near-field illuminates the sample with a high spatial resolution. The excited surface structures transfer back part of the energy to the probe tip via near-field interaction. Then the radiated energy inside the probe is collected through the same optical line, as used for the illumination.
- *Aperture-less.* The last scheme in Fig. 8.16, right, features a probe without an aperture (a metallic needle). Here both the illumination and collection is done in the far-field. The role of the probe is to locally enhance the incoming wave. This configuration corresponds to a broad family of tip-enhanced techniques. One can also add here a class of tips, which combine aperture and aperture-less designs, such as structured apertures or other sophisticated antenna arrangements [37].

Schematic representations of Fig. 8.17 can be quantitatively depicted using numerical solutions of the Maxwell equations for the light propagation. Two common calculation techniques are the finite-different time-domain method (FDTD) and the finite element method (FEM). In Fig. 8.17a result from FEM simulations is shown.
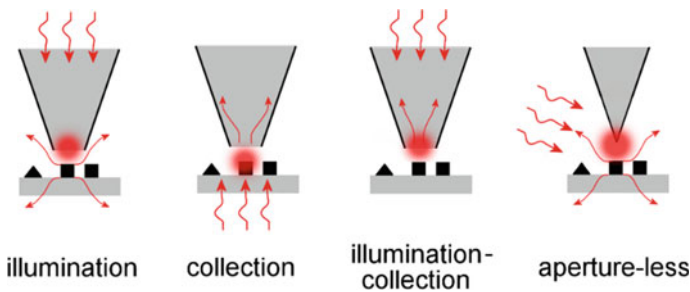

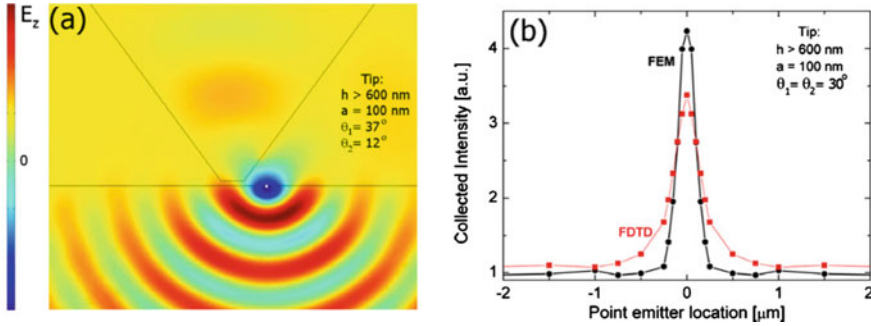**Fig. 8.16**   Different operation modes of the SNOM

Fig. 8.17 Numerical simulations of the light propagation from a point source on the substrate close to a 100 nm aperture: (left) electric field intensity; (right) collected signal as a function of the source position [37]

A point source is placed on a sample surface, representing a single oscillating dipole with electric field polarized out-of-plane [37]. This situation corresponds to the collection mode of the SNOM operation, as shown in Fig. 8.16, second from the left. A metal-protected probe with a ~ 100 nm aperture collects the signal (Fig. 8.17, left). The value of the collected signal can be calculated by integrating the light intensity in the probe, away from the aperture over the whole probe cross-section. A set of calculations for a varying source position was performed (Fig. 8.17, right) and the resulting line-shape represents collected signal for a scanning probe/source.

It is seen that the full-width at half-maximum (FWHM) of the signal profile roughly equals to the aperture width. This result does not depend on the simulation method (FEM or FDTD), confirming that the aperture size indeed defines the spatial resolution. Note that the collected signal will drop quickly with increased separation of the probe from the sample, signifying a near-field type of the energy transfer (cf. Figure 8.15) [38]. The free space wavelength used for simulations was 800 nm.

Historically, first aperture probes were fabricated from tapered optical fibers with metal coating [38, 39]. The development of aperture-free probes followed, where the metal tip itself acts as an antenna for the local field enhancement [40]. This feature of the aperture-less probe is illustrated by numerical simulations in Fig. 8.18. The light intensity distribution (square of the electric field absolute value) is shown for a 700 nm wavelength Gaussian light beam, entering from the left side. It was obtained by solving Maxwell equations by the FEM method. The incoming electric field was polarized along the y-axis. A gold needle (black) is placed in the middle.

It is seen that the needle partially reflects the light, forming a standing wave with the incoming beam (interference pattern to the left). At the tip of the needle one can notice a strong intensity enhancement, which quickly decays within tens of nanometers from the apex.

Indeed, a strong tip curvature results in significant modification of the field intensity distribution around the sharp tip. Similarly to the evanescent way from a subwavelength aperture, as shown in Fig. 8.15, this local enhancement is used to surpass

Fig. 8.18 Light intensity distribution around a metal tip
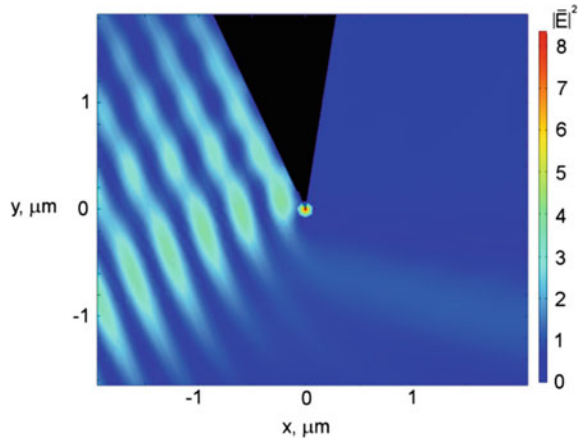
the diffraction limit for imaging and other characterization modalities, as discussed below.

The strong contribution from the scattered light, visible in Fig. 8.18, represents an unwanted parasitic signal. For the successful detection of the near-field signal it has to be distinguished and filtered from scattered part. A number of filtering techniques exist, such as time-tagging, which are frequently used in the rapidly developing area of the tip-enhanced near-field microscopy [41, 42].

Based on the data acquisition modality the SNOM operation can be alternatively categorized as:

- *Imaging.* This regime provides information about spatial distribution of the light absorption or reflection in the sample. It is obtained typically under white light (broadband) irradiation, which is an ordinary mode of the conventional optical microscope. An example, including a pre-fabricated metallic pattern with sub-wavelength features, is shown in Fig. 8.19 [43]. A scanning electron microscope image is included to illustrate real dimensions of the used aluminium pattern (Fig. 8.19a).

It is obvious from Fig. 8.19b, that the diffraction limit prevents resolving sample features in a conventional optical microscope. An illumination SNOM image (Fig. 8.19c), on the other hand, reveals a much better resolution under the white light. Here the aperture SNOM tip was used. The final image in this figure (Fig. 8.19d) is a result of post-processing, where the image was sharpened with the help of a deconvolution algorithm, using known instrument resolution.

The imaging regime is similar to other probe-related topography scanning techniques, such as STM or AFM. The difference is that it is not the tunneling current or the surface forces that form the image, but the optical field. One should keep in mind that the proximity of the tip to the sample may introduce new near-field interaction effects, resulting in image artefacts. So the topographical features may

Fig. 8.19 A sub-wavelength pattern as seen by **a** an SEM **b** an optical microscope **c** SNOM and **d** de-convoluted SNOM [44]
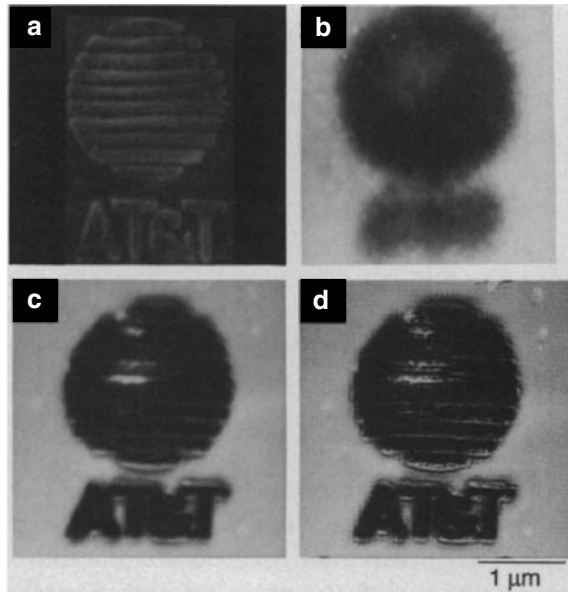
look different in a SNOM image due to optics-related near-field effects, such as polarization-dependence or an edge-related enhancement [38].

- *Photoluminescence.* When probing samples by the optical near-field it is more natural to focus on the light-related signal, which cannot be readily acquired in other probe-based microscopes (STM or AFM). Photoluminescence (PL) is one of such modalities. PL is material-specific light emission excited by an external light source. Detailed description of this effect is given in the micro-photoluminescence chapter. Important aspect to mention here is that the material response can be monitored with sub-wavelength resolution. In this regime the excitation light is of higher energy than the emitted photons. So a wavelength-based discrimination of the signal from the excitation should also be included in the SNOM-PL set-up. Semiconductor micro- and nanostructures are typically addressed by this approach. PL maps of the emission intensity, wavelength and decay times with sub-wavelength resolution can be acquired.
- *Raman Scattering.* Photon energy of the scattered light may be shifted by a certain energy from the incident beam energy. The discrete shift reflects certain vibration modes in the sample, excited by the impinging light. This method is called Raman spectroscopy and the corresponding energy shift is sometimes referred to as a Raman shift. Vibrations of molecules in chemical compounds, or specific phonon modes in solid state samples can be analyzed by this method.

Local enhancement of the light intensity at the aperture-less tip adds high-spatial resolution to this technique, called tip-enhanced Raman scattering (TERS). The value

of the Raman signal enhancement scales as a power of four of the local field enhancement. So the resulting values can be rather high, reaching sometimes millions [42], allowing sensitivity to access even single molecules. For applications in semiconductor micro- and nanostructures this method can yield information about the crystal phases in the material or about local stress at the heterostructure interfaces. The exact energy of the vibrational modes (phonons) is sensitive to the crystal structure configuration. Scattering from individual nanostructures can also be probed thanks to the high spatial resolution.

8.11 Equipment

8.11.1 SNOM Tips

The SNOM operation heavily depends on the tip quality. Indeed, it generates and collects evanescent field, defining spatial resolution of the instrument [45].

For the fabrication of *aperture* SNOM tips an optical fiber is typically used as a starting material. Tapering is often performed by hydrofluoric (HF) acid etching (Fig. 8.20). By slowly retracting a fiber from the HF solution the apex part becomes longer exposed to the etchant (HF chemically decomposes SiO_2, the fiber material). The method allows to control and to vary the apex angle by the retraction speed, HF concentration, etc. A care is needed, however, to protect outer fiber parts from an unwanted exposure. Some non-uniformities and surface defects may also be introduced. Alternatively, pulling a fiber under thermal treatment can result in a tip-shaped glass waveguide [39]. The surface of the glass fiber then becomes more smooth and uniform than after the chemical etching [43]. A combination of these two methods is also used to achieve a high tip quality [46].

After the shaping of the glass fiber a metal deposition on the tapered tip is carried out. It is usually performed on a rotating tip at an angle from behind. This geometry ensures that only a small aperture remains open, thanks to the shadowing effect [39]. A complete deposition over the whole fiber surface with subsequent focused ion beam (FIB) sculpturing to open up the aperture can also be employed [47].
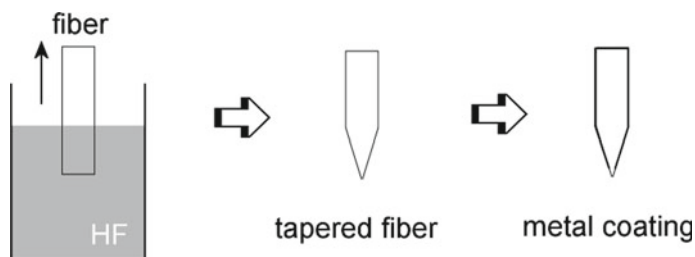

Fig. 8.20 Basic steps for fabrication of an aperture SNOM tip

In Fig. 8.21 scanning electron microscope images of apex areas of different tips are shown. A tapered fiber with a smooth surface prepared by pulling and HF etching is depicted in Fig. 8.21a. In Fig. 8.21b a metal-covered tapered fiber is shown (top view). The uncovered optical fiber part is seen as a round dark aperture of ~ 100 nm diameter. It is clear that the metal deposition layer (Al) consists of grains. Therefore, a thick enough deposition is needed for a good quality surface coverage. The grain-like structure of the coating (Pt–Pd) is also visible in Fig. 8.21c. Here SEM images before and after the metal deposition are overlaid on each other. The optical fiber part can be seen as a tapering brighter part in the middle. To fabricate a ~ 150 nm diameter aperture the whole tip was covered by the metal film and only the tip part was opened with the help of FIB etching.

Aperture-less probes, which are essentially very small needles, are, in general, easier to fabricate and to apply [41, 48]. Metal tips are common, because free electron oscillations can provide a strong local field. Plasmonic effects in metals can further contribute to the field amplification. Semiconducting materials, on the other hand, may have an advantage of well-established processing technology, allowing mass-fabrication of tips. For example, one of the first realizations of the aperture-less SNOM employed a silicon tip with ~ 2 nm tip apex radius [40].

Main tools for the tip preparation are the chemical and/or ion-beam induced etching. Direct sculpturing of gold wires in FIB can form sharp enough curvatures at the tip apex [41]. A more advanced design includes the exploitation of plasmonic resonance effects, such as shown in Fig. 8.22a. Three different diameter gold spheres respond resonantly to the impinging 650 nm laser light and to each other, creating a strong net field enhancement at the apex.

Similarly, very narrow needles can be grown at the conventional tip edges by electron beam induced deposition (EBID), a method similar to the chemical vapor deposition. Electron beam decomposes chemical precursors on the surface, forming a solid compound under the electron beam. Alternatively, thin and long nanostructures, such as carbon nanotubes, can be attached to the tips by FIB or EBID.
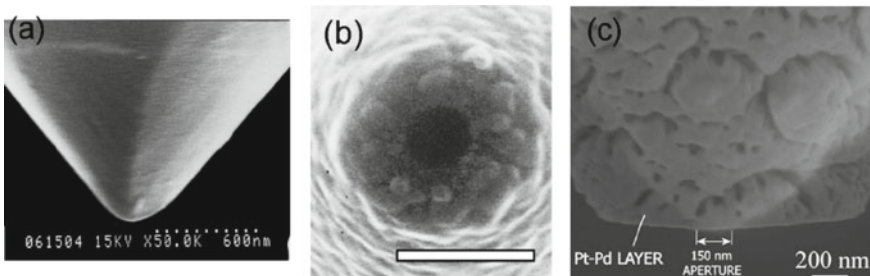

Fig. 8.21 SEM images of aperture SNOM tips: **a** Side view of a tapered fiber. Reprinted with permission from [46]. Copyright (1997) by the American Vacuum Society. **b** Top view of a tip covered by metal, deposited at an angle. Scale bar is 300 nm. Reprinted with permission from [39]. Copyright (2000) by AIP Publishing. **c** Overlaid side views of a metal-covered FIB-sculptured tip [47]
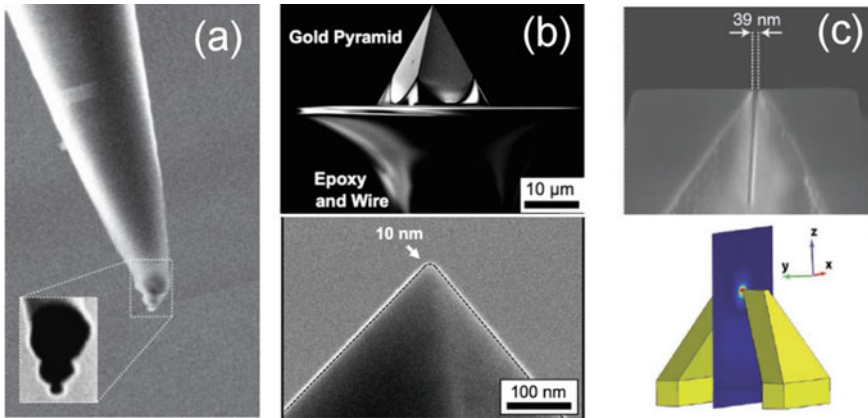
Fig. 8.22 SEM images of aperture-less SNOM probes. **a** Three gold spheres deposited on a dielectric tip. Reprinted with permission from [49]. Copyright (2011) by Springer Nature. **b** Gold pyramids prepared from a mold array in silicon [50]. **c** Hybrid aperture and aperture-less tip. A slit made at the apex with FIB ensures local field enhancement for an aperture SNOM tip, as shown by field strength simulations below [45]

An example of semiconductor processing for the aperture-less tip fabrication is presented in Fig. 8.22a–b. Using lithography and etching of a silicon substrate an array of molds can be prepared. Depositing gold in such pits and lifting it out by an epoxy layer produces metal-covered pyramids with sharp edges, as a high-resolution SEM image below illustrates. This geometry is defined by the mold, which shape is formed by the anisotropic KOH etching of silicon [50].

Taking an aperture probe and enhancing its local field performance by additional modifications is another approach [45]. An example of such a *hybrid* design is shown in Fig. 8.22c. Here a slit was fabricated at the apex of a metal-covered optical fiber to provide extra field enhancement, as simulations below clearly illustrate.

8.11.2 SNOM Set-Up

The whole instrument arrangement can be separated into several main components. Sample approach routine for the tip is typically that of an atomic force microscope (AFM) or of a scanning tunneling microscope (STM). In the former case the tip is attached to a cantilever, whose vibration frequency is monitored on the way to the sample surface. In the close sample proximity, the frequency is altered by the surface, signaling that a working distance reached. In the latter mode a tunneling current between the tip and the substrate is monitored, which becomes measurable at a few nm separations. A conductive tip and a conductive substrate are necessary conditions for this type of sample approach.
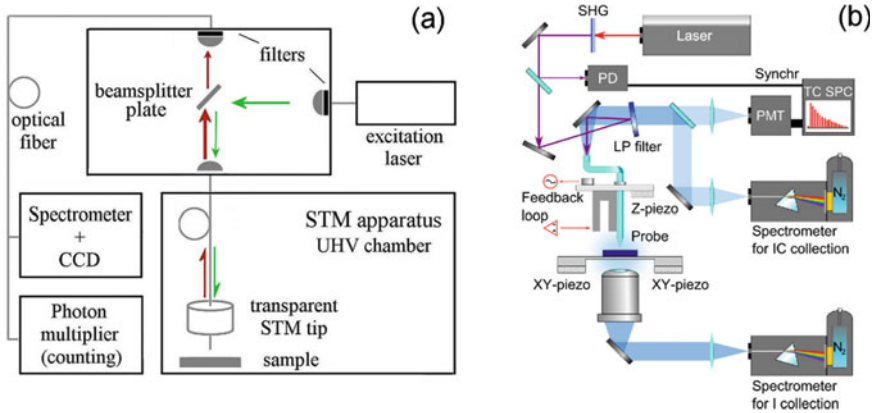
Fig. 8.23   a Schematics of an illumination-collection SNOM set-up based on STM sample approach for cooled samples in vacuum. Reprinted with permission from [47]. Copyright (2009) by the IOP Publishing. **b** Schematics of illumination (I) and illumination-collection (IC) SNOM set-up based on AFM sample approach for samples at ambient [51]

In Fig. 8.23 two SNOM set-ups based on these routines are schematically shown. Optical fibers are attached in both cases to realize the aperture SNOM tip configuration. Additional optical components are integrated for the management of the excitation and collection beams.

A sample cooling function exists in the set-up shown in Fig. 8.23a. For that the sample is mounted in a vacuum chamber, preventing surface condensation when the metal holder is cooled to liquid nitrogen or helium temperatures. Clean surfaces can be prepared under ultra-high vacuum (UHV) conditions by ion sputtering or sample annealing. Those are helpful in acquiring high quality topographical scans and in the reduction of parasitic signal contribution from the surface contaminants. Piezo-driven stages ensure both the sample approach and the in-plane scanning.

Standard optical components, such as lenses, filters (such as a long pass (LP) filter), dichroic mirrors, and ordinary mirrors are used for driving the light to and from the sample. The optical line can be arranged either via free space or through an optical fiber. Spectrometers and photodetectors (PD) are attached for signal analysis; their detailed description is provided in the micro-PL chapter.

Nowadays SNOM instruments are commercially available, complimenting widespread probe-based family of characterization tools, such as AFM or STM.

8.12   Application to Micro- and Nano-Scale Materials

In this section some examples, illustrating different SNOM operation modes, will be presented for micro- and nano-scale materials. Semiconductor and metal nano- and micro-structures are mainly described. An important practical condition for SNOM

operation is flat topography of the samples, so that artefacts related to the unevenness do not obscure/distort material-specific response.

8.12.1 Imaging

Conventional optical imaging is based on the refractive index variations across the sample. Both real (refraction) and imaginary (absorption) parts of the refractive index contribute to the contrast formation in this case. As shown in Fig. 8.19, the metal-dielectric contrast between the stripes of aluminium on glass can be clearly resolved in SNOM-imaging, using white light in illumination mode.

Additionally, within the same material there can be areas (grains or domains) with different optical properties. In particular, their response to the polarized light may vary. One such example would be magnetic domains. The change in the polarization of the reflected light (so called magneto-optical Kerr effect) for them is different than from the rest of the sample. Based on this property, a polarization-resolved SNOM measurements can be used to visualize their presence in the sample with a high spatial resolution [38].

Similarly, an image contrast can be based on the refractive index polarization dependence (birefringence) within the material for non-magnetic samples. An example would be a compound with different crystal phases (metallic and insulating), such as vanadium oxide. Around the phase transition temperature these two phases can actually co-exist, being distributed in different grains. Because the metal–insulator transition is around room temperature this material is often investigated as a model system for such crystal phase transformations.

In Fig. 8.24 an example of a SNOM image of VO_2 microcrystals is shown [52]. Aperture-less SNOM tip under 632 nm laser illumination was applied. Darker and brighter areas can be seen on the image. They correspond to metal and insulator phases of the vanadium oxide. The intensity profile along the white dotted line is shown to the right (blue curve). Clearly the grain imaging with a subwavelength resolution has been achieved.

Due to the different unit cell size for these two phases, the undulations in the topographical scan (black curve) can also be seen. They strongly correlate with the SNOM-derived contrast, confirming phase-related contrast. Thus, with the help of

Fig. 8.24 (left) SNOM image of a VO microcrystal and (right) a line profile of the SNOM intensity and of the topographical height [52]
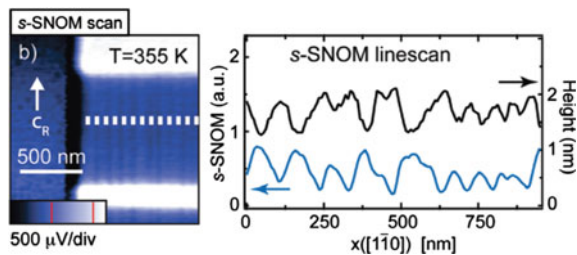
SNOM imaging, the dynamics of the metal phase formation and growth process was demonstrated and analyzed as a function of temperature [52].

8.12.2 Photoluminescence Spectroscopy

Photo-excited light emission from materials carry information about the lowest excited energy state and its relaxation transition strength. Therefore, for semiconductor micro- and nanostructures, SNOM-PL can probe bandgap variations across the sample. These variations may originate from the quantum size effect, such as in quantum wells, nanowires or quantum dots. Chemical composition variations for compound materials may also induce changes in the bandgap.

The distribution of different recombination centers, both radiative and non-radiative, with varying properties, such as lifetime, are also accessible. In case of samples with a relatively high surface nanoparticle density, nanostructures can be resolved and probed individually. While the details of the PL technique are described in the micro-PL chapter, here several examples of the high spatial resolution SNOM-PL are demonstrated.

One of the first successful attempts of a sub-wavelength imaging of nanostructures with a SNOM-PL was presented in [53]. The sample contained single (SQW) and multiple quantum wells (MQW), as well as nanowires of GaAs with AlGaAs barriers. The aperture SNOM was applied in a collection mode under a far-field 750 nm laser excitation (Fig. 8.25, left). The PL signal was detected across the sample cross-section.

The exact emission wavelength of the different components of the sample (SQW, MQW, nanowire) slightly varies, as can be seen in the PL spectrum in Fig. 8.25
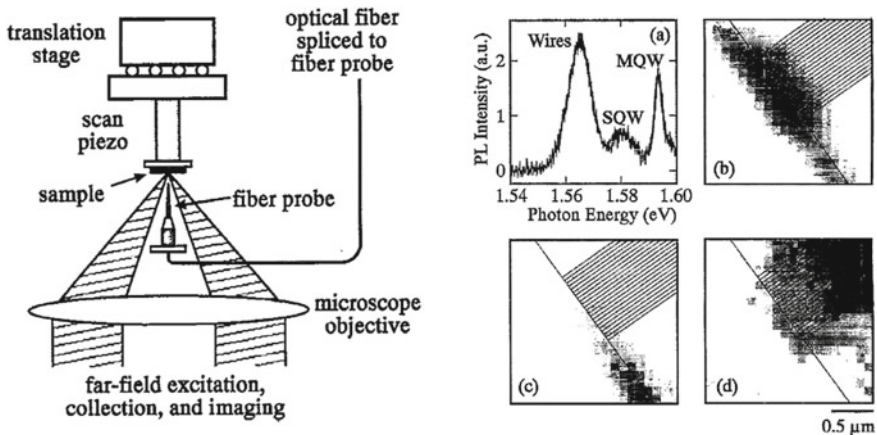

Fig. 8.25 (left) Collection mode of a SNOM-PL. (right) [**a**] PL spectrum and three maps of the PL intensity acquired separately for the three PL peaks: **b** wires, **c** single QW, **d** MQW [53]

(right, a). Therefore spatial maps corresponding to each structure could be acquired by detecting PL of a specific wavelength. In this way a spatial compositional map was obtained with a subwavelength resolution (Fig. 8.25, right, b–d).

8.12.3 Photoluminescence Lifetime

One of the first examples of resolving individual particles (quantum dots), located within a diffraction limited spot (less than a micron apart), was demonstrated in [54]. In the inset of Fig. 8.26 a PL image of a ~ $3 \times 3 \ \mu m^2$ sample surface area is shown. Individual bright spots correspond to the position of self-assembled InGaAs/GaAs quantum dots, grown by molecular beam epitaxy. The sample was cooled to liquid helium temperature after it had been placed in a vacuum chamber. The SNOM-PL data were acquired in the illumination-collection mode. Although the surface density of QDs is not uniform, areas where single particles are resolvable could be found.

QD individual lifetimes could be measured by locating the tip over the object of interest. Single QDs were excited by driving a pulsed laser light of 780 nm through the same optical line. Spectral filtering by a monochromator was applied to isolate a particular energy state, the so-called spectrally resolved lifetime measurements. The signal was detected by an avalanche photodiode, which operation principle is discussed in details in the micro-PL chapter. Synchronization with the excitation pulse and dichroic mirrors ensure separation of the emission and excitation beams. Here, carrier lifetimes of hundreds of ps were found in the individual emitters.
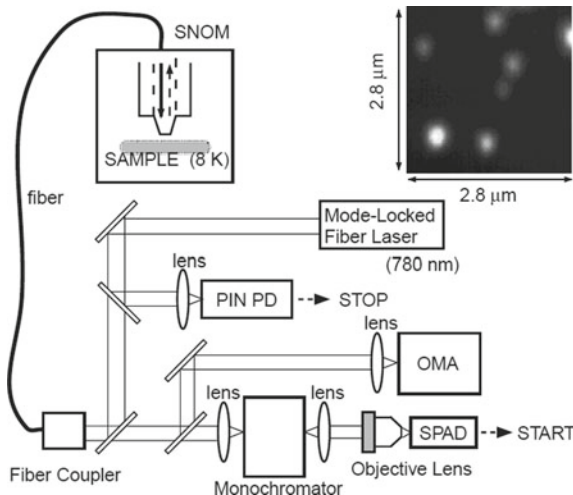

Fig. 8.26 SNOM set-up for low-temperature single-dot PL measurements. Inset is an SNOM-PL image [54]

8.12.4   *Photoluminescence Mapping*

A combination of the spectral and temporal measurements with a high spatial resolution is an ultimate characterization a SNOM-PL system can achieve. An in-plane distribution of the emission peak wavelengths and decay times in the InGaN quantum well was reported in [55, 56]. This material is important for the blue light-emitting diodes. Their emission can be readily converted to the white-light emission by luminescent phosphors; such diodes are already common in general lighting applications. To improve their efficiency and stability detailed information about the structural and related carrier recombination properties is an important feedback to the fabrication process.

Band structure variations at the nanometer scale exist in such QWs due to compositional and stress-related non-uniformities, as well as due to the in-built electric fields. These effects lead to the variations of the bandgap (PL peak position) and lifetime. Therefore, spatial mapping of such parameters can be used as a probe to the local band structure as well as defect monitoring.

In Fig. 8.27a topographical image, a distribution of emission peak wavelength (PL wavelength map), and a PL decay times map are shown for the same area [55]. An illumination-collection mode of a SNOM was applied with a pulsed laser excitation of 380 nm. The collected signal was separated between two arms: one for the lifetime detection by a photomultiplier tube, and the other for spectroscopy analysis. Unlike in the previous example, spectral and temporal information was recorded at every point of the scanned map.

One should note that for such samples it is not only the aperture size, which defines spatial resolution of the measurement system. Carrier diffusion in the material may lead to the broadening of the excitation spot. With diffusion constants being a function of the material and even direction a careful consideration of this effect required for the correct assessment of the system resolution.

From the PL maps shown in Fig. 8.27 one can extract 2D distributions of the radiative and non-radiative recombination lifetime in these quantum wells. This information can be correlated with the PL intensity, linewidth and emission wavelength. Conclusions regarding the structure influence on the light emitting properties can be
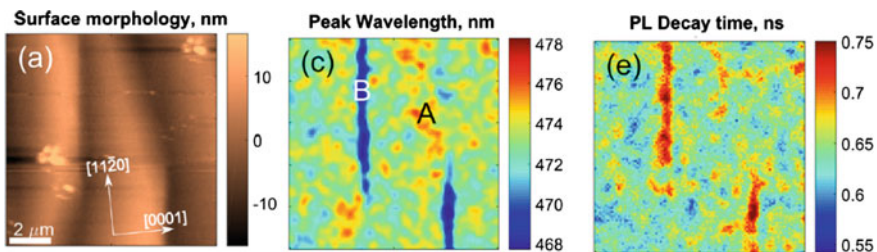

Fig. 8.27   (left) Topological image by the SNOM probe of a InGaN QW (middle) PL peak position maps, and (right) PL decay lifetime maps of the same are [55]

drawn from such comparison. For example, it was found that a reduced planarity of the interfaces leads to a larger carrier separation and a reduced light emission efficiency [55].

8.12.5 Photoluminescence Excitation

In contrast to the ordinary PL spectroscopy, where the emitted light is spectrally analyzed, in the photoluminescence excitation spectroscopy (PLE) the excitation wavelength is a variable parameter. The intensity of the emitted light in this case reveals absorption spectrum of the sample.

For the SNOM set-up and the sample shown in Fig. 8.26 instead of a pulsed excitation laser with a fixed wavelength a tunable laser can be used. Then, the absorption spectra of an individual quantum dot is possible to obtain.

In Fig. 8.28a a PLE spectra from an ensemble of InGaAs QDs is shown [57]. A couple of broad features in the absorption spectrum can be discerned. When a SNOM-PLE is applied, on the other hand (Fig. 8.28b), a fine structure of these broad features becomes visible. So the discrete energy structure of a QD cannot be easily revealed by ensemble measurements due to averaging effects, but it becomes apparent for an instrument capable of single particle interrogation.
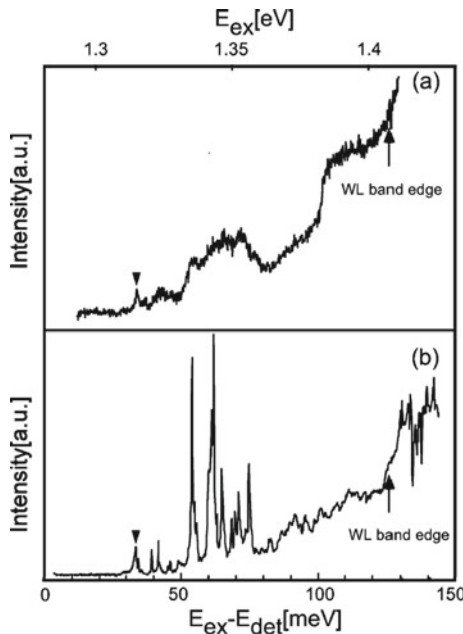

Fig. 8.28 a Far-field PLE spectrum of a QD ensemble. **b** SNOM-PLE data from individual QDs. Energy states above the lowest state are resolved [57]

8.12.6   Raman Scattering

While photoluminescence modality detects material-specific emission, the Raman method is based on the impingent light scattering. Therefore aperture-less probes are typically applied for such measurements. Spectral analysis of the scattered light, however, may be arranged similarly to the SNOM-PL configurations.

Consider an example, shown in Fig. 8.29. Here a graphene monolayer, grown on silicon dioxide substrate was imaged under different laser wavelengths [58]. Graphene layers can serve as transparent electrodes and this 2D materials is currently being actively investigated for applications in nanoelectronics and nanophotonics.

Because the atom vibration energy (phonon modes) are material-specific, scattered light intensity maps taken at different energies will reveal the material contrast. At a particular excitation energy (explicitly indicated in Fig. 8.29b, for every image) the sample materials show stronger or weaker signal. With this imaging capability a coupling of the excitation to voltage-dependent plasmonic modes for graphene was demonstrated [58].

Similarly, ferroelectric domains, featuring different phonon modes, can be imaged at nanoscale with the help of TERS [59]. An electrochemically etched gold tip of the aperture-less SNOM was used to locally enhance the illumination field from the 632 nm laser (Fig. 8.30). A $BaTiO_3$ nanorod was probed, whose topographical scan is presented in Fig. 8.30a.

The area at the top of the nanoparticle appears brighter in the spectra-integrated TERS image (Fig. 8.30b). This fact reflects contribution from more phonon modes, permitted in the a_y domain as opposite to the c-domain of the material. Schematic distribution of the ferroelectric domains extracted from the aperture-less SNOM scans is shown in Fig. 8.30d. Three phonon modes are permitted for the a_y domain (A_1, TO, and E_{TO}), yielding a stronger scattered light intensity than for c-domain, where only two such modes allowed (A_1 and LO).
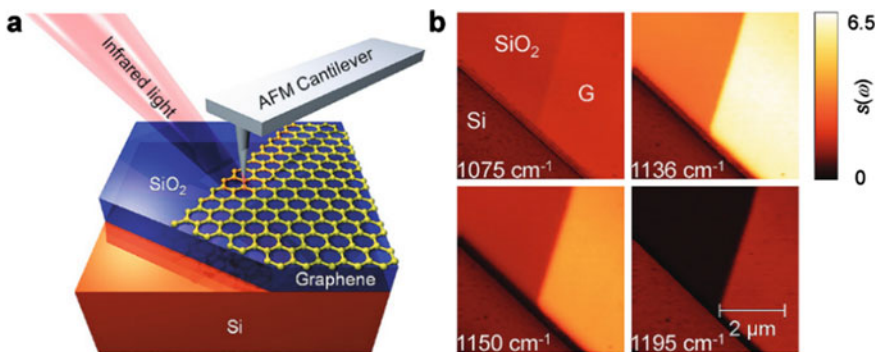

Fig. 8.29   a Schematics of an aperture-less SNOM experiment on graphene. **b** Images at different Raman shift frequencies, showing material contrast [58].
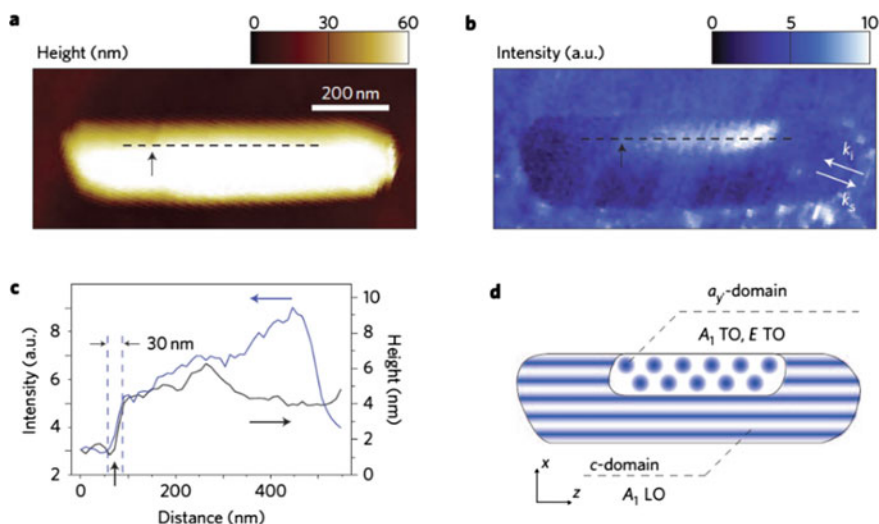
Fig. 8.30 **a** Topographical scan of a BaTiO$_3$ nanorod, **b** TERS map, **c** intensity profiles along the dotted lines, and **d** schematics of domain distribution derived from **b** [59]

References

1. Binnig, G., Rohrer, H., Gerber, C., & Weibel, E. (1982). Surface studies by scanning tunneling microscopy. *Physical Review Letters, 49*(1), 57.
2. Martin, Y., & Wickramasinghe, H. K. (1987). Magnetic imaging by "force microscopy"with 1000 ˚a resolution. *Applied Physics Letters, 50*(20), 1455–1457.
3. Prater, G. B., Hansma, P. K., Tortonese, M., & Quate, C. F. (1991). Improved scanning ion-conductance microscope using microfabricated probes. *Review of Scientific Instruments, 62*, 2634–2638.
4. Schaeffer, T. E., Viani, M., Walters, D. A., Drake, B., Runge, E. K., Cleveland, J. P., Wendman, M. A., & Hansma, P. K. (1997). Atomic force microscope for small cantilevers. *SPIE, 3009*, 48–52.
5. Jagtap, R., et al. (2006). Overview literature on atomic force microscopy (AFM): Basics and its important applications for polymer characterization. *Indian Journal of Engineering and Materials Sciences, 13*, 368.
6. Sadewasser, S., & Glatzel, T. E. (2012). *Kelvin probe force microscopy*, Number 334 in Springer Series in Surface Sciences (1st ed.). Springer. ISBN 978-3-319-75686-8
7. Boisen, A., Hansen, O., & Bouwstra, S. (1996). AFM probes with directly fabricated tips. *Journal of Micromechanics and Microengineering, 6*, 58. https://doi.org/10.1088/0960-1317/6/1/012
8. Weymouth, A. J. (2017). Lateral force microscopy (LFM). *Journal of Physics: Condensed Matter, 29*, 323001.
9. Israelachvili, J. N. (2011). *In practice, the electrostatic force is attractive and is sense for a distance in range of 10nm, intermolecular and surface forces.* Academic Press. ISBN: 9780123919274.
10. Meyer, E. (1992). Atomic force microscopy. *Progress in Surface Science, 41*(1), 3–49.
11. Melitz, W., Shen, J., Kummel, A. C., & Lee, S. (2011). Kelvin probe force microscopy and its application. *Surface Science Reports, 66*(1), 1–27.
12. Weisenhorn, A. L., Hansma, P. K., Albrecht, T. R., & Quate, C. F. (1989). Forces in atomic force microscopy in air and water. *Applied Physics Letters, 54*, 2651–2653.

13. Hartmann, U. (1990). van der Waals interactions in force microscopy. *Advanced Materials, 2*, 594–597.
14. Walczyk, W., & Schonherr, H. (2014). Dimensions and the profile of surface nanobubbles: Tip-nanobubble interactions and nanobubble deformation in atomic force microscopy. *Langmuir, 30*, 11955–11965.
15. Lee, S. W. (2016). Mechanical properties of suspended individual carbon nanotube studied by atomic force microscope. *Synthetic Metals, 216*, 88–92.
16. Becker, R. S., Golovchenko, J. A., & Swartzentruber, B. S. (1987). Atomic-scale surface modifications using a tunneling microscope. *Nature, 325*, 419–421.
17. Sohn, L. L., & Willett, R. L. (1995). Fabrication of nanostructures using atomic-force-microscope-based lithography. *Applied Physics Letters, 67*, 1552–1554.
18. Bouchiat, V., & Esteve, D. (1996). Lift-off lithography using an atomic force microscope. *Applied Physics Letters, 69*, 3098–3100.
19. Piner, R. D., Zhu, J., Xu, F., Hong, S. H., & Mirkin, C. A. (1999). "Dip-pen" nanolithography. *Science, 283*, 661–663.
20. Zhang, F. C., et al. (2010). Mechanical manipulation assisted self-assembly to achieve defect repair and guided epitaxial growth of individual peptide nanofilaments. *ACS Nano, 4*, 5791–5796.
21. Abdellaoui, T., Bardaoui, A., Daoudi, M., & Chtourou, R. (2010). Correlation of atomic force microscopy and photoluminescence analysis of GaAs nanocrystallites elaborated by electrochemical etching of n+type GaAs. *The European Physical Journal Applied Physics, 51*, 20501.
22. Yan, Y., et al. (2010). Top-down nanomechanical machining of three-dimensional nanostructures by atomic force microscopy. *Small (Weinheim an der Bergstrasse, Germany), 6*, 724–728.
23. Chen, T.-H., Hsu, H.-F., & Wu, H.-Y. (2012). Formation of Ni-silicide nanowires on silicon-on-insulator substrates by atomic force microscope lithography and solid phase reaction. *ECS Journal of Solid State Science and Technology, 1*, P90–P93.
24. Temiryazev, A. (2014). Pulse force nanolithography on hard surfaces using atomic force microscopy with a sharp single-crystal diamond tip. *Diamond and Related Materials, 48*, 60–64.
25. Lee, D. H., Kim, C. K., Lee, J.-H., Chung, H.-J., & Park, B. H. (2016). Fabricating in-plane transistor and memory using atomic force microscope lithography towards graphene system on chip. *Carbon, 96*, 223–228.
26. Binnig, G., Quate, C. F., & Gerber, C. (1986). Atomic force microscope. *Physical Review Letters, 56*(9), 930. https://doi.org/10.1103/PhysRevLett.56.930
27. Singh, A. K. (2017). Engineered nanoparticles: Structure, properties and mechanisms of toxicity. *Bulletin, 42*, 75. https://doi.org/10.1557/mrs.2016.320
28. Kajen, R. S., Chandrasekhar, N., Feng, X., Müllen, K., & Su, H. (2011). Vibrational excitations in molecular layers probed by ballistic electron microscopy. *Nanotechnology, 22*(43), 6. https://doi.org/10.1088/0957-4484/22/43/435701
29. Butt, H.-J., Cappella, B., & Kappl, M. (2005). Force measurements with the atomic force microscope: Technique, interpretation and applications. *Surface science reports, 59*(1), 1–152. https://doi.org/10.1016/j.surfrep.2005.08.003
30. Garcıa, R., & Perez, R. (2002). Dynamic atomic force microscopy methods. *Surface Science Reports, 47*(6), 197–301. https://doi.org/10.1016/S0167-5729(02)00077-8
31. Hou, A., Nechay, B. A., & Bloom, D. M. (1996). Scanning probe microscopy for testing ultrafast electronic devices. *Optical and Quantum Electronics, 28*, 819–841. https://doi.org/10.1007/BF00820151
32. Cadena, M. J., Sung, S. H., Boudouris, B. W., Reifenberger, R., & Raman, A. (2016). Nanoscale mapping of dielectric properties of nanomaterials from kilohertz to megahertz using ultrasmall cantilevers. *ACS Nano, 10*, 4062–4071. https://doi.org/10.1021/acsnano.5b06893
33. Nonnenmacher, M., O'Boyle, M., & Wickramasinghe, H. K. (1991). Kelvin probe force microscopy. *Applied Physics Letters, 58*(25), 2921–2923.

34. Garcia-Vidal, F. J., Martin-Moreno, L., Ebbesen, T. W., & Kuipers, L. (2010). Light passing through subwavelength apertures. *Reviews of Modern Physics, 82*, 729–787.

35. Kawata, S., Ohtsu, M., Irie, M. (2002). *Electromagnetism theory and analysis for near-field nano-optics, nano-optics* (pp. 35–59). Springer. ISBN: 978-3-540-45273-7

36. Greffet, J. J., & Carminati, R. (1997). Image formation in near-field optics. *Progress in Surface Science, 56*, 133–237.

37. Sychugov, I., Omi, H., Murashita, T., & Kobayashi, Y. (2008). Modeling tip performance for combined STM-luminescence and aperture-SNOM scanning probe: Spatial resolution and collection efficiency. *Applied Surface Science, 254*, 7861–7863.

38. Betzig, E., & Trautman, J. K. (1992). Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit. *Science, 257*, 189–195.

39. Hecht, B., Sick, B., Wild, U. P., Deckert, V., Zenobi, R., Martin, O. J. F., & Pohl, D. W. (2000). Scanning near-field optical microscopy with aperture probes: Fundamentals and applications. *The Journal of Chemical Physics, 112*, 7761–7774.

40. Zenhausern, F., Oboyle, M. P., & Wickramasinghe, H. K. (1994). Apertureless near-field optical microscope. *Applied Physics Letters, 65*, 1623–1625.

41. Novotny, L., Stranick, S. J. (2006). Near-field optical microscopy and spectroscopy with pointed probes. Annual Review of Physical Chemistry, 303–331.

42. Hartschuh, A. (2008). Tip-enhanced near-field optical microscopy. *Angewandte Chemie International Edition, 47*, 8178–8191.

43. Hoffmann, P., Dutoit, B., & Salathé, R.-P. (1995). Comparison of mechanically drawn and protection layer chemically etched optical fiber tips. *Ultramicroscopy, 61*, 165–170.

44. Trautman, J. K., Betzig, E., Weiner, J. S., DiGiovanni, D. J., Harris, T. D., Hellman, F., & Gyorgy, E. M. (1992). Image contrast in near-field optics. *Journal of Applied Physics, 71*, 4659–4663.

45. Schuck, P.J., Weber-Bargioni, A., Ashby, P.D., Ogletree, D.F., Schwartzberg, A., Cabrini, S. (2013). Life beyond diffraction: opening new routes to materials characterization with next-generation optical near-field approaches. Advanced Functional Materials, *23*, 2539–2553.

46. Murashita, T. (1997). Novel conductive transparent tip for low-temperature tunneling-electron luminescence microscopy using tip collection. *Journal of Vacuum Science and Technology B, 15*, 32–37.

47. Sychugov, I., Omi, H., Murashita, T., & Kobayashi, Y. (2009). Optical and electrical characterization at the nanoscale with a transparent probe of a scanning tunnelling microscope. *Nanotechnology, 20*, 145706.

48. Mauser, N., & Hartschuh, A. (2014). Tip-enhanced near-field optical microscopy. *Chemical Society Reviews, 43*, 1248–1262.

49. Novotny, L., & van Hulst, N. (2011). Antennas for light. *Nature Photonics, 5*, 83.

50. Johnson, T. W., Lapin, Z. J., Beams, R., Lindquist, N. C., Rodrigo, S. G., Novotny, L., & Oh, S.-H. (2012). Highly reproducible near-field optical imaging with sub-20-nm resolution based on template-stripped gold pyramids. *ACS Nano, 6*, 9168–9174.

51. Mensi, M., Ivanov, R., Uzdavinys, T. K., Kelchner, K. M., Nakamura, S., DenBaars, S. P., Speck, J. S., & Marcinkevicius, S. (2018). Direct measurement of nanoscale lateral carrier diffusion: Toward scanning diffusion microscopy. *ACS Photonics, 5*, 528–534.

52. Jones, A. C., Berweger, S., Wei, J., Cobden, D., & Raschke, M. B. (2010). Nano-optical investigations of the metal−insulator phase behavior of individual VO_2 microcrystals. *Nano Letters, 10*, 1574–1581.

53. Grober, R. D., Harris, T. D., Trautman, J. K., Betzig, E., Wegscheider, W., Pfeiffer, L., & West, K. (1994). Optical spectroscopy of a GaAs/AlGaAs quantum-wire structure using near-field scanning optical microscopy. *Applied Physics Letters, 64*, 1421–1423.

54. Ono, M., Matsuda, K., Saiki, T., Nishi, K., Mukaiyama, T., Kuwata-Gonokami, M. (1999). Time-resolved emission from self-assembled single quantum dots using scanning near-field optical microscope. Japanese Journal of Applied Physics Part 2-Letters, *38*, L1460–L1462.

55. Ivanov, R., Marcinkevicius, S., Uzdavinys, T. K., Kuritzky, L. Y., Nakamura, S., & Speck, J. S. (2017). Scanning near-field microscopy of carrier lifetimes in m-plane InGaN quantum wells. *Applied Physics Letters, 110*, 031109.

56. Marcinkevicius, S., Zhao, Y., Kelchner, K. M., Nakamura, S., DenBaars, S. P., & Speck, J. S. (2013). Near-field investigation of spatial variations of (20(2)over-bar(1)over-bar) InGaN quantum well emission spectra. *Applied Physics Letters, 103*, 131116.
57. Toda, Y., Moriwaki, O., Nishioka, M., & Arakawa, Y. (1999). Efficient carrier relaxation mechanism in InGaAs GaAs self-assembled quantum dots based on the existence of continuum states. *Physical Review Letters, 82*, 4114–4117.
58. Fei, Z., Andreev, G. O., Bao, W., Zhang, L. M., McLeod, A. S., Wang, C., Stewart, M. K., Zhao, Z., Dominguez, G., Thiemens, M., Fogler, M. M., Tauber, M. J., Castro-Neto, A. H., Lau, C. N., Keilmann, F., Basov, D. N. (2011). Infrared nanoscopy of dirac plasmons at the graphene–SiO_2 interface. *Nano Letters*, *11*, 4701–4705.
59. Berweger, S., Neacsu, C. C., Mao, Y., Zhou, H., Wong, S. S., & Raschke, M. B. (2009). Optical nanocrystallography with tip-enhanced phonon Raman spectroscopy. *Nature nanotechnology, 4*, 496.