Vilnius University

Faculty of Physics

Laboratory of Atomic and Nuclear Physics

Experiment No. 4

STUDY OF CHARACTERISTIC X-RAYS OF VARIOUS MATERIALS

by Andrius Poškus

(e-mail: andrius.poskus@ff.vu.lt)

2026-11-25

Contents

Objective of the work **1**

1. Tasks **1**

2. Control questions **1**

Literature 1

3. X-ray radiation **2**

3.1. Sources of X-ray radiation 2

3.2. Bremsstrahlung X-ray radiation 2

3.3. Characteristic X-ray radiation 5

3.4. Moseley's law 7

3.5. Fine structure of the characteristic X-ray spectrum 10

4. Interaction of X-ray radiation with matter **12**

4.1. Concept of interaction cross section 12

4.2. Photoelectric effect 13

4.3. Coherent scattering 15

4.4. Compton scattering 17

4.5. Attenuation coefficient 20

5. Properties of Semiconductor Ionizing Radiation Detectors **22**

5.1. Simplified detector model 22

5.2. Detector operation 23

5.3. Pulse height spectra 24

5.4. Detector energy resolution 25

5.5. Operating principle of a semiconductor detector 26

5.6. Electron Energy Levels in a Solid 27

5.6.1. Energy bands. Dielectrics, metals, and semiconductors 28

5.6.2. Impurity energy levels 30

5.6.3. n-type and p-type semiconductors 30

5.7. Potential Barrier at a Semiconductor–Semiconductor or Metal–Semiconductor Junction . . . 32

5.7.1. Built-in potential 32

5.7.2. Depletion layer and its thickness dependence on external voltage 33

5.7.3. Role of the depletion layer in a semiconductor detector 34

5.7.4. Depletion layer formation technology. Si-PIN detectors 35

5.8. Effect of Ionizing Radiation on a Semiconductor 36

5.8.1. Average ionization energy 36

5.8.2. Dynamics of creation and disappearance of free charge carriers 36

5.9. Detector Efficiency 37

6. Diffraction of X-Rays in Crystals **38**

6.1. Crystals, crystal lattice, unit cell 39

6.2. Diffraction of waves in a crystal 40

6.3. Bragg's law 41

6.4. Symmetry of crystals 42

6.5. Miller indices of crystallographic planes and crystallographic axes 45

7. Theory of the Experimental Method **47**

7.1. Measuring the wavelength of X-ray radiation using Bragg's law 47

7.2. Measuring X-ray photon energy using a Si-PIN detector 48

7.3. Detector calibration and response function 49

7.4. Measurement of the X-ray fluorescence spectrum 51

8. Experimental Procedure **53**
8.1. Equipment and measurement procedure 53
8.2. Analysis of Measurement Data 56
8.3. Linear Fitting 63

Objective of the work

Measure the photon energies of characteristic X-ray radiation emitted by an X-ray tube with a molybdenum anode and of X-ray fluorescence of various materials, verify Moseley's law, and measure the Rydberg constant and the screening constant.

1. Tasks

1. Using a semiconductor (Si-PIN) detector, measure the radiation spectrum of an X-ray tube with a molybdenum anode transmitted through a 3 mm aluminum plate.
2. Using a lithium fluoride (LiF) crystal as a monochromator, measure the calibration spectrum of radiation reflected from the crystal.
3. Measure the secondary X-ray radiation (X-ray fluorescence) spectra emitted by various materials (stainless steel, Cu, Zn, GeO_2, KBr, $SrSO_4$, silver wire, tin solder).
4. From the measurement data, calculate the photon energies and wavelengths of the characteristic X-ray K_α lines of ten chemical elements ($_{24}$Cr, $_{26}$Fe, $_{29}$Cu, $_{30}$Zn, $_{32}$Ge, $_{35}$Br, $_{38}$Sr, $_{42}$Mo, $_{47}$Ag, $_{50}$Sn) and of the K_β lines of eight chemical elements from $_{29}$Cu to $_{50}$Sn.
5. Verify whether the dependence of the characteristic X-ray wavelength on the atomic number of the anode material obeys Moseley's law.
6. From the measurement data, calculate the Rydberg constant and the screening constant.
7. Compare the experimental values of the photon energies and the Rydberg constant with the true values and discuss the causes of the discrepancies.

2. Control questions

1. Structure and operating principle of an X-ray tube.
2. Origin and spectrum of bremsstrahlung X-ray radiation.
3. Origin and spectrum of characteristic X-ray radiation.
4. Moseley's law.
5. Interaction of X-ray radiation with matter. Attenuation coefficient.
6. Operating principle of a semiconductor detector. Detector response function and efficiency. Structure of a Si-PIN detector.
7. Diffraction of X-rays in crystals. Bragg's law.
8. Measurement methodology of an X-ray radiation spectrum. Concepts of multichannel analyzer and detector calibration. Factors determining the response function of a Si-PIN detector.

Literature

1. Born M. Atomic Physics. 8th Edition. New York: Dover Publications, 1989. 544 p.
2. Eisberg R., Resnick R. Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles. New York: John Wiley & Sons, 1985. 864 p.
3. Knoll G. F. Radiation Detection and Measurement. 3rd Edition. New York: John Wiley & Sons, 2000. p. 48 - 55, 103 - 116, 353 - 386, 398.
4. Photon cross sections and attenuation coefficients. – https://kaeri.re.kr

3. X-ray radiation

3.1. Sources of X-ray radiation

In 1895, the German physicist Wilhelm Conrad Röntgen, while investigating high-voltage electrical discharges in a vacuum, noticed the glow of a barium platinocyanide screen located near a vacuum tube covered with cardboard. Later it became clear that this glow was caused by electromagnetic waves whose wavelength varies from several tens of nanometers to picometers. These waves were named **X-ray radiation** (or **Röntgen radiation**). X-ray radiation is invisible, but it exposes photographic plates, ionizes gases, and causes certain materials to glow (fluoresce). These properties are utilized in designing various detection instruments.

The role of an X-ray radiation source is most commonly fulfilled by special (X-ray) tubes. A simplified schematic of an X-ray tube is shown in Fig. 1. It is a two-electrode electrovacuum device consisting of a positive electrode – the anode (1), a negative electrode – the cathode (2), and a glass bulb (3). A large potential difference (high voltage) is created between the cathode and the anode. X-ray radiation is produced when fast electrons bombard the surface of the anode. Two types of X-ray tubes are used, differing in the method of generating the electron beam: gas-filled (ion) tubes and electron (thermionic) tubes. In **gas-filled tubes**, free electrons are produced when positive ions of a rarefied gas bombard a cold cathode. These ions are created due to a glow discharge in the rarefied gas. In **electron tubes**, free electrons are produced due to thermionic emission from the cathode (thermionic emission is the liberation of electrons from a metal by heating it). In conventional X-ray tubes, the energy of the electrons exciting the radiation is $(10^4–10^5)$ eV. Particularly short-wavelength X-ray radiation is obtained using electron accelerators. In the latter case, the energy of the electrons generating X-rays can reach ~ 10^9 eV. X-rays are also emitted by radioactive elements decaying via K- or L-electron capture. Many celestial bodies emit X-ray radiation (the Sun, stars). Plasma is also a source of X-ray radiation.

When electrons strike the atoms of the anode material, two types of X-ray radiation are excited: bremsstrahlung and characteristic radiation. **Bremsstrahlung** arises because an electron is decelerated in the anode material, i.e., it moves with acceleration. According to classical electrodynamics, an electrically charged particle moving with acceleration radiates electromagnetic waves. This is bremsstrahlung X-ray radiation. The spectrum of bremsstrahlung X-ray radiation is continuous, meaning the radiation contains all wavelengths within a certain interval (e.g., all wavelengths satisfying the condition $\lambda > 0.1$ nm). **Characteristic radiation** arises because a high-energy electron can knock out an electron from an inner electron shell of an anode material atom. A vacancy then appears in that shell, into which an electron from a higher electron shell jumps. During such a transition, a photon is emitted whose energy is equal to the difference between the energies of the two electron shells. These photons constitute characteristic X-ray radiation. The spectrum of characteristic X-ray radiation is a line spectrum, meaning the radiation consists of several precisely defined wavelengths. These wavelengths depend only on the anode material, i.e., they "characterize" it (hence the name "characteristic" radiation).

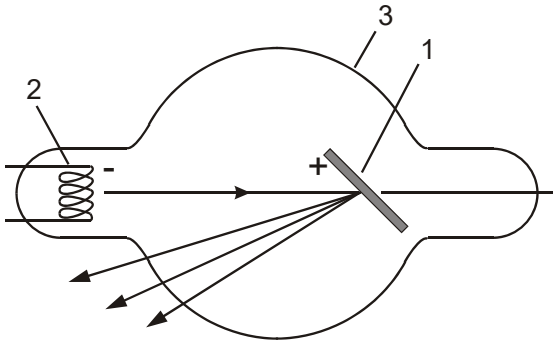


Fig. 1. Schematic of an X-ray tube: 1 – anode, 2 – cathode, 3 – bulb.

A more detailed description of bremsstrahlung and characteristic X-ray radiation is provided below.

3.2. Bremsstrahlung X-ray radiation

An electron in the anode material is decelerated due to the Coulomb interaction with the charged particles of the anode material – atomic nuclei and electrons. This interaction reduces the electron's speed, i.e., imparts acceleration to it. According to classical electrodynamics, when a charged particle moves with acceleration, it radiates electromagnetic waves whose energy flux (amount of energy per unit time) is proportional to the square of the particle's acceleration. The **radiation spectrum** is the dependence of the spectral intensity I_λ on the wavelength λ. The **spectral intensity** is the amount of energy passing through a unit area per unit time within a unit wavelength interval, i.e., $I_\lambda = \mathrm{d}I/\mathrm{d}\lambda$.

Quantum mechanics states that electromagnetic radiation (and hence X-ray radiation) can be described as a stream of elementary particles – **photons**. The energy of a photon is equal to $E = h\nu = hc/\lambda$, where ν is the radiation frequency, h is Planck's constant, and c is the speed of light. Therefore, instead of a unit wavelength interval, a unit energy interval can be used, i.e., the spectral intensity can be defined as $I_E = \mathrm{d}I/\mathrm{d}E$. These two definitions of spectral intensity are related to each other by the following relation:

$$I_E = \frac{\lambda}{E} I_\lambda . \tag{3.2.1}$$

The spectrum of bremsstrahlung radiation is continuous. This is because each electron in the anode material interacts with many atoms, and its interaction force with each atom as well as the duration of that interaction are random. Consequently, the energies of the bremsstrahlung photons emitted during each interaction event are random. The dependence of the bremsstrahlung spectral intensity I_λ on λ for two values of the accelerating voltage is shown in Fig. 2a. This function has a maximum. The decrease in spectral intensity moving away from this maximum toward long and short wavelengths is different. Toward long wavelengths, the curve asymptotically approaches zero as the wavelength increases ($\lambda \to \infty$). Toward short wavelengths, the curve drops steeply and terminates at a certain wavelength λ_{lim}, which defines the **short-wavelength limit** (or **Duane–Hunt limit**) of the spectrum. The value of the short-wavelength limit can be calculated using the law of conservation of energy, according to which the total energy of all bremsstrahlung photons is equal to the reduction in the electron's energy. The smallest possible bremsstrahlung wavelength corresponds to the largest possible energy of a single photon. According to the law of conservation of energy mentioned above, the largest possible bremsstrahlung photon energy is equal to the initial energy of the electron. A photon of such energy would be produced in the highly unlikely event when an electron is completely stopped in a single collision with a material atom. Since in an X-ray tube the kinetic energy of an electron is equal to eU (where e is the elementary charge and U is the accelerating voltage), the maximum bremsstrahlung photon energy is equal to

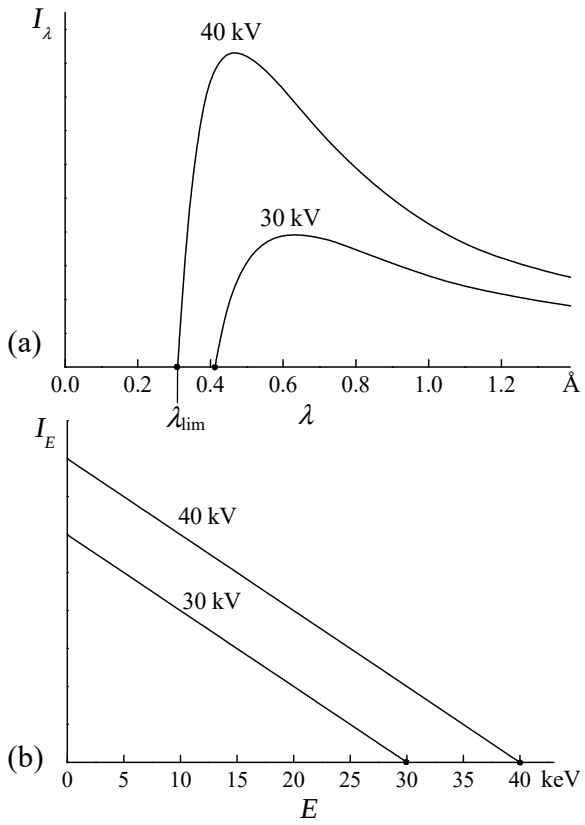


Fig. 2. Bremsstrahlung X-ray spectrum corresponding to two accelerating voltages, 30 kV and 40 kV: (a) dependence of the spectral intensity $I_\lambda = \mathrm{d}I/\mathrm{d}\lambda$ on the radiation wavelength λ, (b) dependence of the spectral intensity $I_E = \mathrm{d}I/\mathrm{d}E$ on the photon energy E.

$$E_{\max} = h\nu_{\max} = eU. \tag{3.2.2}$$

Consequently, the wavelength of the short-wavelength limit is equal to

$$\lambda_{\mathrm{lim}} = \frac{c}{\nu_{\max}} = \frac{ch}{eU}. \tag{3.2.3}$$

Substituting the values of c, e, and h into Eq. (3.2.3), we see that:

$$\lambda_{\mathrm{lim}} = \frac{12.345}{U(\mathrm{kV})}\ \text{Å}.$$

Eq. (3.2.3) can be used to determine Planck's constant h. This is one of the most accurate methods for measuring Planck's constant.

The wavelength λ_{m} corresponding to the maximum of the bremsstrahlung spectrum is related to the short-wavelength limit λ_{lim} as follows:

$$\lambda_{\mathrm{m}} = \frac{3}{2}\lambda_{\mathrm{lim}} \tag{3.2.4}$$

(see, e.g., Fig. 2a). This result follows from **Kramers' law** (named after the Dutch physicist H. A. Kramers, who derived it in 1923), which states that the energy spectral intensity I_E of bremsstrahlung is a linearly decreasing function of the photon energy:

$$I_E \sim E_0 - E, \tag{3.2.5}$$

where E_0 is the initial electron energy (which in an X-ray tube is equal to eU). Two dependences of I_E on energy for accelerating voltages of 30 kV and 40 kV are shown in Fig. 2b. However, Kramers' law (3.2.5) corresponds to an idealized case when certain simplifying assumptions are satisfied, e.g., that the decrease in electron energy is proportional to the penetration depth of the electron into the anode and that the probability of bremsstrahlung per unit path length is constant. In reality, these assumptions do not hold or are satisfied only approximately. Moreover, low-energy photons are strongly absorbed in the walls of the X-ray chamber, so the relative fraction of low-energy photons in the X-ray radiation emerging into the surroundings is reduced compared to the ideal case corresponding to Fig. 2b. In practice, one must also account for additional losses of low-energy photons due to detector imperfection: detectors have a certain "dead" layer (e.g., a protective window, housing, electrode) that photons must traverse before reaching the active medium of the detector; furthermore, the detector electronics have a certain "sensitivity threshold" corresponding to the minimum detectable photon energy. These low-energy photon losses can be approximately accounted for by using I_E of the following form:

$$I_E \sim E(E_0 - E). \tag{3.2.6}$$

The corresponding dependences of I_λ and I_E on λ and E are shown in Fig. 3. Using the approximation (3.2.6), the photon energy corresponding to the maximum of the energy spectrum is equal to half of the initial photon energy E_0:

$$E_\mathrm{m} = \frac{E_0}{2} \tag{3.2.7}$$

(see Fig. 3b), while the wavelength corresponding to the maximum of the wavelength spectrum is equal to

$$\lambda_\mathrm{m} = \frac{4}{3}\lambda_\mathrm{lim} \tag{3.2.8}$$

(see Fig. 3a), i.e., due to additional absorption of low-energy photons, the wavelength corresponding to the maximum of the wavelength spectrum shifts closer to the minimum possible wavelength λ_lim (cf. Eq. (3.2.4)).

It has been experimentally established that the intensity of bremsstrahlung increases with increasing nuclear charge of the anode atoms. It was also established that under identical conditions, the bremsstrahlung intensity of heavy charged particles (e.g., protons) is lower than that of light particles – electrons. Let us explain why this is the case.

When a charged particle (e.g., an electron) moves in the electric field of an atomic nucleus of charge Ze, it is deflected by a force

$$m\mathbf{a} = -e\mathscr{E}, \tag{3.2.9}$$

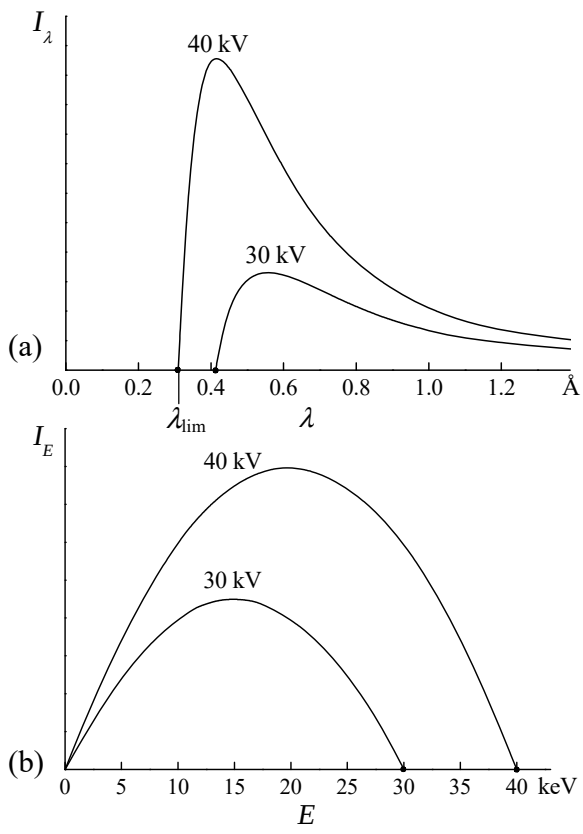


Fig. 3. Bremsstrahlung X-ray spectrum accounting for low-energy photon losses at two accelerating voltages, 30 kV and 40 kV: (a) dependence of the spectral intensity $I_\lambda = \mathrm{d}I/\mathrm{d}\lambda$ on the radiation wavelength λ, (b) dependence of the spectral intensity $I_E = \mathrm{d}I/\mathrm{d}E$ on the photon energy E.

where m and e are the mass and charge of the particle, $\mathbf{a}$ is its acceleration vector, and $\mathscr{E}$ is the electric field strength vector of the nucleus (since the letter "E" is customarily used to denote energy, the field strength is denoted here using a different typeface):

$$\mathscr{E} = \frac{1}{4\pi\varepsilon_0} \cdot \frac{Ze}{r^3}\mathbf{r}, \tag{3.2.10}$$

where **r** is the radius vector of the particle in a coordinate system centered at the nucleus, r is the distance to the nucleus: $r = |\mathbf{r}|$, and $\varepsilon_0 = 8.854 \cdot 10^{-12}$ F/m is the electric constant. Substituting Eq. (3.2.10) into Eq. (3.2.9), we obtain the expression for the acceleration:

$$\mathbf{a} = -\frac{1}{4\pi\varepsilon_0} \cdot \frac{Ze^2}{mr^3}\mathbf{r}. \tag{3.2.11}$$

According to classical electrodynamics, it is proven that a particle moving with acceleration a radiates during time dt an amount of energy

$$\mathrm{d}E = \frac{1}{6\pi\varepsilon_0} \cdot \frac{e^2}{c^3}a^2\mathrm{d}t. \tag{3.2.12}$$

Thus, a particle moving in the electric field of a nucleus radiates during time dt an amount of energy

$$\mathrm{d}E = \frac{1}{96\pi^3\varepsilon_0^3} \cdot \frac{e^6Z^2}{c^3m^2r^4}\mathrm{d}t. \tag{3.2.13}$$

These energy losses decrease rapidly as the mass of the decelerated particle increases ($\sim 1/m^2$). Therefore, the energy losses of light particles (electrons) are much greater than those of heavy particles (protons). Furthermore, it follows from Eq. (3.2.13) that in a target composed of heavy elements, the energy losses of charged particles are greater than in a target composed of light elements ($\sim Z^2$). If a continuous X-ray spectrum is required for experiments, it is advantageous to use heavy elements to excite this radiation.

The flux of bremsstrahlung radiation ($\mathrm{d}E/\mathrm{d}t$) emerging from the anode into the surroundings is much smaller than that following from Eq. (3.2.13), because the majority of the bremsstrahlung energy is absorbed in the anode material and converted into heat.

3.3. Characteristic X-ray radiation

When the energy of the electrons striking the anode exceeds a certain threshold, a line spectrum consisting of separate sharp lines appears on top of the continuous spectrum background (Fig. 4). This radiation is called **characteristic X-ray radiation**, because it "characterizes" the anode material: the wavelength corresponding to each line depends only on the nature of the material and does not depend on the accelerating voltage. As the accelerating voltage increases, the intensity of the lines increases; moreover, new lines may appear in the short-wavelength region of the spectrum. However, the positions of already existing lines in the spectrum do not change.

The schematic diagram of the origin of characteristic X-ray radiation is shown in Fig. 5. An electron incident on the anode (**primary electron**) 1, if its energy is sufficiently high, can transfer its energy to one of the inner electrons 2 of an anode material atom and knock it out of the atom. As a result, the atom is ionized: a vacancy appears in one of the inner electron shells (e.g., in the K shell, as shown in Fig. 5). One of the electrons 3 from outer shells (e.g., an L-shell electron, as shown in Fig. 5) transitions into this vacant state. When electron 3 transitions from a state of energy E_2 to a state of lower energy E_1, the energy difference is radiated in the form of a photon $h\nu = E_2 - E_1$. Since this energy difference is strictly defined, a spectral line of radiation with a precisely defined frequency ν is emitted.

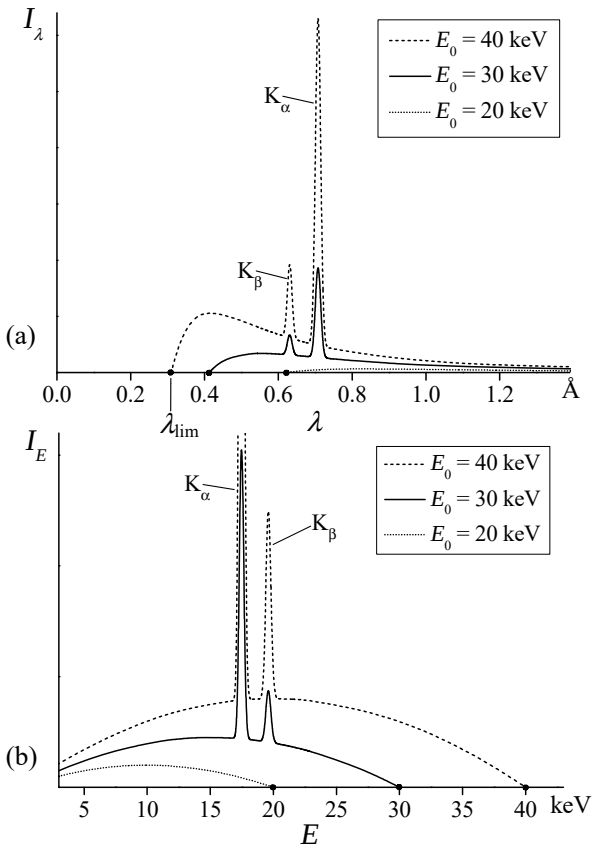


Fig. 4. Total X-ray spectrum (bremsstrahlung + characteristic) at three accelerating voltages, 20 kV, 30 kV, and 40 kV (peak positions correspond to the K_α and K_β lines of molybdenum): (a) dependence of the spectral intensity $I_\lambda = \mathrm{d}I/\mathrm{d}\lambda$ on the radiation wavelength λ, (b) dependence of $I_E = \mathrm{d}I/\mathrm{d}E$ on the photon energy E.

Thus, the conversion of the primary electron's energy into a quantum of characteristic X-ray radiation takes place in two stages:

1) first, the primary electron ionizes the atom, i.e., part of the primary electron's energy is spent on breaking the electron's bond in the atom, while the remaining kinetic energy is distributed between the liberated free electron (**secondary electron**) and the primary electron;

2) since the resulting positive ion is in an unstable excited state, the atomic electrons rearrange among the states while emitting a photon, i.e., the excitation energy is converted into photon energy (a quantum of characteristic X-ray radiation).

In general, characteristic X-ray radiation arises whenever vacancies are created in the inner electron shells of an atom. The method of creating these vacancies described above (using fast electrons) is the most widely used in characteristic X-ray generators. However, the vacancies can be created not only by electrons, but also by other charged particles (e.g., alpha particles or protons). Characteristic X-ray emission caused by bombarding a target with charged particles is called **particle-induced X-ray emission** (abbreviation – PIXE). Furthermore, vacancies can be created by photons (the corresponding characteristic X-ray emission is called **X-ray fluorescence**). Another process during which characteristic X-ray radiation can arise is a process in which a positive ion captures an electron (atomic electron capture). If this ion has a large deficit of electrons, and the captured electron lands in one of the outer electron shells, then characteristic X-ray radiation also occurs. Vacancies also arise due to certain nuclear processes – internal conversion and electron capture by an atomic nucleus.

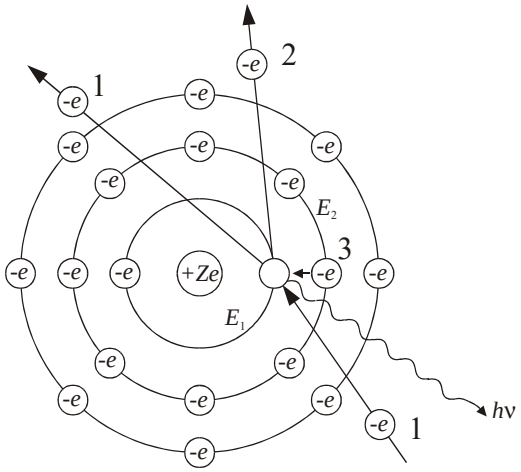


Fig. 5. Schematic of the origin of characteristic X-ray radiation.

The characteristic X-ray spectrum of an element does not depend on whether the element is free or part of a chemical compound. For example, the characteristic X-ray spectra of an iodine atom and an iodine molecule are identical. In this respect, characteristic X-ray spectra differ from optical spectra.

The arrangement of characteristic X-ray lines in spectra is regular: the lines form series located in different regions of the spectrum. The appearance of the characteristic X-ray spectrum is explained as follows. If a primary electron knocks out one of the two deepest inner K-shell electrons, then the resulting vacancy can be filled by an electron from any higher shell L, M, N. The corresponding series of lines is denoted by the letter K – the same letter as the electron shell containing the vacancy. The K-series edge $h\nu_{\text{edge}}$ corresponds to the transition of a free electron located outside the atom ($E = 0$) into the K shell (see Fig. 6). The transition of free electrons, which outside the atom can have any kinetic energy ($E > 0$), into a vacant spot in the K shell is accompanied by a continuous spectrum (adjacent to the K-series edge) of photons with energies $h\nu > h\nu_{\text{edge}}$ (Fig. 6). The K series consists of the shortest-wavelength lines. The series of lines resulting from electron transitions into the L shell is denoted by the letter L, and so forth.

The lines of the characteristic X-ray spectrum are labeled so that the initial and final electron shells are clear. Currently, the two most widespread line notation systems are the so-called Siegbahn notation and the newer system recommended in 1991 by the International Union of Pure and Applied Chemistry (IUPAC). In Siegbahn notation, a line is denoted by the letter K, L, M, etc., indicating the final shell (which contained the vacancy), and a subscript α, β, etc., indicating the initial shell (from which the electron jumped), e.g., "L_α". In the IUPAC system, the final and initial electron shells are indicated by the letters K, L, M, etc., separated by a hyphen. For example, "K–L" denotes a transition from the L shell to the K shell (the corresponding Siegbahn notation is "K_α"). Since the IUPAC notation system is clearer, we will use it henceforth.

As the atomic number of the anode element increases, the number of excited series in the spectrum increases. The X-ray spectra of elements are simple, consisting of a relatively small number of lines. By comparison, optical spectra are usually complex, containing hundreds or thousands of lines. Moreover, the characteristic X-ray spectra of different elements are very similar. The only change observed in characteristic X-ray spectra when moving from light elements to heavier ones is a gradual shift of the lines toward shorter wavelengths. Fig. 7 shows the K series of several elements. The K series consists of three lines: K–L,

K–M, and K–N. The K–L line in this series has the longest wavelength and exhibits a doublet structure, i.e., it consists of two closely spaced lines. The K–M and K–N lines are also doublets, but resolving their fine structure is difficult. The other spectral series of characteristic X-ray radiation of elements (L, M, N) have a more complex structure and likewise consist of a small number of lines.

In the K series, the most intense line is the K–L line, the second most intense is K–M, and the third is K–N, i.e., the intensity of the lines in a series decreases from the first line toward the series edge. The intensity of a line is proportional to the probability per unit time of the corresponding transition. When a vacancy is filled via a quantum transition from a given higher electron shell, the probability is greater the smaller the energy difference between the initial and final shells[1].

When an electron jumps from a higher electron shell into a vacancy in a deeper shell, the released energy may not only be radiated in the form of a photon, but may also be transferred directly to one of the outer (valence) electrons, which is consequently ejected from the atom (thus, the atom loses one more electron). This phenomenon is called the **Auger effect** (named after the French physicist Pierre Auger, who discovered it), and those electrons are called **Auger electrons**. The energy of an Auger electron is equal to the difference between the energy released during the initial transition and the binding energy of the ejected electron. The ratio of the probabilities of the Auger effect and characteristic X-ray photon emission increases with decreasing atomic number Z, because the electron binding energy then decreases. When $Z = 30$ (zinc), the probability of the Auger effect is approximately equal to the probability of characteristic X-ray emission. In materials with higher Z, the probability of characteristic X-ray emission is greater than the probability of the Auger effect. For example, for iodine (I) (where $Z = 53$), the probability of characteristic X-ray emission is 7.3 times greater than the probability of the Auger effect.

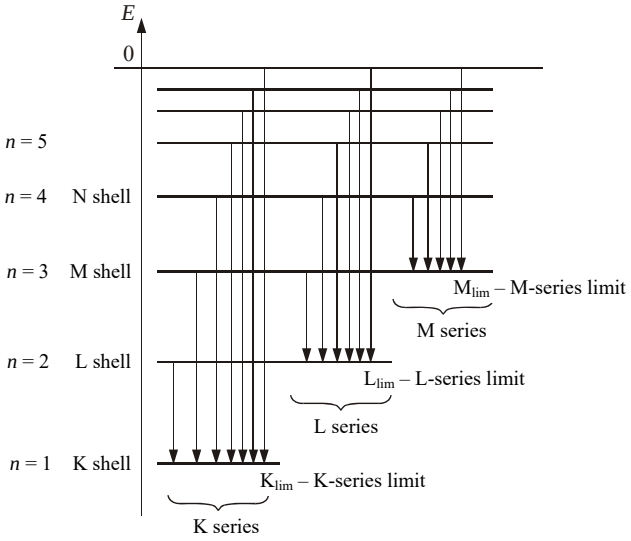


Fig. 6. Schematic of the origin of characteristic X-ray series.

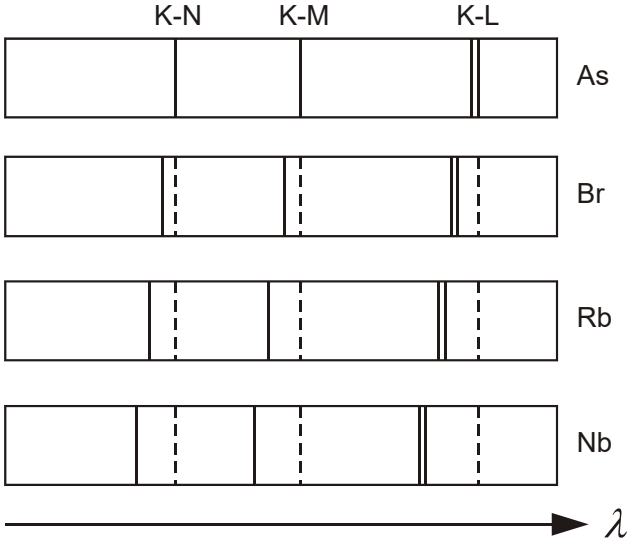


Fig. 7. Characteristic X-ray K series of As, Br, Rb, and Nb.

3.4. Moseley's law

Using the methods of quantum mechanics or the Bohr model of the atom, it is proven that the total mechanical energy of a stationary hydrogen-like atom (i.e., a hydrogen atom or an ion having only one electron) can be equal only to the following values:

$$E_n = -\frac{Z^2}{n^2}\frac{me^4}{8\varepsilon_0^2 h^2} \approx -\frac{Z^2}{n^2}\cdot 13.6\ \text{eV} \quad (n = 1, 2, \dots), \tag{3.4.1}$$

[1] This is because the quantum transition probability per unit time is proportional to the squared modulus of the matrix element of the electron's radius vector corresponding to the given initial and final electron states. This matrix element depends on the spatial overlap of the initial and final wavefunctions: the smaller the spatial region in which both functions differ appreciably from zero, the smaller the matrix element of the radius vector. Since as the shell number increases, the maximum of the electron's radial probability density moves farther away from the nucleus, as the difference in shell numbers corresponding to the initial and final states increases, the spatial overlap of the wavefunctions decreases, and therefore the matrix element of the radius vector and the quantum transition probability also decrease.

where Ze is the nuclear charge, $\varepsilon_0 = 8.854 \cdot 10^{-12}$ F/m is the electric constant, and $m = 9.11 \cdot 10^{-31}$ kg is the electron mass (more precisely, the reduced mass of the electron–nucleus system $m_e M/(m_e + M)$, where m_e is the electron mass and M is the nuclear mass; since $M \gg m_e$, the reduced mass m is slightly smaller than the electron mass m_e). The integer n specifies the electron shell number: $n = 1$ corresponds to the K shell, $n = 2$ corresponds to the L shell, and so on. Consequently, the energy of a photon emitted when an L electron transitions into a vacancy in the K shell is equal to

$$h\nu_{\text{K-L}} = E_{\text{L}} - E_{\text{K}} = \frac{me^4}{8\varepsilon_0^2 h^2} Z^2 \left(\frac{1}{1^2} - \frac{1}{2^2} \right). \tag{3.4.2}$$

However, this formula is strictly applicable only to hydrogen-like atoms. If an atom has more than one electron, then when calculating the transition energy, the mutual interaction of the electrons must be taken into account. An electron transitioning into a vacancy in the K shell is acted upon by the field of the nuclear charge $+Ze$, weakened by the field of the second electron located in the K shell. Therefore, the effective nuclear charge $+Z_1 e$ that determines the electric field in which an electron transitions from the L (M, N) shell into the K shell is smaller than the nuclear charge $+Ze$ by an amount approximately equal to the elementary charge e. That is,

$$Z_1 = Z - \sigma,$$

where σ is the **screening constant**. The screening constant depends on which shell the electron transitions into. If an electron transitions into the K shell, then $\sigma = \sigma_{\text{K}} \approx 1$, and if into the L shell, then $\sigma = \sigma_{\text{L}} \approx 7.4$. The screening constant depends weakly on Z. The stated values of σ are approximate because they were derived on the basis of the simplified Bohr model, according to which electrons orbit the nucleus in trajectories of definite radius. In reality, this is not the case: electrons are statistically distributed in space around the nucleus and form an "electron cloud." Therefore, the deviations of the true screening constants from the stated values can reach several tens of percent.

Taking into account the screening effect mentioned above, the atomic number Z in Eq. (3.4.2) must be replaced by the effective nuclear charge number $Z - \sigma_{\text{K}}$:

$$h\nu_{\text{K-L}} = \frac{me^4}{8\varepsilon_0^2 h^2} (Z - \sigma_{\text{K}})^2 \left(\frac{1}{1^2} - \frac{1}{2^2} \right). \tag{3.4.3}$$

Hence, the wavenumber[2] of the K-L line is

$$k_{\text{K-L}} \equiv \frac{1}{\lambda_{\text{K-L}}} = \frac{\nu}{c} = \frac{me^4}{8\varepsilon_0^2 h^3 c} (Z - \sigma_{\text{K}})^2 \left(\frac{1}{1^2} - \frac{1}{2^2} \right) = R(Z - \sigma_{\text{K}})^2 \left(\frac{1}{1^2} - \frac{1}{2^2} \right), \tag{3.4.4}$$

where R is the **Rydberg constant**, which is defined as follows[3]:

$$R = \frac{me^4}{8\varepsilon_0^2 h^3 c} \approx 1.097 \cdot 10^7 \text{ m}^{-1}. \tag{3.4.5}$$

Expressions for the wavenumbers of other spectral lines of characteristic X-ray radiation can be obtained in the same way. For example, the wavenumber of the K-M line, which is radiated when an M electron ($n = 3$) transitions into a vacancy in the K shell ($n = 1$), is

$$k_{\text{K-M}} = \frac{1}{\lambda_{\text{K-M}}} = R(Z - \sigma_{\text{K}})^2 \left(\frac{1}{1^2} - \frac{1}{3^2} \right); \tag{3.4.6}$$

the wavenumber of the L-M line, corresponding to the transition of an M electron ($n = 3$) into a vacancy in the L shell ($n = 2$), is

$$k_{\text{L-M}} = \frac{1}{\lambda_{\text{L-M}}} = R(Z - \sigma_{\text{L}})^2 \left(\frac{1}{2^2} - \frac{1}{3^2} \right); \tag{3.4.7}$$

for the M-N line, instead of σ_{L}, $1/2^2$, and $1/3^2$, one has σ_{M}, $1/3^2$, and $1/4^2$, respectively, and so on.

---

[2] The wavenumber is more commonly defined with a factor of 2π, i.e., $k = 2\pi/\lambda$. The definition of wavenumber given in this section ($k = 1/\lambda$) is widespread in spectroscopy.

[3] Since the reduced mass m depends on the nuclear mass, the exact value of the Rydberg constant is slightly different for different atoms. The value of the Rydberg constant obtained by replacing the reduced mass m with the exact electron mass is denoted by R_∞, and its value is $R_\infty = 1.09737315685 \cdot 10^7$ m^{-1} (this value corresponds to an atom whose nuclear mass is infinitely large).

From Eqs. (3.4.4), (3.4.6), and (3.4.7), it is evident that the square root of the wavenumber is directly proportional to the atomic number Z. This relationship between the wavenumber k of a characteristic X-ray spectral line and the atomic number was discovered in 1913 by the English physicist Henry Moseley. Therefore, this law is called **Moseley's law**. Moseley's law can be written as follows:

$$Z - \sigma = C\frac{1}{\sqrt{\lambda}} = C\sqrt{k}, \tag{3.4.8}$$

where for the K-L line $C = 1/\sqrt{R[(1/1^2) - (1/2^2)]}$, for the L-M line $C = 1/\sqrt{R[(1/2^2) - (1/3^2)]}$, and so on. Fig. 8 shows the empirical dependence of $1/\sqrt{\lambda}$ on the atomic number Z for the K-L, L-M, and M-N lines. We see that this dependence is linear for elements at the beginning of the periodic table (i.e., when $Z < 70$). It has been established that as the atomic number Z increases, the deviation from this straight line is slight and becomes noticeable only for the M and N series.

Moseley's law made it possible to clarify the physical meaning of an element's atomic number: it is the number of elementary charges in the atomic nucleus. Moseley's law showed that the chemical properties of an element are determined not by the atomic mass, but by the electric charge of the atomic nucleus.

Based on Moseley's law, the atomic number of an element can be determined. To do this, one needs to measure the wavelength of the characteristic radiation and then apply Eq. (3.4.8).

Using the methods of quantum mechanics, the energies of characteristic radiation lines can be calculated much more accurately than according to Moseley's law. Therefore, Moseley's law is no longer used in scientific research. The main phenomena causing the true wavelengths to deviate from the linear dependence (3.4.8) are the following:

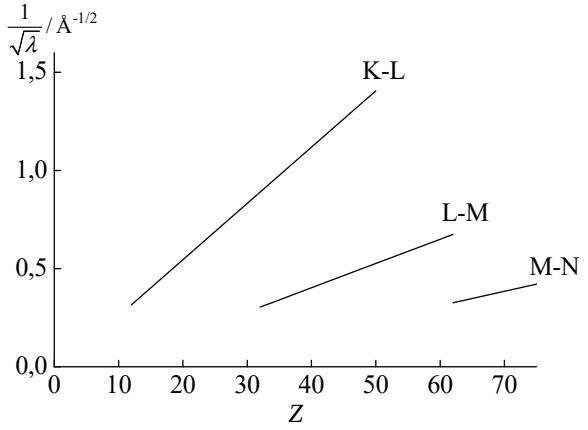


Fig. 8. Moseley diagrams.

1. Since the spatial charge distribution of each electron is continuous, each electron is screened by *all other* electrons of the same atom. Although the strongest screening effect is created by the *s* electrons of the K shell, L-shell electrons, for example, are also screened by electrons of the same L shell, because part of the spatial distribution of every electron (even one located in an outer shell) lies very close to the nucleus, thereby reducing the effective nuclear charge acting on other electrons. For this reason, the screening constant σ is actually not constant and increases with increasing atomic number Z. In addition, since the K_β line arises from electron transitions from the third (M) shell to the first (K) shell, during this transition the electron traverses a larger "cloud" of intermediate electrons than during a transition from the second (L) shell to the first. Therefore, K_β lines correspond to a larger screening constant than K_α lines. Moseley's law ignores this phenomenon (according to Moseley's law, the screening constants of all K-series lines should be equal to 1). For example, when the atomic number Z belongs to the interval $24 \le Z \le 50$, the screening constant σ calculated according to Eq. (3.4.8) for K_α lines varies from roughly 1.2 to 1.7, while the screening constant for K_β lines varies from roughly 1.6 to 2.7. This increase in the screening constant with increasing Z manifests itself in the fact that the inverse square root of the wavelength ($\lambda^{-1/2}$), as a function of the atomic number Z, increases faster than linearly, and the photon energies become larger than those predicted by Moseley's law.

2. As Z increases, the average speed of inner-shell electrons increases. For example, the average speed of *s* electrons in a zinc (Zn, $Z = 30$) atom is 22 % of the speed of light, while the average speed of *s* electrons in a tin (Sn, $Z = 50$) atom is 36.5 % of the speed of light. According to the theory of relativity, as the speed of a particle increases, its mass also increases. Although the speeds mentioned above are several times smaller than the speed of light, the relativistic mass increase is relatively large and manifests itself in noticeable shifts of energy levels. Although Eq. (3.4.1) for the energy levels of a hydrogen-like atom is non-relativistic, on its basis one can understand the influence of the relativistic increase in electron mass m: because of it, the absolute value of the energy increases, i.e., the energy levels become "deeper." Since this shift of inner-shell energy levels is larger than that of outer-shell

ones, the energies of quantum transitions (level differences) increase. Consequently, the effect acts in the same direction as the additional screening effect mentioned earlier: as Z increases, the inverse square root of the wavelength increases faster than linearly.

The Rydberg and screening constants (R and σ, respectively) can be determined by fitting the experimental $\lambda^{-1/2}$ values with a linear function of the atomic number Z (3.4.8). However, since the true dependence is not strictly linear, the values of R and σ determined in this way are inaccurate and the deviation is of a non-random (systematic) nature. As explained earlier, $\lambda^{-1/2}$ increases faster than a linearly increasing function of the atomic number Z. Therefore, at the beginning of the fitting interval, the straight line lies below the true $\lambda^{-1/2}$ values, in the middle of the interval it lies above the true values, and at the end of the interval it again lies below the true values. For the same reason, the values of R and σ calculated by the fitting method depend on the chosen interval of Z values over which the fitting is performed. For example, if the fitting interval is $24 \leq Z \leq 50$, then the value of R obtained in this way is 3–5 % larger compared to the exact Rydberg constant. The value of σ obtained by the fitting method is also inaccurate, because the linear fitting method yields only a single value of σ, whereas in reality different chemical elements correspond to different screening constants (as mentioned, σ increases with increasing atomic number and with the number of the shell from which the electron transition occurs).

3.5. Fine structure of the characteristic X-ray spectrum

Using high-resolution instruments, the fine structure of the characteristic X-ray spectrum was observed: each line consists of two or more closely spaced lines. It was found that the lines of the K series are doublets (i.e., each line of the K series consists of two closely spaced lines). The lines of other series have a more complex structure (multiplets).

The complex structure of characteristic X-ray spectral lines arises due to the multiplicity of atomic energy levels, i.e., the splitting into several closely spaced levels (the cause of this splitting is the spin–orbit interaction). Therefore, the energy levels of a multi-electron atom are classified according to the orbital angular momentum quantum number L, the spin quantum number S, and the total angular momentum quantum number J. The calculation of the angular momenta of an atom's X-ray radiation states is particularly simple because the initial electron shell is completely filled, and the final shell lacks only one electron to be completely filled (see Fig. 5). The spin and orbital angular momenta of a filled subshell (and thus of a filled shell) are equal to zero: the angular momenta of the electrons compensate one another. This is because in a completely filled shell all electrons are "paired": for every electron one can find another electron of the same subshell whose spin and orbital angular momenta exactly compensate the corresponding angular momenta of the first electron. In a shell lacking one electron, there is only one unpaired electron. Therefore, the spin and orbital angular momenta are equal to the corresponding angular momenta of this electron: $S = s = 1/2$, $L = l, J = j$.

In the K shell there are two s electrons, whose principal quantum number is $n = 1$ and azimuthal (orbital) quantum number is $l = 0$. If one s electron is removed, the energy state of the atom corresponds to a vacancy in the K shell, i.e., it is specified by the quantum numbers $n = 1, l = 0, j = 1/2$. The symbol for this level is $1^2S_{1/2}$ (the superscript 2 is the "spin multiplicity," which in this case is equal to $2S + 1 = 2$).

In the L shell there are eight electrons, whose principal quantum number is $n = 2$ and orbital quantum number l is equal to 0 or 1. Of the eight electrons in the L shell, two are in the s subshell, and the remaining six are in the p subshell. If one of these two s electrons is removed, the state of the atom is determined by the vacancy in the s subshell, for which $n = 2, l = 0, j = 1/2$. The symbol for the corresponding energy level is $2^2S_{1/2}$. If one of the six p-subshell electrons is removed in the L shell, the state of the atom is specified by $n = 2, l = 1, j = 1/2$ (symbol $2^2P_{1/2}$) or $n = 2, l = 1, j = 3/2$ (symbol $2^2P_{3/2}$). Therefore, a singly ionized L shell has three energy levels: $2^2S_{1/2}$, $2^2P_{1/2}$, $2^2P_{3/2}$.

In the same way, it can be proven that a singly ionized M shell has five energy levels: $3^2S_{1/2}$, $3^2P_{1/2}$, $3^2P_{3/2}$, $3^2D_{3/2}$, $3^2D_{5/2}$. An ionized N shell has seven energy levels, and so on.

The energy levels of singly ionized K, L, and M shells of an atom are shown in Fig. 9. The arrows indicate the transitions corresponding to the K-L, K-M, and L-M lines. If a transition could occur from any sublevel of the initial shell to any sublevel of the final shell, then the multiplicity of a line would be equal to the product of the multiplicities of both levels. For example, the multiplicity of the K-L line would be 3, K-M would be 5, L-M would be 15, and so on. However, in reality the multiplicity is smaller. This is

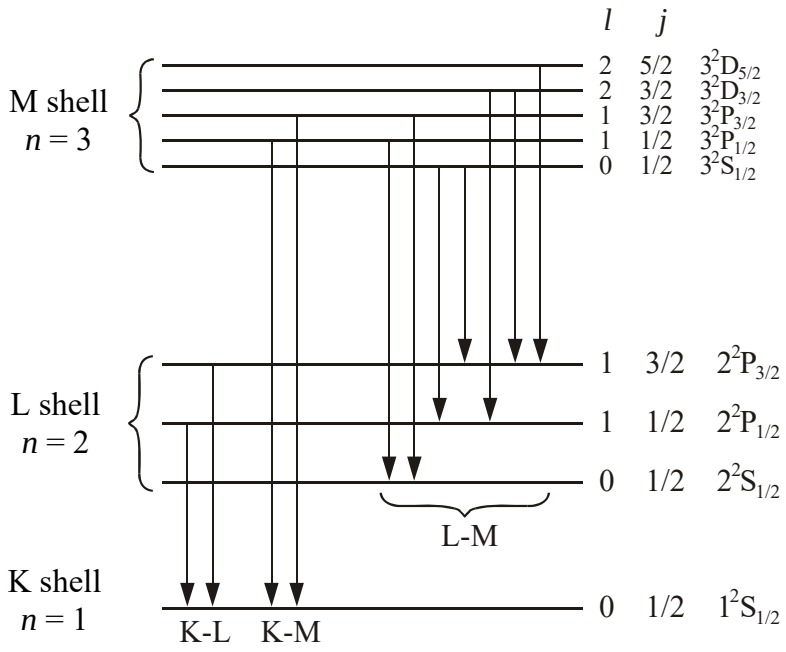


Fig. 9. Fine structure of the characteristic X-ray spectrum.

Table 1. Siegbahn and IUPAC designations of some characteristic X-ray spectral lines

Siegbahn	IUPAC	Siegbahn	IUPAC	Siegbahn	IUPAC	Siegbahn	IUPAC	Siegbahn	IUPAC
$K_{\alpha 1}$	K-L_3	$K^{I}_{\beta 2}$	K-N_3	$L_{\alpha 1}$	L_3-M_5	$L_{\beta 2}$	L_3-N_5	$L_{\beta 5}$	L_3-$O_{4,5}$
$K_{\alpha 2}$	K-L_2	$K^{II}_{\beta 2}$	K-N_2	$L_{\alpha 2}$	L_3-M_4	$L_{\beta 3}$	L_1-M_3	L_{η}	L_2-M_1
$K_{\beta 1}$	K-M_3	$K_{\beta 3}$	K-M_2	$L_{\beta 1}$	L_2-M_4	$L_{\beta 4}$	L_1-M_2	L_l	L_3-M_1

due to selection rules, according to which the changes in the quantum numbers L and J during a quantum transition of an atom can only be equal to $\Delta L = 0, \pm 1$, $\Delta J = 0, \pm 1$. Transitions between the K, L, and M shells satisfying these rules (excluding transitions with $\Delta L = 0$) are shown in Fig. 9. We see that the lines of the K series are doublets, while the structure of the lines of the L series is more complex.

When using the IUPAC notation system, the individual lines of each multiplet are indicated by inserting the number of each shell's sublevel after the letter denoting that shell: "L", "M", "N" (no number is inserted after the letter "K", because the K level is not split). The sublevels are numbered in the same order in which they are arranged in Fig. 9, from bottom to top. For example, the designations of the K-L line are K-L_2 and K-L_3 (in Siegbahn notation, these two lines are denoted by $K_{\alpha 2}$ and $K_{\alpha 1}$, respectively). The Siegbahn designations of some lines are given in Table 1.

In practice, it is easiest to resolve the fine structure of K-L lines. Other lines of the K series may appear to be singlets (single lines), because the spacing between the energy sublevels of the M, N, and higher shells is smaller than the spacing between the energy sublevels of the L shell (see Fig. 9).

The intensity of the K-L_3 (i.e., $K_{\alpha 1}$) component of the K-L line is approximately twice as large as the intensity of the K-L_2 (i.e., $K_{\alpha 2}$) component, because the $P_{3/2}$ level has twice as many quantum states from which quantum transitions can occur as the $P_{1/2}$ level: the $2P_{3/2}$ level has four quantum states, whereas the $2P_{1/2}$ level has two. Therefore, the average photon energy of the K_{α} line is equal to

$$E_{K\alpha} \approx \frac{2E_{K\alpha 1} + E_{K\alpha 2}}{3}, \tag{3.5.1}$$

here $E_{K\alpha 1}$ and $E_{K\alpha 2}$ are the energies of the $K_{\alpha 1}$ and $K_{\alpha 2}$ lines, respectively. The "number of quantum states" mentioned above is the maximum possible number of electrons in the energy level. In a completely filled $P_{3/2}$ level there are four electrons, and in a completely filled $P_{1/2}$ level there are two (a total of six p electrons).

As the energy of the primary electrons increases, the series of characteristic spectral lines appear not all at once, but sequentially. First the long-wavelength series appear, while the shortest-wavelength (highest-energy) K series appears last. A primary electron with energy eU can liberate an electron from an atom only when its kinetic energy is greater than the electron binding energy, which is equal to the difference between

the minimum energy of a free electron and the energy of the electron in the atom. Since the minimum kinetic energy of a free electron is equal to zero, the ionization condition can be written as follows:

$$eU > -E(nlj),$$

where $E(nlj)$ is the electron energy level specified by the triplet of quantum numbers n, l, and j (here $E < 0$).

4. Interaction of X-ray radiation with matter

4.1. Concept of interaction cross section

Any possible consequence of an interaction ("collision") between two particles will be called an "event". The probability of a specific event (e.g., photon absorption due to interaction with an atom or elastic photon scattering due to interaction with an electron) can be expressed using the concept of a cross section: each target particle (e.g., an electron, atom, or nucleus) is replaced by an imaginary flat region perpendicular to the direction of motion of the incident particles, whose area is chosen such that the probability of the given event (e.g., absorption or scattering) coincides with the probability that the incident particle hits this region. The area σ defined in this way is called the **cross section** of that event. Thus, one speaks of "absorption cross section", "scattering cross section", etc. Knowing the cross section, the probability of a given event can be calculated according to the rules of geometric probability in the manner described below.

If the concentration of target particles is n, then in a material layer of area S and infinitesimally small thickness $\mathrm{d}x$, there are $n \cdot S \cdot \mathrm{d}x$ target particles. The total "area" $\mathrm{d}S'$ of these particles, which covers a portion of the area S, is equal to the sum of the cross section areas, i.e., $\mathrm{d}S' = \sigma \cdot n \cdot S \cdot \mathrm{d}x$ (see Fig. 10). Since the momentum of the incident particle is described by a plane de Broglie wave whose squared modulus is uniform at all points on the target surface, there is an equal probability that the particle will hit any point on the surface of area S. Therefore, the probability $\mathrm{d}P$ that an incident particle "hits" any target particle located in the layer of thickness $\mathrm{d}x$ is equal to the ratio of the areas:

$$\mathrm{d}P = \frac{\mathrm{d}S'}{S} = \sigma n \mathrm{d}x. \qquad (4.1.1)$$

This relation can also be regarded as the definition of the interaction cross section σ: the interaction cross section σ is equal to the ratio of the interaction probability $\mathrm{d}P$ to the number of particles per unit area $n \cdot \mathrm{d}x$.

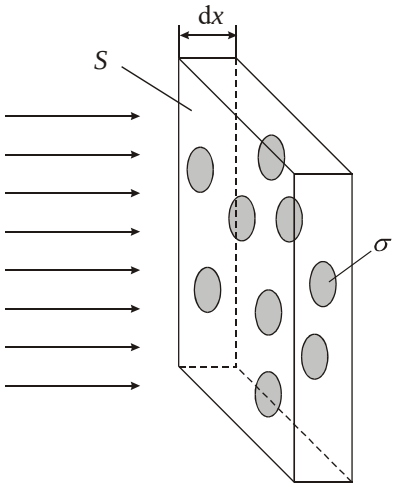


Fig. 10. Explanation of the interaction cross section σ.

Each event is characterized by its own cross section. According to the addition rule for probabilities of mutually exclusive events, the sum of the cross sections of all possible events defines the total particle interaction (collision) cross section:

$$\sigma = \sum_i \sigma_i, \qquad (4.1.2)$$

where σ_i is the collision cross section of the i-th type.

The total (or specific interaction) cross section when several types of targets exist is calculated analogously. Then the interaction cross section is equal to

$$\sigma = \sum_i p_i \sigma_i, \qquad (4.1.3)$$

where p_i is the relative fraction of target particles of the i-th type (i.e., the ratio of the concentration of target particles of that type to the total concentration of target particles), and σ_i is the cross section of the given interaction (or the total cross section of several interactions) when interacting exclusively with target particles of the i-th type.

In atomic and nuclear physics, interaction cross sections are usually expressed in barns (b). $1\ \mathrm{b} = 10^{-28}\ \mathrm{m}^2$.

The product σn of the cross section σ of a given event and the target particle concentration n is called the **macroscopic cross section** (Σ) of that event. Thus, it can be stated that the macroscopic cross section

is the interaction cross section corresponding to a unit volume of the target. According to Eq. (4.1.1), the macroscopic cross section defines the probability of an interaction event of a given type per unit path length of an incident particle:

$$\Sigma = \frac{\mathrm{d}P}{\mathrm{d}x} = \sigma n. \tag{4.1.8}$$

In certain cases, the macroscopic cross section is equal to the component of the radiation attenuation coefficient associated with the given type of interaction (such a case will be discussed in Section 4.5 when speaking about the attenuation of X-ray radiation in matter).

In matter, X-ray radiation is absorbed and scattered. **Photon absorption** is an interaction of a photon with matter during which the photon disappears, and its entire energy is transferred to the atoms of the material or converted into the energy of secondary particles. The primary absorption mechanism of X-ray radiation is the photoelectric effect (it will be described in Section 4.2). During **scattering**, the photon changes its direction of motion and either loses no energy (coherent scattering) or transfers a part of its energy to an atom of the material or a free electron (Compton scattering). These two scattering mechanisms will be described in Sections 4.3 and 4.4, respectively. If a sample is irradiated with a parallel beam of radiation, then when measuring the intensity of the radiation passing through the sample in the *same* direction, scattering and absorption manifest themselves identically: due to both phenomena, the flux of photons in the initial direction decreases. This is precisely how the radiation attenuation coefficient μ is measured. The **attenuation coefficient** is the reciprocal of the material thickness after passing through which the radiation intensity in the initial direction is $e = 2.7183$ times smaller than the initial intensity. Thus, the attenuation coefficient is determined by both absorption and scattering. Therefore, in practice, scattering is often not distinguished from absorption, and the attenuation coefficient is called the **absorption coefficient**. Scattering differs from absorption in that it does not change the *total* flux of photons (in all directions). Scattering is studied by measuring the intensity of radiation in directions that differ from the direction of the primary radiation. However, it is not always accurate to call the radiation measured in this way "scattered" radiation, because part of the recorded radiation may be secondary radiation emitted by excited atoms after a certain time following the transfer of energy to them via interaction with incident particles (see Section 4.2).

4.2. Photoelectric effect

The **photoelectric effect** or **photoelectric absorption** is an interaction of a photon with an atom during which the atom absorbs the *entire* energy of the photon (i.e., the photon ceases to exist), and one of the atomic electrons is ejected from the atom. The electron ejected from the atom is called a **photoelectron**. This phenomenon is sometimes called the **internal photoelectric effect** or **atomic photoelectric effect** to emphasize that the photoelectron usually remains inside the material (the "external photoelectric effect" is the liberation of electrons from a metal into a vacuum when the metal is illuminated with visible or ultraviolet light). From the law of conservation of energy follows the expression for the kinetic energy of the photoelectron:

$$E_{\mathrm{ph}} = h\nu - \varepsilon_{\mathrm{b}}, \tag{4.2.1}$$

where ε_{b} is the binding energy of the corresponding electron shell.

Based on the laws of conservation of energy and momentum, it can be proven that free electrons cannot absorb a photon, i.e., they cannot undergo the photoelectric effect.

From Eq. (4.2.1) it is clear that the photoelectric effect is possible only when $h\nu > \varepsilon_{\mathrm{b}}$. Therefore, the dependence of the photoelectric cross section on photon energy is characterized by sharp jumps corresponding to the binding energies of individual electron shells (K, L, …) (see Fig. 11a). The last (highest-energy) jump corresponds to the liberation of an electron from the innermost (K) electron shell (in Fig. 11a, this jump is at an energy of 1.84 keV, and it is the only one because the L-shell binding energy of a silicon atom is less than 200 eV, whereas the X-axis of the plot in Fig. 11a starts at 1 keV). Such events account for approximately 80 % of all photoelectric effect events if the photon energy $h\nu$ is greater than the binding energy of the K-shell electrons. When $h\nu$ is of the same order of magnitude as the electron rest energy mc^2 (i.e., when $h\nu$ is on the order of several hundred keV), the K-shell photoelectric cross section (m^2) is approximately equal to

$$(\sigma_{\mathrm{ph}})_{\mathrm{K}} \approx 10^{-37}\, Z^5/(h\nu)^{7/2}, \tag{4.2.2}$$

where the photon energy $h\nu$ is expressed in MeV. As can be seen, the photoelectric cross section decreases

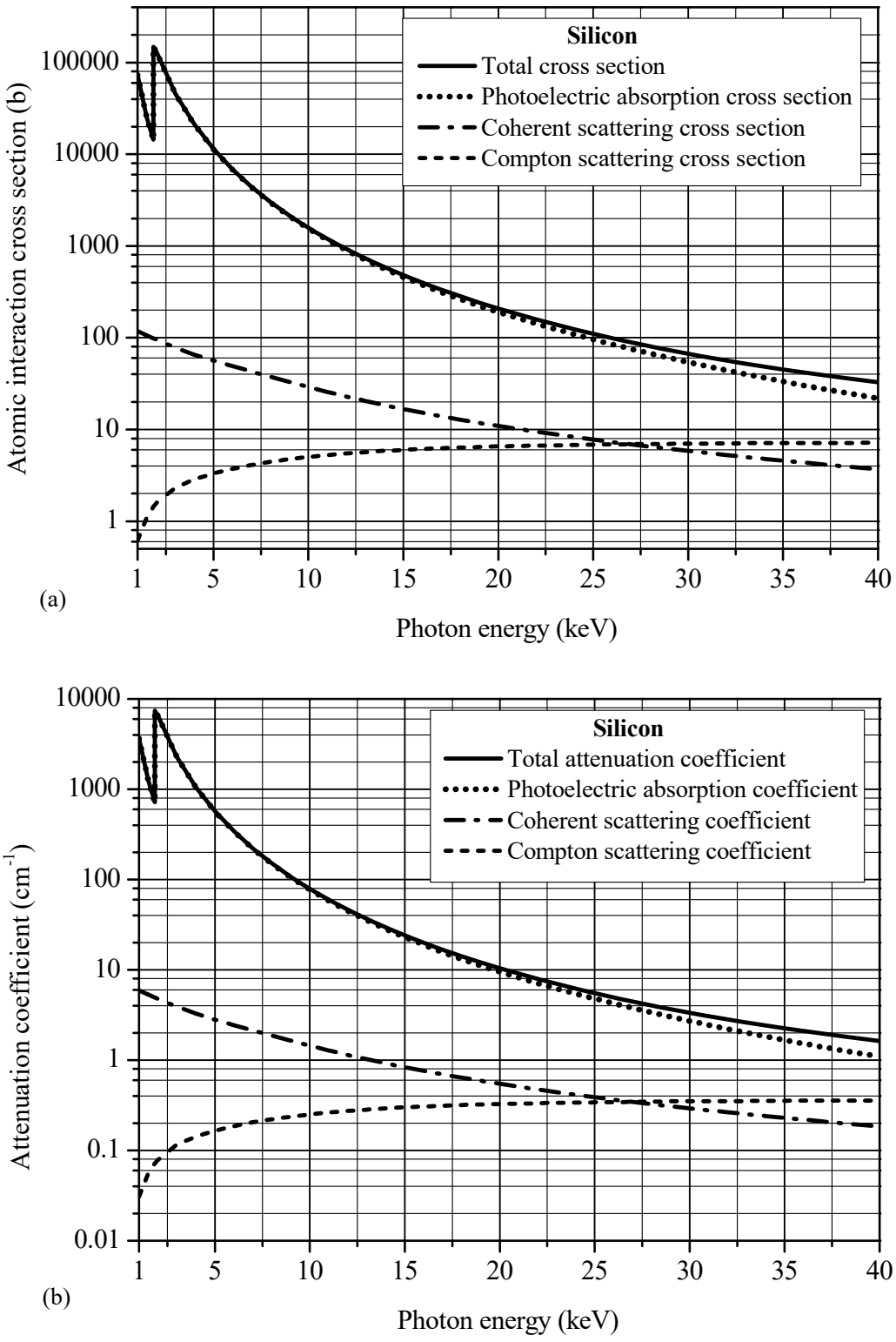


Fig. 11. Dependence of the X-ray interaction cross section with silicon atoms (a) and attenuation coefficient (b) as well as their components on photon energy (from [4]).

very rapidly with increasing photon energy and decreasing nuclear charge. When $h\nu \gg mc^2$, the photoelectric cross section is inversely proportional to the photon energy (i.e., the decrease of the cross section with increasing energy slows down), but it remains directly proportional to Z^5. Therefore, for heavier elements (e.g., lead), the photoelectric effect manifests itself even when the photon energy is close to 5 MeV. The dependence of the L-shell photoelectric cross section on Z and $h\nu$ is similar in form to Eq. (4.2.2), except that the proportionality coefficient is much smaller. Therefore, the photoelectric effect from the L shell becomes the dominant photoelectric process only when the photon energy becomes smaller than the binding energy of the K-shell electrons (then the photoelectric effect from the K shell is impossible, so $(\sigma_{\text{ph}})_{\text{K}}$ drops abruptly

to zero).

The rapid decrease of the photoelectric cross section with increasing photon energy can be explained as follows. As the difference between the photon energy and the electron binding energy increases, the fraction of the photon energy that is converted into the electron's kinetic energy E_{ph} during the photoelectric effect increases, while the fraction spent on breaking the electron's bond with the atom decreases (see Eq. (4.2.1)). That is, the approximate equality $E_{\text{ph}} \approx h\nu$ becomes increasingly accurate. This equality can be exact only when the binding energy ε_{b} is equal to zero, i.e., when the electron is free. In other words, as the photon energy increases, an atomic electron becomes increasingly similar to a free electron from the photon's perspective. However, as mentioned, a free electron cannot absorb a photon. Therefore, the photoelectric cross section decreases with increasing photon energy and in the high-energy region becomes much smaller than the Compton scattering cross section, which also begins to decrease with increasing photon energy (in Fig. 11a this decrease is not visible because the *X*-axis of the plot ends at 40 keV). Similarly, one can explain the increase of the photoelectric cross section as Z increases: as the nuclear charge increases, the electron binding energy in the atom increases, so the difference between the photon energy and the electron binding energy decreases. The mathematical form of the dependencies of the photoelectric and Compton scattering cross sections on photon energy can be explained only by the methods of quantum electrodynamics. Thus, there is no simple explanation.

During the photoelectric effect, the vacancy created in the inner electron shell is filled by an electron from a higher electron shell, emitting a characteristic X-ray photon. A fraction of these secondary characteristic X-ray photons is absorbed in the same material in which the photoelectric effect takes place ("**self-absorption**"). This fraction increases as the atomic number of the material decreases, because the energy $h\nu_{\text{sec}}$ of the secondary characteristic X-ray photons decreases rapidly; according to Moseley's law (3.4.8), it is approximately proportional to Z^2:

$$h\nu_{\text{sec}} \sim Z^2. \tag{4.2.3}$$

Substituting this dependence into Eq. (4.2.2) in place of $h\nu$, we obtain the following dependence on Z:

$$\sigma_{\text{ph}} \sim Z^{-2}. \tag{4.2.4}$$

Consequently, if we have two materials with atomic numbers Z_1 and Z_2, and if $Z_1 < Z_2$, then the characteristic radiation of the atoms of the first material is absorbed more strongly in the first material than the characteristic radiation of the atoms of the second material is absorbed in the second material. It should be borne in mind that the dependencies (4.2.2) and (4.2.4) are very approximate and become unreliable when the photon energy is on the order of 10 keV or less. However, this rapid strengthening of the self-absorption of secondary radiation with decreasing Z is indeed observed in practice.

4.3. Coherent scattering

Scattering of electromagnetic radiation during which its frequency does not change is called **coherent scattering** (in the opposite case, the scattering is called **incoherent scattering**). Such scattering is sometimes called "classical" because, according to classical electrodynamics, the frequency of electromagnetic waves does not change during scattering. Classical electrodynamics states that electromagnetic radiation consists of electric and magnetic fields. In such a wave, the electric field varies according to a harmonic law. In an electric field of strength $\mathscr{E}$, an electron is acted upon by a force $e\mathscr{E}$ (since the letter "E" is customarily used to denote energy, the field strength is denoted here in a different typeface). Since the time dependence of the field strength $\mathscr{E}$ in an electromagnetic wave is harmonic, the force acting on the electron periodically changes direction and thus sets the electron into oscillation. Therefore, an electron acted upon by an electromagnetic wave oscillates at a frequency equal to the frequency of that wave. During oscillatory motion, the electron has a certain non-zero acceleration, because not only the coordinates of the electron change periodically, but also their time derivatives, i.e., the components of the velocity vector. Thus, another law of electrodynamics manifests itself: a charged particle moving with acceleration radiates electromagnetic waves. If the motion is harmonic, then the secondary electromagnetic radiation is of the same frequency as the oscillations of the particle. This is coherently scattered radiation. However, such an explanation applies only when electromagnetic radiation is scattered by *free* electrons (in classical electrodynamics, such scattering is called **Thomson scattering**). The classical explanation of coherent scattering due to interaction with *atomic* electrons is based on the assumption that the atomic electrons behave as a spatial distribution

of negative charge (a "cloud") that can oscillate around the point-like positive nuclear charge at the center without changing its internal configuration. That is, the electromagnetic wave interacts with the system of atomic electrons as a whole. Unlike in the case of a free electron, the oscillating system in this case has its own natural frequency, which is higher the stronger the Coulomb bond of the electrons to the nucleus. Therefore, the dependence of the oscillation amplitude on frequency has a different form than in the case of free electrons: it is resonant, i.e., the amplitude of forced oscillations is greatest when the wave frequency is equal to the natural frequency. However, as in the case of free electrons, another fundamental conclusion remains valid: the frequency of forced oscillations is equal to the frequency of the external force (the electric field), so the electrons oscillating around the nucleus emit secondary electromagnetic radiation of the same frequency. Such coherent scattering (due to interaction with atoms) is called **Rayleigh scattering**.

According to quantum mechanics, the scattering of photons by free electrons is always incoherent, because during it the photon energy always changes. That is, coherent scattering is possible only when interacting with atomic electrons. Moreover, according to quantum mechanics, during coherent scattering, not only does the energy remain unchanged, but also the quantum state of the atomic electron. Constancy of the quantum state is a stricter requirement than constancy of energy, because a single energy level corresponds to several different quantum states (e.g., states of a hydrogen atom with the same principal quantum number n but different azimuthal quantum numbers l have the same energy and are expressed by Eq. (3.4.1)). The quantum explanation of coherent scattering is based on the uncertainty principle, which states that the radius vector and momentum vector of an atomic electron in a stationary state are indeterminate (they can be equal to any values from 0 to infinity), although their probability densities decrease rapidly with increasing radial coordinate or absolute value of momentum. The simplest formulation of this principle is the **Heisenberg inequality**: the uncertainties (standard deviations) of a particle's coordinate and momentum are related by the inequality:

$$\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}, \tag{4.3.1}$$

where $\hbar = h/(2\pi) \approx 1.05 \cdot 10^{-34}$ J · s is the reduced Planck constant (analogous relations hold for the y and z components as well). For example, if a particle's coordinate x is precisely determined ($\Delta x = 0$), then according to Eq. (4.3.1) $\Delta p_x = \infty$, i.e., the x-component of momentum is completely indeterminate. The uncertainty of the radial coordinate r (i.e., the distance between the nucleus and the electron) of an atomic electron is close to the statistical mean of the radial coordinate $\langle r \rangle$, and the uncertainty of the magnitude of its momentum p is close to the average momentum of the electron in the atom. The probability that the electron's radial coordinate r or momentum p will become much larger than the corresponding mean is very small, yet remains non-zero for any values of r and p. Therefore, even for an atom in the ground state (when the electron energy is strictly defined), regardless of the magnitude of the incident photon's wave vector $\mathbf{k}$, there exist possible values of the electron's momentum vector that satisfy the condition of coherent scattering: the quantum state of the atomic electron does not change during coherent scattering. This condition can be formulated as a requirement on the momentum vector $\mathbf{p}$ of the atomic electron. To do this, one must use the spherical symmetry of the atom. Because of this property, the wave function of an atomic electron (more precisely, its squared modulus, which has the meaning of probability density) is also a spherically symmetric function. The argument of the wave function can be either the electron's radius vector $\mathbf{r}$ (this is the "coordinate representation") or its momentum vector $\mathbf{p}$ ("momentum representation"). The spherical symmetry mentioned above is characteristic of both coordinate wave functions $\psi(\mathbf{r})$ (if $\mathbf{r}$ is measured from the center of the atom) and momentum wave functions $\phi(\mathbf{p})$. When discussing momentum transfer during scattering, it is more convenient to use momentum wave functions. Due to this spherical symmetry, processes during which the quantum state of an atomic electron does not change are those in which only the direction of the electron's momentum vector changes, but not its magnitude. We will denote the momentum transferred by the photon to the atomic electron ("**momentum transfer**") by $\mathbf{q}$:

$$\mathbf{q} = \mathbf{k} - \mathbf{k}', \tag{4.3.2}$$

here $\mathbf{k}$ and $\mathbf{k}'$ are the initial and final photon momenta, respectively. Since the scattering is coherent, $|\mathbf{k}'| = |\mathbf{k}|$, but the vectors $\mathbf{k}$ and $\mathbf{k}'$ have different directions, so $|\mathbf{q}| > 0$. If the initial momentum vector of the atomic electron is $\mathbf{p}$, then during scattering it becomes equal to $\mathbf{p} + \mathbf{q}$. As mentioned, $|\mathbf{p} + \mathbf{q}|$ must be equal to $|\mathbf{p}|$, i.e.,

$$(\mathbf{p} + \mathbf{q})^2 = \mathbf{p}^2$$

or

$$2\mathbf{p}\cdot\mathbf{q} = -|\mathbf{q}|^2,$$
$$\mathbf{p}\cdot\frac{\mathbf{q}}{|\mathbf{q}|} = -\frac{|\mathbf{q}|}{2}. \tag{4.3.3}$$

In momentum space, this equality is satisfied by all points belonging to the plane that is perpendicular to the vector $\mathbf{q}$ and passes through the point $-\mathbf{q}/2$. Consequently, that plane corresponds to the "portion" of the electron momentum distribution that interacts with the photon during coherent scattering with momentum transfer $\mathbf{q}$.

According to quantum mechanics, the coherent scattering cross section is proportional to $|F(\mathbf{q})|^2$, where $F(\mathbf{q})$ is the **overlap integral**:

$$F(\mathbf{q}) = \int \phi^*(\mathbf{p})\phi(\mathbf{p}+\mathbf{q})\,\mathrm{d}^3\mathbf{p} \tag{4.3.4}$$

(the integration is performed over all momentum space). $|\phi(\mathbf{p})|^2$ is a spherically symmetric momentum probability density centered at $\mathbf{p} = \mathbf{0}$, while $|\phi(\mathbf{p}+\mathbf{q})|^2$ is a distribution of the exact same shape "shifted" to be centered at $\mathbf{p} = -\mathbf{q}$. As $|\mathbf{q}|$ increases, these two distributions "overlap" less, so the integrand of Eq. (4.3.4) decreases, $|F(\mathbf{q})|$ decreases, and the coherent scattering cross section decreases. The region of maximum overlap lies halfway between the two centers (just like the plane mentioned above) and is symmetric with respect to that plane. Therefore, the overlap integral (4.3.4) can be interpreted as a "filter" that "selects" from all possible electron momentum vector values of the state $\phi(\mathbf{p})$ the subset that participates in coherent scattering with momentum transfer $\mathbf{q}$.

As the photon energy decreases or the atomic number of the scattering material increases, the coherent scattering cross section (probability) increases. Coherent scattering becomes the dominant scattering mechanism when the photon energy is less than 10 keV (the higher the atomic number of the element, the wider the energy interval over which coherent scattering dominates). For example, during the scattering of X-rays with a wavelength of 50–250 pm (5–25 keV photon energy), coherent scattering usually dominates. Since these wavelengths are close to the distances between neighboring atoms in solids, X-ray diffraction in crystals can be easily realized at such wavelengths (see Section 6). During the scattering of medium- and high-energy (> 500 keV) gamma quanta in materials of low atomic number (e.g., in aluminum), coherent scattering practically does not manifest itself, and it can be assumed that Compton scattering is the sole scattering process.

4.4. Compton scattering

From the standpoint of quantum mechanics, a scattering event is a collision of two particles—a photon and an electron, or a photon and an atom. If a photon collides with a free electron, such an interaction is elastic, because during it the sum of the kinetic energies of the particles does not change. The energies and momenta of the particles after the collision are determined by the laws of conservation of energy and momentum. From these laws it follows that due to electron recoil, the scattered photon transfers part of its energy and momentum to the electron. Therefore, the energy and frequency of the photon decrease (so the scattering is incoherent). This phenomenon, described in 1922 by the American physicist A. Compton, became one of the cornerstones of quantum mechanics, as it showed that electromagnetic radiation possesses corpuscular properties. Such scattering, during which the frequency of the radiation decreases, is called **Compton scattering**, and the decrease in frequency (and the corresponding increase in wavelength) is called the **Compton effect**. Most often, the photon transfers a significant fraction of its energy to the electron. Those fast electrons produced by Compton scattering are called **Compton recoil electrons**. They lose energy in matter due to Coulomb interaction with atomic electrons and nuclei, as well as due to bremsstrahlung.

When a photon interacts with a polyatomic molecule or a solid, another incoherent scattering process is possible—**Raman scattering**. During Raman scattering, the energy of a photon can both decrease and increase. Quantum mechanics explains Raman scattering as follows. Part of the photon's energy can be converted into vibrational energy of the atoms of the molecule or solid, or into rotational energy of the molecule (then the photon energy decreases). The reverse process is also possible: a photon can gain energy from a molecule that is in an excited vibrational or rotational state (then the photon energy increases, while the vibrational or rotational energy of that molecule decreases by the same amount). There exists an approximate classical explanation of Raman scattering that applies only to certain specific cases. Thus, strictly speaking,

incoherent scattering of electromagnetic radiation is possible not only from the standpoint of quantum mechanics, but also from the standpoint of classical physics. Hereafter, we will disregard Raman scattering, because: 1) its intensity is many times smaller than the intensity of coherent scattering, 2) it is significant only at low photon energies (on the order of electron-volts), and 3) it manifests itself only when radiation interacts with a polyatomic system (and not with isolated atoms).

The Compton effect is the primary factor responsible for the energy transfer of hard X-rays and medium-energy gamma radiation (i.e., electromagnetic radiation whose quantum energy is between 0.1 MeV and 1 MeV) to materials composed of atoms of light elements (e.g., water and organic compounds). In practice, the Compton effect is applied when setting up shielding against gamma radiation using low-density materials (concrete, bricks, iron, etc.), in which absorption at the initial photon energy is weak. Specifically, as the photon energy decreases due to multiple collisions with electrons, its probability of absorption (photoelectric effect) increases.

Photons can be scattered not only by free electrons, but also by atoms (more precisely, by atomic electrons). If part of the photon's energy is converted into internal energy of the atom during such scattering, the scattering is also called Compton scattering. Thus, Compton scattering due to interaction of a photon with an atom (rather than with a free electron) is inelastic. During Compton scattering, the atom can be either ionized or merely excited (without ionization). In either case, almost the entire decrease in the photon's energy is equal to the increase in the internal energy of the atom (the recoil kinetic energy of the atom is relatively small and can be neglected). If the atom is ionized, then (neglecting the recoil energy of the atom) the energy loss of the photon consists of two parts: the kinetic energy of the Compton recoil electron and the binding energy of that electron in the atom (**electron binding energy** is the work required to remove an electron from a given electron shell).

The role of the electron binding energy or atomic excitation energy during Compton scattering becomes important only when the photon energy is not much greater than the average excitation energy $\bar{I}$ of the atom. However, then the probability of Compton scattering becomes much smaller than the probabilities of other interaction processes—the photoelectric effect and elastic scattering. Therefore, in practice, Compton scattering needs to be taken into account only when the photon energy is much greater than $\bar{I}$. In such a case, during Compton scattering, the atom is almost always ionized, and scattering proceeds almost identically to scattering by a free electron. The average excitation energies of atoms are less than 1 keV. Therefore, when atomic electrons scatter photons whose energy is greater than 100 keV, the free-electron approximation can be applied. Furthermore, since the K-shell electron binding energy is much smaller than the incident photon energy, the probability of the photoelectric effect is small, so in such a case it can be stated that quanta whose energy is less than 1 MeV interact with matter solely via scattering.

The limiting case where the classical description of scattering applies is the scattering of visible light or lower-frequency electromagnetic radiation by free electrons. The increase in radiation wavelength caused by the Compton effect does not depend on the radiation frequency and is less than 5 pm. Thus, when the radiation wavelength exceeds 100 nm, the *relative* increase in wavelength caused by the Compton effect becomes almost unnoticeable, so the scattering can be considered coherent (Thomson) scattering.

Hereafter, in analyzing Compton scattering, we will rely on the assumption that photons are scattered by free electrons (rather than atoms), because, as mentioned, in most cases encountered in practice, the electron binding energy in the atom does not have a major influence on Compton scattering. Due to Compton scattering, instead of the primary photon of energy $h\nu$, a scattered photon of energy $h\nu' < h\nu$ appears, while the electron that scattered the photon acquires kinetic energy $E_C = h\nu - h\nu'$.

Using the laws of conservation of energy and momentum, one can obtain the relation between the scattered photon energy $h\nu'$, the incident photon energy $h\nu$, and the photon scattering angle θ, as well as the relation between the Compton recoil electron energy E_C and its emission angle φ (the angles θ and φ are measured from the initial direction of photon motion):

$$\begin{cases} h\nu' = \dfrac{h\nu}{1 + (1 - \cos\theta)\gamma}, \\ E_C = \dfrac{2h\nu\gamma}{1 + 2\gamma + (1+\gamma)^2 \operatorname{tg}^2\varphi} = h\nu \dfrac{\gamma(1 - \cos\theta)}{1 + \gamma(1 - \cos\theta)}, \\ \operatorname{tg}\left(\dfrac{\theta}{2}\right) = \dfrac{\operatorname{ctg}\varphi}{1+\gamma}, \quad \text{where } \gamma \equiv \dfrac{h\nu}{m_0 c^2}. \end{cases} \tag{4.4.1}$$

Since the radiation wavelength λ and frequency ν are related by the expression $\nu = c/\lambda$, the first equation of system (4.4.1) can be written as follows:

$$\Delta\lambda = \lambda_C(1 - \cos\theta), \tag{4.4.2}$$

where $\Delta\lambda$ is the increase in photon wavelength due to Compton scattering, θ is the scattering angle, and λ_C is the electron **Compton wavelength**: $\lambda_C = h/(m_0c) = 2.4263$ pm. Eq. (4.4.2) is called the **Compton formula**.

From the Compton formula (4.4.2) it follows that the Compton effect strengthens with increasing scattering angle θ (see Fig. 12). In other words, as the scattering angle increases, the energy transferred by the photon to the electron (recoil energy) increases. The maximum possible increase in wavelength corresponds to scattering at an angle $\theta = \pi$ (when the direction of photon motion reverses to the opposite direction): then $\Delta\lambda = 2\lambda_C = 4.8526$ pm. In the perpendicular direction ($\theta = \pi/2$), $\Delta\lambda$ coincides with λ_C.

From the second equation of system (4.4.1) it follows that the Compton recoil electron energy is maximum when $\varphi = 0$ (i.e., when $\theta = \pi$), and this maximum energy is equal to

$$E_{C\,max} = \frac{h\nu}{1 + \dfrac{m_0c^2}{2h\nu}}. \tag{4.4.3}$$

When the photon energy is greater than 1 MeV, it can be stated that

$$E_{C\,max} \approx h\nu - \frac{m_0c^2}{2} \approx h\nu - 0.25\ \text{MeV}, \tag{4.4.4}$$

and the average recoil electron energy is approximately equal to half of the incident photon energy.

The energy spectrum of Compton recoil electrons is continuous: a photon can transfer any fraction of its energy to the electron—from zero to the maximum energy specified by Eq. (4.4.3).

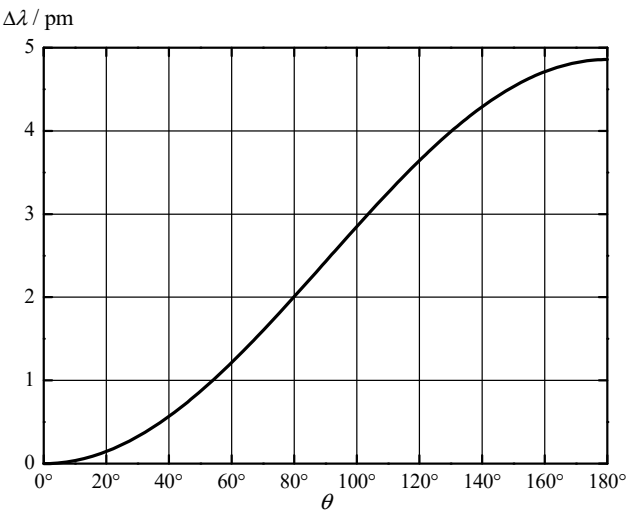


Fig. 12. Dependence of the wavelength change due to the Compton effect on the scattering angle.

The Compton scattering cross section depends on the photon energy, and the angular distribution of the scattered radiation intensity is not symmetric with respect to the $\theta = \pi/2$ plane (at small scattering angles, the probability is greater than at large scattering angles). In this respect, Compton scattering differs from classical scattering. According to the quantum theory of Compton scattering, which was developed in 1928 by Swedish physicist O. Klein and Japanese physicist Y. Nishina, the dependence of the Compton scattering cross section on photon energy is described by the formula

$$\sigma = \frac{3}{4}\sigma_T\left\{\frac{1+\gamma}{\gamma^2}\left[\frac{2(1+\gamma)}{1+2\gamma} - \frac{1}{\gamma}\ln(1+2\gamma)\right] + \frac{1}{2\gamma}\ln(1+2\gamma) - \frac{1+3\gamma}{(1+2\gamma)^2}\right\}, \tag{4.4.5}$$

here γ is the ratio of the initial photon energy $h\nu$ to the electron rest energy $m_0c^2 = 511$ keV:

$$\gamma = \frac{h\nu}{m_0c^2}, \tag{4.4.6}$$

and σ_T is the Thomson scattering cross section:

$$\sigma_T = \frac{e^4}{6\pi\varepsilon_0^2c^4m^2} \approx 6.6 \cdot 10^{-29}\ \text{m}^2. \tag{4.4.7}$$

As the parameter γ decreases (i.e., as the initial photon energy $h\nu$ decreases), σ increases, approaching σ_T (see Fig. 13).

Limiting cases:

1. When $\gamma \ll 1$, $\sigma \approx \sigma_T(1 - 2\gamma)$. That is, when the photon energy is much smaller than 0.5 MeV, the Compton scattering cross section decreases linearly with increasing photon energy.
2. When $\gamma \gg 1$, $\sigma \approx \dfrac{3}{8}\sigma_T\dfrac{1}{\gamma}\left(\dfrac{1}{2} + \ln 2\gamma\right)$. Consequently, when the photon energy is much greater than 0.5 MeV, the Compton scattering cross section is inversely proportional to the photon energy.

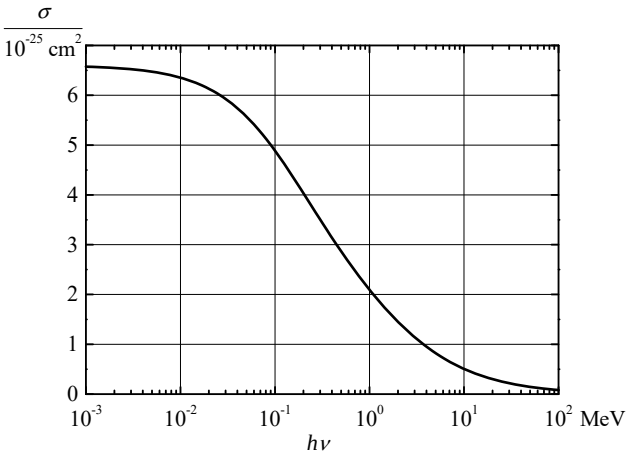


Fig. 13. Dependence of the Compton scattering cross section on the initial photon energy (Eq. (4.4.5)).

Eq. (4.4.5) was derived using the methods of relativistic quantum mechanics and taking into account the interaction of the electron spin and its magnetic moment with the electromagnetic field (i.e., with photons). Thus, it is impossible to provide a simple explanation of such an energy dependence. According to Eq. (4.4.5) (Fig. 13), the scattering cross section of photons ($h\nu$ = 662 keV) emitted by the cesium isotope ^{137}Cs is $\sigma \approx 2.5 \cdot 10^{-29}$ m^2, i.e., 2.6 times smaller than the Thomson scattering cross section σ_T (4.4.7). Accordingly, the mean free path of photons of this energy in matter is 2.6 times greater than that predicted by classical theory. For example, in water or biological tissue, it is equal to $l \approx 13$ cm, i.e., a photon of 0.66 MeV energy travels an average distance of 13 cm between collisions with electrons in such a medium.

Up to this point, the scattering cross section has been defined for a single electron. However, in practice, the scattering cross section is most often defined for a single atom. Since, at typical photon energies, Compton scattering is caused by the interaction of a photon with a single electron, the Compton scattering cross section calculated for a single atom is equal to the product of the number of electrons in the atom (Z) and the single-electron scattering cross section σ:

$$\sigma_C = Z\sigma. \tag{4.4.8}$$

Thus, the atomic Compton scattering cross section is directly proportional to the number of electrons in the atom. The dependence of σ_C on energy is quite complex. As can be seen from Fig. 11a, when the photon energy is low (much smaller than the electron rest energy m_0c^2 = 511 keV), σ_C increases with increasing energy. When the photon energy is on the order of m_0c^2 or greater, σ_C decreases (in Fig. 11a this decrease is not visible because the X-axis of the plot ends at 40 keV). The decrease of σ_C as the photon energy approaches zero is related to the fact that at low photon energies, photons are scattered not by free electrons, but by atoms (if they were scattered by free electrons, then as the photon energy approached zero, σ_C would increase toward the value $Z\sigma_T$, where σ_T is the Thomson scattering cross section (4.4.7)).

4.5. Attenuation coefficient

Based on the general expression for interaction probability (4.1.1), one can express the dependence of X-ray or gamma radiation intensity I on the thickness x of a material layer (although "intensity" means energy flux density, the same equality also describes the dependence of particle flux density on x if all photons have the same energy). Suppose that a narrow parallel beam of monoenergetic photons falls on a material layer of thickness x (see Fig. 14). As shown in Fig. 14, in the material a photon can be absorbed due to the photoelectric effect (then the photon disappears), or it can be scattered due to coherent or Compton scattering (then the photon is deflected away from the detector). The photons that reach the detector are those that passed through the material without interacting with the atoms of the material even once. [For comparison: if the radiation beam consisted of heavy charged particles, then the detector would register all particles whose range is greater than the thickness of the material layer, even though each such particle in the material interacted with many atoms.] The probability dP of a single photon interacting with a material layer of thickness dx is given by Eq. (4.1.1). By the definition of probability, the interaction probability dP is equal to the ratio of the number of photons that interacted with the material to the total number of photons that struck it ("incident" photons). Since the radiation intensity is proportional to the number of photons striking the material per unit time, and the decrease in intensity is proportional to the number of photons that interacted with the material, in calculating the probability dP, in the numerator of the fraction instead of the number of collisions we can write the decrease in intensity (i.e., the change with a minus sign) $-$dI, and in the denominator instead of the number of photons striking the material layer we can write the intensity I of the radiation striking the material layer. Therefore, on the left-hand side of Eq. (4.1.1), the quantity dP can

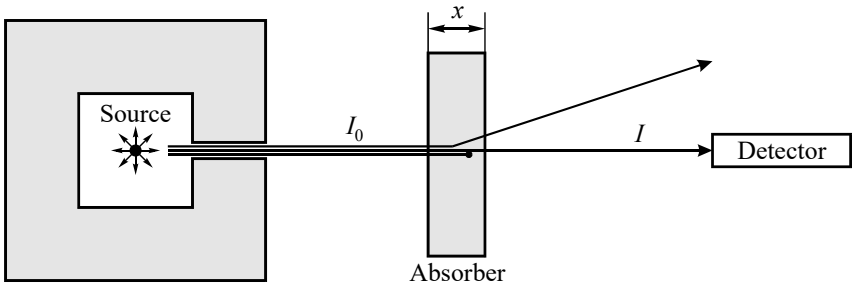


Fig. 14. Experiment for studying the absorption of X-ray or gamma radiation in matter. A narrow parallel beam of radiation falls on a layer of material ("absorber") of thickness x. In that layer, some photons are absorbed or scattered. All remaining photons (those that did not interact with the material) reach the detector.

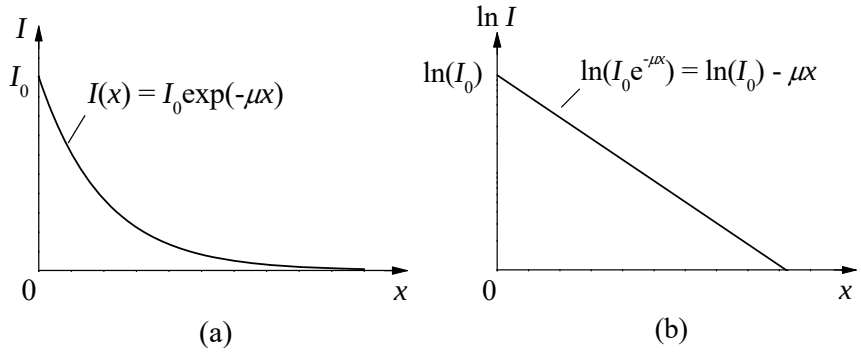


Fig. 15. Dependence of the radiation intensity (a) and its logarithm (b) on the thickness of the material layer.

be replaced by the relative decrease in intensity $(-\mathrm{d}I/I)$ upon passing through a layer of thickness dx:

$$-\frac{\mathrm{d}I}{I} = \sigma \cdot n_\mathrm{a} \cdot \mathrm{d}x, \tag{4.5.1}$$

where n_a is the number of atoms per unit volume. Integrating Eq. (4.5.1), we obtain:

$$I(x) = I_0 \mathrm{e}^{-\sigma n_\mathrm{a} x}, \tag{4.5.2}$$

where I_0 is the initial intensity. This relation can be written in the following form:

$$I(x) = I_0 \mathrm{e}^{-\mu x}, \tag{4.5.3}$$

where μ is the **attenuation coefficient**:

$$\mu = \sigma \cdot n_\mathrm{a}. \tag{4.5.4}$$

Thus, the radiation intensity in the initial direction decreases exponentially with increasing thickness of the material layer (see Fig. 15).

The cross section σ for the interaction of X-ray radiation with atoms of matter is equal to the sum of the cross sections of three distinct processes:

$$\sigma = \sigma_\mathrm{ph} + \sigma_\mathrm{coh} + \sigma_\mathrm{C}, \tag{4.5.5}$$

where σ_ph, σ_coh, and σ_C are the cross sections of the photoelectric effect, coherent scattering, and Compton scattering, respectively. Therefore, the attenuation coefficient μ of X-ray radiation can also be expressed as the sum of three attenuation coefficients corresponding to the three processes—the photoelectric effect, coherent scattering, and Compton scattering:

$$\mu = \mu_\mathrm{ph} + \mu_\mathrm{coh} + \mu_\mathrm{C}. \tag{4.5.6}$$

The expressions for the coefficients μ_ph, μ_coh, and μ_C are obtained by substituting the corresponding cross section into Eq. (4.5.4) in place of σ. When the photon energy is on the order of 10 keV or less, the dominant type of interaction is the photoelectric effect (e.g., see Fig. 11b).

Since the attenuation coefficient μ is proportional to the interaction cross section (see Eq. (4.5.4)), the dependence of μ on photon energy has the exact same form as the dependence of the interaction cross section (see Fig. 11b). Since the attenuation coefficient μ depends on the photon energy, the exponential attenuation law (4.5.3) holds only in the case of monochromatic (single-energy) radiation.

5. Properties of Semiconductor Ionizing Radiation Detectors

5.1. Simplified detector model

An ionizing radiation detector is a device used to detect ionizing radiation and convert its energy into another form of energy that can be measured. In order to detect radiation, it must interact with the active material of the detector. The types and physical mechanisms of these interactions depend on the nature of the radiation and the energy of its particles. Different types of detectors "respond" to different types of interactions. Most often, radiation is detected via the ionization phenomenon (i.e., the liberation of electrons from the atoms of a material when incident radiation particles interact with them). The result of the interaction is a measurable detector *signal*. Different detectors use different signals: these may be electric current pulses, particle tracks in a nuclear emulsion, and so on. The methods for recording and measuring these signals are also very diverse. Before discussing specific types of detectors, let us examine a hypothetical detector operating according to the following general model:

1) radiation creates free charge carriers in the active material of the detector;

2) under the action of an electric field created inside the detector, the charge carriers move and induce an electric current in the load circuit.

Such a detector model is suitable for analyzing gas-filled and semiconductor detectors.

Suppose that the interaction of a single particle with the detector manifests itself by the appearance in the detector volume of an equal number of positive and negative free charge carriers, whose total charges are $+Q$ and $-Q$, respectively. More briefly, this statement can be formulated as follows: "a charge Q appears in the detector volume." The time over which this charge appears is the time during which the incident particle (and the secondary charged particles created by the interaction of the incident particle with the detector material) loses all of its energy. This time is on the order of 10^{-9} s in gases and on the order of 10^{-12} s in solids. In most cases, this time is so short that the charge Q can be considered to be created practically instantaneously. Suppose the charge Q is created at the instant $t = 0$. We will call this event an "interaction event," even though at the microscopic level it consists of many individual interaction events (e.g., a single incident particle can ionize tens of thousands of atoms). Subsequently, the created charge is collected. For this purpose, an electric field is used, which forces the created positive and negative charge carriers to move in opposite directions. The charge collection time depends on the type of detector. For example, in gas-filled detectors, the charge collection time can reach several milliseconds, whereas in semiconductor detectors this time is only on the order of 10^{-8} s. This time depends on the mobility of the charge carriers and on the distance they must travel to the detector electrodes. As the charge carriers move toward the detector electrodes, the charge induced on those electrodes changes, i.e., an electric current flows in the load circuit connected to those electrodes. This current is equal to the conduction current in the detector volume.

Thus, the detector can be modeled as a current source that generates a current pulse $i(t)$ every time a particle interacts with the active material of the detector. The amplitudes (heights) and durations (widths) of different pulses can vary depending on the characteristics of the interaction events that caused them (see Fig. 16). The duration of each current pulse is equal to the charge collection time. We will denote this duration by t_c (see Fig. 16). The integral of the current pulse with respect to time is equal to the total created charge Q:

$$\int_0^{t_c} i(t)\,\mathrm{d}t = Q. \tag{5.1.1}$$

If the frequency of interaction events is high, some current pulses may overlap with one another. Hereafter, we will assume that interaction events are relatively rare, so that pulses do not overlap in time. It should be kept in mind that the time intervals between current pulses are random, because the arrival of radiation particles at the detector is a random process described by the Poisson distribution.

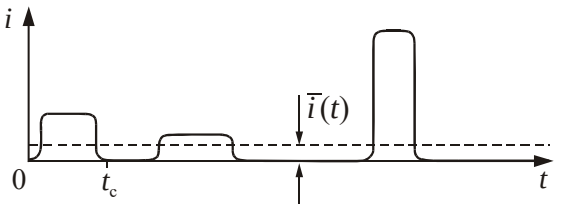


Fig. 16. Examples of detector current pulses. The dashed line indicates the time average of the current.

5.2. Detector operation

A detector can operate in pulse mode or in current mode. **Pulse mode** measuring devices connected to the detector are designed to separately record *each* current pulse caused by an interaction event in the detector. In this way, individual particles interacting with the detector are recorded. When measuring the energy of individual particles (radiation spectroscopy), pulse mode is always used. In this case, the integral of each current pulse (5.1.1) is measured, i.e., the total charge Q, which is directly related to the energy loss of the particle in the detector. Another case where pulse mode is required is counting particles (i.e., current pulses) regardless of the value of Q. Such measurements are useful when only the average rate of particles striking the detector needs to be known, rather than the energies of individual particles (an example is measuring the activity of a radioactive source). When the rates of interaction events are very high, pulse mode can be difficult or impossible to implement, because the intervals between current pulses may become too small to distinguish the pulses, or adjacent pulses may overlap with one another. In such cases, **current mode** is used, in which only the average current over a time much longer than the intervals between interaction events is measured. Only pulse mode will be discussed further below.

In pulse mode, the detector's current pulse is converted into a voltage pulse, which is then amplified and recorded or analyzed. Many detectors include a special device that converts the detector's current pulse into a voltage pulse. This device is called a **preamplifier** (this name reflects the fact that the amplitude of the preamplifier's voltage pulse is often too small to be analyzed directly, so the pulse is further amplified). In a simplified analysis, the device connected to the detector electrodes (e.g., a preamplifier or an amplifier) can be replaced by an equivalent parallel RC circuit as shown in Fig. 17. In this schematic, R is the input resistance of the circuit, and C is the sum of the detector capacitance and the input capacitance of the device connected to its electrodes (we will call the capacitance C the "equivalent capacitance" of the detector). The measured signal is the voltage drop $U(t)$ across the resistor R (the "output voltage" of the equivalent RC circuit). Hereafter, we will also refer to this voltage pulse as the "measured signal," "measured pulse," or "measured voltage." This pulse must be distinguished from the detector's current pulse: we will refer to the "detector current" as the "input current" $i(t)$ of the equivalent RC circuit mentioned above (preamplifier) (see Fig. 17). The general expression for the pulse $U(t)$ for any shape of the current $i(t)$ is

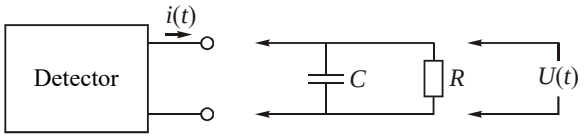


Fig. 17. Simplified equivalent circuit of detector pulse-mode operation.

$$U(t) = \frac{1}{C}\int_0^t i(t')\exp\left(\frac{t'-t}{RC}\right)\mathrm{d}t'. \tag{5.2.1}$$

In practice, the inequality $RC \gg t_c$ usually holds. Then, during charge collection (while $t < t_c$), the exponential factor in the integrand of Eq. (5.2.1) is close to unity, so

$$U(t) \approx \frac{1}{C}\int_0^t i(t)\,\mathrm{d}t \qquad (0 \le t \le t_c). \tag{5.2.2}$$

Therefore, such detector operation (with a large time constant RC) can be called "current-integrating operation." During charge collection, the voltage $U(t)$ increases. This part of the voltage pulse is called the "leading edge" of the pulse.

The physical meaning of Eq. (5.2.2) is as follows. With a large time constant, the electric current flowing through resistor R during charge collection is much smaller than the detector current $i(t)$. This means that almost all of the latter current is spent on charging the capacitance C. Therefore, during charge collection, the charge accumulated on this capacitance is equal to the integral of the current $i(t)$ with respect to time. By the definition of capacitance, the measured voltage is equal to the ratio of this charge to the capacitance C.

After a time t_c from the interaction event, the charge accumulated on the capacitance C is equal to the charge Q collected in the detector, and the measured signal is equal to its maximum value

$$U_{\max} = \frac{Q}{C}, \tag{5.2.3}$$

where Q is defined by Eq. (5.1.1) (for a rectangular current pulse, $Q = i_0 t_c$). Subsequently, the capacitance C begins to discharge through the load resistance R, and the voltage drop $U(t)$ begins to decrease exponentially:

$$U(t)|_{t>t_c} = U_{max} \exp\left(-\frac{t - t_c}{RC}\right). \tag{5.2.4}$$

This part of the voltage pulse is called the "trailing edge" of the pulse. If the interval between pulses is sufficiently long, then $U(t)$ has time to decrease to zero before the next interaction event.

Thus, in pulse mode, the output signal of the detector consists of a train of pulses, each pulse of which reflects the interaction of a single particle with the active material of the detector. Under normal operating conditions (when pulses do not overlap with one another and are not lost due to other reasons), the average frequency of those pulses is equal to the rate of interaction events in the detector. Moreover, the amplitude of each pulse reflects the charge that was created in the detector during the corresponding interaction event. As we will see later, one of the methods for studying radiation properties is based on analyzing the pulse height distribution (histogram). For example, if Q is directly proportional to the energy of the incident particle, then the pulse height distribution reflects the energy distribution (energy spectrum) of the incident particles.

Pulse-mode operation of a detector provides much more information about radiation than current-mode operation, in which all information about the amplitudes of individual detector current pulses is lost. Therefore, pulse mode is used in practice more frequently than current mode.

5.3. Pulse height spectra

As mentioned in Section 5.2, in pulse mode the amplitude of each pulse is proportional to the charge created in the detector during the corresponding interaction event (see Eq. (5.2.3)). Having measured a large number of such pulses, one would notice that their amplitudes are not identical. The pulse height distribution can reflect both the energy spectrum of the incident particles and fluctuations in the detector's response to particles of a definite energy. Therefore, the pulse height distribution is frequently used when studying incident radiation or the operation of the detector itself.

The most commonly used method of representing a pulse height distribution is the so-called **differential pulse height spectrum**. An example of such a spectrum is shown in Fig. 18a. The pulse height (expressed in volts, V) is plotted along the horizontal axis. The ratio dn/dH of the number of pulses dn whose amplitudes belong to an infinitesimally narrow interval to the width of that interval dH (units: reciprocal volts, V^{-1}) is plotted along the vertical axis. The number of pulses $n(H_1 < H < H_2)$ whose amplitudes belong to the interval from H_1 to H_2 is determined by integrating the differential pulse height spectrum from H_1 to H_2:

$$n(H_1 < H < H_2) = \int_{H_1}^{H_2} \frac{dn}{dH}\, dH. \tag{5.3.1}$$

This integral is represented by the hatched area in Fig. 18a. The total number of pulses n_0 is determined by integrating the entire spectrum:

$$n_0 = \int_0^{\infty} \frac{dn}{dH}\, dH. \tag{5.3.2}$$

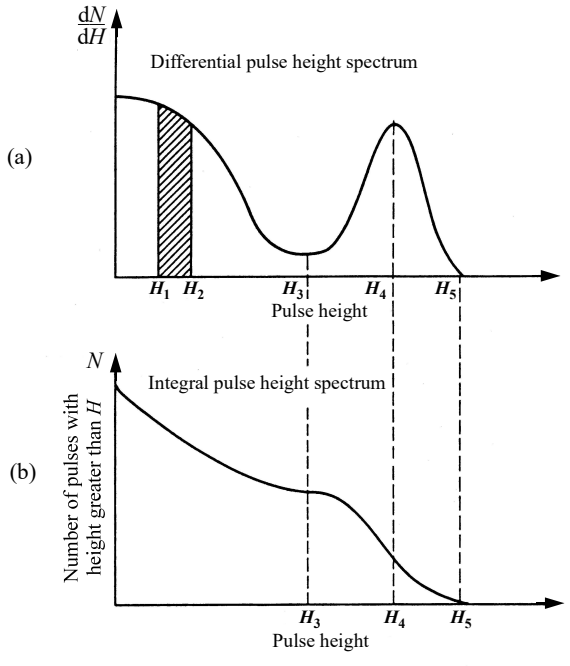


Fig. 18. Examples of differential (a) and integral (b) pulse height spectra.

The maximum pulse height is indicated by the abscissa of the point at which the spectrum drops to zero (e.g., in Fig. 18a, the maximum amplitude is H_5). The maxima of the spectrum (e.g., at point H_4) correspond to the most probable amplitudes, i.e., amplitudes that are recorded particularly frequently. The minima of the spectrum (e.g., at point H_3) correspond to the least probable amplitudes, i.e., amplitudes that are recorded particularly rarely. Any physical interpretation of a pulse height spectrum is related to

calculating the integrals (i.e., areas) of the spectrum. The ordinate ($\mathrm{d}n/\mathrm{d}H$) acquires physical meaning only after multiplying it by a small change in the abscissa (ΔH): $(\mathrm{d}n/\mathrm{d}H) \cdot \Delta H \approx \Delta n$, where Δn is the number of pulses with amplitudes between H and $H + \Delta H$.

The same information conveyed by a differential pulse height spectrum can be expressed using a so-called **integral pulse height spectrum**. The integral pulse height spectrum corresponding to the differential spectrum shown in Fig. 18a is presented in Fig. 18b. When constructing an integral spectrum, the pulse height is likewise plotted along the abscissa axis, while the number of pulses whose amplitude is greater than a given value (abscissa) is plotted along the ordinate axis:

$$n(H) = \int_H^\infty \frac{\mathrm{d}n}{\mathrm{d}H}\,\mathrm{d}H. \tag{5.3.3}$$

$n(H)$ is a monotonically decreasing function because, if $H_2 > H_1$, then the number of pulses with amplitudes greater than H_2 is always smaller than the number of pulses with amplitudes greater than H_1. The value of the integral spectrum at the point $H = 0$ gives the total number of pulses n_0. The integral pulse height spectrum becomes zero at the point corresponding to the maximum amplitude (in Fig. 18b, at point $H = H_5$).

Differential and integral pulse height spectra provide identical information about the amplitude distribution and are uniquely related to one another. The value of the differential spectrum corresponding to any amplitude H is equal to the absolute value of the derivative (slope) of the integral spectrum corresponding to the same amplitude H. Maxima of the differential spectrum correspond to the steepest slope of the integral spectrum (Fig. 18, point H_4). Minima of the differential spectrum correspond to the shallowest slope of the integral spectrum (in the case of Fig. 18, point H_3). Small changes in a spectrum are easier to observe using a differential spectrum; therefore, in practice, the differential spectrum is used more frequently than the integral one.

5.4. Detector energy resolution

Radiation detectors are frequently used to measure the energy spectrum of radiation. Such measurements belong to **radiation spectroscopy**. In this section, we will discuss some general characteristics of detectors that are important in the field of radiation spectroscopy.

Suppose that particles of a definite energy E_0 strike the detector. In the ideal case, the charge Q collected at the detector electrodes should be directly proportional to the energy E_0 of the particle that created this charge. Therefore, according to Eq. (5.2.3), the pulse height of the detector should also be directly proportional to E_0. However, in reality, this is not the case. Even when the energies of all particles striking the detector are strictly identical, different particles produce pulses of slightly different heights. This may be due to fluctuations in the number of created charge carriers, the loss of those charge carriers as they move toward the electrodes, and the generation of extraneous charge carriers (unrelated to the detected radiation).

This uncertainty in pulse heights is customarily described by the so-called **detector response function**—the differential pulse height spectrum obtained when particles of a definite energy strike the detector. We will denote the response function corresponding to a given particle energy E_0 by $G(H; E_0)$. The pulse height H is the argument of the response function, while the energy E_0 of the radiation particles is a parameter of the function. This function has the shape of a Gaussian function:

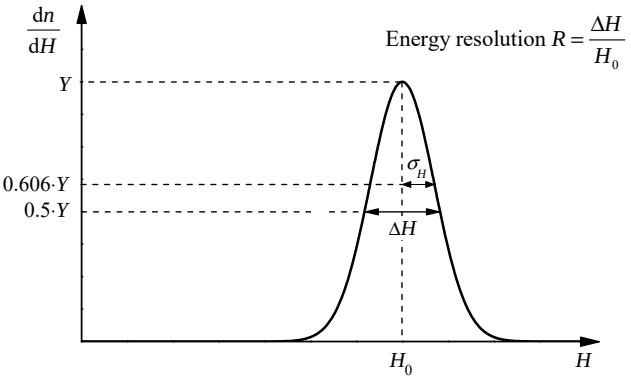


Fig. 19. Detector response function and definition of energy resolution.

$$G(H; E_0) = \frac{n_0}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(H - H_0(E_0))^2}{2\sigma^2}\right), \tag{5.4.1}$$

where H_0 is the average pulse height (which depends uniquely on E_0), σ is the standard deviation of the pulse height, and n_0 is the total number of pulses (the integral of the response function from $-\infty$ to $+\infty$). The detector response function is shown in Fig. 19.

In detectors designed to study radiation spectra, the ratio of the average pulse height H_0 to the particle energy E_0 must be constant:

$$H_0 = \text{const} \cdot E_0. \tag{5.4.2}$$

Thus, information about the energy of a particle is provided by the *average* pulse height H_0. The uncertainty in pulse heights determines the measurement error of this average pulse height. This error is customarily expressed by the so-called **energy resolution** of the detector:

$$R = \frac{\Delta H}{H_0}, \tag{5.4.3}$$

where ΔH is the width of the response function at half maximum (see Fig. 19). Since the response function is a Gaussian function (5.4.1), this width is equal to

$$\Delta H = 2.35\sigma.$$

Fig. 20 shows two response functions corresponding to the same radiation energy. Since the energy is identical, both response functions have maxima at the same point H_0 (which indicates the average pulse height). If an equal number of pulses n_0 was recorded in both cases, then the *areas* (integrals) of both peaks are also equal. However, it is obvious that the *widths* of the peaks are different. A larger peak width corresponds to poorer energy resolution. A large peak width means that the fluctuations in pulse height were large, even though the exact same amount of energy was transferred to the detector's active medium during each interaction event. As these fluctuations decrease, the width of the response function narrows and the response function approaches a Dirac δ-function. Evidently, as the width of the response function decreases, it becomes easier to distinguish fine details in the radiation energy spectrum.

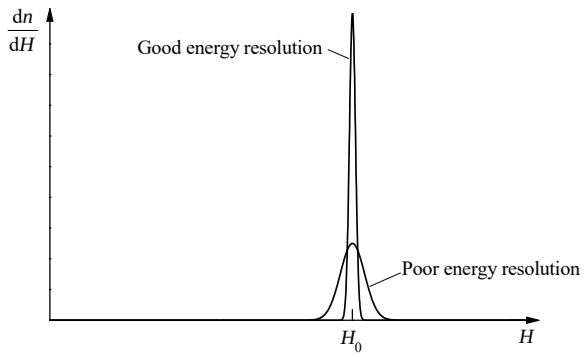


Fig. 20. Examples of detector response functions. In one case, the detector's energy resolution is relatively good, while in the other case it is poor.

Energy resolution is most frequently expressed as a percentage. For semiconductor detectors used in α-particle spectroscopy, the energy resolution is less than 1 %. For scintillation detectors used in γ-radiation spectroscopy, the energy resolution is 5–10 %. For Si-PIN detectors used in X-ray spectroscopy, the energy resolution is 3–10 %. The smaller the energy resolution, the easier it is for the detector to resolve two closely spaced spectral lines of radiation. If the radiation consists of particles of two energies, then the smallest difference between their energies at which it is still possible to separate the two peaks in the measured spectrum is approximately equal to RE_0, where E_0 is the average of those two energies.

5.5. Operating principle of a semiconductor detector

Semiconductor detectors have the best parameters among all currently existing ionizing radiation detectors. Their advantages compared to gas-filled detectors are: (a) much better energy resolution (due to the larger number of charge carriers created), (b) a 2–3 orders of magnitude greater mass of material in the sensitive volume, resulting in higher detection efficiency.

To a first approximation, a semiconductor detector can be regarded as an ionization chamber with a solid dielectric between the electrodes. Just as in a gas ionization chamber, ionizing radiation creates pairs of positive and negative charge carriers, which, under the action of an external electric field, drift toward the corresponding electrodes and induce an electric current in the load resistor (see Fig. 21). In a solid, the charge carrier mobility μ is defined as follows:

$$v^+ = \mu^+ \mathscr{E}, \qquad v^- = \mu^- \mathscr{E}, \tag{5.5.1}$$

where v^+ and v^- are the drift velocities of the positive and negative charge carriers, and $\mathscr{E}$ is the electric field strength.

The active material of the detector must meet the following requirements:

a) small average energy required to create a single charge-carrier pair;

b) high charge-carrier mobility;

c) absence of charge-carrier recombination and trapping;

d) high resistivity.

Let us discuss each of these requirements in more detail.

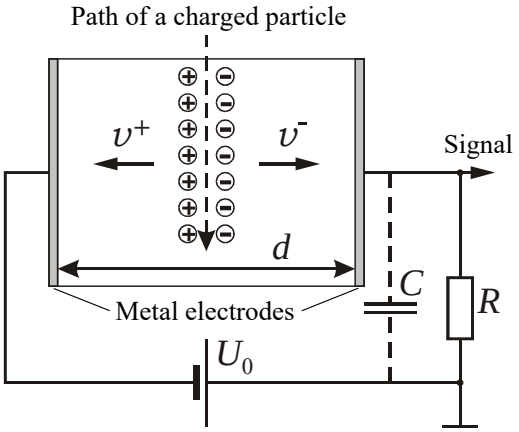


Fig. 21. Connection schematic of a "solid-state ionization chamber."

Small average energy required to create a single charge-carrier pair. The smaller the average energy W expended in creating a single charge-carrier pair, the more charge carriers are created by a particle of a given energy. Consequently, the pulse amplitude increases (see Eq. (5.2.3)). In addition, the relative fluctuations of this pulse decrease, because they are inversely proportional to the square root of the number of charge carriers. In semiconductor detectors, W is an order of magnitude smaller than in gas-filled detectors, and two orders of magnitude smaller than in scintillation detectors.

High charge-carrier mobility is necessary so that the charge carriers created by ionizing radiation reach the electrodes as quickly as possible. Then the pulse duration is short, and the detector's speed is high. If the mobilities of charge carriers of both signs are close to each other, then the pulse amplitude does not depend on the point where primary ionization occurred (this condition is necessary for the pulse amplitude to uniquely determine the energy of the particles).

Absence of charge-carrier recombination and trapping. Charge-carrier recombination manifests itself by fewer charge carriers reaching the electrodes than were created by ionizing radiation. That is, the pulse amplitude decreases, its statistical fluctuations increase, and the average amplitude ceases to be proportional to the initial number of ion pairs. In addition, in solids, charge carriers can be captured in **traps**—regions of the crystal where the periodicity of the crystal lattice is disrupted. If trapping is permanent, a space charge forms inside the dielectric volume, whose electric field opposes the external electric field. The space charge arises because charge carriers of both signs manage to shift toward the electrodes (in opposite directions) before being captured in traps. Due to this phenomenon, the total electric field can decrease so much that the detector stops operating. If trapping in traps is temporary (i.e., if the captured charge carriers are thermally released after a short time), then it manifests itself by lengthening the charge collection time and increasing the statistical fluctuations of this time (i.e., degrading the timing parameters of the detector).

High resistivity is required so that almost the entire measured electric current is caused by the charge carriers created by the ionizing action of radiation, rather than by charge carriers thermally generated within the semiconductor volume. The same requirement can be formulated differently: the generation of charge carriers in the volume of the active material must be as weak as possible. The current caused by thermally generated charge carriers in the volume of the active material is called the **leakage current**. This current adds to the current caused by ionizing radiation, making the latter difficult to measure accurately.

In nature, there is no homogeneous material that simultaneously possesses such high resistivity and meets the other three requirements mentioned above. Some insulators have sufficiently high resistivity (up to $10^{14}\ \Omega \cdot \text{m}$), small W, and high μ, but they exhibit a high probability of charge-carrier trapping. Two semiconductors—silicon (Si) and germanium (Ge)—best satisfy all four requirements, although their resistivity is several orders of magnitude lower than the minimum permissible value.

The vast majority of existing semiconductor detectors are made of germanium or silicon. To reduce the leakage current, semiconductor detectors use p–n junctions or surface-barrier junctions. In the next section, several statements of semiconductor physics are presented that will be needed when discussing the properties of such junctions.

5.6. Electron Energy Levels in a Solid

In a solid, the mutual binding energy of atoms, ions, or molecules is much greater than the kinetic energy of their thermal motion; therefore, solids are characterized by the property of maintaining their volume and

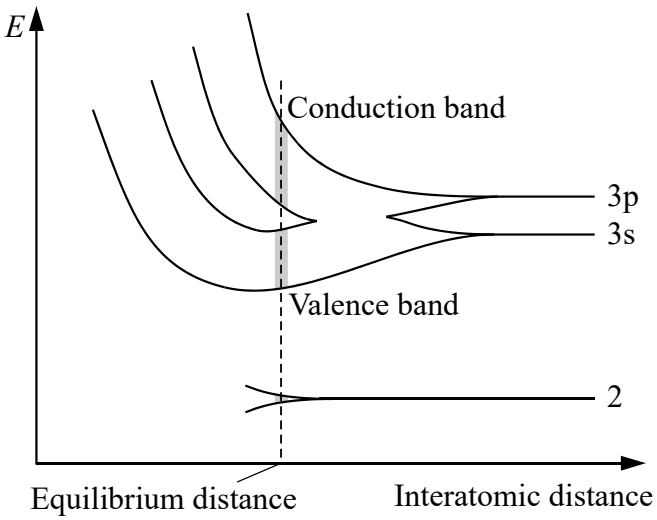


Fig. 22. Splitting of atomic energy levels into energy bands in a solid (silicon).

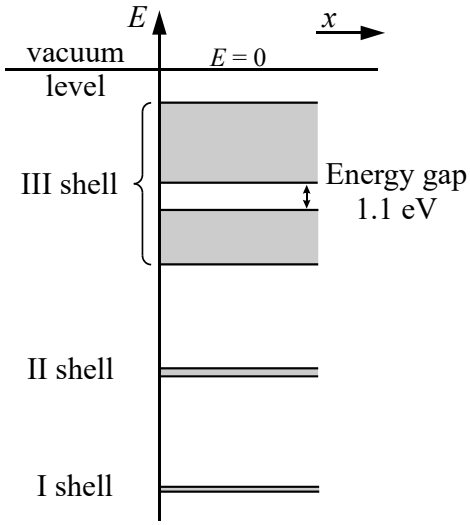


Fig. 23. Energy band diagram of silicon.

shape. Just as in molecules, attractive and repulsive forces act simultaneously between atoms in a solid, so there exists an arrangement of atoms or ions for which the total interaction energy is at a minimum. This is the equilibrium arrangement of atoms. At the position of minimum energy, the mutual arrangement of atoms repeats periodically in space, i.e., it is characterized by three-dimensional periodicity. Solids in which atoms are arranged periodically are called **crystals**. The mathematical model of the periodic arrangement of atoms in a crystal is called a **crystal lattice**.

Below are several statements of solid-state physics that are important for understanding the operation of semiconductor detectors.

5.6.1. Energy bands. Dielectrics, metals, and semiconductors

When a solid is formed from isolated atoms, the atomic energy levels split into **energy bands** (see Fig. 22).

In the lowest-energy (ground) state, the lowest energy bands are completely filled, while the highest energy bands are completely empty. In addition, a partially filled band may exist between the completely filled and empty bands. The highest of the energy bands that are completely filled in the ground state of a solid is called the **valence band**, and the adjacent energy band above it is called the **conduction band**. The gap between these bands is called the **forbidden energy band**, or **band gap**, or **energy gap** (see Fig. 23).

The valence band is occupied by electrons that form chemical bonds between atoms. Electrons located in the conduction band are free. When an external electric field is present, free electrons begin to move directionally, i.e., they determine the electrical conductivity of the material. Directional motion of electrons is associated with a change in their momentum, i.e., a change in their state of motion. This means that such motion requires empty nearby energy levels into which an electron can transition. Such energy levels exist only when the energy band in which the electron is located is only partially filled. Therefore, under normal conditions, the electrical properties of a material are determined solely by the electrons located in the conduction and valence bands: all lower bands are completely filled.

Due to the thermal motion of atoms, quantum transitions of electrons between energy bands occur continuously. As the temperature decreases, the rate of these transitions decreases. At absolute zero temperature, the energy of the body is at a minimum. Then all energy levels below a certain threshold energy are completely filled, and all higher levels are empty. The electrical properties of a material depend strongly on whether, at absolute zero temperature, the conduction band is empty or partially filled. The following cases are possible:

1. If the conduction band of a material in its lowest-energy state is partially filled, then the material's electrical conductivity is high (i.e., its resistivity is low) and depends weakly on temperature. Such materials are called **metals**. Thus, from the standpoint of band theory, a **metal** is a material whose conduction band is partially filled when the material is in its lowest-energy state.

2. If the conduction band is empty at absolute zero temperature, then the material's electrical conductivity is equal to zero (i.e., its resistivity is infinitely high). Such materials are called **dielectrics**. Thus, from the standpoint of band theory, a **dielectric** is a crystal that has no electrons in the conduction band. For example, silicon at extremely low temperature is a dielectric. Many crystals (e.g., diamond) are dielectrics at room temperature (300 K) as well.

Some materials are dielectrics when the temperature is close to absolute zero, but at room temperature their conductivity is relatively high. This is because, as temperature increases, the number of electrons excited from the valence band into the conduction band due to thermal motion grows. Using methods of quantum statistics, it can be proven that in a pure dielectric, the number of electrons in the conduction band is proportional to $\exp(-E_g/2kT)$. Here E_g is the width of the band gap (the subscript "g" comes from the word "gap"), $k = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann constant, and T is the absolute temperature. As we can see, the number of free electrons increases rapidly with increasing temperature and decreasing E_g. The main difference between dielectrics and **semiconductors** is that in semiconductors the band gap width is relatively small ($\approx$ 1 eV or less), so at room temperature the electrical conductivity is much higher than that of dielectrics (though much lower than that of metals). The band gap width of silicon is 1.1 eV, so silicon is a semiconductor. Two other semiconductors frequently encountered in practice are germanium (Ge; band gap width 0.7 eV) and gallium arsenide (GaAs; band gap width 1.4 eV). The band gap width of diamond (C) is 5.4 eV, so diamond is practically an ideal dielectric.

If a semiconductor is pure, then the number of unoccupied states in the valence band is equal to the number of free electrons. In an external electric field, these unoccupied states behave like positively charged free particles, i.e., they move in the direction of the field. These free quasiparticles are called **holes**. The word "quasiparticle" reflects the fact that a hole is not a particle in the literal sense of the word, yet it can be assigned quantities customarily used to describe the motion of "real" particles (e.g., mass, velocity, acceleration, momentum, kinetic energy). The state of a hole describes the collective state of motion of the electrons participating in chemical bonds.

Thus, in semiconductors there are two types of free charge carriers—electrons (in the conduction band) and holes (in the valence band). The resistivity of semiconductors is usually between 10^{-4} $\Omega\cdot$m and 10^{3} $\Omega\cdot$m (i.e., the electrical conductivity is between 10^{-3} $(\Omega\cdot\text{m})^{-1}$ and 10^{4} $(\Omega\cdot\text{m})^{-1}$). For comparison, the resistivity of metals is less than 10^{-7} $\Omega \cdot$ m, and the resistivity of dielectrics is greater than 10^{14} $\Omega \cdot$ m. The main parameters of the most widely used semiconductors—silicon and germanium—are given in Table 2.

Table 2. Parameters of pure silicon and germanium (from [3])

Parameter	Si	Ge
Atomic number	14	32
Atomic mass	28.09 g/mol	72.60 g/mol
Mass numbers of stable isotopes	28; 29; 30	70; 72; 73; 74; 76
Density (at 300 K)	2330 kg/m^3	5320 kg/m^3
Atomic concentration	$4.96 \cdot 10^{28}$ m^{-3}	$4.41 \cdot 10^{28}$ m^{-3}
Relative permittivity	11.7	16
Band gap width at 300 K	1.115 eV	0.665 eV
Band gap width at 0 K	1.165 eV	0.746 eV
Free electron concentration (300 K)	$1.5 \cdot 10^{16}$ m^{-3}	$2.4 \cdot 10^{19}$ m^{-3}
Resistivity (300 K)	2300 $\Omega \cdot$ m	0.47 $\Omega \cdot$ m
Electron mobility (300 K)	0.135 m^2/(V $\cdot$ s)	0.39 m^2/(V $\cdot$ s)
Hole mobility (300 K)	0.048 m^2/(V $\cdot$ s)	0.19 m^2/(V $\cdot$ s)
Electron mobility (77 K)	2.1 m^2/(V $\cdot$ s)	3.6 m^2/(V $\cdot$ s)
Hole mobility (77 K)	1.1 m^2/(V $\cdot$ s)	4.2 m^2/(V $\cdot$ s)
Average energy per electron-hole pair (300 K)	3.62 eV	—
Average energy per electron-hole pair (77 K)	3.76 eV	2.96 eV

5.6.2. Impurity energy levels

The electrical conductivity of pure semiconductors is low because they contain relatively few free electrons and holes. For example, at room temperature in pure silicon, the concentrations of electrons and holes are equal to approximately 10^{16} m^{-3}.

The number of electrons and holes in a semiconductor can be changed by replacing a fraction of the semiconductor atoms with impurity atoms. Introducing impurity atoms into a semiconductor in order to alter its electrical properties is called **doping** the semiconductor. Let us examine the role of impurity atoms.

Impurity atoms create discrete energy levels within the band gap of the semiconductor (see Fig. 24). We will call these levels **impurity levels**. Impurity levels differ from the energy levels that form the energy bands not only in their position on the energy band diagram, but also in the fact that the charge carrier of each impurity level is localized in space, i.e., bound to the corresponding impurity atom. It can happen that impurity levels are located close to the edge of one of the bands. For example, phosphorus atoms in the band gap of silicon create levels that lie at a distance of 0.045 eV from the conduction band edge. Boron atoms in silicon create levels that lie at almost the same distance from the valence band edge (see Fig. 24). If the distance between the impurity levels and the edge of either band is on the order of kT or less, then those levels can exchange electrons with the nearest energy band. If an impurity level is close to the conduction band edge, this means that the impurity atom easily loses one of its valence electrons, becoming a positive ion (see Fig. 25b). Thus, in this case, the difference between the conduction band edge and the impurity level determines the ionization energy of the impurity atom. The electrons lost by such impurity atoms become free, i.e., they transition into the conduction band. Therefore, such impurity atoms are called **donors**. In silicon and germanium, the role of a donor is most often played by phosphorus or arsenic. A characteristic feature of these elements is that their valence (number of valence electrons) is one greater than the valence of silicon and germanium (phosphorus and arsenic are pentavalent, whereas silicon and germanium are tetravalent). Therefore, once chemical bonds are formed between the impurity atom and four neighboring silicon or germanium atoms, one electron of the impurity atom remains "unpaired." The bond of this electron to the atom is relatively weak, so it is easily liberated.

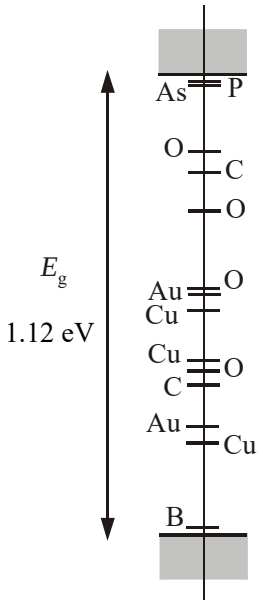


Fig. 24. Impurity energy levels in silicon.

If an impurity level is close to the valence band edge, this means that the impurity atom easily accepts one electron from the valence band, becoming a negative ion (see Fig. 25c). Such impurity atoms are called **acceptors**. The ionization energy of an acceptor is the difference between the impurity level and the valence band edge. In silicon and germanium, the role of an acceptor is most often played by boron. A characteristic feature of boron is that its valence is one less than the valence of silicon and germanium. Boron has three valence electrons, which form covalent bonds with three neighboring Si or Ge atoms. One electron is missing to form the fourth bond, i.e., one unoccupied electronic state exists at that site. Since the energy of this state is only slightly higher than that of the electrons forming bonds between Si or Ge atoms, an electron from one of the neighboring bonds between Si or Ge atoms can be excited into that state. In this way, a hole appears that can move through the semiconductor.

5.6.3. n-type and p-type semiconductors

Doping with donors increases the number of free electrons in a semiconductor, while doping with acceptors increases the number of holes. However, using methods of quantum statistics, it can be proven that in thermodynamic equilibrium, the product of the concentrations of both types of charge carriers is practically independent of the type and quantity of impurity atoms and is equal to

$$pn = n_i^2 = N_c N_v \exp\left(-\frac{E_g}{kT}\right), \tag{5.6.1}$$

where n is the free electron concentration, p is the hole concentration, n_i is the concentration of free electrons and holes in an intrinsic (pure) semiconductor (the subscript "i" originates from the term "intrinsic"), N_c is

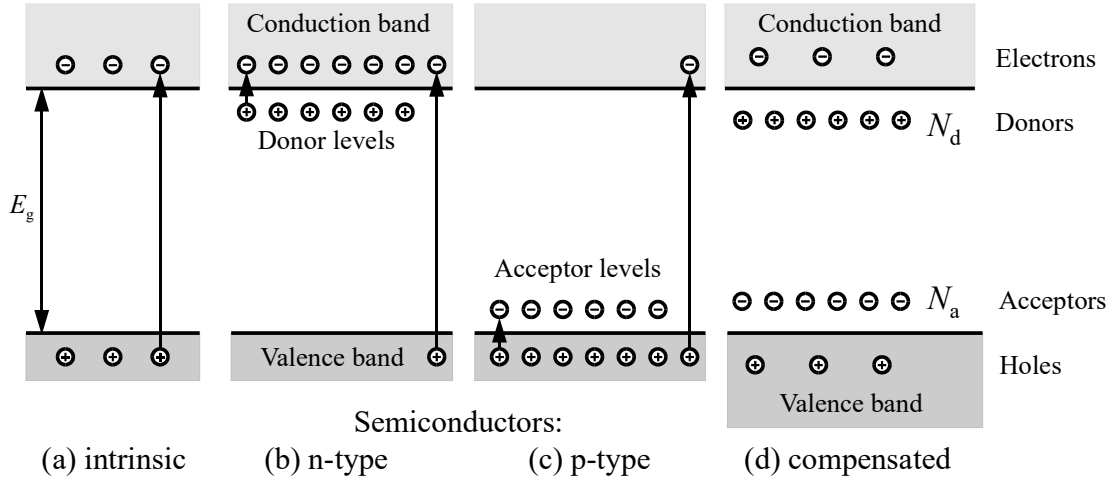


Fig. 25. Energy band diagrams of intrinsic, n-type, p-type, and compensated semiconductors.

the effective density of states in the conduction band, and N_v is the effective density of states in the valence band. N_c and N_v are material parameters that depend weakly on doping (they are on the order of 10^{26} m^{-3}). According to Eq. (5.6.1), when the number of charge carriers of one type increases, the number of charge carriers of the other type decreases by the same factor. Semiconductors in which the number of free electrons is much greater than the number of holes are called **n-type semiconductors**, and semiconductors in which the number of holes is much greater than the number of free electrons are called **p-type semiconductors**.

Electrons in an n-type semiconductor or holes in a p-type semiconductor are called **majority charge carriers**. Holes in an n-type semiconductor or electrons in a p-type semiconductor are called **minority charge carriers**.

At room temperature, the vast majority of donors and acceptors are ionized. Since their concentration usually exceeds n_i by several orders of magnitude, in an extrinsic (doped) semiconductor almost all charge carriers of either type (electrons or holes) originate from the ionization of impurity atoms. Thus, it can be stated that in an n-type semiconductor, the electron concentration is equal to the donor concentration, and in a p-type semiconductor, the hole concentration is equal to the acceptor concentration.

The best donors for silicon are phosphorus and arsenic, and the best acceptor is boron, because the ionization energies of these atoms in a silicon crystal are the smallest (see Fig. 24). In addition, these elements dissolve easily in silicon, and their concentration can reach 10^{26} m^{-3}.

Frequently, donor and acceptor impurities exist simultaneously. If the concentrations of donors and acceptors coincide and the impurity atoms are completely ionized, then the effect is as if electrons from the donor atoms transitioned into the acceptor atoms (see Fig. 25d). That is, the concentrations of free electrons and holes remain of the same order of magnitude as in an intrinsic semiconductor, no matter how high the concentration of impurities is. Thus, it can be stated that in this case the donors "compensate" the effect of the acceptors (or vice versa). Therefore, semiconductors containing both acceptor and donor impurities simultaneously are called **compensated semiconductors**. In practice, compensation is usually only partial, i.e., the donor and acceptor concentrations differ, so the concentration of free charge carriers in compensated semiconductors is much higher than their concentration in an intrinsic semiconductor.

Thus, by means of doping, it is possible to increase the number of electrons or holes in a semiconductor and thereby increase its electrical conductivity. However, the primary benefit of doping is that by creating an n-region in a p-type semiconductor or a p-region in an n-type semiconductor, a so-called **p–n junction** is formed at the boundary between the p- and n-regions. In the p–n junction region, there exists a transition layer in which the concentrations of electrons and holes are much lower than the concentration of majority charge carriers in each of the semiconductors forming the junction. Therefore, the resistivity of that layer is much higher than the bulk resistivity of either semiconductor. This property of a p–n junction is utilized in semiconductor ionizing radiation detectors. The properties of p–n junctions are described in the next section.

5.7. Potential Barrier at a Semiconductor–Semiconductor or Metal–Semiconductor Junction

A **junction** is a contact between two materials that have different energy band structures or differ in the types and concentrations of impurity atoms. A **p–n junction** is a junction between p-type and n-type semiconductors. By attaching electrical contacts to the semiconductors between which a p–n junction is formed, a semiconductor device called a **p–n diode** is obtained. In practice, p–n junctions are most often fabricated by introducing donors into a p-type semiconductor or by introducing acceptors into an n-type semiconductor. That is, on both sides of the junction we have the same semiconductor (most often silicon or germanium), but with different types of impurities. Below, we briefly describe the properties of a p–n junction.

5.7.1. Built-in potential

In an n-type semiconductor, the electron concentration is much higher than in a p-type semiconductor, whereas in a p-type semiconductor, the hole concentration is much higher than in an n-type semiconductor. Therefore, diffusion of electrons from the n-region into the p-region and diffusion of holes from the p-region into the n-region occur across the p–n junction. Correspondingly, in the region of the n-type semiconductor adjacent to the contact plane, the electron concentration decreases, while in the region of the p-type semiconductor adjacent to the contact plane, the hole concentration decreases (see Fig. 26a). Meanwhile, the donor and acceptor ions are immobile: they occupy crystal lattice sites, and their concentration on both sides of the junction remains unchanged (see Fig. 26b). Consequently, an uncompensated space charge of impurity ions arises in the junction region: a positive space charge of donor ions on the n-type semiconductor side and a negative space charge of acceptor ions on the p-type semiconductor side (see Fig. 26c). This space charge creates an electric field that opposes further redistribution of charge carriers. That is, alongside charge carrier diffusion, charge carrier drift occurs in the opposite direction. Under conditions of thermodynamic equilibrium, the diffusion and drift carrier fluxes compensate each other exactly, and the net current is equal to zero.

Thus, a potential difference U_b, called the **contact potential difference** or **built-in potential**, arises between the n and p regions. In silicon p–n junctions, the built-in potential is close to 0.7 V, whereas in germanium it is approximately 0.3 V.

When an electric field is present, the electron energy levels change. This is because the expression for the total energy of an electron acquires a term $-e\varphi$, where φ is the electric field potential. This term represents the potential energy of an electron in an electric field. Therefore, the bottom edge of the conduction band E_c and the top edge of the valence band E_v (like any other electron energy level

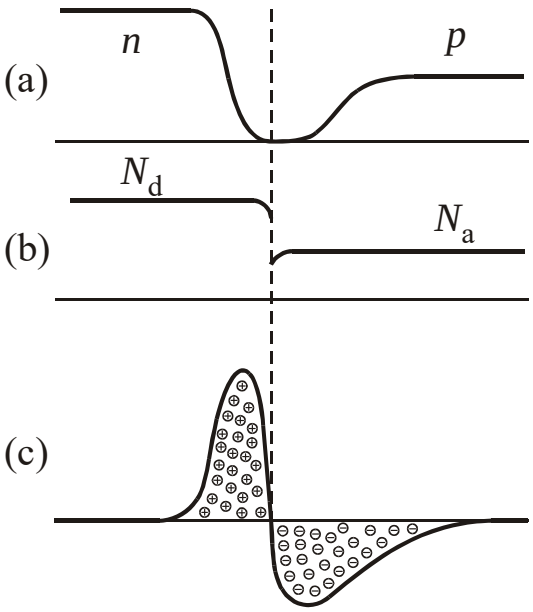


Fig. 26. Coordinate dependence of the properties of an equilibrium p–n junction: (a) electron and hole concentrations; (b) ionized impurity atom concentration; (c) space charge density.

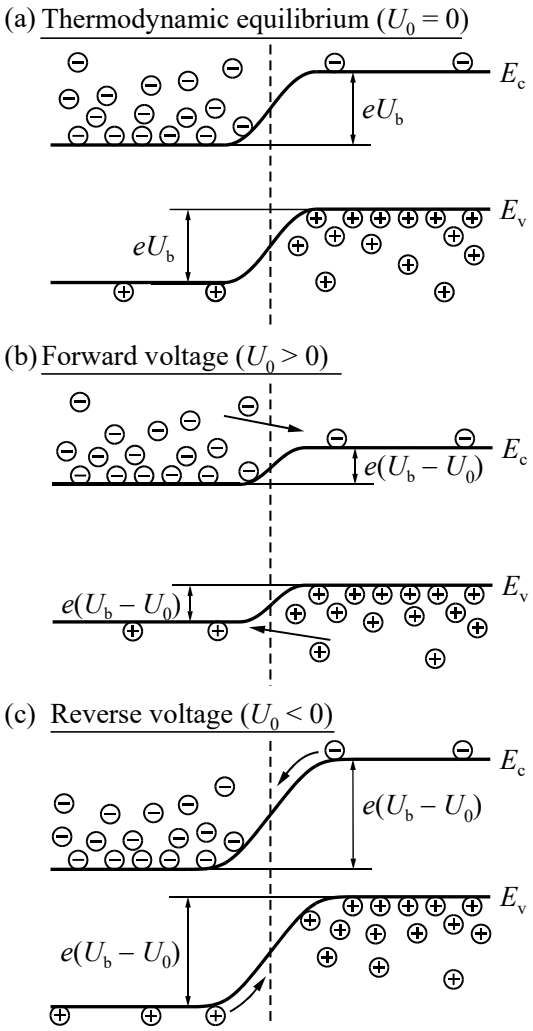


Fig. 27. Energy band diagrams of a p–n junction.

in a semiconductor) can be expressed as follows:

$$E_c = E_{c0} - e\varphi, \tag{5.7.1a}$$

$$E_v = E_{v0} - e\varphi, \tag{5.7.1b}$$

where E_{c0} and E_{v0} are the conduction and valence band edges in the absence of an electric field. When an electric field is present, the potential φ depends on the spatial coordinate, so the electron energy levels also begin to depend on the coordinate. This dependence has the shape of an "inverted" coordinate dependence of the potential. As shown in Fig. 27a, due to the built-in potential, a **potential barrier** exists in the p–n junction region, which must be overcome by electrons moving from the n-region into the p-region and by holes moving from the p-region into the n-region. The height of this potential barrier is equal to eU_b. Only those charge carriers whose kinetic energy is greater than the barrier height can pass over this barrier.

Since the kinetic energy of charge carriers is mentioned here, it is necessary to define the meaning of kinetic energy from the standpoint of the band model. The kinetic energy of a free electron or a hole is the difference between the energy level occupied by the charge carrier and the corresponding band edge. The kinetic energy of a free electron is measured upward from the bottom edge of the conduction band, whereas the kinetic energy of a hole is measured *downward* from the top edge of the valence band. This is because a hole is a quasiparticle whose energy state must describe the energy state of a system of real particles—the electrons forming covalent chemical bonds. The reference direction for electron energy is directed "upward," i.e., as an electron's energy increases, it transitions to higher energy levels. This means that as the energy of an electron system increases, vacancies (unoccupied energy levels) "sink" downward. Therefore, when replacing bound electrons with free quasiparticles—holes—the reference direction for the kinetic energy of these quasiparticles must also be redefined. The reference direction for hole kinetic energy is directed "downward," i.e., as a hole's kinetic energy increases, it transitions to *lower* electron energy levels.

5.7.2. Depletion layer and its thickness dependence on external voltage

In the space-charge region, the free carrier concentration is much lower than in the neutral regions. This is because a strong electric field (on the order of 10^6 V/m) exists in this region, which very rapidly "sweeps out" from this region the charge carriers that arise due to thermal generation or enter from the neutral regions. Therefore, this region is called the **depletion layer** (or depletion region). In theoretical analysis, one can rely on the assumption that there are *no* free charge carriers at all in the space-charge region, i.e., charges exist only in the form of immobile donor and acceptor ions. Such an approximation is called the **depletion approximation**. Fig. 28 shows the coordinate dependences of the space charge density (a), electric field strength (b), and potential (c) of a p–n junction (the symbol "n^+" indicates that the donor concentration in the n-region is much higher than the acceptor concentration in the p-region). These dependences were obtained by applying the depletion approximation. Consequently, the space charge density has points of discontinuity, and the electric field strength has points where its slope changes abruptly (kinks). In reality, the distributions of these quantities are smooth (e.g., the coordinate dependence of the space charge density is more similar to Fig. 26c).

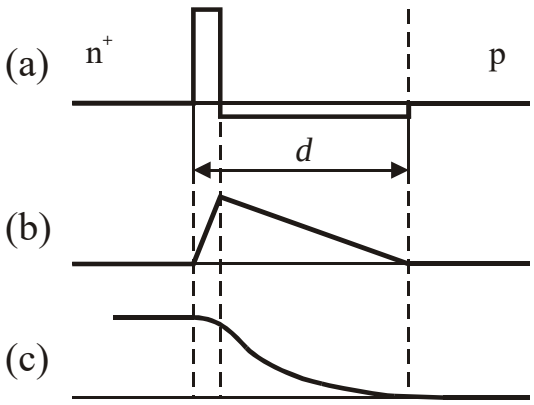


Fig. 28. Coordinate dependence of the space charge density (a), electric field strength (b), and potential (c) of a p–n junction in the depletion approximation, when the donor concentration in the n-region is much higher than the acceptor concentration in the p-region.

The potential barrier height and the depletion layer thickness change when a voltage source is connected to the p–n junction. Then the potential difference between the n and p regions becomes equal to $U_b - U_0$, and the potential barrier height becomes equal to $e(U_b - U_0)$; here U_0 is the external voltage (see Fig. 27b and Fig. 27c). If the negative terminal of the voltage source is connected to the n-region, a **forward bias** (forward voltage) is said to be applied to the p–n junction; in the opposite case, a **reverse bias** (reverse

voltage) is said to be applied. The depletion layer thickness d is equal to

$$d = \sqrt{\frac{2\varepsilon_0 \varepsilon (N_\mathrm{d} + N_\mathrm{a})}{eN_\mathrm{d}N_\mathrm{a}}(U_\mathrm{b} - U_0)}, \tag{5.7.2}$$

where N_d is the donor concentration in the n-region, N_a is the acceptor concentration in the p-region, ε_0 is the electric constant ($\varepsilon_0 = 8.854 \cdot 10^{-12}$ F/m), and ε is the relative permittivity (dielectric constant) of the semiconductor (for silicon $\varepsilon = 11.7$, for germanium $\varepsilon = 16$). If the external voltage is a forward bias, then $U_0 > 0$, and if it is a reverse bias, then $U_0 < 0$.

From Eq. (5.7.2), it follows that a reverse bias increases d, whereas a forward bias decreases d. This is also evident from Fig. 27 (the depletion layer is the region where the band edges depend on the spatial coordinate). This dependence of the depletion layer thickness on the external voltage is easy to understand considering that the resistance of the depletion layer is much greater than the resistance of the neutral regions. Therefore, the entire external voltage drops across the depletion layer, while the electric field in the neutral regions is practically equal to zero. This means that outside the boundaries of the depletion layer there is no space or surface charge that could create an electric field. Thus, even when an external voltage is applied, the sole source of the electric field remains the donor and acceptor ions in the depletion layer. Since the potential difference $(U_\mathrm{b} - U_0)$ between the neutral n and p regions is equal to the integral of the electric field strength with respect to the coordinate, it is clear that when the external voltage U_0 changes, the electric field strength in the depletion layer must change. Consequently, the total charge of donor and acceptor ions accumulated in the depletion layer must change. Since the ion concentration is constant, their number can change only when the thickness of the depletion layer changes.

Most often in practice, the doping of a p–n junction is strongly asymmetric, i.e., the concentration of one type of impurity is much higher than that of the other type. The thickness d of the depletion layer is determined by the smaller of the two impurity concentrations. This is because the total charge of acceptor ions in the depletion layer is always exactly opposite to the total charge of donor ions (the condition of electrical neutrality). As shown in Fig. 28a, as the acceptor concentration decreases, the portion of the depletion layer belonging to the p-region increases (so that the total negative charge remains the same). Therefore, when $N_\mathrm{a} \gg N_\mathrm{d}$, it can be stated that the *entire* depletion layer belongs to the n-region, and when $N_\mathrm{d} \gg N_\mathrm{a}$, it can be stated that the *entire* depletion layer belongs to the p-region. The latter case corresponds to Fig. 28. If $N_\mathrm{d} \gg N_\mathrm{a}$, then the term N_a in the numerator of Eq. (5.7.2) for the thickness d can be neglected. We obtain:

$$d \approx \sqrt{\frac{2\varepsilon_0 \varepsilon}{eN_\mathrm{a}}(U_\mathrm{b} - U_0)}. \tag{5.7.3}$$

5.7.3. Role of the depletion layer in a semiconductor detector

Since there are almost no free charge carriers in the depletion layer of a p–n junction (e.g., see Fig. 26a), its resistivity is high. Thus, by using a p–n junction instead of a homogeneous semiconductor, the main drawback of semiconductors—low resistivity—is eliminated, while preserving all the advantages of semiconductors mentioned in Section 5.5.

The electric field existing in the depletion layer makes it possible to detect ionizing particles even when the external voltage is equal to zero: free charge carriers generated by the action of ionizing radiation move in this field and create a signal in the external circuit. In this case, the neutral portions of the n and p regions, where the charge carrier concentration is high, act as electrodes. However, in the absence of an external voltage source, the thickness of the depletion layer is very small ($< 10^{-4}$ cm). This means, first, that the sensitive volume of the detector is small, and second, that the capacitance of the sensitive region of the detector is large (it is inversely proportional to the thickness of the depletion layer). A large capacitance means that the signal amplitude is small (see Eq. (5.2.3)). Therefore, the thickness of the depletion layer of a semiconductor detector must be as large as possible. According to Eq. (5.7.3), the thickness d of the depletion layer can be increased in two ways: by increasing the reverse bias U_0 and by decreasing the impurity concentration in the lightly doped region. Therefore, in semiconductor detectors, a relatively high reverse bias (around 1000 V) is applied to the p–n junction, and the role of the lightly doped region is played by an almost intrinsic (or almost exactly compensated) semiconductor. Since the reverse bias is always much greater than the built-in potential U_b, the term U_b in Eq. (5.7.3) for the depletion layer thickness can

be neglected:

$$d \approx \sqrt{\frac{2\varepsilon_0 \varepsilon}{eN_a}|U_0|}. \tag{5.7.4}$$

From Eq. (5.6.1), it follows that in thermodynamic equilibrium (when the external voltage is equal to zero), the total concentration of electrons and holes in the depletion layer cannot become smaller than the carrier concentration in an intrinsic semiconductor. Thus, as long as there is no external reverse bias, a p–n junction has no advantage over an intrinsic semiconductor of the same thickness. However, when a reverse bias is applied, another property of a p–n junction manifests itself: in the n-region the hole concentration is very low, and in the p-region the electron concentration is very low (much lower than in an intrinsic semiconductor). Therefore, the current caused by the passage of electrons from the p-region into the depleted layer and the passage of holes from the n-region into the depleted layer ("extraction" of minority carriers by the electric field) is very small, and the total current is determined by the charge carriers that are thermally generated in the depletion layer. This current ("reverse current") is relatively small and depends weakly on the reverse bias. Meanwhile, the current pulse that arises due to the absorption of a particle (e.g., an X-ray photon) in the depletion layer depends strongly on the reverse bias U_0: the amplitude (height) of that pulse is proportional to $|U_0|$, the pulse duration (width) is inversely proportional to $|U_0|$, and the integral (area) does not depend on $|U_0|$, because it is determined by the total number of electron-hole pairs created by the absorption of the particle, which depends only on the energy of the absorbed particle. Therefore, by increasing the reverse bias, one can achieve a pulse amplitude much greater than the constant reverse current, thereby improving the "signal-to-noise ratio."

5.7.4. Depletion layer formation technology. Si-PIN detectors

One of the drawbacks of the depletion layer of a p–n junction is that the electric field inside it is non-uniform (in Fig. 28b it is evident that the coordinate dependence of the field has a triangular shape). Moreover, the increase in the depletion layer thickness with increasing reverse bias (see Eq. (5.7.4)) slows down the growth of the electric field, i.e., the electric field strength grows more slowly than linearly (more precisely, the maximum electric field is proportional to $|U_0|^{1/2}$). To obtain a uniform electric field and a constant depletion layer thickness, the role of the depletion layer in semiconductor spectrometers is most often played by a layer of intrinsic semiconductor to which electrodes satisfying the following requirement are attached: the positive electrode must not inject holes, and the negative electrode must not inject electrons into the semiconductor. In other words, each electrode must be "blocking" for the charge carriers on which the electrostatic force directed from the electrode into the semiconductor acts (otherwise we would have a so-called "ohmic" contact, which presents no barrier to the movement of charge carriers through it). Then the reverse current is purely generation-driven and sufficiently small (see the last paragraph of Section 5.7.3). For example, in silicon detectors, such "blocking" contacts are formed by making one electrode out of aluminum and the other out of gold, and connecting the positive terminal of the voltage source to the aluminum electrode and the negative terminal to the gold electrode. The role of a blocking contact can also be played by a very thin surface layer of doped semiconductor, which can be formed, for example, by implanting the required impurity ions using an accelerator or by introducing them via thermal diffusion (in this way, a thin n-type region is formed on one surface of the semiconductor layer, and a thin p-type region on the other side). This type of three-layer semiconductor structure is called a "PIN junction" (the letter "I" comes from the term "intrinsic"). Silicon photodetectors based on a PIN junction are called "Si-PIN detectors." In manufacturing a Si-PIN detector, an intrinsic Si wafer with an impurity concentration of only 10^{11}–10^{12} cm^{-3} and a thickness of (100–300) μm is used. Phosphorus ions are introduced near one surface of that wafer (thus forming a thin n-type region near that surface), and boron ions near the other surface (thus forming a thin p-type region near that surface). At a sufficiently high reverse bias U_0, the electric field in the intrinsic semiconductor layer is almost uniform, and its strength is approximately equal to

$$\mathscr{E} \approx \frac{|U_0|}{d}, \tag{5.7.5}$$

where d is the thickness of the intrinsic semiconductor layer. The reverse bias used in Si-PIN detectors is (50–200) V.

5.8. Effect of Ionizing Radiation on a Semiconductor

5.8.1. Average ionization energy

When a charged particle passes through a semiconductor, electron–hole pairs are created along the particle's track. These charge carriers may be formed due to direct interaction with the particle or due to interaction with high-energy electrons that the particle liberated from the valence and deeper bands. Regardless of the physical mechanism by which the electrons and holes are created, the most important quantity for practical applications of detectors is the average energy loss of the incident particle corresponding to a single created electron–hole pair. This energy, which we will denote by the letter W, is practically independent of the energy and type of incident radiation[4]. Therefore, from the number of electron–hole pairs, one can determine the energy of the particle (provided the particle was completely stopped or absorbed within the active volume of the detector). The energy W is roughly 4 times larger than the width of the energy gap, i.e., approximately $4E_g$. This is because a large fraction of the energy loss is expended not on creating electrons and holes, but on other processes (e.g., exciting atomic vibrations).

The number of holes is always equal to the number of electrons, regardless of whether the semiconductor is intrinsic or doped. Furthermore, the number of created free charge carriers, the particle range, and other quantities that reflect the interaction of radiation with matter are practically independent of the semiconductor conductivity type (n or p) and of the impurity concentration in it, because that concentration is always much smaller than the concentration of semiconductor (e.g., silicon or germanium) atoms. Therefore, for example, the interaction of γ radiation with intrinsic, n-type, and p-type silicon layers of the same thickness will be completely identical.

The main advantage of semiconductor detectors is the small energy W defined earlier. For silicon or germanium, the value of W is approximately 3 eV (see Table 2 in Section 5.6.1), i.e., roughly 10 times smaller than the analogous parameter for gases. This means that for the same particle energy loss, the number of electron–hole pairs created in a semiconductor is roughly 10 times larger than the number of ion pairs created in gases. A larger number of charge carriers is advantageous in two respects. First, the limiting (minimum) energy resolution, which is inversely proportional to the square root of the number of created charge carriers, decreases. This factor is most important at intermediate and high particle energies (> 100 keV), because then the actual energy resolution is often close to the limiting one. Second, the pulse amplitude increases, which is proportional to the created charge (see Eq. (5.2.3)). This factor is most important for lower-energy radiation, when the energy resolution is limited by the electronic noise of the amplifier. As the number of charge carriers increases, the signal-to-noise power ratio increases, and the energy resolution improves.

5.8.2. Dynamics of creation and disappearance of free charge carriers

The first stage of the interaction of an ionizing particle with a semiconductor is shown in Fig. 29a. In this stage, electrons appear in unoccupied energy bands, and vacancies (holes) appear in occupied bands. This stage lasts approximately 10^{-12} s (this is the typical time over which a particle transfers its energy to the material).

In the second stage, the fast electrons and holes lose kinetic energy, i.e., electrons "drop" to the bottom of the conduction band, while holes "float up" to the top of the valence band. Vacancies that appear in deeper bands are filled due to electron transitions from higher energy bands. During this process, photons are created that can be absorbed in the material, generating additional electron–hole pairs. In addition, electrons in the conduction band and holes in the valence band possess high kinetic energy, part of which they

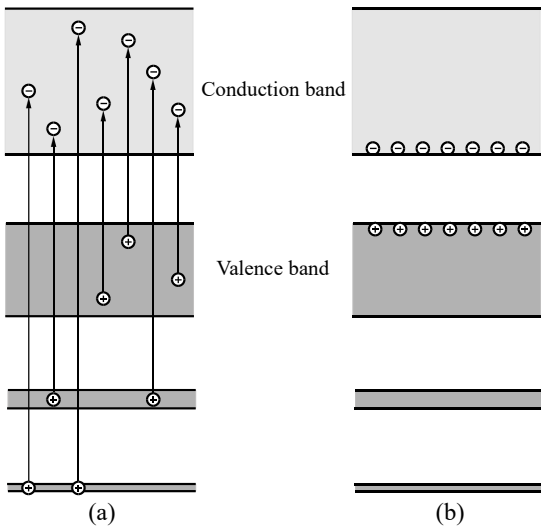


Fig. 29. Energy diagram of electron–hole pair creation in a semiconductor due to interaction with a charged particle.

[4] Precise measurements show that W depends on the type of radiation, although this dependence is weak. For example, the relative difference between the W values for α particles and protons in silicon does not exceed 2.2 %.

lose by creating further electron–hole pairs. This is the so-called **impact ionization**. Thus, in the second stage, the number of free electrons and holes increases. At the end of this stage, the energy distribution of all electrons and holes is the same as under conditions of thermodynamic equilibrium, i.e., electrons occupy states at the bottom edge of the conduction band, and holes occupy states at the top edge of the valence band (see Fig. 29b). This process is called **carrier thermalization** (the term "thermalization" refers to the transition of particles into a state of thermal equilibrium). The duration of this stage is also on the order of 10^{-12} s.

The third stage is the recombination of the created excess charge carriers. Over time, the number of excess charge carriers decreases exponentially:

$$N = N_0 \exp(-t/\tau_r); \tag{5.8.5}$$

here N_0 is the number of charge carriers created at the instant $t = 0$, and τ_r is the **mean carrier lifetime**. The physical meaning of τ_r is the average time from the creation of a charge carrier until its disappearance due to recombination.

Charge carriers most often recombine not directly, but via **recombination centers**—energy levels located deep within the energy gap that can "capture" electrons from the conduction band or holes from the valence band. Since such a level is far from the band edges, the probability of thermal excitation of the captured charge carrier back into the conduction or valence band is small. That charge carrier can only recombine with a free charge carrier of the opposite sign.

The value of τ_r depends on the temperature and the number of recombination centers. Obviously, a semiconductor detector can operate well only when the charge collection time at the electrodes is much shorter than τ_r (so that the number of charge carriers does not have time to decrease noticeably due to recombination while the charge carriers move toward the electrodes). Therefore, semiconductor detectors use materials whose τ_r is relatively large. For example, for good-quality germanium, τ_r = (50–200) μs.

5.9. Detector Efficiency

Detector efficiency is the ratio of the number of recorded particles to the number of particles striking the detector. More precisely, this defines the *intrinsic detector efficiency*. The ratio of the number of recorded particles to the total number of emitted particles is called the *absolute detector efficiency*. Absolute efficiency is used less frequently than intrinsic efficiency because absolute efficiency depends on the measurement geometry (e.g., on the distance between the radiation source and the detector). Intrinsic efficiency depends only on the internal properties of the detector. For example, when the distance between the radiation source and the detector is changed, the intrinsic efficiency does not change (the numbers of recorded particles and particles striking the detector change by the same factor, so their ratio remains unchanged). Hereafter, only intrinsic efficiency is discussed and the word "intrinsic" is omitted.

If the active material of the detector has the shape of a flat layer, and the beam of X-ray or gamma radiation is perpendicular to its surface, then the detector efficiency can be calculated approximately using the concept of the attenuation coefficient μ (see Section 4.5). According to the exponential attenuation law expression (4.5.3), the fraction of photons that did not interact with a layer of material of thickness d is equal to $e^{-\mu d}$, where d is the layer thickness. Consequently, the fraction of photons that interacted with the layer (the detector efficiency) is equal to

$$\varepsilon = 1 - e^{-\mu d}. \tag{5.9.1}$$

Since the attenuation coefficient μ increases as the photon energy decreases (see, e.g., Fig. 11b), the efficiency ε also increases. However, in practice, absorption always occurs in a "dead" layer of the detector, i.e., in materials located between the active material layer and the detector surface onto which the radiation falls. This may be the material of the detector housing or window, the neutral semiconductor layer (e.g., the n or p region of a Si-PIN detector), and so on. Absorption in the dead layer reduces the detector efficiency. Moreover, as the photon energy decreases, absorption in the dead layer increases, so the fraction of photons reaching the active material decreases. This latter fraction is an additional factor in the detector efficiency expression (5.9.1) that is not accounted for by that formula. Therefore, the dependence of detector efficiency on photon energy has a maximum and approaches zero as the photon energy becomes greater than the energy of the maximum (see Fig. 30). It is evident that the detector whose efficiency is shown in Fig. 30 is unsuitable for detecting photons whose energy is less than 2 keV. The upper energy limit of photons recorded

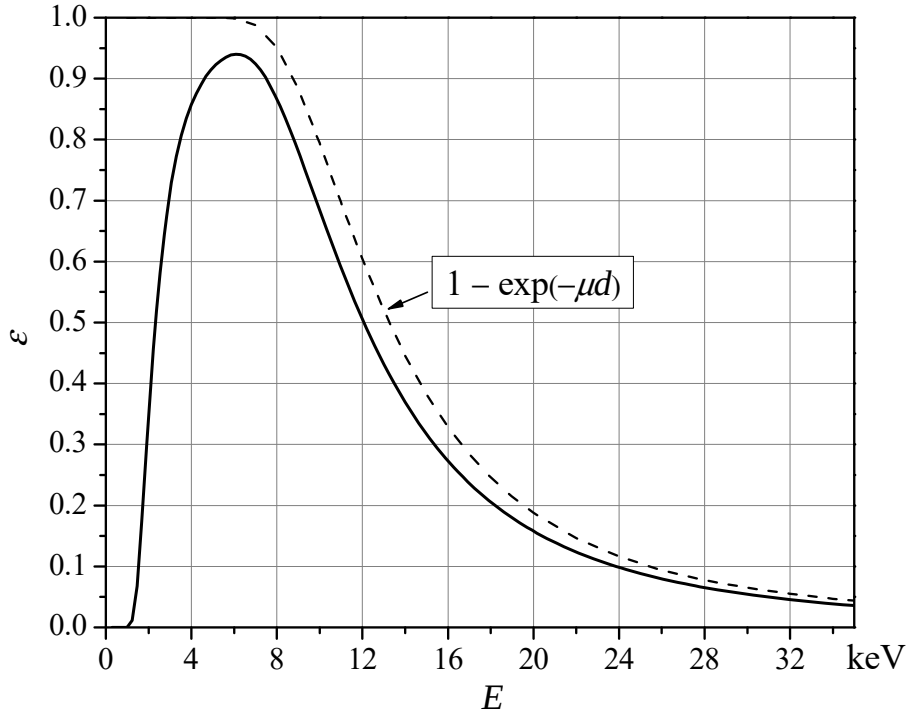


Fig. 30. Dependence of the efficiency of the Si-PIN detector used in this experiment on photon energy (solid line) and the theoretical curve calculated according to Eq. (5.9.1) (dashed line). The thickness of the detector's "i" region (intrinsic semiconductor layer) is d = 200 μm = 0.02 cm, and the dependence of the silicon attenuation coefficient μ on E is given in Fig. 11b.

by this detector is approximately 60 keV, although detecting photons with energies greater than 35 keV is practically difficult due to the very low efficiency.

6. Diffraction of X-Rays in Crystals

When X-ray radiation with photon energy on the order of ~10 keV or less interacts with matter, a relatively important role is played by coherent scattering. In this energy range, the coherent scattering cross section is much smaller than that of the photoelectric effect, yet larger than or of the same order of magnitude as the Compton scattering cross section (see, e.g., Fig. 11). Coherent scattering is special in that waves scattered by different atoms can "reinforce" one another through interference. That is, the wave properties of photons manifest themselves here (unlike in the photoelectric effect and Compton scattering, where the interaction process can be described under the assumption that a photon is a point particle). It should be kept in mind that interference occurs not between different photons, but between components of the wavefunction of a *single* photon corresponding to its scattering at different atoms (the superposition principle). The interference of waves scattered at different points in space is called **diffraction**. Although quantum mechanics fully explains all regularities of photon diffraction, such an explanation is more complicated than, for example, explaining the diffraction of material particles, because the concept of spatial probability density cannot be applied to photons. That is, the question "what is the probability that at time t a photon was located in a given spatial volume element dV?" is meaningless. One can only speak of the probability of *detecting* a photon in a given volume element, which depends on the spatial distribution and time variation of the photon's electromagnetic field. It is simpler to explain the diffraction of electromagnetic radiation using classical electrodynamics.

In an electromagnetic wave, the electric field varies in space and time according to a harmonic law:

$$\mathscr{E}(x,t) = \mathscr{E}_0 \cos\left[2\pi\left(\frac{x}{\lambda} - \nu t\right)\right],$$

where x is the coordinate along the direction of wave propagation, t is time, λ is the wavelength, and ν is the frequency, which is equal to $\nu = c/\lambda$, where c is the speed of light. Interference occurs because, when two waves are superposed, it may happen that at certain points in space the electric fields of both waves have the same direction, i.e., the waves reinforce each other, whereas at other points in space the electric fields have

opposite directions, i.e., the waves weaken each other. Diffraction of visible light can be investigated using a diffraction grating (a system of identical parallel narrow slits). When a diffraction grating is illuminated with monochromatic light and a screen is placed behind the grating, instead of a geometrically exact image of the slits, equidistant parallel bright and dark fringes are obtained, the width of which depends on the wavelength and on the distance between adjacent slits (the period of the diffraction grating). Therefore, knowing the period of the grating and having measured the width of the diffraction fringes, one can determine the wavelength. And conversely: knowing the wavelength, one can determine the period of the grating from the width of the diffraction fringes. However, a sharp diffraction pattern is formed only when the period of the grating is of the same order of magnitude as the wavelength: if the period is much larger than the wavelength, a geometrically exact image of the slit system is obtained, whereas if it is much smaller, uniform illumination is obtained. The wavelength of characteristic X-ray radiation is on the order of angstroms (1 Å = 1 angstrom = 10^{-10} m), i.e., on the order of atomic dimensions. It would be difficult to fabricate a diffraction grating with such a period. Fortunately, natural diffraction gratings exist in nature whose period is on the order of angstroms. These are **crystals**—solids in which the arrangement of atoms, molecules, or ions in space is ordered and periodic. Diffraction of X-rays in crystals began to be studied in the early second decade of the 20th century. Some regularities of this diffraction are described below.

6.1. Crystals, crystal lattice, unit cell

A **crystal** is a solid whose structural elements (atoms, molecules, or ions) are arranged in an orderly manner throughout all space. This orderliness of the atomic arrangement manifests itself in the fact that crystals have the shape of regular polyhedra. Single crystals are called **monocrystals** (single crystals), while bodies composed of a large number of small crystals arranged randomly in space are called **polycrystals**. For example, the structure of metals is most often polycrystalline.

If we assign a point in space to each structural element of a crystal, the set of such points forms a crystal lattice. Thus, a **crystal lattice** is the set of points arranged in infinite space (for convenience of analysis), each of which corresponds to a structural element of the crystal. These points are called **lattice points**, and the structural element of the crystal is called the **basis** of the lattice.

For example, the structural elements of a sodium chloride (NaCl) crystal are identically oriented pairs of adjacent Na and Cl atoms (see Fig. 31a). Thus, the basis of the NaCl crystal lattice is a pair of Na and Cl atoms. To construct the lattice of this crystal, a point in space must be assigned to each such pair. The only requirement is that the position of this point relative to the structural element be identical for all structural elements. For example, the lattice points may be chosen at the Cl atoms. The corresponding crystal lattice is shown in Fig. 31b. However, the lattice points can equally well be chosen at the Na atoms, at the centers of identically oriented pairs of adjacent Na and Cl atoms, and so forth.

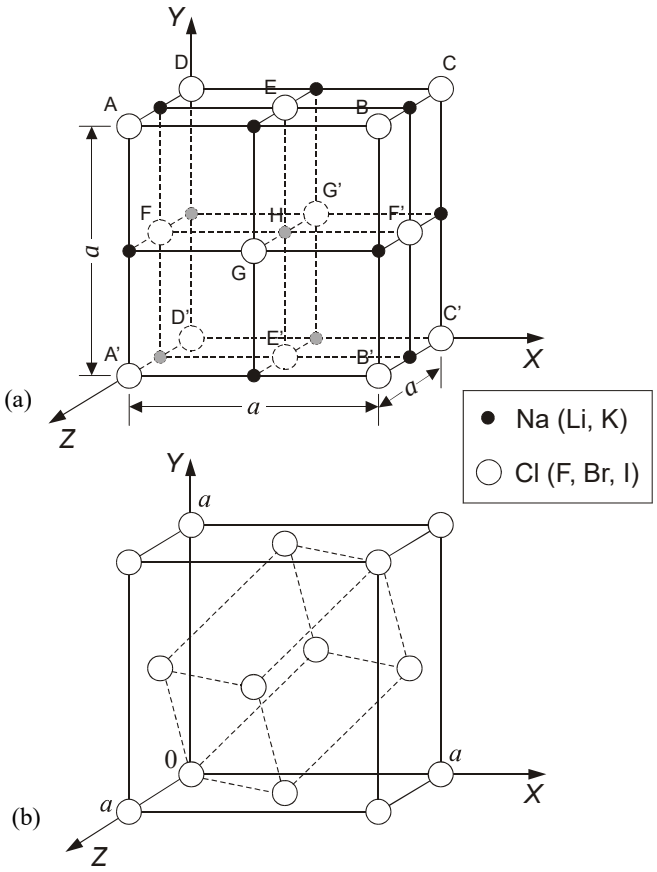


Fig. 31. (a) Arrangement of atoms in alkali metal halide crystals (e.g., NaCl, LiF, etc.). (b) Face-centered Bravais lattice unit cell (solid lines) and primitive cell (dashed lines).

The smallest distance along a given direction between two lattice points is called the **lattice constant** or **lattice period** along that direction. For example, for the NaCl crystal lattice shown in Fig. 31b, the periods along the X, Y, and Z directions are identical and equal to $a = 5.64 \cdot 10^{-10}$ m = 5.64 Å. The spatial region inside a parallelepiped formed on the basis of three lattice periods is called a **unit cell**. For example, the cube shown in Figs. 31a and 31b is

a unit cell. The parallelepiped shown by dashed lines in Fig. 31b is also a unit cell. Thus, the edge lengths of a unit cell (the lattice periods) and the angles between them depend on the chosen directions. A unit cell of minimum volume is called a **primitive cell** (or **simple cell**). For example, the unit cell shown by dashed lines in Fig. 31b is primitive.

If any point of a crystal lattice is chosen, the arrangement of all remaining lattice points around it is identical, regardless of the chosen point. Therefore, a single unit cell provides complete information about the structure of the crystal: the entire crystal can be constructed by placing identical unit cells side by side.

A unit cell is always constructed so that a lattice point lies at each of its vertices. Consequently, a unit cell always contains at least one lattice point, because a parallelepiped has eight vertices, and each lattice point located at a vertex of the cell belongs to eight adjacent cells (i.e., $1/8$ of a lattice point belongs to one cell). In a primitive cell, lattice points are located only at the vertices of the cell, i.e., a primitive cell contains only one lattice point. In general, lattice points can be located not only at the vertices of the cell, but also inside the cell or on the interior regions of its faces (as in Fig. 31b). A lattice point located in the interior region of a cell face belongs to the boundary between two adjacent cells (i.e., $1/2$ of a lattice point belongs to one cell). Therefore, in general, the number of lattice points in a unit cell is calculated using the formula

$$N = 1 + \frac{1}{2}N_\mathrm{p} + N_\mathrm{t}, \tag{6.1.1}$$

where N_p is the number of lattice points on the surface of the cell (more precisely, in the interior regions of its faces), and N_t is the number of lattice points inside the volume of the cell. When the NaCl unit cell is constructed as shown in Fig. 31b, $N_\mathrm{p} = 6$ (six lattice points in the interior regions of the cell faces), and $N_\mathrm{t} = 0$ (there are no lattice points inside the volume of the cell). According to Eq. (6.1.1), in this case the unit cell contains 4 lattice points, i.e., 4 Na atoms and 4 Cl atoms.

6.2. Diffraction of waves in a crystal

The primary cause of diffraction is the coherent scattering of waves in matter (**coherent scattering** is scattering during which the wave frequency does not change). When a wave of a definite direction strikes a crystal, **scattered waves** of the same frequency arise, propagating in various directions. Waves scattered at different points of the crystal are coherent, i.e., they have a constant phase difference that depends on the scattering direction, and they have the same polarization. As is known from wave optics, coherent waves interfere, i.e., when two coherent waves are superposed, the total intensity can differ from the sum of the intensities of the two waves depending on their phase difference. When calculating the directions of interference maxima and minima of waves scattered in a crystal, it suffices to consider only waves scattered at *equivalent* points of the crystal (i.e., points whose surroundings are identical), because the amplitudes of waves scattered at those points are identical, so they can exactly "cancel" each other due to interference. The set of equivalent points of a crystal (e.g., atoms of a single species) forms the so-called **crystal lattice**, and those points are called **lattice points**. If at all lattice points the phases of the waves scattered in a given direction coincide (more precisely, differ by $2\pi n$, where n is an integer), then the amplitudes of the waves scattered in that direction from all points add up, i.e., the waves "reinforce" one another due to mutual interference. Such directions are called **directions of diffraction maxima**. For example, in Fig. 32, the electric field direction of the electromagnetic wave $\mathrm{A_1B_1}$ scattered at lattice point $\mathrm{A_1}$ coincides at all instants of time with the electric field direction of the wave $\mathrm{A_2B_2}$ scattered in the same direction at the adjacent lattice point $\mathrm{A_2}$ (the phase difference is equal to 2π), so both waves $\mathrm{A_1B_1}$ and $\mathrm{A_2B_2}$ reinforce each other. Since all lattice points are spaced at equal intervals from one another, at the next lattice point $\mathrm{A_3}$ the amplitude of the wave $\mathrm{A_3B_3}$ scattered in the same direction also adds to the amplitude of wave $\mathrm{A_1B_1}$, and so forth. Therefore, the direction $\mathrm{A_1B_1}$, specified by the angle α, corresponds to a diffraction maximum.

It should be noted that the condition for a diffraction maximum is a condition on the angle α (see Fig. 32). That is, it does not matter which points of the crystal lattice are treated as scattering centers; what matters is only that the scattering direction satisfies the condition for a diffraction maximum. This is because the distance between the scattering region (i.e., the part of the crystal onto which the primary radiation falls) and the point of intersection of the beams of waves scattered at different atoms where interference of the scattered waves is observed (e.g., a photographic film) is always much larger than the dimensions of the scattering region. Thus, although those scattered beams are not strictly parallel (otherwise they would not intersect), this deviation from parallelism is so small that it can usually be neglected. Therefore, when investigating

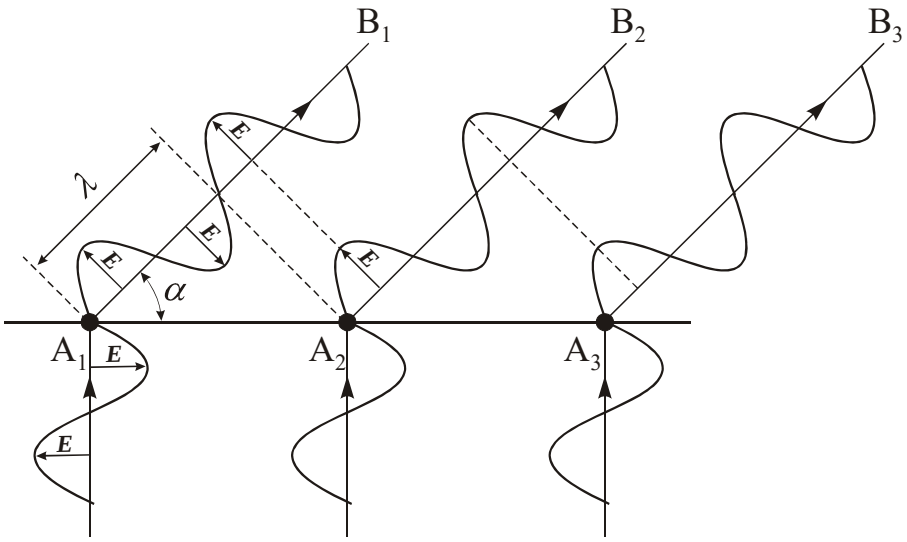


Fig. 32. Formation of a diffraction maximum.

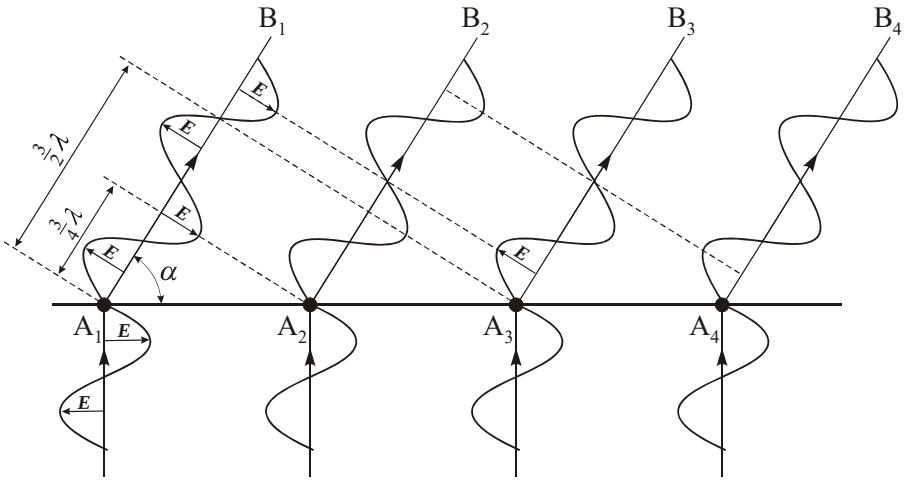


Fig. 33. Formation of a diffraction minimum.

diffraction in crystals, one can state that all rays intersecting at a given point are parallel, i.e., propagate in the same direction as shown in Fig. 32.

Usually, there are several directions of diffraction maxima. For all other directions, for every lattice point one can find another lattice point at which the phase of the scattered wave is opposite, i.e., all lattice points can be grouped into pairs such that in each pair the waves scattered in the given direction cancel each other upon superposition. For example, in the case of Fig. 33, along the direction A_1B_1, the electric field vector of the electromagnetic wave scattered at lattice point A_1 is at all instants opposite to the electric field vector of the wave scattered in the same direction at lattice point A_3, so both scattered waves weaken each other. Analogously, the waves A_2B_2 and A_4B_4 also cancel each other upon superposition, and so forth. Therefore, along the direction A_1B_1, specified by the angle α, zero radiation intensity is observed in this case (**diffraction minimum**). From this explanation, it is evident that the most important factor in diffraction is the periodic spatial arrangement of identical atoms. If the atoms were arranged randomly, then regardless of the chosen scattering direction and regardless of how the atoms were grouped into pairs, in some pairs of atoms the waves would weaken one another, while in others they would reinforce one another, so the scattered radiation would be uniformly distributed in all directions (without sharp maxima and minima).

6.3. Bragg's law

As mentioned, the diffraction pattern formed due to the diffraction of waves in a crystal is a consequence of the interference of scattered spherical waves whose centers coincide with the lattice points. However, we obtain the exact same pattern if we assume that the incident wave is *reflected* from crystallographic planes (this is the name given to planes in which crystal lattice points are located). Using this visualization,

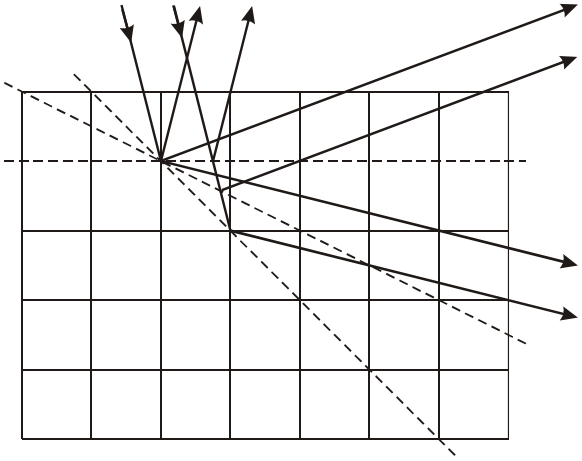


Fig. 34. Reflection of a wave from planes in which crystal lattice points are located (crystallographic planes). The lattice points are represented by grid vertices, and the crystallographic planes by dashed lines.

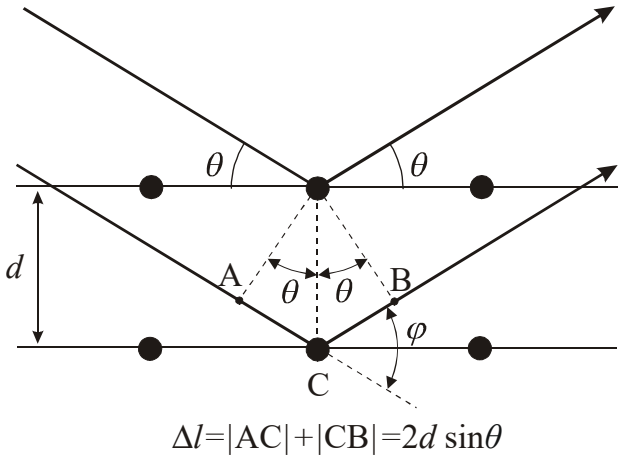


Fig. 35. Calculation of the path difference Δl of waves reflected from parallel crystallographic planes: θ – glancing angle, φ – scattering angle ($\varphi = 2\theta$).

a diffraction pattern is a consequence of the interference of waves reflected from various crystallographic planes. Through any point of a crystal lattice, many crystallographic planes can be drawn (see Fig. 34), and each plane will reflect the wave in such a direction that the angle of reflection equals the angle of incidence. However, one must take into account that reflection in a given direction occurs not from a single crystallographic plane passing through a given lattice point, but from all other crystallographic planes that are parallel to that plane. It may happen that for every crystallographic plane that reflects the wave in a given direction, one can find another parallel plane from which the reflected wave has the opposite phase (more precisely, the phase difference between these two reflected waves is equal to $(2n+1)\pi$, where n is an integer). Then all waves reflected in the given direction cancel each other in pairs, and zero intensity is observed in that direction (a diffraction minimum). If the phase difference of waves reflected from adjacent parallel crystallographic planes is a multiple of 2π, then all waves reflected in that direction reinforce one another, and a diffraction maximum is observed. Since a phase change of 2π corresponds to a path length equal to one wavelength, the condition for a diffraction maximum can be formulated as follows: the directions of diffraction maxima are those directions for which the path difference between waves reflected from adjacent parallel crystallographic planes is equal to an integer number of wavelengths (see Fig. 35). The mathematical expression of this condition, derived in 1912 by the English physicist W. L. Bragg and in 1913 by the Russian physicist G. V. Wulff, is called **Bragg's law** (or the **Bragg–Wulff condition**):

$$2d \sin\theta = n\lambda \qquad (n = 0, 1, 2, \dots), \tag{6.3.1}$$

where d is the distance between adjacent planes in a given family of crystallographic planes, and θ is the angle between the incident ray and the crystallographic plane (see Fig. 35). The angle θ is called the **glancing angle** (or Bragg angle). The number n is called the **order of reflection**.

6.4. Symmetry of crystals

The symmetry of a body is characterized by **symmetry operations**—rotations, reflections, translations, and other operations of a body after which the body is in a position indistinguishable from its initial position. Bodies of finite dimensions do not possess translational symmetry. If a body (e.g., a crystal) is shifted by any distance, part of the body ends up in a region of space that was empty prior to the shift, while another region of space becomes empty; therefore, the position of the body after the shift is distinguishable from its position prior to the shift. Only an object of infinite dimensions, such as a crystal lattice, can possess translational symmetry.

With each symmetry operation is associated a geometric element called a **symmetry element**, e.g., an axis of rotation, a plane of reflection, a center of inversion, or a direction of translation.

The symmetry elements of a crystal are planes of symmetry, axes of symmetry, a center of inversion, and axes of rotary-reflection (improper rotation). The symmetry elements of a crystal (a body of finite dimensions) are simultaneously symmetry elements of the crystal lattice (an object of infinite dimensions). However, a crystal lattice also possesses axes of translation. In addition, a crystal lattice may possess glide planes and screw axes.

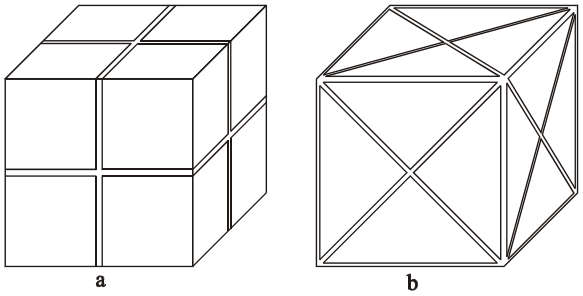


Fig. 36. Nine planes of symmetry of a cube (indicated by double lines), of which three are principal (a) and six are diagonal (b).

A **plane of symmetry** is a plane that divides a figure into two equal parts such that one part is the mirror image of the other (see Fig. 36).

An **axis of symmetry** is a straight line about which, when a figure is rotated by a certain angle, it comes into a position indistinguishable from its initial position (see Fig. 37). The number of coincidences of the figure's positions when rotated about the axis by an angle of 360° defines the order of the axis. In other words, the order of symmetry is the ratio of 360° to the smallest angle of symmetric rotation. For example, if the smallest angle of symmetric rotation is equal to 90°, then the order of the symmetry axis is equal to 4. The symmetry axes of a crystal (and of a crystal lattice) can only be of the second, third, fourth, and sixth orders.

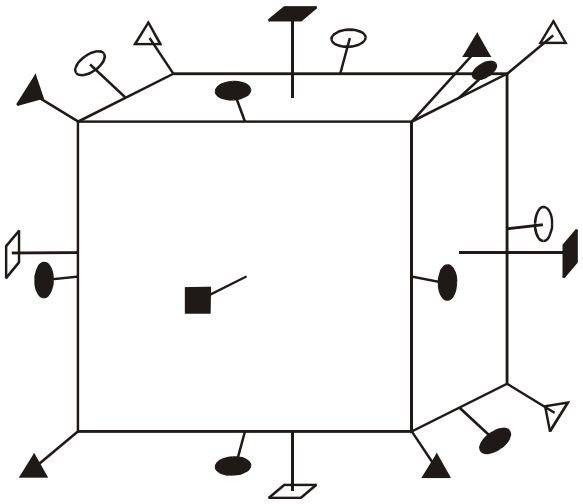


Fig. 37. Fourth- (■), third- (▲), and second-order (⬮) symmetry axes of a cube.

A **center of inversion** is a point through which, upon inverting the other points of a figure, the figure comes into a position indistinguishable from its initial position (see Fig. 38).

An **axis of rotary-reflection** (improper rotation axis) is a straight line about which, after rotating a figure by a certain angle and reflecting it in a plane perpendicular to the axis of rotation, the figure comes into a position indistinguishable from its initial position.

A **translation axis** is a direction along which, when a crystal lattice is shifted by a certain distance (the lattice period), the lattice comes into a position indistinguishable from its initial position.

A **glide plane** is a plane in which, upon reflecting a crystal lattice and shifting the lattice parallel to that plane by half a lattice period, the lattice comes into a position indistinguishable from its initial position.

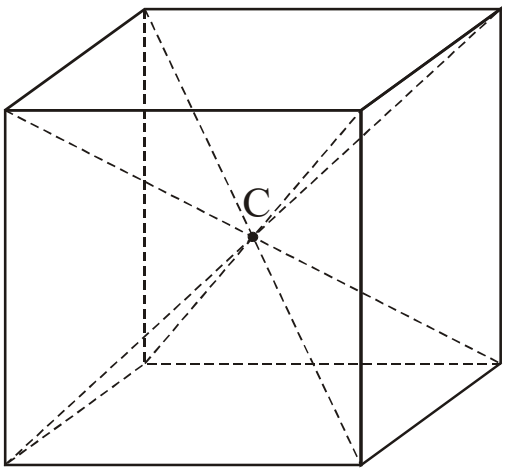


Fig. 38. Center of inversion of a cube.

A **screw axis** is a direction about which, upon rotating a crystal lattice by a certain angle and shifting it along that direction by a certain distance smaller than the lattice period, the lattice comes into a position indistinguishable from its initial position.

The first four symmetry operations defined above (reflection in a plane, rotation about an axis, inversion, and rotary-reflection) share one common property: during them, at least one point in space remains fixed. During inversion, the center of inversion does not move; during rotation, all points belonging to the axis of rotation do not move; and during reflection, all points belonging to the plane of symmetry do not move. Therefore, this type of symmetry is called **point symmetry**. The set of point symmetry operations that cannot be obtained by successively performing several other operations forms a **point symmetry group**. For example, the point symmetry elements of a cube (see Figs. 36–38) are: nine planes of symmetry (three parallel to the faces and six passing through the diagonals of opposite faces), three fourth-order axes of symmetry (passing through the centers of opposite faces), four third-order axes of symmetry (passing through

opposite vertices), six second-order axes of symmetry (passing through the midpoints of opposite edges), and a center of inversion. Any other point symmetry operation of a cube can be obtained by successively performing two or more of the 23 operations mentioned.

The point of intersection of all point symmetry elements is called the **center of symmetry**. The center of symmetry of a cube coincides with its center of inversion (see Fig. 38).

The point symmetry of any crystal is described by one of the 32 point symmetry groups. Accordingly, crystals are divided into 32 **crystal symmetry classes**. These 32 symmetry classes are, in turn, grouped into seven **crystal systems**. This grouping is based on the fact that the unit cell of a crystal can always be constructed so that the point symmetry elements of the unit cell coincide with the point symmetry elements of the entire crystal lattice. A unit cell of minimum volume whose point symmetry elements are simultaneously the point symmetry elements of the entire crystal lattice is called a **Bravais cell**, in honor of the 19th-century French physicist A. Bravais. Having constructed a unit cell in this manner, all possible forms of it can be classified according to the ratios of its edge lengths and the angles between its edges. This classification is presented in Table 3. Here the angle between edges a and b is denoted by γ, the angle between edges b and c by α, and the angle between edges a and c by β (see column II of Fig. 39).

Table 3. Crystal systems

Crystal system	Bravais cell edge lengths	Angles between Bravais cell edges
Triclinic	$a \neq b \neq c \neq a$	All angles unequal and not equal to 90°
Monoclinic	$a \neq b \neq c \neq a$	$\alpha = \beta = 90°$, $\gamma \neq 90°$
Orthorhombic	$a \neq b \neq c \neq a$	$\alpha = \beta = \gamma = 90°$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$
Trigonal	$a = b = c$	$\alpha = \beta = \gamma \neq 90°, 60°, 120°$
Hexagonal	$a = b$	$\alpha = \beta = 90°$, $\gamma = 60°, 120°$
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°, 60°$

The hexagonal crystal system differs from the other crystal systems in that its Bravais cell is not a parallelepiped, but a hexagonal prism (see Fig. 39).

The Bravais cell of a NaCl crystal is shown in Fig. 31b by solid lines. This cell has the shape of a cube, because a NaCl crystal belongs to the cubic crystal system.

In general, a Bravais cell is not a unit cell of minimum volume (i.e., it is not a primitive cell). This is because the point symmetry of a primitive cell is often lower than the point symmetry of the entire lattice (so it does not satisfy the requirements imposed on a Bravais cell). For example, the primitive cell of a NaCl crystal, shown in Fig. 31b by dashed lines, is not cubic and does not possess most of the point symmetry elements of the crystal lattice[5]. In such cases, a Bravais cell contains more than one lattice point. The additional lattice points may be located at the centers of the faces of the Bravais cell or at the center of the cell. According to the arrangement in the cell of the lattice points that do not lie at the vertices of the Bravais cell, four types of Bravais cells are distinguished:

1) **primitive Bravais cell** – lattice points are located only at the vertices of the cell;

2) **body-centered Bravais cell** – an additional lattice point is located at the center of the cell;

3) **base-centered Bravais cell** – additional lattice points are located at the centers of two opposite faces of the cell;

4) **face-centered Bravais cell** – additional lattice points are located at the centers of all faces of the cell.

Crystal lattices of each crystal system are classified according to the centering type of their Bravais cell. In this way, 14 Bravais lattices can be constructed, differing from one another either in crystal system or in centering type (this was proven in 1848 by A. Bravais, after whom these lattices are named). Four Bravais

[5] In the primitive cell shown by dashed lines in Fig. 31b, all edges have equal length ($a/\sqrt{2}$), yet the angles between them are not equal to 90°. Specifically, each of the six faces of this primitive cell is a rhombus with two 60° angles and two 120° angles. For example, consider the two faces that each have one vertex on the bottom face of the cubic Bravais cell and one on its top face: the interior angles at those two vertices are equal to 60°, whereas the angles at the other two vertices (located on the lateral faces of the Bravais cell) are equal to 120°.

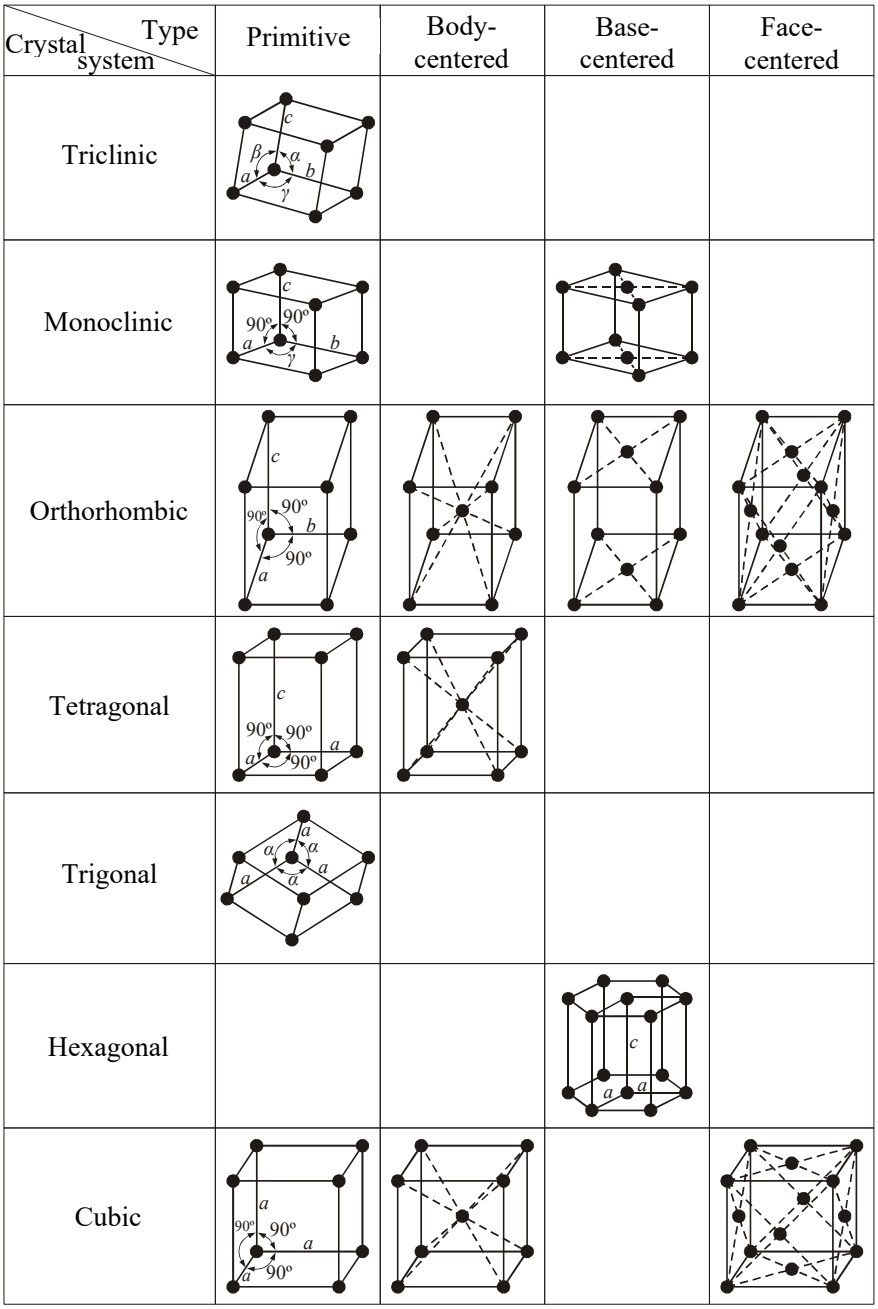


Fig. 39. Unit cells of the Bravais lattices.

lattices belong to the orthorhombic crystal system, three to the cubic system, two each to the monoclinic and tetragonal systems, and one each to the triclinic, trigonal, and hexagonal systems. All of these Bravais lattices are shown in Fig. 39. The Bravais lattice of alkali metal halides (e.g., NaCl) is face-centered cubic (see Fig. 31b). The Bravais lattices of most metals are of the same type. However, the basis of a metal lattice consists of only a single atom (rather than a pair of different atoms, as in alkali metal halides).

When speaking about the symmetry of a crystal lattice, one must also take into account symmetry operations during which no point in space remains fixed, i.e., translations, reflections with glide, and screw rotations. These symmetry operations together with the point symmetry operations form the **space symmetry group** of the crystal. In total, there are 230 space symmetry groups.

6.5. Miller indices of crystallographic planes and crystallographic axes

A **crystallographic plane** is any plane containing three non-collinear points of a Bravais lattice. All parallel crystallographic planes form a **family of crystallographic planes**. Each point of a crystal lattice belongs to an infinite number of different crystallographic planes (see the dashed lines in Fig. 34).

A **crystallographic axis** is any one of the three directions that are parallel to the edges of a Bravais cell.

In crystal structure analysis, crystallographic axes play the role of coordinate axes.

The orientation of any crystallographic plane can be specified by the three intercepts A, B, and C that the plane cuts off on the crystallographic axes (see Fig. 40a). The lengths of the intercepts A, B, and C can be expressed in terms of the quantities A/a, B/b, and C/c, where a, b, and c are the lattice periods (the edge lengths of the Bravais unit cell). In the case of Fig. 40a, these ratios are equal to 2, 8, and 4, respectively. In crystal structure analysis, it is more convenient to specify the orientation of a crystallographic plane not by the ratios A/a, B/b, and C/c, but by their reciprocals divided by their greatest common divisor. These three numbers are called the Miller indices of the crystallographic plane. Thus, the **Miller indices of a crystallographic plane** are three integers having no common divisor that are proportional to the reciprocals of the coordinates of the points where the crystallographic plane intersects the crystallographic axes. The Miller indices of a crystallographic plane are written in parentheses: (hkl). For example, the Miller indices of the plane shown in Fig. 40a are (412). If any Miller index is negative, a minus sign is written above the number, e.g., $(1\bar{1}1)$. For planes that are parallel to a crystallographic axis, the corresponding Miller index is equal to zero. For example, in Fig. 31a, the Miller indices of the plane $AB'C'$ are (110) (this plane is parallel to the Z axis), and the Miller indices of the plane $A'B'E$ are (012) (this plane is parallel to the X axis). The planes (100), (010), and (001), which are simultaneously parallel to two crystallographic axes, are called **principal crystallographic planes**. For example, in Figs. 31a,b, the principal crystallographic planes are parallel to the coordinate planes $x = 0$, $y = 0$, and $z = 0$.

Miller indices are used not only to specify the orientation of crystallographic planes, but also to specify a direction in a crystal (e.g., a symmetry axis). The **Miller indices of a direction** are found as follows. Through the origin of coordinates, a straight line parallel to the given direction is drawn, and the coordinates x, y, and z of any point belonging to that line are found (in the case of an oblique coordinate system, the coordinates of a point are the oblique projections of the point's position vector onto the coordinate axes, rather than the perpendicular projections). These coordinates are divided by the corresponding lattice periods. Those ratios x/a, y/b, and z/c are divided by their greatest common divisor. The resulting numbers are the Miller indices of the given direction. The Miller indices of a direction are written in square brackets: $[hkl]$. For example, the Miller indices of the direction shown in Fig. 40b are $[121]$. The Miller indices of the crystallographic axes are $[100]$, $[010]$, and $[001]$.

In a crystal of the cubic system, the Miller indices of a crystallographic plane coincide with the Miller indices of the direction perpendicular to it (this rule does not hold for crystals of other systems).

Using Miller indices, one can concisely specify a crystallographic plane or the direction of a symmetry axis. For example, the Miller indices of the planes and axes of symmetry of a crystal of the cubic system are:

- planes of symmetry – (100), (010), (001), (110), $(1\bar{1}0)$, (101), $(10\bar{1})$, (011), $(01\bar{1})$;
- fourth-order axes of symmetry – $[100]$, $[010]$, $[001]$;
- third-order axes of symmetry – $[111]$, $[11\bar{1}]$, $[1\bar{1}1]$, $[1\bar{1}\bar{1}]$;
- second-order axes of symmetry – $[110]$, $[1\bar{1}0]$, $[101]$, $[10\bar{1}]$, $[011]$, $[01\bar{1}]$.

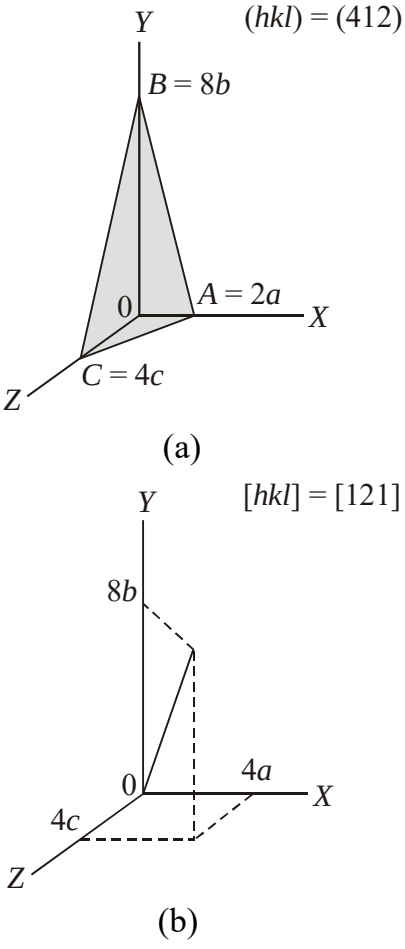


Fig. 40. Examples of calculating the Miller indices of a plane (a) and of a direction (b).

We see that the Miller indices of all planes of symmetry coincide with the Miller indices of the fourth- and second-order axes of symmetry. Consequently, the planes of symmetry of the cubic system are planes perpendicular to the fourth- and second-order axes of symmetry.

The set of crystallographically equivalent planes (i.e., planes that are mapped into one another under crystal symmetry operations) is denoted in curly braces: $\{hkl\}$. For example, the set of planes $\{100\}$ consists of the three families of planes parallel to the faces of a cube: (100), (010), and (001).

7. Theory of the Experimental Method

7.1. Measuring the wavelength of X-ray radiation using Bragg's law

From Bragg's law (6.3.1), it follows that the glancing angle θ and the reflection order n uniquely determine the photon wavelength λ:

$$\lambda = 2d\frac{\sin\theta}{n} \quad (n = 0, 1, 2, \dots). \tag{7.1.1}$$

Thus, the X-ray spectrum can be measured as follows. A crystal is used whose surface is parallel to a well-known family of crystallographic planes. The crystal is most often cut so that these planes are parallel to a unit cell face, because then the number of atoms per unit area of the crystallographic plane is greatest, and therefore the reflection is most intense. The crystal and the reflected-radiation detector are rotated simultaneously such that the angle between the incident radiation and the crystal surface is equal to the angle between the crystal surface and the line passing through the crystal and the detector (see Fig. 41). That angle is the glancing angle θ. Then the detector registers only radiation whose wavelength satisfies Bragg's law (7.1.1). By systematically varying the glancing angle θ up to a certain maximum value (e.g., up to 80°), the intensity ΔI (more precisely, a quantity proportional to it—the number of photons per unit time) of the radiation reflected from the crystal in that direction is measured. The dependence of this intensity on the angle θ is uniquely related to the dependence of the spectral intensity I_λ of the radiation on the wavelength λ. Let us find this relationship. Since the X-ray beam incident on the crystal is not strictly parallel, and the dimensions of the detector's entrance window are finite, the detector registers radiation reflected not at a strictly defined angle θ, but radiation reflected within a certain angular interval of width $\Delta\theta$. Therefore, the wavelength of the radiation striking the detector (after reflection from the crystal) is likewise not strictly defined, but is distributed within a certain interval of width $\Delta\lambda$. If the intervals $\Delta\theta$ and $\Delta\lambda$ are sufficiently small, they are related to each other in the same way as the corresponding differentials. We obtain this differential relation by differentiating Bragg's law (7.1.1):

$$\Delta\lambda \approx 2d\frac{\cos\theta}{n}\,\Delta\theta. \tag{7.1.2}$$

By the definition of spectral intensity I_λ, $\mathrm{d}I = I_\lambda\,\mathrm{d}\lambda$. Consequently,

$$\Delta I \approx I_\lambda^{(d,n)}\Delta\lambda \approx I_\lambda^{(d,n)}\frac{2d}{n}\cos\theta\,\Delta\theta, \tag{7.1.3}$$

where $I_\lambda^{(d,n)}$ is the spectral intensity of the radiation of wavelength λ resulting from n-th order reflection from a given family of crystallographic planes (specified by the interplanar spacing d). It is proportional to the spectral intensity I_λ of the incident radiation:

$$I_\lambda^{(d,n)} = a_\lambda^{(d,n)} I_\lambda, \tag{7.1.4}$$

where $a_\lambda^{(d,n)}$ is the "reflection coefficient," which specifies the relative fraction of energy of the incident radiation of wavelength λ reflected from the given family of crystallographic planes (since radiation is not only reflected, but also absorbed). Substituting Eq. (7.1.4) into Eq. (7.1.3) and solving for I_λ, we obtain:

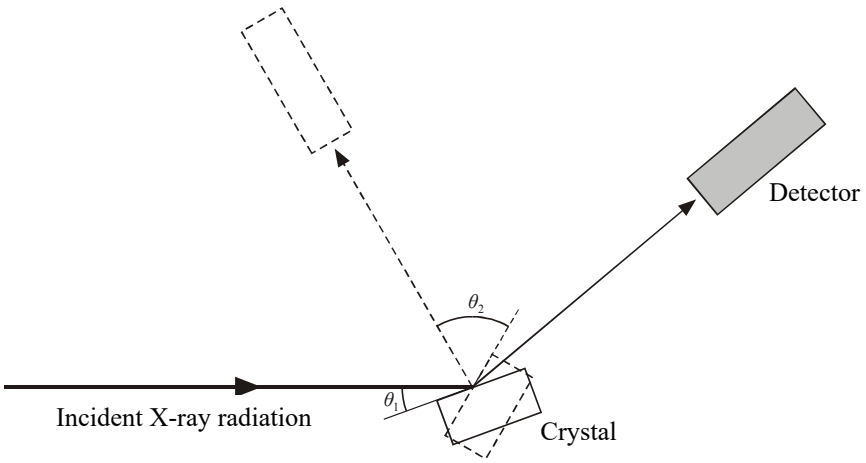


Fig. 41. Measurement of an X-ray spectrum using the Bragg condition: the crystal and the detector are rotated simultaneously so that only radiation reflected from the crystal surface enters the detector.

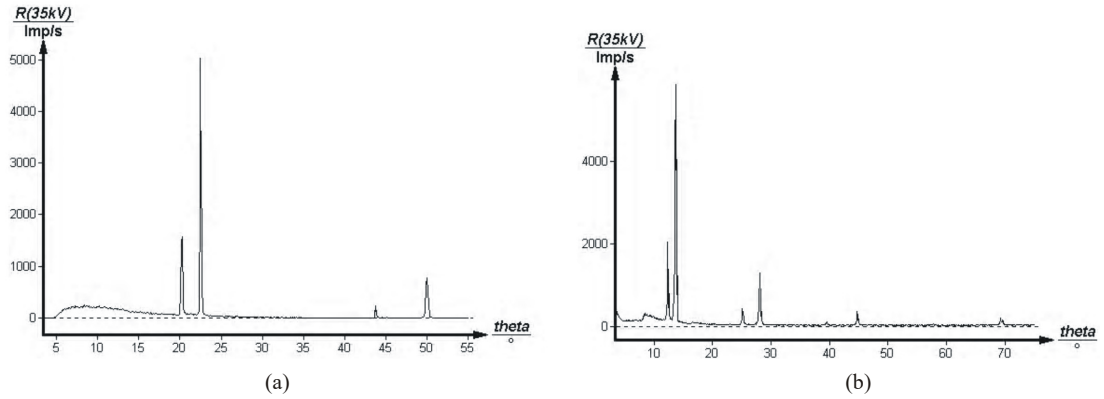


Fig. 42. Dependence of the count rate of a detector registering X-ray radiation reflected from a crystal surface on the glancing angle θ: (a) using a LiF (100) crystal; (b) using a KBr (100) crystal. In both cases, an X-ray tube with a copper anode is used (in the case of the molybdenum anode used in this experiment, the peak positions are different and correspond to smaller angles than in this example).

$$I_\lambda \approx \frac{\Delta I}{\Delta\theta} \cdot \frac{n}{2a_\lambda^{(d,n)} d \cos\theta}. \tag{7.1.5}$$

If the angular interval $\Delta\theta$ "seen" by the detector is identical for all values of θ, then the spectral intensity of the radiation is proportional to the measured intensity ΔI multiplied by the second fraction in Eq. (7.1.5). The quantity $a_\lambda^{(d,n)}$ in the denominator of that fraction is not known in this experiment. However, in this experiment it is not needed, because the objective of this work is not to measure the spectral intensity I_λ of the radiation, but to measure its wavelength (or photon energy). According to Eq. (7.1.5), sharp maxima ("peaks") in the dependence $I_\lambda(\lambda)$ (see Fig. 4) must correspond to sharp maxima in the dependence $\Delta I(\theta)$ (see Fig. 42), and the values of the angle θ corresponding to those peaks must be related to the corresponding wavelengths according to Bragg's law (7.1.1):

$$\sin\theta = \frac{n\lambda}{2d} \quad (n = 0, 1, 2, \dots). \tag{7.1.6}$$

Moreover, it follows from this equality that, in general, a given peak of the function $I_\lambda(\lambda)$ (i.e., a given value of λ) corresponds to several peaks of the function $\Delta I(\theta)$ (i.e., several values of the angle θ): the first peak corresponds to $n = 1$, the second corresponds to $n = 2$, and so forth. For example, Fig. 42a shows the dependence of the count rate on the angle θ when using a lithium fluoride (LiF) crystal cut so that the Miller indices of the crystal surface are (100) (then the distance between adjacent crystallographic planes parallel to the crystal surface is equal to half the lattice constant, i.e., $d = a/2 = 2.014 \cdot 10^{-10}$ m). Here we see two pairs of peaks: the first pair corresponds to $n = 1$ (first-order reflection), and the second corresponds to $n = 2$ (second-order reflection). The higher (right-hand) peak in each pair corresponds to the K_α line, and the left-hand peak in each pair corresponds to the K_β line. Fig. 42b shows an analogous dependence when using a potassium bromide (KBr) crystal whose surface Miller indices are also (100). In a KBr crystal, the distance between adjacent atoms is larger than in a LiF crystal, so the distance between adjacent crystallographic planes is larger ($d = 3.29 \cdot 10^{-10}$ m). Correspondingly, the number of "repeating" peaks (or groups of peaks) also increases: in Fig. 42b we already see not two, but four pairs of peaks (reflection orders $n = 1; 2; 3; 4$). In the direction of increasing angle θ, the peaks are arranged as follows: K_β ($n = 1$), K_α ($n = 1$), K_β ($n = 2$), K_α ($n = 2$), and so on.

7.2. Measuring X-ray photon energy using a Si-PIN detector

In this work, the photon energy spectrum of X-ray radiation is measured. A semiconductor detector (specifically, a Si-PIN detector) is used for the measurements, which generates a voltage pulse each time a photon is absorbed in the intrinsic semiconductor layer (the "i" region of the PIN junction; the structure of a Si-PIN detector was explained in Section 5.7.4). The amplitudes of different pulses are not strictly identical even when the energies of the photons that caused those pulses are strictly equal. Therefore, the peaks of the measured spectrum are relatively broad. In radiation spectroscopy, the width of spectral peaks is customarily measured at half the peak height. A peak width expressed in this way is denoted by the abbreviation

FWHM ("full width at half-maximum"). For example, when the photon energy is 6.4 keV, the width of the corresponding peak obtained using the Si-PIN detector of this experiment is 0.55 keV (the corresponding energy resolution is equal to $0.55/6.4 \approx 8.6$ %). However, if the number of peaks in the spectrum is small and adjacent peaks do not overlap (or their overlap is weak), then the energy measurement error can be much smaller than the FWHM value. For this, each peak must be symmetric with respect to its maximum (e.g., Gaussian in shape), and the number of registered photons (i.e., detector pulses) in the peak region must be sufficiently large—such that the peak is smooth. Then the position of the peak maximum (average pulse amplitude), which is proportional to the photon energy (see Eq. (5.4.2)), can be measured quite accurately. However, it is impossible to observe the fine structure of K_α lines (see Section 3.5) in this way, because the energy difference between the $K_{\alpha 1}$ and $K_{\alpha 2}$ lines is less than 0.3 keV if $Z < 54$. Consequently, the $K_{\alpha 1}$ and $K_{\alpha 2}$ lines correspond to a single peak in the experimental spectrum.

The pulse height spectrum is measured with a special device—a **pulse height analyzer**, also called a **multichannel analyzer** (MCA). A multichannel analyzer can be visualized as a set of counters with a common input, each of which counts only those pulses whose amplitudes belong to a specific narrow interval called a **channel**. The channels have equal width, do not overlap with one another, and there are no gaps between them. Therefore, when a pulse arrives at the input of the multichannel analyzer whose amplitude lies between the lower edge of the first channel (i.e., the minimum voltage that the analyzer can register) and the upper edge of the last channel (i.e., the maximum voltage), that amplitude necessarily belongs to one (and only one) channel. That is, such a pulse is registered by one of the counters mentioned above. Thus, a multichannel analyzer sorts pulses according to their amplitudes. In this way, the **pulse height spectrum** is measured—the amplitude function $\delta n/\delta H$, where δn is the number of pulses corresponding to each channel, and δH is the channel width. It is convenient to choose the unit of amplitude measurement equal to the channel width. Then $\delta H = 1$, $\delta n/\delta H = \delta n$, and the amplitude H is equal to the channel number.

7.3. Detector calibration and response function

In order to use a detector to measure photon energies, it must be calibrated. The purpose of **calibration** is to determine the proportionality coefficient in the relationship between photon energy and average pulse amplitude (see Eq. (5.4.2)). To do this, one must measure the average amplitude H_0 of the detector pulses when photons of known energy E_0 strike the detector. Then the energy E of photons that produce pulses of amplitude H can be calculated as follows:

$$E = \frac{E_0}{H_0} \cdot H. \tag{7.3.1}$$

If H has the meaning of a channel number (a dimensionless integer), and the first channel corresponds to zero energy, then instead of Eq. (7.3.1) one must use

$$E = \frac{E_0}{(H_0 - 1)} \cdot (H - 1). \tag{7.3.2}$$

In this work, the role of the calibration source is fulfilled by X-ray tube radiation reflected from a LiF crystal. Specular reflection from the crystal surface in a given direction (specified by the glancing angle θ) is possible only if the wavelength satisfies Bragg's law (7.1.1). Therefore, knowing the distance d between adjacent crystal atoms (more precisely, between adjacent atomic planes) and the reflection order n, one can calculate the wavelength λ using Eq. (7.1.1). Since the crystal lattice of LiF is face-centered cubic, and the Miller indices of the crystal surface (see Section 6.5) are (100), we have $d = a/2$, where a is the lattice constant of the crystal, i.e., the edge length of the unit cell (see Fig. 31). In the case of LiF, $d = 2.014 \cdot 10^{-10}$ m.

An X-ray tube emits continuous-spectrum radiation, i.e., the spectrum contains all wavelengths starting from the short-wavelength limit (3.2.4), or all photon energies from zero up to eU, where U is the anode voltage of the X-ray tube (an example of the spectrum was presented in Fig. 4). Therefore, at *any* value of the glancing angle θ (see Fig. 41), the detector will register radiation of the corresponding wavelength (7.1.1), provided that wavelength is greater than the short-wavelength limit (3.2.4). Since $E \sim H$, the value of E_0 in the calibration relation (7.3.1) or (7.3.2) can be arbitrary. Consequently, the glancing angle θ can be chosen freely, and the anode material of the X-ray tube is irrelevant (the short-wavelength limit depends only on the anode voltage, not on its chemical composition). However, to improve accuracy, the highest possible intensity of reflected radiation is required, because:

1) higher intensity allows a larger number of photons to be registered in the same time, thereby reducing statistical errors (i.e., the spectrum peak becomes smoother);

2) the ratio of the peak height to the "background" caused by extraneous pulses increases (extraneous pulses arise, for example, from Compton scattering of photons in the crystal or from radiation scattering in the crystal holder and surrounding objects).

Therefore, it is advisable to choose the glancing angle θ so that it corresponds to one of the peaks in the spectrum. For example, if the anode material is copper, then reflections of the K_α and K_β spectral lines of characteristic radiation occur at four glancing angles (see Fig. 42a): 20.3° (K_β, $n = 1$), 22.5° (K_α, $n = 1$), 43.8° (K_β, $n = 2$), 49.9° (K_α, $n = 2$). Moreover, in this way one can not only calibrate the Si-PIN detector, but also measure the energies of the K_α and K_β lines of the anode material: it suffices to find the values of the angle θ corresponding to the maxima of the reflected intensity.

In this work, an X-ray tube with a molybdenum (Mo) anode is used. Fig. 43 shows the experimental pulse height distribution of the Si-PIN detector when photons of the K_α line of characteristic X-ray radiation of Mo strike the detector after first-order reflection ($n = 1$) from the family of crystallographic planes of LiF whose Miller indices are (100). The corresponding experimental value of the glancing angle θ is $10.1° \pm 0.1°$, the wavelength calculated according to Bragg's law (7.1.1) is equal to $\lambda = (0.706 \pm 0.007) \cdot 10^{-10}$ m, and the corresponding photon energy is equal to (17.5 ± 0.2) keV.

If the detector were ideal, and only photons of the Mo K_α spectral line struck the detector, then in Fig. 43 we would see only a single narrow peak at 17.5 keV energy, while on both sides of it the spectrum would be equal to zero. One of the reasons why a broad spectrum of reduced-amplitude pulses appears is the recombination of electron–hole pairs generated in the active material of the detector and the trapping of charge carriers at semiconductor defects that act as carrier "traps." The probability of recombination increases if the photoelectric effect occurs in a region of weaker electric field, e.g., near the edge of the depletion layer. These phenomena reduce the number of charge carriers reaching the detector electrodes, thereby reducing the amplitudes of some pulses. Another reason why lower-amplitude pulses can arise is *partial* absorption of photon energy in the active material of the detector. This is due to Compton scattering. It manifests itself when a photon transfers a random fraction of its energy to the material, depending on the scattering angle (see Eqs. (4.4.1) and (4.4.2)). If, after Compton scattering (which may be multiple), the photon escapes from the detector, or if a photon that has already undergone Compton scattering in other components of the apparatus (e.g., in the detector housing or the LiF crystal) interacts with the active material of the detector, then the amount of energy transferred to the active material of the detector is smaller than

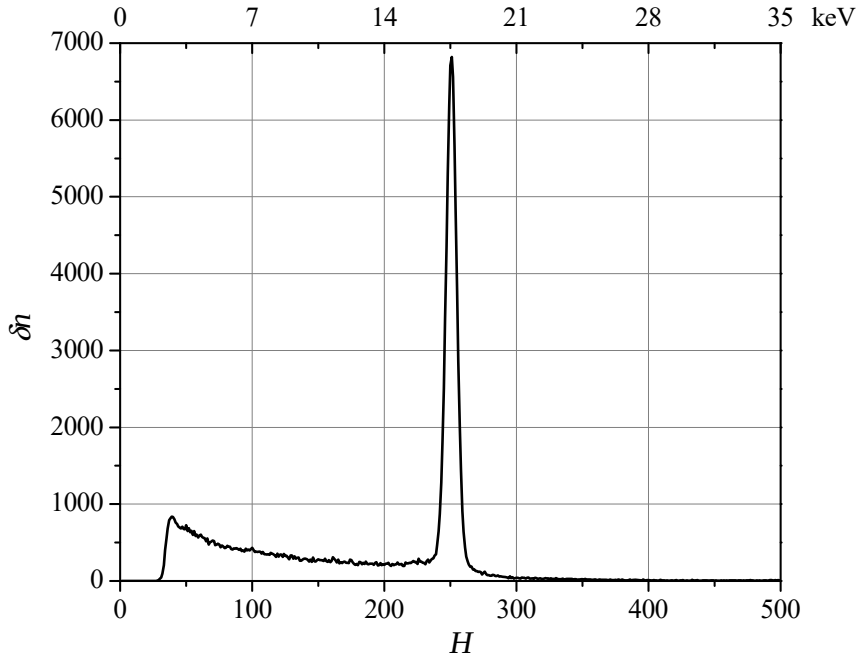


Fig. 43. Pulse height distribution of the Si-PIN detector when photons of the K_α line of characteristic X-ray radiation of Mo strike the detector. The channel number H is plotted along the bottom axis, and the photon energy along the top axis.

the initial photon energy. Since this energy reduction is random, a continuous "Compton spectrum" appears to the left of the pulse height spectrum peak.

The abrupt drop in the number of registered photons when the photon energy becomes less than 3 keV (see Fig. 43) is explained by the decrease in detector efficiency (see Fig. 30).

7.4. Measurement of the X-ray fluorescence spectrum

X-ray fluorescence (XRF) is secondary characteristic X-ray radiation that arises when a primary X-ray photon liberates an electron from an inner shell of an atom of a material. The primary X-ray radiation may be bremsstrahlung or characteristic radiation. As mentioned in Section 3.3, any process that creates vacancies in the inner electron shells of an atom is accompanied by characteristic X-ray emission, because those vacancies are filled by electrons from higher electron shells of the same atom. During these quantum transitions of electrons, photons are emitted whose energy is equal to the energy difference between the electron shells (characteristic X-ray radiation). Therefore, an XRF spectrum is a line spectrum, i.e., composed of individual peaks. For a given series of lines (e.g., the K series) to be visible in the XRF spectrum, it suffices that the primary radiation contain photons whose energy is greater than the binding energy of the corresponding inner electron shell. For example, if the radiation spectrum contains photons whose energy is greater than the K-shell binding energy of the sample material's atoms, then the K_α and K_β lines will be visible in the XRF spectrum (the K_γ line is very close to the K_β line and its intensity is much smaller than that of the K_α and K_β lines; therefore, the K_γ line is visible only when using equipment of exceptionally high quality and under exceptionally good measurement conditions). XRF radiation is approximately isotropic (correlation with the direction of the primary photons is weak).

The measurement geometry is the same as in Fig. 41, except that the value of the angle θ is unimportant in this case. θ must be large enough so that primary X-ray radiation from the X-ray tube does not enter the detector. The samples in which X-ray fluorescence occurs are in the form of metal plates or powders. Since the samples are not monocrystalline, constructive interference of coherently scattered photons is impossible. That is, there are no directions along which diffraction maxima are observed. Scattered radiation is uniformly distributed in all directions (just like XRF radiation).

In this experiment, the anode voltage of the X-ray apparatus (i.e., the electron accelerating voltage in the X-ray tube) is 35 kV. This means that the maximum photon energy of the primary bremsstrahlung radiation is 35 keV, and all series of lines corresponding to binding energies smaller than 35 keV must be visible in the XRF spectrum (the electron shell binding energy is the maximum possible energy of the corresponding line series). Therefore, the K lines of all elements in the periodic table up to xenon (Xe) inclusive ($Z \leq 54$) are visible in the XRF spectrum. Binding energy is directly proportional to Z^2 (see Section 3.4). For elements with $Z > 54$, the K-shell binding energy (the "**K-edge**") is greater than 35 keV. For example, for lead (Pb, $Z = 82$), the K-shell binding energy is 85.7 keV. Therefore, in the XRF spectra of elements with $Z > 54$, only the L_α and L_β lines resulting from the liberation of an electron from the L shell can be visible. L-series lines can also be visible in the XRF spectra of lighter atoms, although their detection is hindered by their low energy. For elements up to silver (Ag) inclusive ($Z \leq 47$), the L-shell binding energy is less than 3 keV, and the efficiency of the Si-PIN detector corresponding to such low photon energies is too low to register them (see Fig. 30). For the same reason, K-series lines of elements up to potassium (K) inclusive ($Z \leq 19$) are not registered. When the energy of XRF photons exceeds approximately 20 keV, the intensities of the pulse height spectrum peaks corresponding to the K_α and K_β lines also decrease very rapidly with increasing XRF photon energy. This is due to two factors:

1) the decrease in detector efficiency as the energy of the registered photons increases (see Fig. 30),

2) the decrease in the number of primary photons whose energy is greater than the K-shell binding energy of the sample atoms as this binding energy increases.

The latter factor becomes obvious from the theoretical spectrum of radiation from an X-ray tube with a Mo anode, which is shown in Fig. 44. For example, the K-edge of copper (Cu, $Z = 29$) is 9.0 keV. In the spectrum of Fig. 44, 99.98 % of the photons have energies greater than 9 keV, i.e., almost all primary photons can initiate the K_α and K_β lines of the copper XRF spectrum. The K-edge of silver (Ag, $Z = 47$) is 25.5 keV. In the spectrum of Fig. 44, only 14.3 % of the photons have energies greater than 25.5 keV, i.e., on average only one in seven primary photons has sufficient energy to initiate the K_α and K_β lines of the silver XRF spectrum.

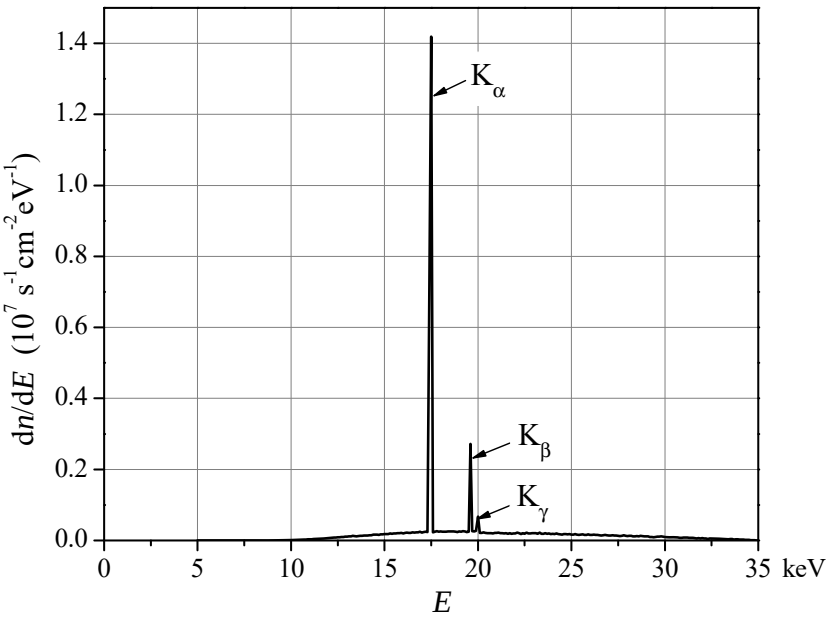


Fig. 44. Approximate energy spectrum of radiation from the X-ray apparatus with a Mo anode used in this work at a distance of 2–5 cm from the diaphragm, when the anode voltage is 35 kV and the anode current is 1 mA. The spectral flux density of photons, expressed in billions of photons per second per square centimeter per 100 eV energy interval, is plotted along the *Y* axis.

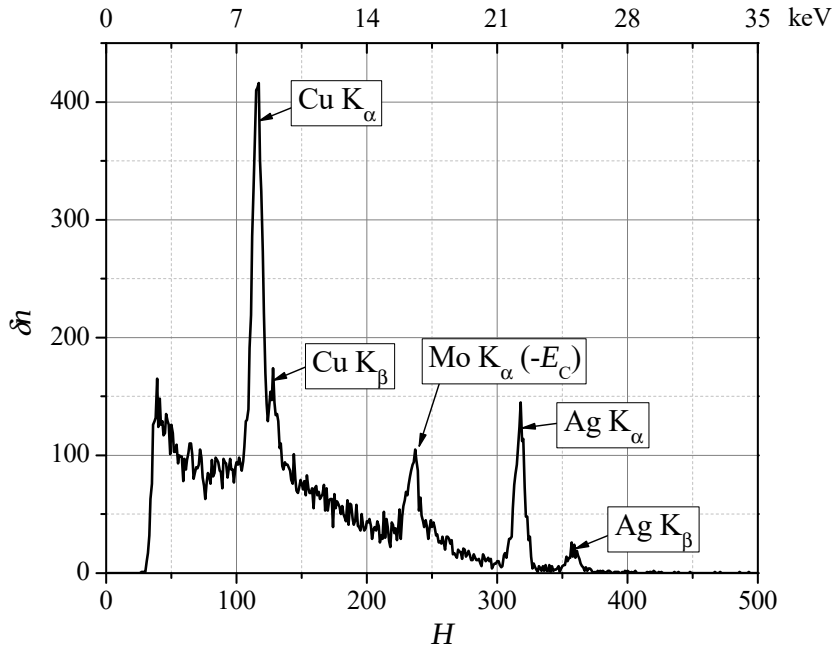


Fig. 45. X-ray fluorescence spectrum of a silver wire coil 160 cm long and 0.6 mm thick, measured with a Si-PIN detector. The mass fractions of Ag and Cu in the wire are 92.5 % and 7.5 %, respectively, and the atomic fractions are 88 % and 12 %. The angle between the primary radiation beam and the line connecting the irradiated spot of the sample to the detector is 150°. The central peak is the K_α line of the primary (molybdenum) X-ray radiation, whose energy is reduced from 17.44 keV to 16.4 keV due to Compton scattering of photons at an angle of 150° in the plastic box containing the silver wire coil (this energy reduction is the Compton recoil electron energy E_C, given by the second equation of system (4.4.1)).

An example of an experimental XRF spectrum is shown in Fig. 45. In this case, the sample was a coil of silver wire 160 cm long and 0.6 mm thick. The mass fractions of Ag and Cu in the wire are 92.5 % and 7.5 %, respectively, and the corresponding atomic fractions are 88 % and 12 %. Although the amount of Cu in the sample is much smaller than the amount of Ag, the intensity of the Cu K_α and K_β peaks in the XRF spectrum is greater than that of the Ag K_α and K_β peaks (see Fig. 45). As explained in the previous paragraph, this is because there are far fewer primary radiation photons with sufficient energy to excite the K-series lines of Ag than photons with sufficient energy to excite the K-series lines of Cu, and because the

detector efficiency is much lower when registering Ag K-series photons, whose energy is 22–25 keV, than when registering Cu K-series photons, whose energy is 8–9 keV (the dependence of efficiency on photon energy is shown in Fig. 30).

In the case of Fig. 45, the angle between the primary radiation beam and the line connecting the irradiated spot of the sample to the detector was 150°. Since the wire coil is placed in a plastic box whose wall thickness is 1 mm, the K_α line of the primary (molybdenum) X-ray radiation is also visible in Fig. 45, with its energy reduced from 17.44 keV to 16.4 keV due to Compton scattering of photons at an angle of 150° in the plastic layer (this energy reduction is the Compton recoil electron energy E_C, given by the second equation of system (4.4.1)).

8. Experimental Procedure

8.1. Equipment and measurement procedure

This experiment uses an educational X-ray apparatus manufactured by the German company “Phywe Systeme”. The experimental setup consists of:

1) an X-ray apparatus with a molybdenum (Mo, $Z = 42$) X-ray tube (X-ray tube anode voltage adjustment range: 0 to 35 kV; anode current adjustment range: 0 to 1 mA),
2) a goniometer (an accessory of the X-ray apparatus designed for investigating X-ray spectra and scattering),
3) a computer with the “*MCA Lab*” software, designed for displaying and analyzing spectra,
4) a Si-PIN detector with a multichannel analyzer (in a single housing), connected to the computer via a USB port,
5) a lithium fluoride (LiF) crystal with a holder,
6) a set of X-ray fluorescence samples (plates of various metals and plastic vials containing powders of various inorganic compounds) and a sample holder.

The X-ray apparatus is shown in Fig. 46. On its left side is the X-ray tube unit with a collimator aperture; to the right is the goniometer, on the right edge of which the detector is visible.

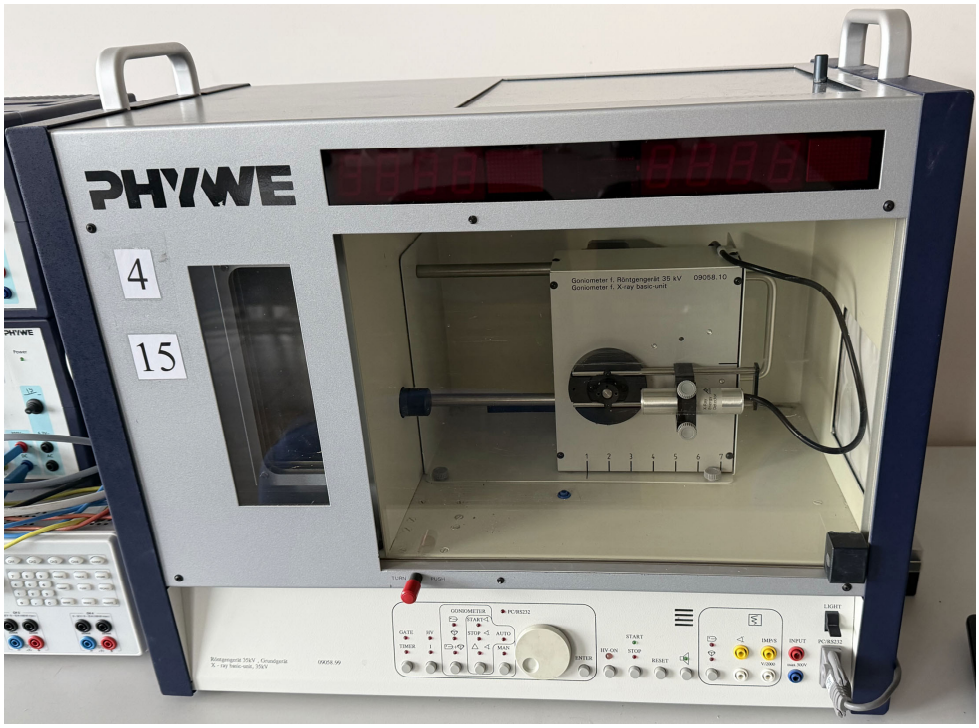


Fig. 46. X-ray apparatus with goniometer and detector (the X-ray tube is on the left, and the detector is on the right).

Measurement procedure:

1. Ensure that there is no stand for vials inside the chamber of the X-ray apparatus. If it is present, remove it from the chamber. In addition, if the sample holder or the LiF crystal is mounted in the goniometer, remove it as well.

2. Turn on the X-ray apparatus and the computer. When the X-ray apparatus is turned on, the goniometer automatically rotates the detector to the initial position corresponding to a zero glancing angle θ (this position is shown in Fig. 46).

3. Using the controls of the X-ray apparatus located at the bottom of its front panel (see Fig. 46), set the X-ray tube anode voltage to **35** kV and the anode current to **0.1** mA.

4. Even such a low anode current is too high to direct the X-ray beam straight into the detector. A Si-PIN detector can record spectra without distortion only up to approximately one thousand photons per second, whereas in the primary radiation beam several hundred thousand photons per second would strike the detector window. Moreover, such a high radiation intensity can damage the detector. Therefore, place a 3 mm thick aluminum plate in front of the X-ray tube aperture, which will reduce the radiation flux by two orders of magnitude (since the detector records only a few percent of all photons incident on its window, the count rate will decrease to a few hundred photons per second).

5. Launch the "*MCA Lab*" program. The channel number ("*MCA Channel*") must be plotted on the *X* axis. If the *X*-axis label is "*Energy E, keV*", click the toolbar button  (see Fig. 47).

6. Turn on the X-ray tube anode voltage. The X-ray tube begins emitting X-rays. **Caution!** If the primary X-ray beam is directed into the detector window, the X-ray tube anode voltage may be turned on only when the aluminum filter is placed between the detector and the aperture (see step 4). An excessively high incident radiation intensity can damage the detector.

7. In the "*MCA Lab*" program, set the spectrum measurement duration (**3** min).

8. Measure the detector pulse height spectrum when radiation from the X-ray tube with a Mo anode transmitted through a 3 mm thick aluminum layer is incident on the detector. An example of the spectrum is shown in Fig. 47 (in this example, the measurement duration was longer—10 min).

9. After completing the measurement, turn off the X-ray tube accelerating voltage.

10. Using the "*MCA Lab*" menu command "*File* > *Save As...*", save the spectrum data to a file with the extension ".mca" (this file can later be opened in "*MCA Lab*").

11. Remove the aluminum filter. Enable manual goniometer angle adjustment.

12. Mount the LiF crystal in the goniometer and rotate it together with the detector so that the glancing angle θ is equal to **9.5**° (see Fig. 41). To rotate the crystal and the detector, use the controls of the X-ray apparatus located at the bottom of the front panel.

13. Turn on the X-ray tube anode voltage. Start pulse counting in the "*MCA Lab*" program. Increase the glancing angle θ from **9.5**° to **10.5**° in steps of 0.1° and find the value of θ corresponding to the highest count rate. The current count rate is displayed in the "*MCA Lab*" window below the graph, before the words "*Counts/s*" (e.g., see Fig. 48). Record the value of θ mentioned above, return the crystal and detector to that position, clear the current spectrum (using the menu command "*File* > *New*", declining to save the current spectrum), and perform a single spectrum measurement of **3** min duration. Since only radiation of a single wavelength is reflected from the crystal (see Bragg's law (7.1.1)), a single tall peak corresponding to the first-order reflection of the Mo K_α line from the LiF crystal should be visible in the spectrum (e.g., see Fig. 48). After completing the measurement, turn off the anode voltage. Save the spectrum data to a file (see step 10).

14. Remove the LiF crystal from the goniometer. In place of the crystal, mount the X-ray fluorescence ("XRF") sample holder in the goniometer.

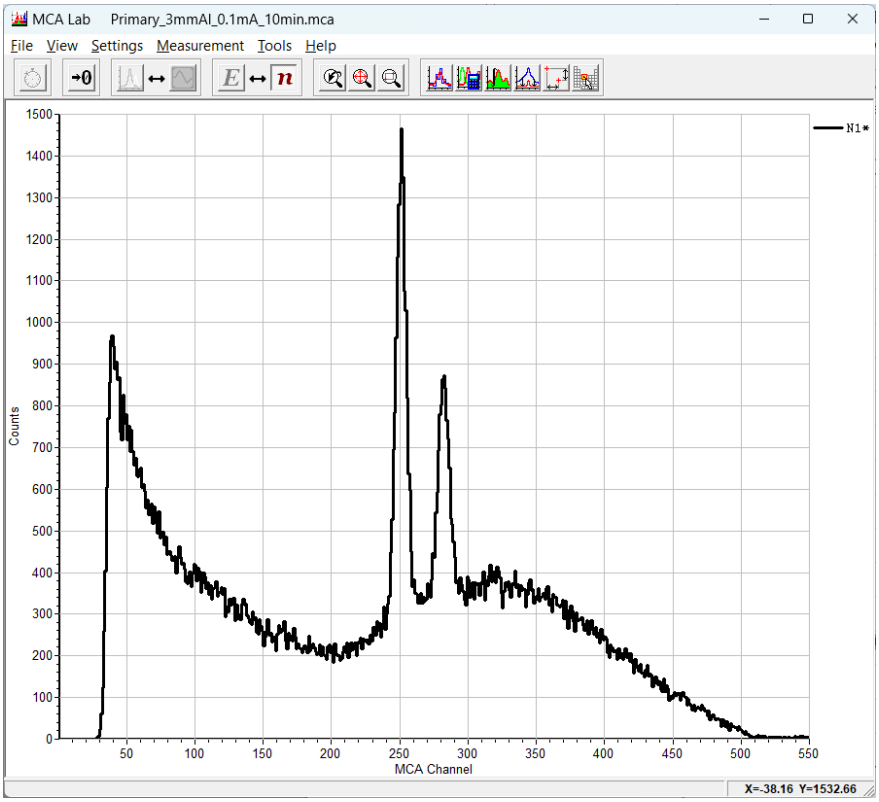


Fig. 47. Example of the primary spectrum measured using a 3 mm thick aluminum filter.

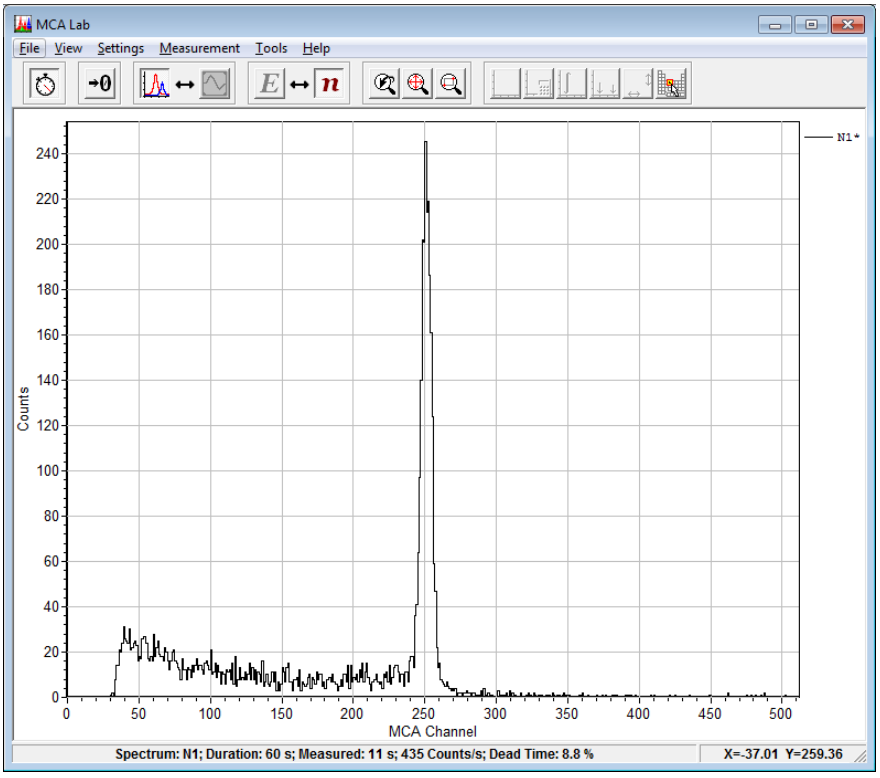


Fig. 48. Example of a measured spectrum of radiation from an X-ray tube with a Mo anode reflected from a LiF crystal.

15. Rotate the sample holder and the detector so that the glancing angle θ is equal to 75° (then the angle between the primary radiation beam and the direction of the recorded radiation is equal to $2\theta = 150°$).

16. Place the XRF samples one after another into the holder and perform one XRF spectrum measurement of 3 min or 10 min duration with each of them. If a sample is in powder form, place the plastic vial with the powder onto the stand right next to the sample holder. Using the "*MCA Lab*" menu command "*File*

> *Save As...*", save each spectrum's data to a file (see step 10). The anode currents and measurement durations corresponding to the different samples are listed in Table 4.

Table 4. XRF samples and spectrum measurement conditions

No.	Sample	Anode current (mA)	Measurement duration (min)
1	Stainless steel plate (≈ 70 % Fe + Cr)	0.75	3
2	Copper (Cu) plate	0.5	3
3	Zinc (Zn) plate	0.5	3
4	Germanium dioxide (GeO_2) powder	0.75	3
5	Potassium bromide (KBr) powder	1	3
6	Strontium sulfate ($SrSO_4$) powder	1	3
7	Silver wire (92.5 % Ag + 7.5 % Cu)	1	10
8	Tin solder (> 97 % Sn + Cu)	1	10

Upon completing the measurements, turn off the X-ray apparatus. Save all ten `*.mca` files created while performing steps 10, 13, and 16 to a USB flash drive or store them in cloud storage (e.g., via e-mail). In addition, using the "*MCA Lab*" menu command "*File* > *Export Graph* > *Save as type*", save each graph as a "Windows Metafile" format file (`*.wmf`). Insert these files into a single Word document (at least two graphs per page). Print that file (you may use the printer available in the laboratory). The sheets with the printed graphs must be signed by the lab supervisor or laboratory assistant.

A separate description of the measurement procedure is located next to the measurement equipment, which is much more detailed than the one presented here. This description must be used during the measurements. After completing the measurements, leave it next to the measurement equipment. When preparing for the experiment, it is not necessary to know all the details of the measurement procedure. If the measurement instructions given earlier differ from those provided in the detailed description, follow the detailed description during the measurements.

8.2. Analysis of Measurement Data

The measurement data are stored in `*.mca` files that were created in the laboratory using the program *"MCA Lab"*. These are text files that can be opened with any text editor, e.g., Windows *"Notepad"*. At the beginning of the file, there are several lines containing various information about the spectrum measurement (e.g., measurement time, total detector dead time, etc.). The spectrum data are located in the last 1024 lines, one line for each of the 1024 channels. Each of those lines begins with the channel number (from 1 to 1024), followed by the corresponding Y values (detector pulse counts), of which there are as many as there are spectra saved in that file. One can copy those lines into any other program (e.g., *"Excel"*, *"Origin"*, etc.) and then use that program for data analysis.

It is most convenient to analyze the measurement data using the program *"MCA Lab"*, which can be downloaded from the website of the German company *"PHYWE"* (link: https://phywe-itemservice.s3.eu-central-1.amazonaws.com/sites/DMS-Phywe/PROD/de-DE/item/phy_itemsoftware/09/09165-88/MCALab.zip). Below is the procedure for analyzing measurement data using the program *"MCA Lab"*.

1. Using Bragg's law (7.1.1), calculate the wavelength corresponding to the glancing angle θ that was measured in step 13 of the measurement procedure (see Section 8.1). The LiF crystallographic planes whose Miller indices are (100) are separated from each other by a distance $d = 2.014 \cdot 10^{-10}$ m. Since the first-order reflection was measured, in Bragg's law $n = 1$. Calculate the corresponding photon energy. The photon energy must be expressed in kiloelectronvolts (keV).

2. Launch the program *"MCA Lab"*. Using the menu command *"File* > *Open"* (or by pressing the *"F3"* key), open the `*.mca` file containing the spectrum data of radiation reflected from the LiF crystal (an example is shown in Fig. 49). Initially, the channel number (*"MCA Channel"*), rather than energy, must be plotted along the X axis. If the X axis is labeled *"Energy E, keV"*, then click the toolbar button n above the graph (see Fig. 49). The limits of the X or Y axis can be changed by right-clicking on

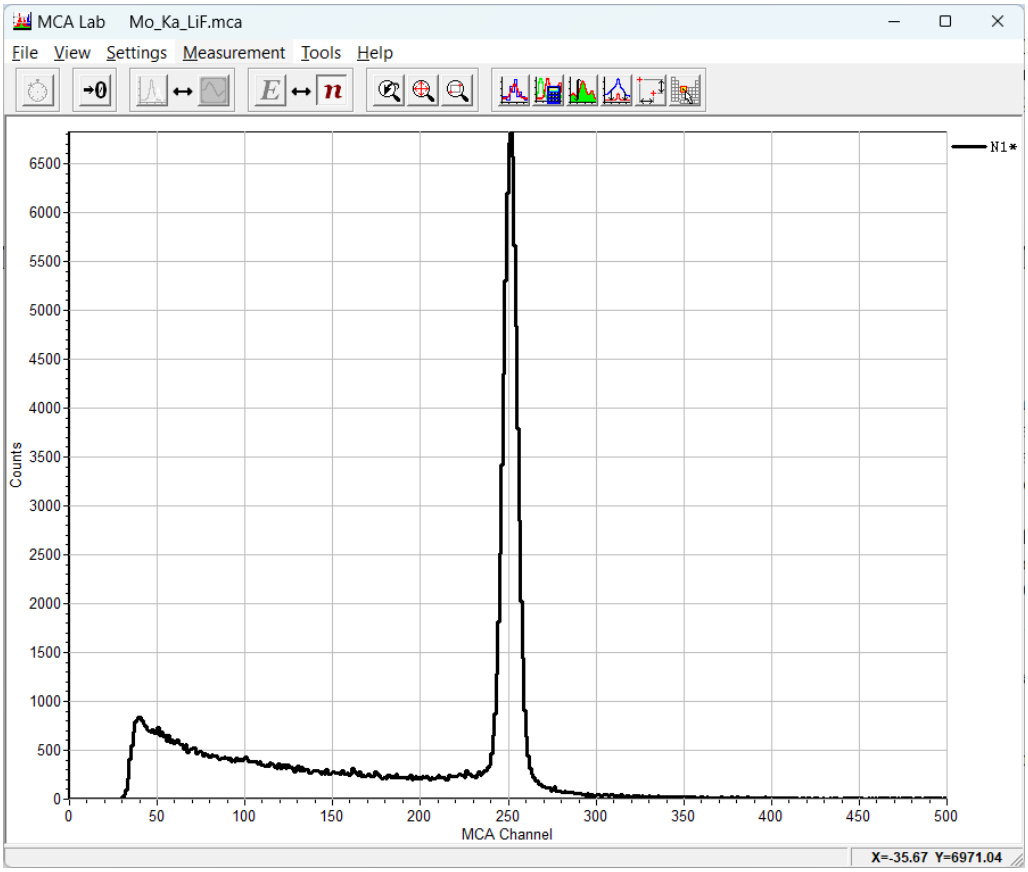


Fig. 49. Example of the *"MCA Lab"* program window. The graph shows the spectrum of the Mo K_α line.

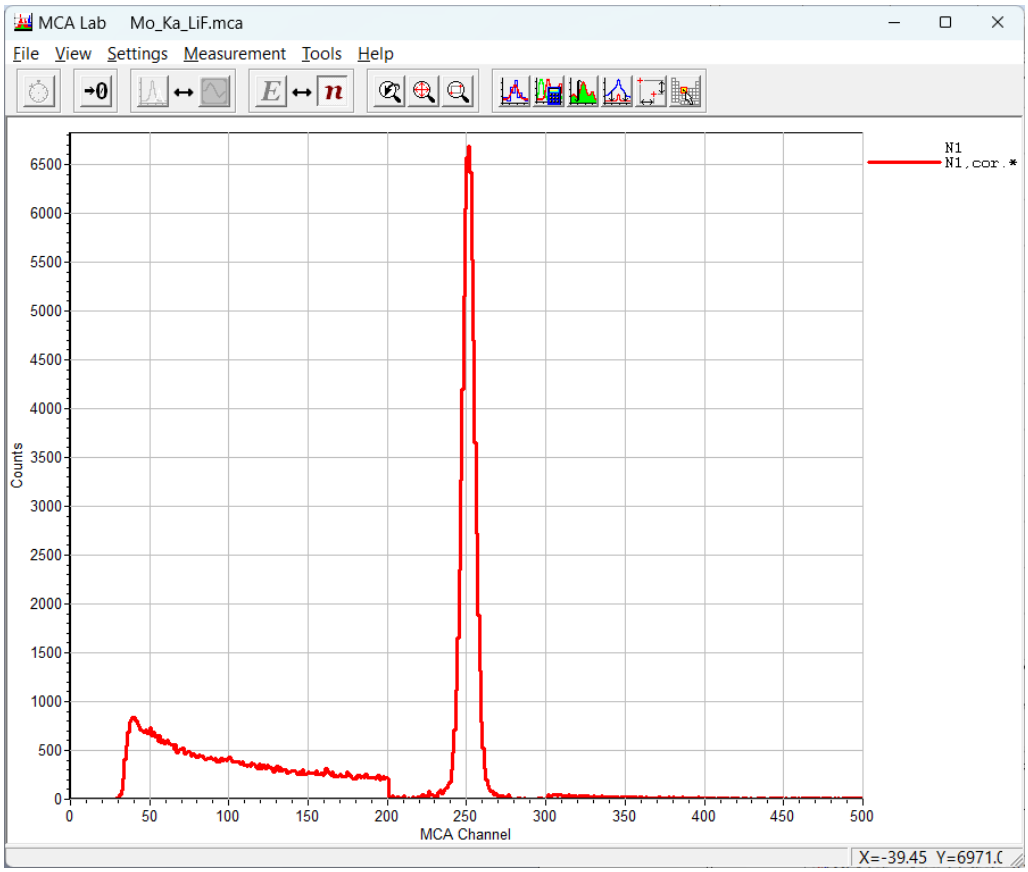


Fig. 50. Spectrum obtained by subtracting the background from the original Mo K_α spectrum (Fig. 49) in channels No. 200–300.

the *X* or *Y* axis. To set the "optimal" limits, i.e., *Y*-axis limits corresponding to the largest count in the channels, select the menu command *"View > Zoom All"*, click the toolbar button , or use the key combination *"Alt" + "A"*.

3. In order to determine the position of the peak in the spectrum (i.e., the central channel of the peak) as accurately as possible, fit it with a Gaussian curve:

   (a) First, subtract the "background" arising from extraneous photons and detector imperfection (the origin of the background is explained in Section 7.3). To do this, select the menu command *"Tools > Baseline Correction"* or click the toolbar button . This is the fourth of the last six buttons on the toolbar (see Fig. 49). After performing this action, the instruction *"Click to set points. Right mouse button = ready."* appears in the status bar of the program (below the graph). This means that you must left-click at several points of the spectrum that lie on both sides of the peak, outside the peak region (the program will automatically connect those points with a broken straight line and subtract that polyline from the spectrum). After marking the last "background" point, click the right mouse button. A new spectrum then appears on the graph, obtained by subtracting the background from the original spectrum. For example, the red line in Fig. 50 was obtained by marking four background points at channels No. 200, 220, 280, and 300. The name of that spectrum in the legend is formed by appending the text *"_cor."* to the name of the original spectrum (from the word *"corrected"*). The corrected spectrum automatically becomes "active" (this is the name given to the spectrum on which all mathematical operations are performed). An asterisk appears next to the name of the active spectrum in the legend (see Fig. 50). Since only the corrected spectrum will be used for further analysis, the original spectrum can be hidden. To hide a spectrum (without deleting its data), right-click on the spectrum name in the legend, uncheck the *"Show in Diagram"* box, and click the *"Close"* button.

   (b) Next, select the menu command *"Tools > Gauss-Fit"* or click the toolbar button . This is the first of the last six buttons on the toolbar (see Fig. 49). After performing this action, the instruction *"Click to set peak positions. Right mouse button = ready."* appears in the status bar of the program (below the graph). This means that you must left-click at the channels corresponding to the peaks being fitted. Since in this case only one peak is being fitted, it suffices to click at a single channel. The *X* coordinate of that point must be chosen so that it falls within the region of the peak (not necessarily its center); the *Y* coordinate does not matter. Then right-click anywhere on the graph area (this signals the program to proceed to the next step). A new instruction then appears in the status bar: *"Click on the fitting interval."* This means that you must mark the first and last channels of the fitting interval. To do this, first left-click on the graph area at the *X* value corresponding to the start of the fitting interval (the *Y* value does not matter), and then at the *X* value corresponding to the end of the fitting interval. Both of these channels must lie between the two adjacent points that were selected in the previous step when defining the background interval. The program will then automatically perform the fit and display the fitting curve (a Gaussian function with a constant offset). The name of that curve in the legend begins with the word *"Fit"*, followed by the fit number. For example, the blue line in Fig. 51 was obtained by marking channels No. 230 and 270.

   (c) The fitting Gaussian function has four parameters: a constant offset, peak height, centroid (i.e., the *X* value corresponding to the peak maximum), and standard deviation (i.e., half-width of the peak at 61 % of its height). Their values can be viewed by right-clicking on the name of the corresponding Gaussian curve in the legend. In the *"Spectrum Settings"* dialog box that opens, those four parameters are listed in the *"Comments"* field and are denoted by *"B"*, *"A1"*, *"μ1"*, and *"σ1"*, respectively (see Fig. 52). The only one of those four parameters needed in this laboratory experiment is the centroid (*"μ1"*). For example, in Fig. 52, the centroid is equal to 250.69.

4. Having the centroid (central channel number) calculated in step 3(c), one can calibrate the detector, because the energy corresponding to that channel is known (it was calculated in step 1 using Bragg's law). Calibration is performed according to Eq. (7.3.2). To do this in the program *"MCA Lab"*, click the toolbar button  (this click will "work" only if the channel number *"MCA Channel"* is plotted

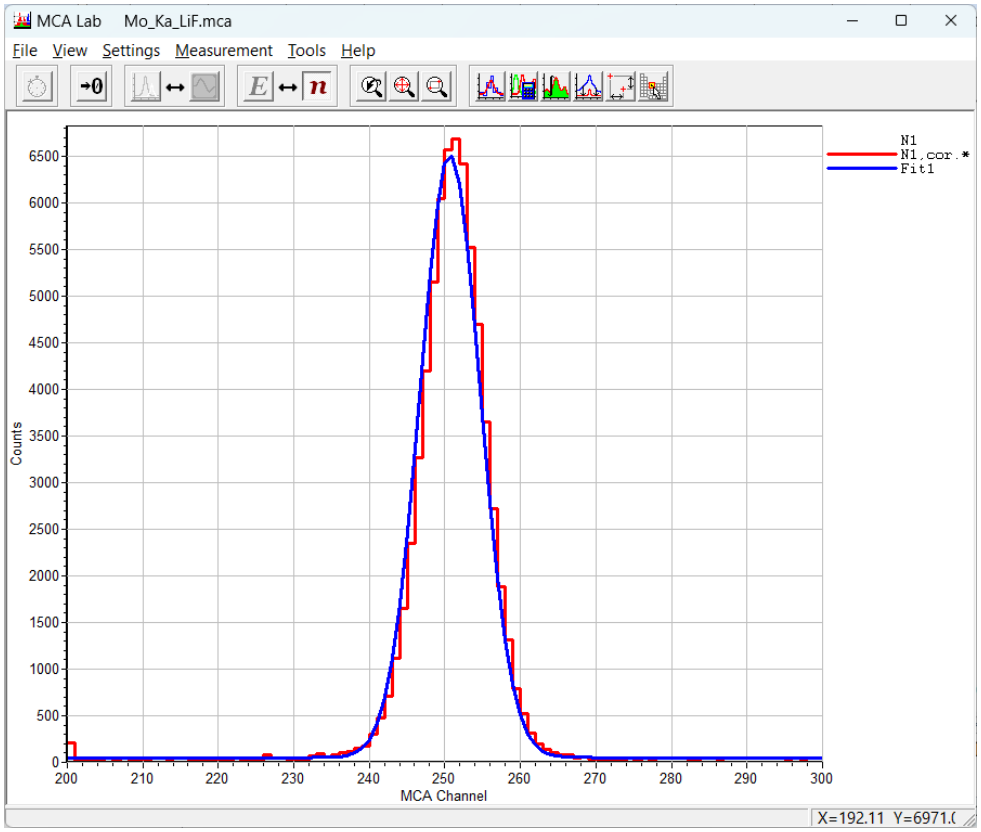


Fig. 51. The blue line represents the Gaussian function fitting the spectrum of Fig. 50 between channels No. 230 and 270. The program *"MCA Lab"* shifts experimental histograms by half a bin width to the right, so the true *X* value lies not at the center of a bar, but at its left edge. However, fitting curves use the true *X* values, which lie at the left edges of the bars. Therefore, if the fit is accurate, the fitting curve passes through the upper-left corners of the bars and appears shifted to the left relative to the experimental spectrum.

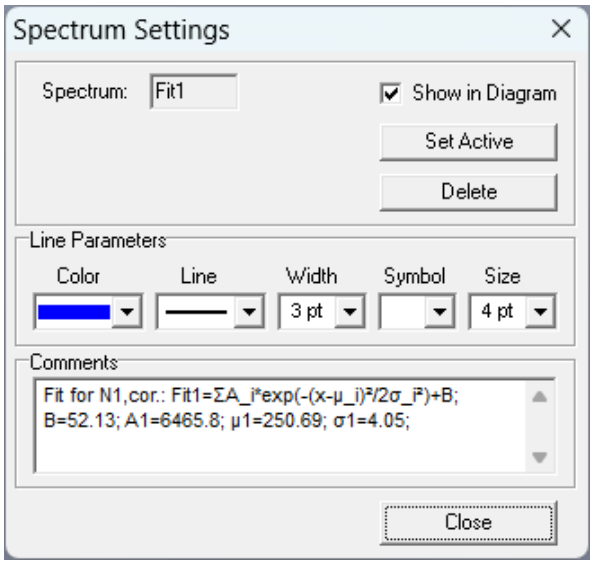


Fig. 52. *"Spectrum Settings"* dialog box displaying the parameters of the fitting Gaussian curve.

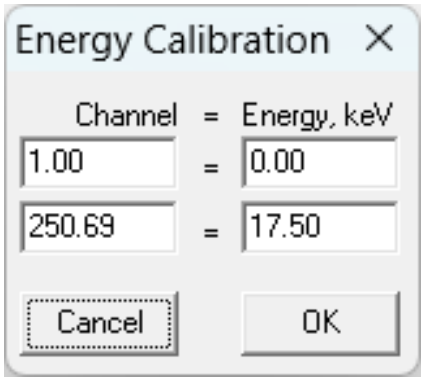


Fig. 53. *"Energy Calibration"* dialog box, in which energy values are assigned to any two channels.

along the *X* axis of the graph). The *"Energy Calibration"* dialog box then opens, in which energy values can be assigned to any two channels (see Fig. 53). Since the first channel always corresponds to zero energy, enter the number *"1"* in the top *"Channel"* field and the number *"0"* in the top *"Energy"* field. In the two bottom fields, enter the previously calculated centroid and the corresponding energy. Then click the *"OK"* button. The *X* axis of the graph changes: instead of channel numbers, the

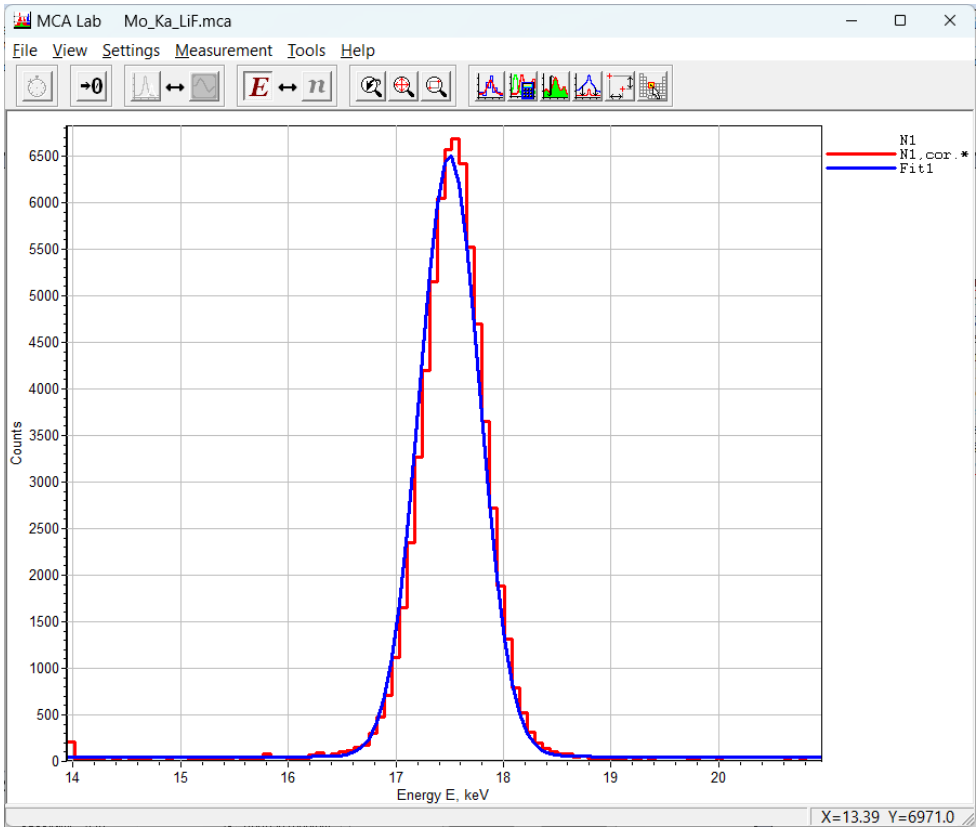


Fig. 54. *"MCA Lab"* program window after calibration (the previous window is shown in Fig. 51).

corresponding energy values are shown, and the *X*-axis title becomes *"Energy E, keV"* (see Fig. 54).

5. In each of the remaining nine spectra (one primary radiation spectrum and eight X-ray fluorescence spectra), two peaks are fitted: in the XRF spectrum of steel, these are the Fe and Cr K_α peaks (in this spectrum, K_β peaks are not visible because the K_β lines of low-atomic-number elements are too close to the K_α lines to be resolved), whereas in the other eight spectra, these are the K_α and K_β peaks of the element with the highest atomic number. Energy calibration does not need to be repeated, because the program saves it automatically. Therefore, the centroid values displayed in the *"Spectrum Settings"* dialog box (see Fig. 52) will have the meaning of energies. Both peaks must be fitted simultaneously, i.e., as the sum of two Gaussian functions. Therefore, when performing step 3(b), mark not one peak, but two peaks, and choose the boundaries of the fitting interval so that both peaks lie entirely within that interval. Since the peaks are close to each other, when performing step 3(a), the "background" must be marked before the first peak and after the second peak, but not between them. The exceptions are the silver and tin spectra, in which the K_α and K_β peaks practically do not overlap, so the background may also be marked between those peaks. In the *"Spectrum Settings"* window (see Fig. 52), the centroids of both peaks are denoted by *"$\mu 1$"* and *"$\mu 2$"*, where the number *"1"* or *"2"* is the ordinal number of the peak depending on the order in which the peaks were marked when performing step 3(b). The "first" peak is the one that was marked earlier (thus, the peak centroids are not necessarily numbered in order of increasing energy). **Note:** In spectra obtained using samples contained in plastic vials (i.e., powders and silver wire), a weak peak corresponding to Compton scattering of Mo K_α radiation in the plastic layer may be visible (see, e.g., Fig. 45). When the intensities of the main peaks are particularly low (e.g., in the case of tin), a peak corresponding to coherent scattering of Mo K_α radiation may also be visible next to them.

6. Display all ten experimental spectra in graphs (a separate graph for each spectrum). In the graph showing the spectrum of radiation reflected from the LiF crystal, the channel number must be plotted along the *X* axis, whereas in the other nine graphs, the photon energy must be plotted along the *X* axis. The graphs must contain two curves: the raw data (without background subtraction) and the sum of the fitting Gaussian functions (with background subtracted).

7. Tabulate the photon energy and wavelength values corresponding to the centroids of all peaks (ten K_α peaks and eight K_β peaks).

8. Plot the dependence of the square root of the reciprocal wavelength on the atomic number Z (i.e., the Moseley diagram) on a graph as individual points (without connecting them with a line). The Moseley diagrams corresponding to the K_α and K_β lines must be plotted on the same graph.

9. Fit each of the two dependences with the theoretical function (3.4.8), i.e.,

$$\frac{1}{\sqrt{\lambda}} = \frac{1}{C}Z - \frac{\sigma}{C} = A + BZ, \tag{8.2.1}$$

where σ is the screening constant, the coefficient $1/C$ is defined in Section 3.4 ($1/C$ is directly proportional to $R^{1/2}$), and the line coefficients A and B are defined as follows:

$$B = 1/C, \tag{8.2.2}$$

$$A = -\sigma/C = -\sigma B. \tag{8.2.3}$$

The fitting lines must be displayed on the same graph as the experimental Moseley diagrams. Linear fitting can be performed using any data analysis program or "by hand"—using the formulas provided in Section 8.3. Then, using Eqs. (8.2.2) and (8.2.3) and the expression for the coefficient C given in Section 3.4, calculate the Rydberg constant R and the screening constant σ, as well as their uncertainties. As explained in Section 8.3, the standard deviations of the line coefficients (which are expressed by Eqs. (8.3.5a) and (8.3.5b)) must be multiplied by the Student's t-coefficient corresponding to a 95 % confidence level and $n - 2$ degrees of freedom, where n is the number of fitted points. Since in this case $n = 10$ or 8, the number of degrees of freedom is equal to 8 or 6. Student's t-coefficients are listed in Table 6.

Since the Rydberg constant R and the screening constant σ are expressed in terms of the line coefficients A and B, the uncertainty bounds of the constants R and σ must be calculated using the general error propagation rules for functions of one or several variables. In the case of several variables, the rule is as follows: if a quantity f is a function of independent arguments $x_1, x_2, \dots, x_N$, then the uncertainty of the quantity f is equal to

$$\Delta f = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 (\Delta x_1)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 (\Delta x_2)^2 + \cdots + \left(\frac{\partial f}{\partial x_N}\right)^2 (\Delta x_N)^2}, \tag{8.2.4}$$

where $\Delta x_1, \Delta x_2, \dots, \Delta x_N$ are the uncertainties of the quantities $x_1, x_2, \dots, x_N$. Multiplication by the Student's t-coefficient must be performed only once. That is, if the Student's t-coefficient was used when calculating $\Delta x_1, \Delta x_2, \dots, \Delta x_N$, then Δf must not be multiplied by the Student's t-coefficient again.

10. Compare the experimental Rydberg and screening constants with their values given in the literature (see Section 3.4). Discuss the causes of the discrepancies.

11. Compare the experimental energies of the K_α and K_β lines with the exact values given in the literature. Calculate the maximum and root-mean-square relative deviations. The "relative" deviation of the experimental energy E_{exp} from the true energy E_{t} is equal to $(E_{\text{exp}} - E_{\text{t}})/E_{\text{t}}$. The exact energies of the K_α and K_β lines are given, for example, in Table 5. That table lists the energies of both components of the K_α line. The higher energy corresponds to the $K_{\alpha 1}$ component (also denoted K-L_3), and the lower energy corresponds to the $K_{\alpha 2}$ component (also denoted K-L_2). The average energy of the K_α line must be calculated according to Eq. (3.5.1). If most of the relative deviations have the same sign (in other words, if the errors are not random, but systematic) or if there are deviations that do not follow the general trend (e.g., are much larger or of the opposite sign compared to the majority), then possible causes of those systematic errors should be discussed. For example, the experimental energies of some peaks may be slightly shifted due to their overlap with other nearby peaks (in such a case, the specific extraneous peaks affecting the measurement results should be indicated).

Table 5. Photon energies of K–L and K–M lines (keV)[6]

No.	Elem.	K-L_3	K-L_2	K-M	No.	Elem.	K-L_3	K-L_2	K-M
3	Li	0.0543			48	Cd	23.1736	22.9841	26.0955
4	Be	0.1085			49	In	24.2097	24.0020	27.2759
5	B	0.1833			50	Sn	25.2713	25.0440	28.4860
6	C	0.2770			51	Sb	26.3591	26.1108	29.7256
7	N	0.3924			52	Te	27.4723	27.2017	30.9957
8	O	0.5249			53	I	28.6120	28.3172	32.2947
9	F	0.6768			54	Xe	29.7790	29.4580	33.6240
10	Ne	0.8486	0.8486		55	Cs	30.9728	30.6251	34.9869
11	Na	1.0410	1.0410	1.0711	56	Ba	32.1936	31.8171	36.3782
12	Mg	1.2536	1.2536	1.3022	57	La	33.4418	33.0341	37.8010
13	Al	1.4867	1.4863	1.5575	58	Ce	34.7197	34.2789	39.2573
14	Si	1.7400	1.7394	1.8359	59	Pr	36.0263	35.5502	40.7482
15	P	2.0137	2.0127	2.1391	60	Nd	37.3610	36.8474	42.2713
16	S	2.3078	2.3066	2.4640	61	Pm	38.7247	38.1712	43.8260
17	Cl	2.6224	2.6208	2.8156	62	Sm	40.1181	39.5224	45.4130
18	Ar	2.9577	2.9556	3.1905	63	Eu	41.5422	40.9019	47.0379
19	K	3.3138	3.3111	3.5896	64	Gd	42.9962	42.3089	48.6970
20	Ca	3.6917	3.6881	4.0127	65	Tb	44.4816	43.7441	50.3820
21	Sc	4.0906	4.0861	4.4605	66	Dy	45.9984	45.2078	52.1190
22	Ti	4.5108	4.5049	4.9318	67	Ho	47.5467	46.6997	53.8770
23	V	4.9522	4.9446	5.4273	68	Er	49.1277	48.2211	55.6810
24	Cr	5.4147	5.4055	5.9467	69	Tm	50.7416	49.7726	57.5170
25	Mn	5.8988	5.8877	6.4905	70	Yb	52.3889	51.3540	59.3700
26	Fe	6.4038	6.3908	7.0580	71	Lu	54.0698	52.9650	61.2830
27	Co	6.9303	6.9153	7.6494	72	Hf	55.7902	54.6114	63.2340
28	Ni	7.4782	7.4609	8.2647	73	Ta	57.5320	56.2770	65.2230
29	Cu	8.0478	8.0278	8.9053	74	W	59.3182	57.9817	67.2443
30	Zn	8.6389	8.6158	9.5720	75	Re	61.1403	59.7179	69.3100
31	Ga	9.2517	9.2248	10.2642	76	Os	63.0005	61.4867	71.4130
32	Ge	9.8864	9.8553	10.9821	77	Ir	64.8956	63.2867	73.5608
33	As	10.5437	10.5080	11.7262	78	Pt	66.8320	65.1120	75.7480
34	Se	11.2224	11.1814	12.4959	79	Au	68.8037	66.9895	77.9840
35	Br	11.9242	11.8776	13.2914	80	Hg	70.8190	68.8950	80.2530
36	Kr	12.6490	12.5980	14.1120	81	Tl	72.8715	70.8319	82.5760
37	Rb	13.3953	13.3358	14.9613	82	Pb	74.9694	72.8042	84.9360
38	Sr	14.1650	14.0979	15.8357	83	Bi	77.1079	74.8148	87.3430
39	Y	14.9584	14.8829	16.7378	84	Po	79.2900	76.8620	89.8000
40	Zr	15.7751	15.6909	17.6678	85	At	81.5200	78.9500	92.3000
41	Nb	16.6151	16.5210	18.6225	86	Rn	83.7800	81.0700	94.8700
42	Mo	17.4793	17.3743	19.6083	87	Fr	86.1000	83.2300	97.4700
43	Tc	18.3671	18.2508	20.6190	88	Ra	88.4700	85.4300	100.1300
44	Ru	19.2792	19.1504	21.6568	89	Ac	90.8840	87.6700	102.8500
45	Rh	20.2161	20.0737	22.7236	90	Th	93.3500	89.9530	105.6090
46	Pd	21.1771	21.0201	23.8187	91	Pa	95.8680	92.2870	108.4270
47	Ag	22.1629	21.9903	24.9424	92	U	98.4390	94.6650	111.3000

[6] Data from https://xdb.lbl.gov/Section1/Table_1-2.pdf.

8.3. Linear Fitting

The objective of fitting a straight line ("linear fitting") is to calculate the coefficients A and B of the equation of a straight line

$$y = A + B \cdot x \tag{8.3.1}$$

and their uncertainties by the method of least squares. The essence of the method of least squares is as follows. Suppose we have a set of measurement data consisting of argument x values $x_1, x_2, \dots, x_{n-1}, x_n$ and the corresponding values of the function $y(x)$; n is the number of measurements. Let us denote the values of the function by $y_1, y_2, \dots, y_n$. The theoretical y value corresponding to a given argument value x_k is a function of the unknown coefficients A and B (see Eq. (8.3.1)), so we can write $y(x_k) = y(x_k; A, B)$ $(k = 1, 2, \dots, n)$. The problem of calculating the coefficients A and B is formulated as follows. The most probable values of the unknown coefficients A and B are those that correspond to the absolute minimum of the expression

$$F(A,B) \equiv \sum_{k=1}^{n} \left[y(x_k; A, B) - y_k \right]^2 . \tag{8.3.2}$$

Expression (8.3.2) is the sum of the squares of the deviations of the theoretical values from the measured values (hence the name "method of least squares"). This expression always has a minimum at some strictly defined values of A and B. However, even if the form of the theoretical function $y(x)$ corresponds exactly to the true relationship between the measured quantities y and x, these optimal values of A and B, which correspond to the minimum of the sum of squares F, do not necessarily coincide with the true values of A and B. This may happen, for example, due to measurement errors. By the method of least squares, one can calculate only the most probable values of the coefficients A and B.

Everything stated above about the method of least squares applies not only when the theoretical function $y(x)$ is a straight line. Regardless of the form of this function and the number of unknown coefficients, an expression of the form (8.3.2) must be minimized. However, when $y(x)$ is a straight line, this problem can be solved analytically (i.e., A and B can be expressed by elementary algebraic expressions), whereas in the case of a nonlinear function, this problem can be solved only numerically (by the method of successive approximations, using a computer).

If $y(x)$ is the linear function (8.3.1), then the sum of squares (8.3.2) has the following form:

$$F(A,B) \equiv \sum_{k=1}^{n} (A + Bx_k - y_k)^2 = nA^2 + B^2 \sum_{k=1}^{n} x_k^2 + \sum_{k=1}^{n} y_k^2 + 2AB \sum_{k=1}^{n} x_k - 2A \sum_{k=1}^{n} y_k - 2B \sum_{k=1}^{n} x_k y_k. \tag{8.3.3}$$

As is known from mathematical analysis, at the minimum point of a function of several variables, its partial derivatives with respect to all variables are equal to zero. Setting the partial derivatives of expression (8.3.3) with respect to A and B equal to zero yields a system of two linear algebraic equations whose unknowns are the coefficients A and B. The solution to this system of equations is

$$B = \frac{n \sum\limits_{k=1}^{n} x_k y_k - \left(\sum\limits_{k=1}^{n} x_k \right) \left(\sum\limits_{k=1}^{n} y_k \right)}{n \sum\limits_{k=1}^{n} x_k^2 - \left(\sum\limits_{k=1}^{n} x_k \right)^2}; \tag{8.3.4a}$$

$$A = \frac{1}{n} \sum_{k=1}^{n} y_k - \frac{B}{n} \sum_{k=1}^{n} x_k. \tag{8.3.4b}$$

The coefficient B is called the directional coefficient or the "slope" of the line. The coefficient A specifies the y value when $x = 0$ (the intercept). The root-mean-square deviations ("empirical standard deviations") of the coefficients A and B are calculated using the formulas

$$\Delta A = \sqrt{\frac{F_{\min}}{n(n-2)} \left(1 + \frac{\bar{x}^2}{D_x} \right)}, \tag{8.3.5a}$$

$$\Delta B = \sqrt{\frac{F_{\min}}{n(n-2)D_x}}, \tag{8.3.5b}$$

where $F_{\min}$ is the minimum value of the sum of squares (8.3.3) (i.e., the sum of squares when the coefficients B and A are equal to their optimal values (8.3.4a) and (8.3.4b)), $\bar{x}$ is the mean of the argument values:

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k, \tag{8.3.6}$$

and D_x is the variance of the argument values:

$$D_x = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{\sum_{k=1}^{n} x_k^2}{n} - \bar{x}^2. \tag{8.3.7}$$

If the true relationship between the measured quantity y and the argument x is linear, and the measurement errors are independent and distributed according to a Gaussian distribution whose width does not depend on the point index, then multiplying the quantities ΔA and ΔB by the Student's t-coefficient corresponding to $n - 2$ degrees of freedom and a chosen confidence level α yields the half-width of the confidence interval ("error bound") of the corresponding coefficient. The center of the confidence interval coincides with the experimental value (e.g., the coefficients A and B calculated according to Eqs. (8.3.4a,b)). If the measurement errors are random and distributed according to a Gaussian distribution, then it can be stated that there is a probability α that the true value belongs to the interval defined in this way (which is precisely why this interval is called a "confidence interval"). In other words, having performed a large number N of identical experiments under identical conditions and having calculated the bounds of the corresponding confidence interval for each experiment, the true value will belong to the confidence interval in αN cases and will not belong to it in the remaining $(1 - \alpha)N$ cases. When determining the Student's t-coefficient, when there are two unknown coefficients, the number of degrees of freedom is equal to $n - 2$, rather than n, because the number of *independent* points is two less than the total number of points (since the conditions (8.3.4a) and (8.3.4b) exist). Thus, the intervals of values of the line coefficients A and B to which the true (unknown) values of those coefficients belong with probability α are

$$A' - t_{\alpha,n-2} \cdot \Delta A < A < A' + t_{\alpha,n-2} \cdot \Delta A, \tag{8.3.8a}$$

$$B' - t_{\alpha,n-2} \cdot \Delta B < B < B' + t_{\alpha,n-2} \cdot \Delta B, \tag{8.3.8b}$$

where A' and B' are the values obtained by the fitting method (they are given by Eqs. (8.3.4a,b)), $t_{\alpha,n-2}$ is the Student's t-coefficient corresponding to the confidence level α and $n - 2$ degrees of freedom, and ΔA and ΔB are expressed by Eqs. (8.3.5a) and (8.3.5b). As can be seen from the inequalities (8.3.8a,b), the error bound of coefficient A or B is $t_{\alpha,n-2} \cdot \Delta A$ or $t_{\alpha,n-2} \cdot \Delta B$, respectively. Measurement results are customarily presented as follows:

$$A = A' \pm t_{\alpha,n-2} \cdot \Delta A, \tag{8.3.9a}$$

$$B = B' \pm t_{\alpha,n-2} \cdot \Delta B. \tag{8.3.9b}$$

The notation (8.3.9a,b) is equivalent in meaning to the notation (8.3.8a,b).

It should be borne in mind that the statistical interpretation of the confidence interval presented above is meaningful only when measurement errors are random. If the errors are systematic (e.g., when a nonlinear experimental dependence is fitted with a straight line), then the interval calculated according to Eqs. (8.3.8a,b) is no longer a "confidence" interval in the true sense of the word (although it may be called so). Then an interval calculated in this way can be used only as a formal precision criterion when comparing results of different experiments performed under identical conditions (i.e., a narrower confidence interval corresponds to better measurement precision).

When analyzing measurement data, the confidence level $0.95 = 95\ \%$ is most frequently used. When the number of measurements n is large, the Student's t-coefficient corresponding to a 95 % confidence level is approximately equal to 2. The values of Student's t-coefficients corresponding to various confidence levels and various numbers of degrees of freedom are given in Table 6.

Most data analysis programs feature a linear fitting function. For example, the program **Origin** or the Excel function LINEST calculates both the coefficients (8.3.4a,b) and the root-mean-square deviations (8.3.5a,b). However, those programs do not multiply those deviations by the Student's t-coefficient. The final fitting results should not depend on the program used, because all programs use the exact same formulas, i.e., Eqs. (8.3.4a,b) and (8.3.5a,b).

Table 6. Student's *t*-coefficients corresponding to various confidence levels and various numbers of degrees of freedom

Degrees of freedom \ Confidence level	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	**12.71**	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	**4.303**	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	**3.182**	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	**2.776**	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	**2.571**	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	**2.447**	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	**2.365**	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	**2.306**	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	**2.262**	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	**2.228**	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	**2.201**	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	**2.179**	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	**2.160**	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	**2.145**	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	**2.131**	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	**2.120**	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	**2.110**	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	**2.101**	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	**2.093**	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	**2.086**	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	**2.080**	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	**2.074**	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	**2.069**	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	**2.064**	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	**2.060**	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	**2.056**	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	**2.052**	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	**2.048**	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	**2.045**	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	**2.042**	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	**2.021**	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	**2.009**	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	**2.000**	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	**1.990**	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	**1.984**	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	**1.980**	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	**1.960**	2.326	2.576	2.807	3.090	3.291