DE GRUYTER

GRADUATE

Hartmut Zabel

MEDICAL PHYSICS

VOLUME 2: RADIOLOGY, LASERS, NANOPARTICLES AND PROSTHETICS

Hartmut Zabel

Medical Physics

De Gruyter Graduate

Hartmut Zabel

Medical Physics

Volume 2: Radiology, Lasers, Nanoparticles and Prosthetics

DE GRUYTER

Physics and Astronomy Classification Scheme 2010
87.59.B-, 87.53.-j, 87.53.Bn, 87.57.uq, 87.50.-a, 87.53.-j, 87.55.-x, 87.56.-v, 87.56.bg, 75.75.Fk, 78.67.Bf, 87.55.N-, 42.62.Be, 87.63.lt, 79.20.Ds, 87.53.Jw, 82.50.-m, PET, 87.57.uk, SPECT, 87.57.uh

Author
Prof. Dr. Dr. h. c. Hartmut Zabel
Ruhr-University Bochum
Institute for Experimental Physics
44780 Bochum
hartmut.zabel@ruhr-uni-bochum.de

ISBN 978-3-11-055310-9
e-ISBN (PDF) 978-3-11-055311-6
e-ISBN (EPUB) 978-3-11-055317-8

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

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

Cover image: Photoprofi30/iStock/Getty Images Plus
Typesetting: PTP-Berlin, Protago-TEX-Production GmbH, Berlin
Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

These volumes are dedicated to my grandchildren
Hanna, Henriette, Katharina, Niklas, Raphael, and Sonja.
Now it’s your turn again.

Preface

According to the American Association of Physicists in Medicine, Medical Physics is an applied branch of physics concerned with the application of concepts and methods of physics to diagnostics and therapeutics of human diseases. In recent years many academic institutions have started offering Medical Physics in their Bachelor, Master, and PhD programs in response to the expanding job market and increasing demand for professionals in this discipline. The present Volumes 1 and 2 on Medical Physics are introductory textbooks intended to guide bachelor students during their first semesters through the broad range of topics relevant in this field and providing a first overview. There are many excellent books available that specialize on more specific topics of Medical Physics, such as imaging and radiation therapy. However, there is rarely a single text covering the entire field, including the physics of the body, imaging methods, radiology, laser methods, nanomedicine, and prosthetics. Because of the multitude of topics to be covered and the rapid development of the field chances are that such a script might be incomplete or partially outdated by the time of publication. Nevertheless, these two volumes try to provide a reasonably balanced overview as well as guidance to more specialized and more comprehensive literature. Each chapter is rounded off with a summary, references, hints for further reading and useful webpages, and with questions and answers to the respective chapters. Historical remarks are kept at a minimum to keep the text at a reasonable total length.

This textbook does not provide any medical interpretation or any medical judgement. It is entirely focused on methods and procedures from a physical point of view. Furthermore, this text is neither an introduction into elementary physics nor into medical physiology. It tries to connect the principles of physics with the functionality of the body and with physical methods used for diagnostics and therapeutics. Furthermore, medical physics is located somewhere in between medical sciences and engineering with regard to diagnostic and therapeutic equipment. As such medical physics acts as a mediator between these disciplines. While the physical principles of medical equipment and instrumentation are introduced, for more detailed engineering aspects of hardware and software developments references are provided for further reading. Throughout the text it is assumed that the reader has a basic understanding of physics corresponding to about one to two years of college physics.

Physics and medicine have their own distinct terminology. Here we will mainly use a “physics language”, but medical and physiological terms will not be avoided. In fact it is important for students of medical physics to know at least the most important terms in order to be able to communicate with physicians in medical practice. Therefore in these volumes medical and physiological terms will be introduced and defined whenever they are used. In addition, a list of acronyms and definitions is provided in the appendix. In physiology it is still common practice to use non-MKS (SI) units, such as mmHg for pressure. In this text we strictly adhere to the SI system with one

DOI 10.1515/9783110553116-001

exemption: for practical reasons we make use of the energy unit ‘electron volt’ (eV) in the context of electromagnetic and particle radiation.

The present Volume 2 contains 15 chapters divided into three parts. The first seven chapters in Part A are devoted to an understanding of various types of radiation, how it is generated, how it interacts with matter, and how healthy tissue can be protected against unsafe radiation levels. The last three chapters of this part provide an introduction to the three main imaging modalities using ionizing radiation: x-rays, γ-rays, and positron annihilation. Part B starts in Chapter 8 with a very short introduction to the cell cycle, cancer development, and to the main concepts that radiation treatment of carcinomas are based on. Chapter 8 is followed by three chapters discussing external beam radiation therapy with x-rays, protons, and neutrons, and one chapter on radiation therapy in direct contact with cancerous tissue (brachytherapy). The last three chapters in Part C complement methods and treatments described in the previous chapters. Lasers introduced in Chapter 13 are not only used for various types of spectroscopies but also for medical treatments in oncology, dermatology, and ophthalmology. The application of functionalized nanoparticles described in Chapter 14 is intended for improved targeting of cancerous tissue, enhanced imaging contrast, and for more local treatment capabilities. Nanomedicine is a highly active and fast developing field. Research and developments in nanomedicine carry huge promise for future targeted cures. The last chapter of this volume is devoted to various types of prosthesis for limbs, stenosis, and for the visual and auditory senses.

Each chapter is concluded by a summary of the main points. A number of contextual questions pertaining to each chapter are listed in the appendix. They are complemented by corresponding answers as a guide and for providing an incentive for further studies.

As for the first volume, I hope that also the second volume may turn out to be a useful companion to students enrolled in a course on Medical Physics during their first semesters.

Bochum and Mainz, January 2017

Acknowledgments

Writing this text would not have been possible without the help of many. First, I am highly thankful to authors and publishers who make their work freely accessible in open access journals, and to the organization Wikipedia including their contributors who do an amazing job for the benefit of those who are eager to know. Furthermore, I am deeply indebted to colleagues and experts who devoted their precious time to critically reading and correcting various chapters. In particular I would like to thank, in alphabetic order, Professor Ping Ao and Dr. Xiaomei Zhu, Shanghai Jiao Tong University, for sharing their deep insight into the causes and development of cancer; Dr. Andrea Denker, Helmholtz Zentrum Berlin, for pointing out intricacies of proton therapy and eliminating misconceptions; Professor Helmut Ermert, University of Erlangen, for sharing his expertise of ultrasonic imaging; Professor Ulf Eysel, Ruhr-University Bochum, for critically reading and correcting the chapter on the mechanism of vision; Professor Werner Havers, University Clinic Essen, for improving the chapter on x-ray radiotherapy; Dr. Margot Jonas, University Clinic Knappschaft Bochum, for valuable suggestions concerning scintigraphy and positron annihilation tomography, and for sharing most instructive images on the comparison of x-ray imaging versus PET; Professor Jan Meijer, University Leipzig, for important amendments to the chapter on dosimetry; Professor Werner Meyer and Dr. Gerhard Reicherz, Ruhr-University Bochum, for their critical reading of and valuable contributions to the chapter on magnetic resonance imaging; Professor Winfried Petry and Professor Franz Wagner, Technical University Munich, for their modifications of the chapters on neutron radiotherapy and on nuclei and isotopes, and for providing proper numbers of neutron energies and dose distributions; Professor Franz Pfeiffer, Technical University Munich, for his improvements to the chapter on x-ray radiography and sharing images on phase contrast imaging; Professor Lutz Pott, Ruhr-University Bochum, for critically reading and rectifying the chapter on electrophysical aspects of the heart; Professor Katharina Theis-Bröhl, University Bremerhaven, for valuable suggestions towards improvement of the chapter on x-ray generators; Dr. Sebastian Tripple, CFEL Hamburg, for his expert input and revisions of the chapter on laser applications. Last but not least, I am grateful to my wife Rosemarie for reading and correcting all chapters with respect to spelling and solecisms. For any remaining inaccuracies or ambiguities I take the sole responsibility.

Furthermore, I would like to thank Professor Mathias Kläui for his hospitality and that provided by the Mainz Graduate School of Excellence at the Johannes Gutenberg University Mainz while this text was being written. My hearty thanks also go to Astrid Seifert and Nadja Schedensack of de Gruyter Verlag, who patiently waited for the submission of this manuscript, and Anne Hirschelmann and her staff, who did a superb job on layout and graphical enhancement of the figures.

DOI 10.1515/9783110553116-002

Contents

Part A: **Radiography**

Part B: **Radiotherapy**

Part C: **Diagnostics and therapeutics beyond radiology**

Part A: **Radiography**

1 X-ray generation

1.1 Introduction

In 1895 Wilhelm Conrad Röntgen discovered a new type of radiation, which he called x-rays. He was uncertain what they were, but he noticed that they were able to penetrate opaque matter. To demonstrate this he took x-ray images of his wife's hand, the first x-ray images ever, which made him and the new method instantly famous worldwide. Since then x-rays have been known to the general public mainly for their medical use. Röntgen was the first awardee of the Physics Nobel Prize in 1901. However, medical x-ray imaging is only one of many other uses of x-rays, the others include x-ray scattering, x-ray spectroscopy, and x-ray microscopy in all fields of science and technology. X-rays from low energies to high energies are so omnipresent that a world without x-rays is hard to imagine. Without x-rays we probably would not know about the helical structure of DNA, the complex folding of proteins such as myoglobin and hemoglobin, the rich structure and functionality of ribonucleic acid (RNA), and many others.

X-rays are electromagnetic (EM) waves with energies ranging from 50 eV up to several MeV, corresponding to wavelengths λ from 25 nm (50 eV) down to 0.0012 nm (1 MeV). The conversion factor derived from the equation for the energy of photons $E = hf = hc/\lambda$ is:

$$\lambda(\text{nm}) = \frac{1240\,\text{eV nm}}{E(\text{eV})}$$

where $h = 6.623 \times 10^{-34}$ Js is the Planck constant, $c = 299\,792\,458\,\text{m/s} \approx 3 \times 10^{8}\ \text{m/s}^{-1}$ is the vacuum velocity of EM waves, and f is the frequency.

At the lower energy end x-rays overlap with far ultraviolet radiation. At the upper energy scale they overlap with γ-radiation. It is not primarily the energy or the respective wavelength that characterizes x-rays; it is the method by which x-rays are produced. Three kinds of x-ray production can be distinguished:

1. *bremsstrahlung*, radiation produced by deacceleration of high energy electrons;
2. *characteristic radiation*, occurring after excitation of core shell electrons of atoms;
3. *synchrotron radiation* emitted by radial acceleration of electrons in high energy storage rings.

One can simplify this characterization and reduce it to two main effects: acceleration and excitation. Bremsstrahlung and synchrotron radiation are due to acceleration of electrons: deacceleration (stopping) of high energy electrons in a target in one case and radial acceleration of electrons on a circular orbit in the other case. Characteristic radiation is due to excitation of atomic core shell electrons. γ-radiation, although overlapping in energy with x-rays, is reserved as decay product of radioactive isotopes.

DOI 10.1515/9783110553116-003

For medical x-ray diagnostics (radiography) and for x-ray cancer treatment (radiotherapy) only bremsstrahlung is used. Their specifications are, however, very different. Hence the same x-ray equipment cannot be employed for both applications. X-ray diagnostics requires bremsstrahlung up to about 150 keV, whereas x-ray radiotherapy entails x-ray energies up to 25 MeV. In this chapter we will introduce basic concepts of x-ray production for both applications, radiography and radiotherapy. X-ray radiography is then presented in Chapter 5 and x-ray radiotherapy is discussed in Chapter 9.

1.2 X-ray production

1.2.1 Bremsstrahlung

Standard *x-ray tubes*, which are frequently used in research laboratories, use a high voltage difference between cathode and anode for accelerating free electrons over a short distance in vacuum. These x-ray glass tubes are permanently evacuated and sealed off like traditional light bulbs. The principle features of such a tube are shown in Fig. 1.1. The cathode is connected to a high negative voltage supply between −10 kV and −100 kV and the anode is grounded. A tungsten filament in the cathode is heated to very high temperatures just below melting temperature of the wire in order to generate free electrons via thermionic emission. The current going through the filament of the cathode I_{cathode} controls the filament temperature and consequently also the rate of electrons emitted into the vacuum tube. Electrons entering at the high negative potential into the vacuum are immediately accelerated towards the anode. A cap with a small aperture surrounding the filament, called a Wehnelt cylinder, acts as an electrostatic lens that keeps the electrons from straying. Obviously there is a close connection between the current heating the filament I_{cathode} and the anode current I_{anode} hitting the target.

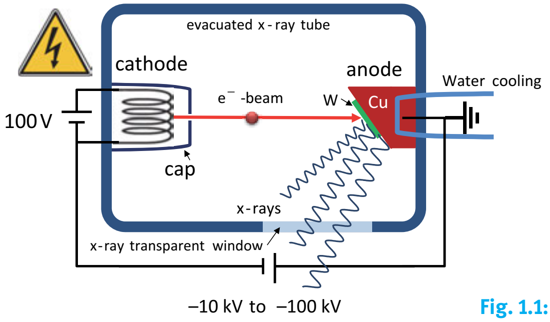


Fig. 1.1: Schematic of an x-ray tube.

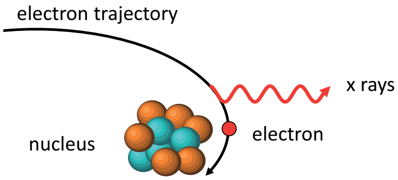


Fig. 1.2: Generation of bremsstrahlung in the Coulomb potential of target nuclei.

The anode (target) consists of a water cooled Cu block coated with some metal film. Tungsten is usually chosen for medical applications. In the target material the free electrons are rapidly retarded by being trapped in the Coulomb potential of the target nuclei (see Fig. 1.2). While circulating around target nuclei, the electrons emit electromagnetic dipole radiation. They may emit just one photon, carrying away the total kinetic energy of the electron, or they may split their energy into several photons.

The deacceleration of high energy electrons in the anode causes emission of x-ray photons with a very broad spectral distribution of energies or wavelengths. This radiation is called *bremsstrahlung*. The German expression, common in the literature, can be translated as “deacceleration radiation”. Bremsstrahlung is polychromatic, lacks phase coherence, and is divergent. Such a radiation source is known as an incoherent broadband photon source, sometimes also referred to as “white x-ray source” in analogy to traditional light bulbs. The total integrated intensity over all energies is proportional to the product of the anode current I_{anode} and the square of the voltage difference V_{acc} between cathode and anode (accelerating voltage):

$$I_{\text{photon}} \propto I_{\text{anode}} \times V_{\text{acc}}^2 .$$

The intensity per energy interval versus photon energy for four different accelerating voltages is plotted in Fig. 1.3. The intensity of the unfiltered radiation can be approximated by a linear equation, known as Kramer’s equation (dashed lines in Fig. 1.3):

$$I_{\text{photon}}(E) = kZ(E_{\max} - E) .$$

Here Z is the atomic number of the target material, k is a constant, and $E_{\max}$ is the cut-off energy:

$$E_{\max} = hf_{\max} = eV_{\text{acc}} .$$

This cut-off energy is defined by the maximum energy that a photon can attain when the complete kinetic energy of the accelerated electron is converted into electromagnetic wave energy. The SI unit of the accelerating voltage is Volts (V) or kilovolts (kV). The unit for the maximum voltage applied across an x-ray tube is termed in medical literature *kilovoltage peak* (kVp). The accelerating voltage V_{acc} determines the hardness of the radiation. The higher the potential difference is between cathode and anode, the higher the energy of the x-ray photons that are produced. High energy photons penetrate matter better than low energy photons.

For taking x-ray radiographs of the body an Al filter is always inserted that absorbs the low energy part of the x-ray bremsstrahlung spectrum. The low energy x-ray photons would be preferentially absorbed in the skin and enhance the dose without contributing to the image quality. The modified bremsstrahlung spectrum with an Al filter of 2 mm thickness inserted is shown in Fig. 1.3 by the bold lines for several kVp.

When operating an x-ray machine, the user has control over kVp, the tube current, and the exposure time. Typical values for taking an x-ray radiograph are: voltage between 80–140 kVp, anode current of 500 mA, and exposure time of 50–100 ms. This yields a tube energy of 2–7 kW s. Production of x-ray photons by means of bremsstrahlung is not very efficient. 99 % of the electron energy is lost in heat production and only 1 % is converted into photons. Therefore x-ray tubes must be cooled very well.

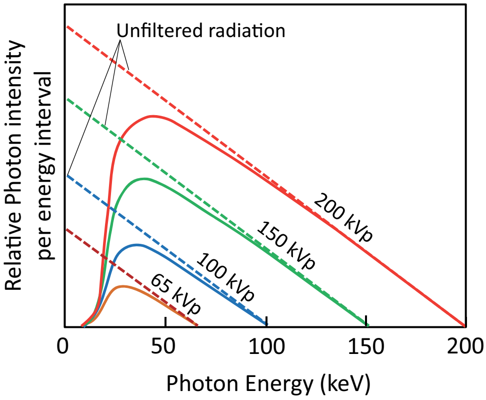


Fig. 1.3: Bremsstrahlung spectrum for four different accelerating voltages and assuming a tungsten target. Note the intensity increase with increasing kVp and the linear intensity drop terminating at the maximum photon energy. The dashed lines refer to unfiltered x-ray intensity, the bold lines to x-ray intensity with a 2 mm Al filter inserted.

1.2.2 Characteristic radiation

When electrons strike a target, a second process of x-ray generation takes place: *characteristic radiation* is generated simultaneously to bremsstrahlung, i.e., characteristic for the target material used. The incoming electron may collide with an electron of the inner atomic shell, for instance of the K-shell. If the energy of the incoming electron E_{in} is higher than the binding energy of the core shell electron E_{K}, it may kick out the core electron leaving a hole in the K-shell, as depicted in Fig. 1.4. This core hole is sequentially filled with an electron from the next higher shell, and so on until after a cascade of transitions the hit atom comes back to the ground state. When an electron in an atom makes a transition from a higher to a lower shell, it will emit the difference in binding energy in terms of electromagnetic waves, i.e., via dipole radiation. This radiation is called *characteristic x-ray radiation*.

As an example we assume that the kinetic energy of the incoming electron E_{in} is high enough to kick out one of the two electrons residing in the K-shell, schematically shown in Fig. 1.4. The former K-electron will leave the atom and thermalize on the path

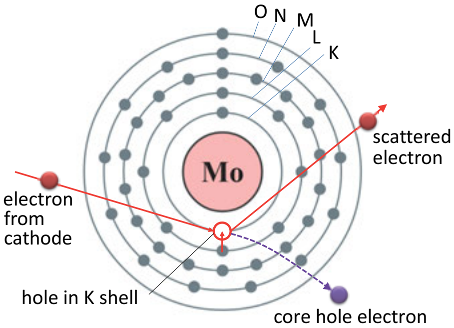


Fig. 1.4: A highly energetic incident electron arriving from the cathode kicks out a core electron in a target (anode) atom, creating a hole in the K-shell. An electron from the L-shell fills the hole and simultaneously emits an x-ray photon.

through the solid target. The scattered electron has the final energy:

$$E_{\mathrm{final}} = E_{\mathrm{in}} - (E_{\mathrm{K}} + E_{\mathrm{kin}}),$$

where E_{in} is the kinetic energy of the cathode electron in the target, E_{K} is the binding energy of the K-shell electron and E_{kin} is its kinetic energy after collision. The hole in the K-shell is filled with an electron from the L-shell, giving off its extra energy as a photon with energy:

$$E_{\mathrm{K}\alpha}^{\mathrm{photon}} = hf_{\mathrm{K}\alpha} = E_{\mathrm{K}} - E_{\mathrm{L}}.$$

Here E_{K} and E_{L} are the binding energies of electrons in the K- and L-shells, respectively.

Clearly, this is an oversimplified scenario because it does not take into account the dipole selection rules and the fine structure of the L-shell. Due to spin-orbit coupling, the L-shell splits up into three sublevels: $2s_{1/2}$, $2p_{1/2}$, and $2p_{3/2}$. The selection rule for dipole transitions requires that the change of the orbital angular momentum Δl must be ±1 and that the spin angular momentum change Δs must be 0. This excludes a dipole transition from $2s_{1/2}$ to 1s. However, the other two transitions $2p_{1/2} \rightarrow 1s$ and $2p_{3/2} \rightarrow 1s$ are dipole allowed. Their spectroscopic terms are $K_{\alpha 2}$ and $K_{\alpha 1}$, respectively. There are also transitions from the M-shell to the K-shell, labeled $K_{\beta 2}$ and $K_{\beta 1}$, etc. A generic energy level scheme of allowed x-ray emission lines is shown in Fig. 1.5. It is clear that while the general transition process is the same for all atoms, the transition energies are specific for the atom in question. Furthermore, as the transitions are within inner core levels, which do not take part in chemical binding, the x-ray energies are fingerprints for the atoms independent of the chemical environment in which they sit. This is at least true for heavy atoms. In light atoms energy shifts due to the chemical environment can be observed with high energy resolution x-ray spectroscopy.

Figure 1.6 shows the combined x-ray emission spectrum for an Mo target. It consists of the continuous bremsstrahlung spectrum, which is cut off at the minimum wavelength $\lambda_{\mathrm{min}} = 0.35$ nm, corresponding to the maximum energy transfer at 35 kVp. Superimposed on the bremsstrahlung spectrum are sharp intensity lines due to characteristic $K_{\alpha 1}$ (17.479 keV), $K_{\alpha 2}$ (17.374 keV) and K_{β} (19.61 keV) x-ray transitions. The intensity of these lines is much higher than the bremsstrahlung over the same spectral range. Measured and calculated bremsstrahlung spectra can be found in [1].

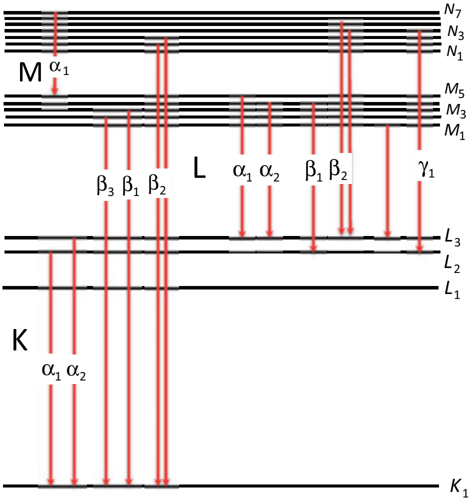


Fig. 1.5: Allowed dipole transitions for characteristic x-ray radiation. Adapted from *X-ray Data Booklet* (Center for X-ray Optics and Advanced Light Source Lawrence Berkeley National Laboratory).

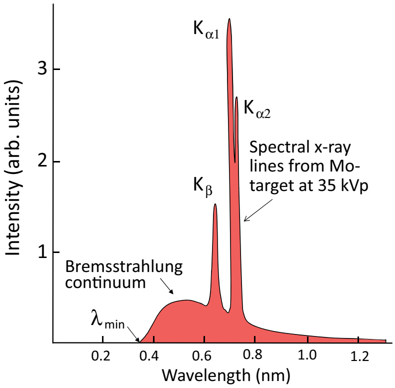


Fig. 1.6: Typical x-ray spectrum for an Mo target. The characteristic x-ray radiation is superposed on the bremsstrahlung spectrum.

1.2.3 Characteristic x-ray energies

Neglecting the spectroscopic splitting of the K-, L-, M-, etc. lines, the x-ray photon energy can be estimated in analogy to the Bohr model for the energy levels in hydrogen atoms. However, one has to take into account that after creating a hole in the K-shell by electron impact, the total charge of the remaining electrons is $Z - 1$, where Z is the atomic number. With this modification the empirical *Moseley's law* for the photon energy reads:

$$E_{\mathrm{i}\to\mathrm{f}} = hf_{\mathrm{i}\to\mathrm{f}} = R_{\mathrm{H}}(Z-1)^2\left(\frac{1}{n_{\mathrm{f}}^2} - \frac{1}{n_{\mathrm{i}}^2}\right).$$

Here n_{i}, n_{f} are the main quantum numbers of the atomic shells for the initial and final state of the electron transition, R_{H} is the Rydberg constant of the hydrogen atom, which is 16.6 eV. For the K_α transition this equation becomes:

$$E_{K_\alpha} = R_\mathrm{H}(Z-1)^2\left(\frac{1}{1^2}-\frac{1}{2^2}\right) = R_\mathrm{H}(Z-1)^2\frac{3}{4}.$$

In the case of Mo with $Z = 42$ we estimate for the K_α radiation an energy of 17.146 keV, which is close to the observed energy of 17.480 keV.

1.2.4 Energy dispersive x-ray chemical analysis

The element specificity of x-ray emission lines is a very useful property that can be utilized for chemical analysis. Let's assume that the electrons do not hit a target of well-known chemical composition, but of an unknown specimen. All atoms will be excited to emit characteristic x-ray radiation, supposing the incident electron energy is high enough to excite the inner K-shell electrons. We only need a detector to analyze the energy of the emitted characteristic radiation and then we can identify different elements in the specimen. Furthermore, the intensities of the characteristic radiation lines are proportional to the concentration of the elements in the target. After proper calibration with standards, the chemical composition of specimens can be derived. Figure 1.7 shows an example where we notice the K_α lines of Fe, Al, and to a lesser extent also from Si. In this example a thin film of Fe and Al was deposited on a silicon substrate. The instrument that provides such information is called a *Scanning Electron Microscope* (SEM) featuring an energy dispersive x-ray detector. For high spatial resolution, a finely focused electron beam with constant incident energy is scanned over the surface of a specimen and the emitted characteristic x-ray radiation is used for a spatially resolved chemical analysis. SEMs belong into any modern analytic laboratory and they are often used in forensic medicine for identifying inorganic chemicals such as those from gunpowder or explosives.

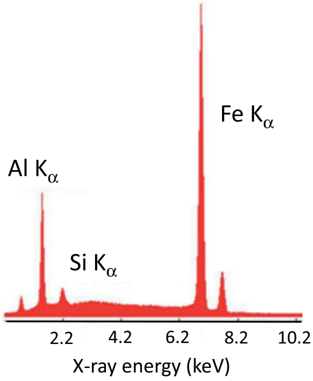


Fig. 1.7: X-ray spectrum from a scanning electron microscope equipped with an energy dispersive detector.

1.2.5 Final remarks

We have now seen that x-ray intensity can be produced by bremsstrahlung and by core electron excitation. Bremsstrahlung generates a smooth distribution of radiation from low to high x-ray photon energies whereas characteristic radiation shows sharp intensity lines with high intensity at specific energies. For radiography and for radiotherapy only the bremsspectrum is useful while the characteristic spectrum may be regarded as a disturbance that cannot be avoided completely. However, to minimize its effect on radiographs, the characteristic lines should be shifted to the high energy end of the bremsstrahlung spectrum. This can be achieved by coating the Cu target with a tungsten film. Tungsten is a metal with a high atomic number ($Z = 74$) that features characteristic x-ray lines at about 60 keV. When using accelerating voltages of 60–80 kVp, the characteristic tungsten K_α and K_β x-ray emission lines do not dominate the bremsstrahlung spectrum. One exemption is *mammography*, i.e., the radiography of the female breast. In mammography lower energy x-rays are preferred because of their higher absorption in soft tissue. This yields better contrast for slight changes in electron densities due to cancerous tissue. For mammography usually molybdenum targets are employed. The characteristic Mo K_α radiation at a wavelength of 0.071 nm and x-ray energy of 17.44 keV is used for imaging, the respective Mo x-ray spectrum is shown in Fig. 1.6.

1.3 X-ray generators

1.3.1 X-ray tubes

In medical applications the exposure time for taking a single x-ray image is rather short, in the order of a few milliseconds for a single image. Furthermore there is a sufficient time lap between one exposure and the next. Therefore heating or even melting of the anode is less likely compared to x-ray tubes which are in continuous operation. For this reason x-ray tubes for medical applications have a different design than standard laboratory x-ray tubes, a typical cross section is reproduced in Fig. 1.8. The cathode filament is moved off-center and the electron beam hits the rim of a rotating wedge-shaped anode, attached to a rotor of an electrical motor. The rotor is driven in a magnetic field of a stator outside the vacuum sealed tube. Thus the “hot” spot on the anode is spread out over the target rim. The target wedge has an inclination angle of 12°, permitting an x-ray radiation cone with an opening angle of about 24°. This wide opening angle is useful for large area exposures applied for thorax radiographs. Otherwise the beam can be shaped to the desired size with the help of diaphragms. Note that in medical x-ray tubes the cathode is grounded and the anode is at high potential, contrary to commercial laboratory or industrial x-ray tubes. The physics of bremsstrahlung generation remains, though, the same.

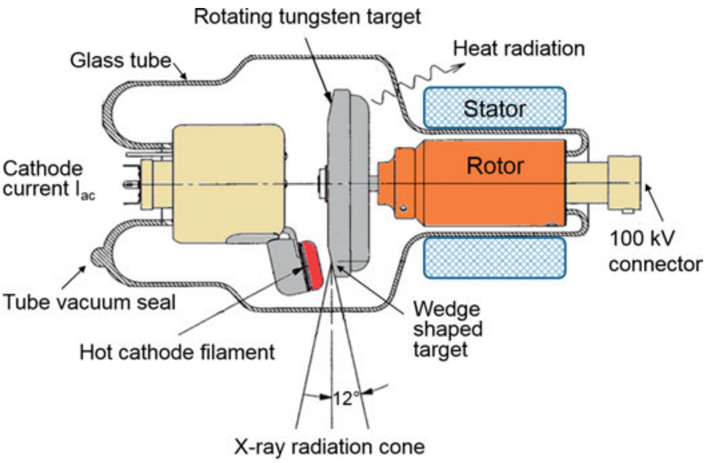


Fig. 1.8: Cross section of a typical x-ray tube for radiography. The glass tube is evacuated. The rotor with the target attached is driven by an external stator. The hot target emits radiation to the environment but is not actively being cooled.

In Fig. 1.9 a real tube together with housing and radiation shield is seen. These tubes can be operated up to about 150 kVp and they are used for short exposure times. For longer exposure times, the air cooling power would not be sufficient. From an accelerator point of view, these standard x-ray tubes are called *direct voltage accelerators*.

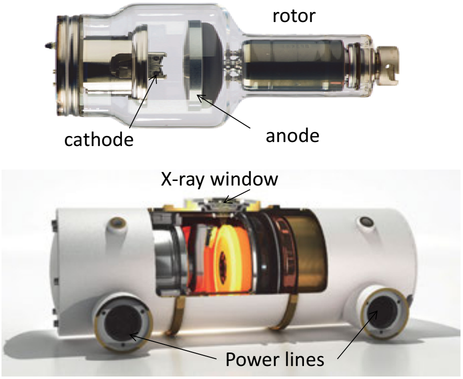


Fig. 1.9: Rotating anode x-ray tube for radiography. Top panel: glass housing of cathode and rotating anode, powered by a stator and rotor. Bottom panel: metal case for protection, containing openings for power lines and x-ray transparent window. Cut-away opening is an artist's view of the glowing target at about 2500 °C during operation (adapted from www.oem-products.siemens.com/x-ray-tube).

For taking computed tomography (CT) images (see Chapter 5) or for taking videos with x-rays (cineradiography), longer exposure times are often needed. Considering the heating and cooling curves of medical x-ray tubes such as the one shown in Fig. 1.9, a maximum average power consumption can be specified for a finite exposure time. Let's assume that 12 kW for a total exposure time of 10 s would be a safe operation. However, we need 40 kW during single exposures. This condition can be met by a pulsed time structure of x-ray exposure. If the pulse duration is taken as $\Delta t = 4$ ms and the repeat frequency or frame rate $\mathrm{RF} = 60\,\mathrm{s}^{-1}$, then the duty time is 240 ms and the duty factor DF is 24 % compared to a continuous exposure. If the peak power is chosen as $P_\mathrm{p} = 80\,\mathrm{kW} \times 500\,\mathrm{mA} = 40\,\mathrm{kW}$, then the average power is $P_\mathrm{a} = P_\mathrm{p} \times \mathrm{DF} = 9.6\,\mathrm{kW}$, which is below the specified safety limit. Pulsing the x-ray power allows to go to higher kVp yielding sharper images in shorter time. A frame rate of $60\,\mathrm{s}^{-1}$ is higher than required for a video, but guarantees sharp pictures of moving organs like the heart. The estimate made here solely refers to the safe operation of the x-ray tube and does not take into account radiation safety aspects with respect to the patient. Radiation safety issues are discussed in Chapter 4.

An innovative new concept for a medical rotating x-ray tube for single shot radiographs as well as CT scans has recently been presented by the manufacturer Siemens. The so called Straton rotating x-ray tube is a step forward in rotating x-ray tube designs for medical applications, overcoming a number of limitations of the standard rotating x-ray tubes. The cross section of the Straton tube is shown in Fig. 1.10 (left panel) together with an image of the bare tube in the right panel. Anode and cathode are fixed inside of an evacuated metal housing and are rotated by an external motor. X-rays from the anode penetrate through a thin Al window that simultaneously cuts

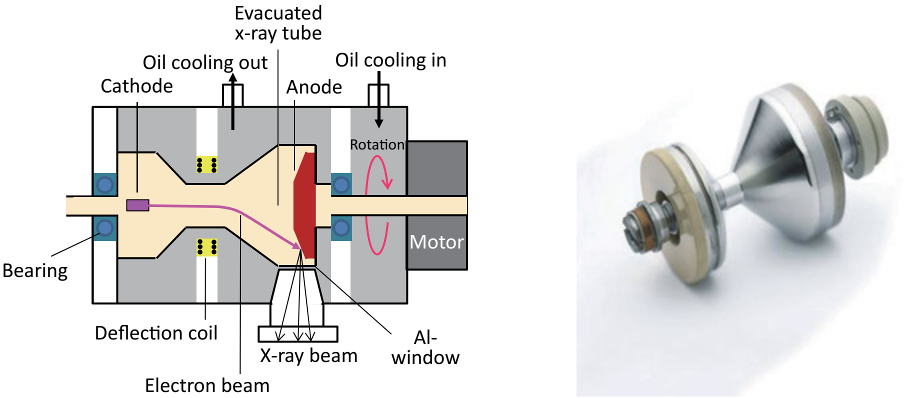


Fig. 1.10: Straton sealed x-ray tube immersed in an oil trough for cooling. Left panel: schematic of the working principle. The entire tube is rotated including the target. The cathode is located in the center and the electron beam is deflected to the target by an external magnetic field; right panel: photograph of the rotating and sealed x-ray tube (adapted from www.siemens.com/press).

off the soft end of the x-ray spectrum. The tube is immersed in an oil trough for cooling. The electron beam from the rotating cathode in the center of the tube to the rotating anode is dynamically deflected by an external magnetic field that can also shape the beam and control the focal spot size. Furthermore, with this design the beam can be kicked back and forth with high frequency between two focal spots of different target materials. Because of the oil cooling, there is no limitation on exposure times.

1.3.2 Linear accelerators

Radiotherapy requires different types of x-ray machines for the production of much harder x-ray bremsstrahlung than for radiography. In the past, high voltage x-ray tubes for peak voltages of up to 250 kVp were used. Since the 1960s the sealed-off tubes have been replaced by linear electron accelerators with peak energies up to 25 MeV. These high MeV x-ray energies produce a completely different type of tissue irradiation than the lower energy x-ray tubes did in the past, as we will discuss in Chapters 3 and 9. The technology of linear accelerators for medical applications originates from high energy physics and has been steadily improved over the past years to make them shorter and more lightweight by using steeper voltage gradients in the accelerating cavities.

Figure 1.11 shows a block diagram of the most important components of a *linear accelerator* (linac), consisting of klystron, pulse modulator, electron source, microwave cavities for electron acceleration, bending magnet, target, and treatment head. For clarity, power supply and vacuum pumps are omitted. Klystrons produce the microwave power for accelerating pulsed electron bunches, which are additionally modulated in the pulse modulator. The klystron can be replaced by a magnetron for lower power applications between 1–6 MeV, but for higher energies klystrons are preferred. In both cases the aim is to produce electron bunches from a DC electron source. The electron source has a similar structure as in x-ray tubes: electrons are emitted into vacuum by thermionic extraction from a hot filament.

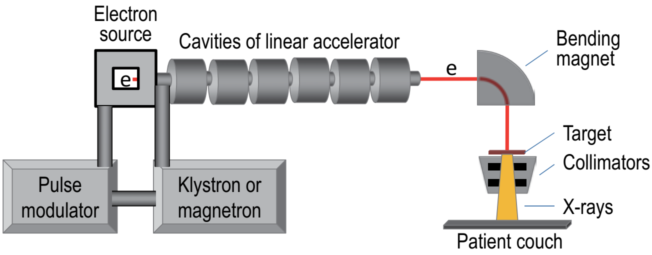


Fig. 1.11: Block diagram of a linear accelerator for radiation therapy.

The linac consists of a linear sequence of pillbox-like cavities (Fig. 1.12 (c)) [2]. In each cavity several microwave standing modes can be excited. The one which is used most frequently is the so called TM_{010} accelerating mode outlined schematically in Fig. 1.12. Panel (a) gives a side view and panel (b) is a view in the direction of the electron path. The TM_{010} mode is characterized by a longitudinal electric field parallel to the cylinder axis of the cavity, which accelerates the electrons entering through a central hole. A curling magnetic field around this axis causes ohmic dissipation that needs to be cooled off. The letters TM refer to the fact that the magnetic field is oriented perpendicular to the accelerating current, the first index indicates the azimuthal dependence of the magnetic field, the second index the radial dependence, and third index the longitudinal component. Thus in TM_{010} mode the magnetic field has only a radial component.

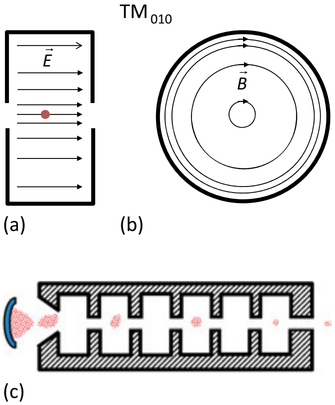


Fig. 1.12: Electrical and magnetic field components for the TM_{010} cavity mode. (a) Side view of a cavity; (b) view along the electron path; (c) sequence of cavities in a linac with the electron source to the left.

Electron acceleration does not depend on the voltage difference as in a standard tube, but on the phase of the voltage in the cavity that the electron experiences when entering the cavity at a velocity v. Obviously, kicking the electron in a phase just beyond the maximum voltage amplitude is most effective, which is equivalent to pushing a child on a swing just after the return point from maximum amplitude. The microwave cavities convert the RF energy from the klystron/magnetron into high energy electron bunches. Once the high energy electrons exit the linac, they may immediately hit a target for producing x-ray bremsstrahlung. In most cases the horizontal electron beam is first turned by a bending magnet from horizontal to vertical orientation (see Fig. 1.11) and then the beam hits a tungsten target like in an x-ray tube. The bending of the electron beam is achieved by a sector magnet turning the electrons either by 90° or by 270°. Linacs are some 1–2 meters long depending on the maximum energy. Traditionally linacs are mounted horizontally inside of a gantry together with magnet sector and collimator head as can be seen by the artist's view in Fig. 1.13. The entire gantry can then be rotated about the patient couch at the center. In the most advanced applications, known as *cyberknife*, the linac is short and lightweight producing 4–8 MeV electrons. It is integrated into a robotic arm that can be positioned anywhere around the patient. In this design the magnet sector for bending the electron beam is no longer required, making the linac even shorter and lighter. Cyberknife technology is described in more detail in Chapter 9. By hitting a tungsten target an intense high energy photon beam is produced. The energy is so high that it can easily penetrate a thin tungsten target, while the electrons are stopped. The x-ray beam is shaped by collimators to the specified physical target volume to be exposed to radiation.

The high electron and photon energies produced by medical linacs are difficult to handle. Firstly, they require extensive radiation shielding. The electron beam as well as the photon beam cause a number of side effects, such as neutron production. Neutrons, in turn, cause further problems via activating materials when captured. In the end, when turning off a linac there will be an "afterglow" of radiation from the material that was exposed to high energy radiation. This radiation consists of fast neutrons and gamma spectral lines. Therefore it is necessary to take extra measures for radiation shielding to protect patients and clinical staff.

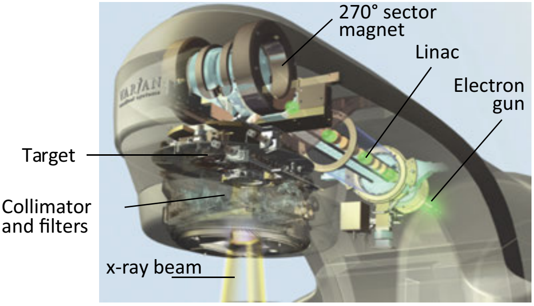


Fig. 1.13: Linac inside a gantry for x-ray production (courtesy of Varian Medical Systems International AG, Switzerland; all rights reserved).

1.3.3 Synchrotron radiation

The third type of machine that produces x-rays is an electron *synchrotron*. First, electrons are brought up to speed with the help of a linac and a booster ring. Once the electron energy is ramped up, the electrons are injected in time-separated bunches into a storage ring. A lattice of magnets keeps the electrons on track in a narrow circular tube, which is evacuated to pressures of less than 10^{-9} mbar. Ultrahigh vacuum is required to avoid collisions of electrons with residual gas atoms that shorten the storage lifetime. A schematic outline of a synchrotron is shown in the inset of Fig. 1.15. While electrons in a circular orbit are constantly accelerated towards the center (radial acceleration), they emit electromagnetic radiation like in an antenna. At low nonrelativistic energies the radiation profile has a typical dipole distribution shown in the left panel of Fig. 1.14. However, in storage rings electrons have a speed close to the speed of light, i.e., they are highly relativistic. Then for an observer in the rest frame

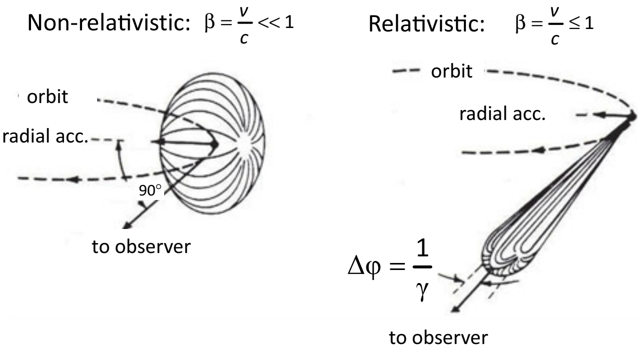


Fig. 1.14: Radiation profile for a classical electron on a circular orbit and for a relativistic electron in a synchrotron storage ring.

the radiation profile becomes distorted, radiating into a narrow radiation cone with an opening angle $\Delta\phi = 1/\gamma$ in the direction of the traveling electrons, as sketched in the right panel of Fig. 1.13. Since $\gamma = E/m_0c^2$, where E_{kin} and m_0c^2 are the kinetic energy and rest energy of the orbiting electron, respectively, and assuming for E_{kin} a typical energy of 6 GeV in a storage ring, we find for the opening angle:

$$\gamma = \frac{E_{\mathrm{kin}}}{m_0c^2} \approx \frac{6\,\mathrm{GeV}}{0.5\,\mathrm{MeV}} = 12 \times 10^3; \quad \Delta\phi = 5/1000°.$$

In this narrow radiation cone the intensity is extremely high, many orders of magnitude higher than in conventional x-ray tubes.

The energy spectrum of synchrotron radiation is continuous and very broad, similar to a bremsstrahlung spectrum with a cut-off that depends on E_{kin} and the radius R of the storage ring. Radius R of the orbit and magnetic field B of the bending magnets are related according to: $R = 3.3E_{\mathrm{kin}}[\mathrm{GeV}]/B[\mathrm{T}]$. In Fig. 1.15 a typical spectrum is shown for $E_{\mathrm{kin}} = 1$ GeV and a magnetic field $B = 1.2$ T. The maximum intensity is reached at the critical energy E_{c} indicated by a blue arrow. The critical energy is related to the kinetic energy and the magnetic field of the bending magnets via: $E_{\mathrm{c}}[\mathrm{keV}] = 0.665E_{\mathrm{kin}}^2[\mathrm{GeV}]B[\mathrm{T}]$. Inserting numbers from above, a critical energy $E_{\mathrm{c}} = 0.8$ keV is obtained. The critical and cut-off energy increases with the square of E_{kin} and hard radiation of interest for medical applications can be gained at $E_{\mathrm{kin}} \approx 6$ GeV and higher.

The radiation spectrum of storage rings can be modified by using special magnetic insertion devices, which wiggle or undulate the electrons on their orbit to give them short periodic kicks for additional acceleration and increased intensity. However, a discussion of such devices is beyond the scope of this text. Synchrotron radiation is sometimes used for cancer irradiation, but more importantly it has been used for x-ray radiography with phase contrast enhancement, presented in Section 5.3.4. A major use of synchrotron radiation is structural investigations of materials including biomolecular crystallography. Further information on basics, methods, and applications of synchrotron radiation can be found in [3].

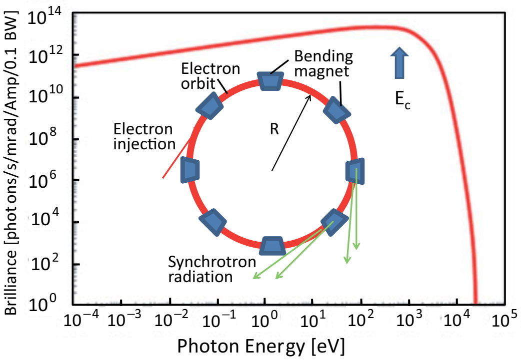


Fig. 1.15: Energy spectrum of synchrotron radiation for 1 GeV electrons and 1.2 T field of the bending magnets. The inset shows an outline of an electron storage ring (synchrotron), where bending magnets keep the electrons on a circular orbit.

1.4 Summary

1. X-rays are produced by accelerating electrons in a static potential gradient between cathode and anode inside an evacuated tube followed by a sudden deacceleration in a target.
2. X-rays produced by accelerating of electrons and stopping in a target consist of a continuous bremsstrahlung spectrum and a characteristic spectrum of several discrete atomic emission lines.
3. The continuous bremsstrahlung spectrum is cut off on the high energy side when the kinetic energy of electrons is completely converted in electromagnetic radiation.
4. The integrated intensity of the bremsstrahlung spectrum scales linearly with the cathode current and quadratically with the accelerating voltage.
5. The characteristic spectrum is specific for the target material used. The emission lines correspond to differences in the binding energies of core atomic levels.
6. The characteristic spectrum can be used for chemical analysis of the target material.
7. X-ray tubes for radiography contain rotating targets for distributing the heat load from the electron beam.
8. X-ray tubes for radiography can only be used for short periods of time. When taking CT images or cineradiographies, a pulsed time structure with low duty factor has to be applied.
9. Medical x-ray tubes for radiography operate up to about 140 kV.
10. For radiotherapy very high energy radiation is required in the range of MeV.
11. High energy electrons up to 25 MeV can be produced with a linear accelerator.
12. Linear accelerators are mounted in gantries which can be rotated about a target at the isocenter.
13. A continuous spectrum of x-ray photons with very high intensity can also be produced by electrons in a synchrotron storage ring.
14. Synchrotron radiation is mainly used for x-ray scattering and x-ray spectroscopy of materials, including biomaterials.
15. All beam shaping of x-ray beams is achieved by collimators.

References

[1] Birch R, Marshall M. Computation of bremsstrahlung X-ray spectra and comparison with spectra measured with a Ge(Li) detector. Phys Med Biol. 1979; 24: 505–517.
[2] Wangler TP. RF linear accelerators. 2nd edition. Wiley & Sons; 2008.
[3] Mobilio S, Boscherini F, Meneghini C, editors. Synchrotron radiation: basics, methods, and applications. Springer Verlag; 2015.

Further reading

Eisberg R, Resnik R. Quantum physics of atoms, molecules, solids, nuclei, and particles. 2nd edition. John Wiley and Sons; 1985.
Bushberg JT, Seibert JA, Leidholdt EM Jr, Boone JM. The essential physics of medical imaging. 3rd edition. Lippincott Williams & Wilkins, Wolter Kluwer; 2012.
Dance DR, Christofides S, Maidment ADA, McLean ID, Ng KH, editors. Diagnostic radiology physics – A handbook for teachers and students. IAEA; 2014.
Wille, K. The physics of particle accelerators. Oxford University Press; 2000.

Useful websites

www.antonine-education.co.uk/Pages/Physics_5_Options/Medical_Physics/MED_07/med_phys_7.htm
http://hyperphysics.phy-astr.gsu.edu/hbase/hph.html

2 Nuclei and isotopes

2.1 Introduction

In Chapters 5 to 12 nuclear methods in medicine are discussed either for imaging (scintigraphy, SPECT, PET) or for radiation treatment of cancerous tissues (proton and neutron irradiation, brachytherapy). It is therefore appropriate to first introduce some basic properties of nuclides and isotopes, and in particular of radioactive isotopes which are used in nuclear medicine. This chapter is not intended to replace a textbook on nuclear physics. But it provides sufficient background information for better understanding the subsequent chapters. Handling of radiation in general and application of radioactive isotopes also requires a detailed knowledge of radiation dose and radiation safety, which are topics of Chapter 4.

Atoms have an electronic shell of orbiting electrons and a *nucleus* at the center. The term nucleus should not be mixed up with the nucleus of a biological cell. Although the names are identical, from the context it should be clear whether an atomic nucleus or a cell nucleus is meant. The nucleus consists of Z *protons* and N *neutrons* bound together by the strong nuclear force. Another term for nucleus is *nuclide*. Both terms are used in this chapter synonymously. The plural form of nucleus is nuclei, the plural of nuclide is nuclides. Individual protons and neutrons together are called *nucleons*. They are part of the hadron family, which are particles composed of *quarks*. More specifically, nucleons are *baryons*, as they always contain three quarks. Protons have two up quarks with charge $+2/3e$ and one down quark with charge $-1/3e$, where e is the elementary electrical charge. Therefore protons have in total one positive elementary charge $+e$, which has the same magnitude but opposite sign as electrons in atomic shells. Neutrons have one up and two down quarks. Neutrons are neutral since their quark charge sums up to zero. The sum of protons and neutrons in nuclides is the *mass number* $A = Z + N$. The number of protons Z in nuclides is also known as the *atomic number*. In charge-neutral atoms the number of protons equals the number of electrons. Charged atoms are called *ions*. They are either *cations* if the number of protons is larger than the number of electrons, or *anions* in the opposite case. Both, protons and neutrons, have an angular momentum (spin) of $S = 1/2\hbar$ with an associated magnetic moment. The magnetic moment of protons is important for magnetic resonance imaging, which is discussed in Chapter 15/Vol. 1. Neutrons also have a magnetic moment, which is utilized in research for probing magnetic materials. Protons and neutrons are heavier than electrons by a factor of roughly 2000 (exactly 1836 for protons and 1838 for neutrons). Their combined mass in nuclides determines the *atomic mass*.

DOI 10.1515/9783110553116-004

2.2 Isotopes

In light nuclei the number of neutrons N and the number of protons Z is roughly equal: $N \approx Z$. One exception is hydrogen that contains only one proton but no neutron. Nuclides with the same number of protons Z but different number of neutrons N are called *isotopes*. The general nomenclature of isotopes is: ${}^{A}_{Z}X$. Here X stands for the chemical symbol, A is the mass number and Z is the atomic number. The chemical symbol X together with the atomic number Z is redundant and therefore Z is often omitted. The number of neutrons follows from $N = A - Z$. Isotopes have the same atomic number Z, but different atomic mass numbers A.

As an example we discuss the isotopes of hydrogen. Hydrogen ${}^{1}_{1}\mathrm{H}$ with $Z = 1$, $N = 0$ and deuterium (D) ${}^{2}_{1}\mathrm{H}$ with $Z = 1$, $N = 1$ are temporally stable isotopes. Hydrogen has another isotope: tritium (T) ${}^{3}_{1}\mathrm{H}$ with $Z = 1$, $N = 2$. The isotopes of hydrogen are the only ones that have their own chemical symbols: H for hydrogen, D for deuterium, and T for tritium. The bare nuclei of these three isotopes are called proton, deuteron, and triton. All other isotopes are only designated by their nuclear notation. Tritium (or triton) is not stable but decays in time by converting one neutron into a proton, simultaneously emitting an electron. The decay product is a stable ${}^{3}_{2}\mathrm{He}$ isotope with the same mass number as tritium but different Z. The decay is formally written similar to a chemical reaction:

$${}^{3}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + \mathrm{e}^{-} + \bar{\nu}_{\mathrm{e}}.$$

Where $\bar{\nu}_{\mathrm{e}}$ is an undetectable electron antineutrino, which we will not discuss any further. The half-life time of the decay is 12.3 years, i.e., after 12.3 years half of the tritium isotopes are converted into ${}^{3}\mathrm{He}$. This example of tritium decay teaches us many important facts about nuclei. First, isotopes become unstable if the number of neutrons is either too large or too small. Second, radioactive decay transforms an unstable isotope into a temporally stable one, the decay time is characteristic for the isotope. Third, radioactive decay occurs by the emission of charged particles that changes the atomic number Z. As we will see later, radioactive decay may require a cascade of transitions before finally a stable isotope is reached and often γ-radiation is emitted alongside, which brings isotopes from an excited intermediate nuclear state into the ground state.

Figure 2.1 shows the stable isotopes of light elements in terms of number of protons versus number of neutrons. The ratio N/Z is very close to one. From this chart we recognize that the elements Be, F, Na, and Al have only one stable isotope.

For the heavy elements displayed in the isotope chart of Fig. 2.2, also known as the Segrè chart, we notice that the valley of stability (black dots) leans over to a neutron excess and ends at the element lead ${}^{208}_{82}\mathrm{Pb}$. For heavy nuclei, the ratio N/Z bends over from 1 to 1.5. This is because with increasing atomic number the repulsive Coulomb interaction between the protons weakens the binding energy of nuclei, which needs to be counterbalanced by an increasing number of neutrons. All isotopes heavier than ${}^{208}_{82}\mathrm{Pb}$ are unstable and decay over time. The decay schemes are marked in blue for electron emission, red for positron emission, and yellow for α-particle emission.

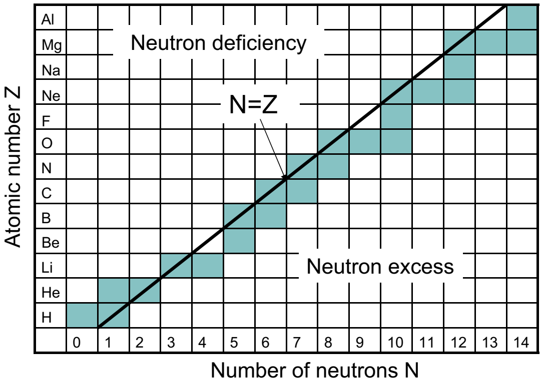


Fig. 2.1: Chart of stable isotopes for the light elements.

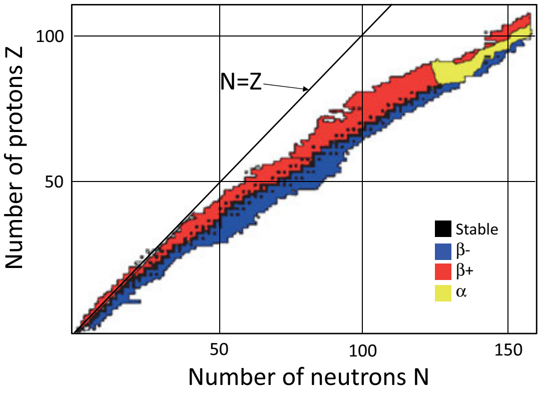


Fig. 2.2: Chart of stable isotopes (black dots) and for radioactive isotopes (in color). Isotopes which decay by electron emission are colored in blue, isotopes which decay by positron emission are colored in red. Isotopes which decay by alpha mission are indicated in yellow color. Note that stable isotopes lie at or below the line $N = Z$, for heavier nuclei beyond $Z = 12$, $N > Z$.

There are two types of *radioactive isotopes* also called *radioisotopes*: those that occur naturally and those that are produced artificially. The naturally occurring radioisotopes can be further subdivided into those that have been around since the birth of the earth, and those that are continuously being reproduced by the environment. The former are known as *terrestrial radioisotopes* and are long lived with half-life times in the order of the age of the earth. Typical examples are ^{40}K, ^{232}Th, ^{235}U, ^{238}U. The latter, termed *cosmic radioisotopes*, are continuously generated by cosmic rays and are relatively short lived, such as ^{14}C, ^{3}H, ^{7}Be. Radioisotopes can also be produced artificially, for instance by proton bombardment in an accelerator or by neutron capture in a fission reactor. These topics are discussed in more detail in Section 2.5.

2.3 Atomic mass and atomic weight

A, N, and Z are integer numbers. When associating them with a mass unit for neutrons and protons, the total mass of a nucleus should also be an integer in terms of this unit. However, this is not the case because of the strong nuclear binding energy. The total mass of a nucleon is smaller than the sum of the individual neutron m_n and proton masses m_p:

$$m(A, Z) = Zm_\mathrm{p} + Nm_\mathrm{n} - B/c^2 .$$

The mass deficiency is due to the binding energy B of nuclei. The binding energy varies with the mass number A. No distinction is made between protons and neutrons, because the strong nuclear force is independent of charge. For light nuclei, the binding energy per nucleon B/A first increases and reaches a maximum of 8.6 MeV at the mass number $A \approx 58$ corresponding to Fe. For higher mass numbers the average binding energy drops off again. From the plot of B/A versus A in Fig. 2.3 we also recognize that in He the binding energy is particularly high compared to neighboring isotopes, which explains the formation of ^{4}He embryos in heavy nuclei before α-decay actually takes place. Furthermore, from this plot it becomes clear that either by fusion of light elements to heavier nucleons or by fission of heavy elements into smaller fragments energy can be gained. The fusion process occurs in the sun and in all other stars; the fission process takes place in nuclear power plants and in neutron research reactors.

Because of the dependence of the nuclear mass on the atomic number A, a unit for the atomic mass had to be defined independent of the energy variation. The *unified atomic mass unit* or short *atomic mass unit* with the symbol u is defined as the mass of the isotope ^{12}C divided by its mass number 12: $1\,\mathrm{u} = 1m_\mathrm{u} = m(^{12}_{6}\mathrm{C})/12$. In SI units u has the mass of 1.6605×10^{-27} kg and the energy equivalent of $m_\mathrm{u}c^2 = 931.494$ MeV.

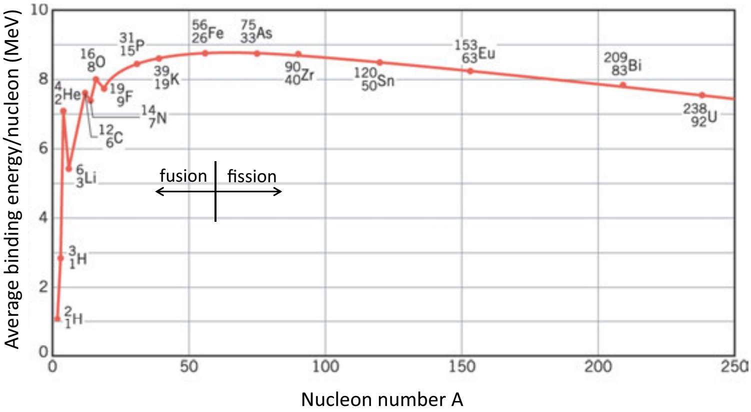


Fig. 2.3: Binding energy of nucleons as a function of atomic number *A* (adapted from physics. stackexchange.com and licensed under Creative Commons).

In terms of atomic mass units, the mass of a proton is $m_p = 1.007276\,u$, the mass of a neutron is $m_n = 1.008664\,u$. The free neutron is slightly heavier than the proton and therefore decays by β^--emission with a mean lifetime of 881.5 ± 1.5 s into a proton, which is a stable particle.

By definition ^{12}C has the unified atomic mass of 12 u. Taking this atomic mass in grams (here 12 g), provides a definition for the number of atoms per atomic mass, which is the Avogadro number N_{Av}. In words: 12 g of isotopically pure ^{12}C contains one mole of atoms or $N_{Av} = 6.022 \times 10^{23}$ particles. One mole of hydrogen or 1 g of protons contains the same number of particles: $N_{Av} = 1\text{ g}/1\text{ u}$. The mass of isotope clean substances is always close to an integer multiple of u. For example, the atomic mass of ^{56}Fe is close to 56 u, etc.

The *atomic weight* is also quoted as the gram equivalent per mole of atoms, but in contrast to the atomic mass an average over all naturally occurring and stable isotopes is taken into account with their respective abundance. Typically, the atomic weight numbers are not close to multiples of u. For instance, the atomic weight of carbon is 12.011. From the 15 known isotopes of carbon only two are stable: ^{12}C with an abundance of 0.989 and ^{13}C with an abundance of 0.011. The weighted sum $12 \times 0.989 + 13 \times 0.011 = 12.011$ is the atomic weight of carbon. For an isotopic mixture the atomic weight expressed in grams contains again one mole of atoms.

2.4 Nuclear decay

Four types of radioactive decay are known, transforming unstable radioisotopes into stable ones.

2.4.1 Electron emission (β^-)

Nuclei exhibiting a neutron excess are unstable. In the nucleus one of the neutrons is converted into a proton by emission of a negatively charged electron, called β^--emission, and an antineutrino. After emission, the positive charge of the nucleus increases by one ($Z \rightarrow Z + 1$), therefore the daughter nucleus is a different chemical element on the right side of the mother nucleus in the chemical table and has a different chemical symbol. The decay has the general nuclear reaction form:

$$^{A}_{Z}X \rightarrow {}^{A}_{Z+1}Y + \beta^- + \bar{\nu}_e + \text{energy}.$$

Decays by electron emission do not change the atomic number. Such decays are called *isobaric transitions*. However, the ratio N/Z decreases bringing the daughter nucleus closer to the valley of stability. The antineutrino together with the electron carry away the total binding energy, the partition of the energy between electron E_e and antineutrino E_ν is not fixed. Therefore the energy spectrum of emitted electrons in β^--decay is very broad, stretching from zero to a maximum energy $E_{\max}$ where the kinetic energy of the antineutrino is negligible. The interaction of neutrinos with matter is basically zero and the rest mass is much smaller than that of electrons. Neutrinos became noticeable mainly because of the broad electron energy distribution, which invokes the existence of another particle for fulfilling energy and momentum conservation.

An example for electron (β^-)-emission is the decay of the terrestrial radioisotope ^{40}K: $^{40}_{19}\text{K} \rightarrow {}^{40}_{20}\text{Ca} + {}^{0}_{-1}\text{e}^- + \bar{\nu}_e$. In the nuclear chart in Fig. 2.4 β^--decays are always directed from the bottom upwards into the valley of stable isotopes.

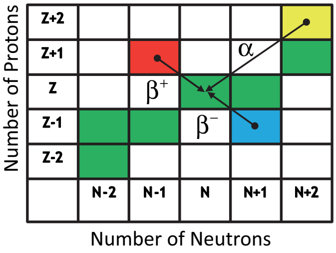


Fig. 2.4: Decay scheme for α- and β-radiation.

2.4.2 Positron emission (β^+)

Nuclei exhibiting a neutron deficiency are unstable. In the nucleus one of the protons is converted into a neutron by emission of a positively charged electron, called positron or β^+. After emission, the positive charge of the nucleus decreases by one. The decay has the general nuclear reaction form:

$$^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + \beta^+ + \nu_e + \text{energy}.$$

Positron decay is also an isobaric transition, changing the chemical element but not the atomic number. The ratio N/Z increases, balancing the neutron deficiency. In the nuclear chart in Fig. 2.4, β^+-decay is always directed from above towards the valley of stable isotopes. Because of the participation of neutrinos in the decay, the energy distribution of positrons is also very broad, similar to the case of β^--decay.

An example for β^+-decay is the artificial radioisotope ^{18}F with a half-life of about 110 minutes: $^{18}_{9}\text{F} \rightarrow {}^{18}_{8}\text{O} + {}^{0}_{1}\text{e}^+ + \nu$.

Positrons are antiparticles to electrons. They do not occur naturally, they can only be found as decay products of radioisotopes. In isolation (vacuum) positrons can be kept for a long time. However, as soon as they come close to an electron, positron and electron will annihilate both together by emission of two γ-photons in opposite directions, a process that is referred to as electron-positron annihilation. The γ-photons have an energy corresponding to the rest mass of the electron-positron pair and therefore the energy is $2m_ec^2 = 2 \times 0.511\,\text{MeV}$. The correlated γ-γ emission is used in positron emission tomography (PET) for imaging the distribution of nuclear isotopes in the body. The radioisotope ^{18}F is one of the favored isotopes for PET, further discussed in Chapter 7.

2.4.3 Electron capture (EC)

In some cases the conversion from a proton into a neutron is achieved by capturing an electron from the K-shell of an atom. This leaves a hole in the K-shell similar to the excitation of characteristic x-ray radiation by kicking out an electron from the K-shell, as discussed in Chapter 1. The hole in the K-shell is filled by an outer electron and the energy difference is emitted as an x-ray photon. Therefore isotopes that show electron capture (EC) decay are also called x-ray emitters. Formally, the decay can be described by the reaction:

$$^{A}_{Z}X + \mathrm{e}^- \rightarrow {}^{A}_{Z-1}Y + \nu_\mathrm{e} + hf_{\text{x-ray}}\,.$$

An example is the decay of ^{40}K, which was presented before. A different decay mode by electron capture is: $^{40}_{19}\mathrm{K} + {}^{0}_{-1}\mathrm{e}^- \rightarrow {}^{40}_{18}\mathrm{Ar} + \nu_\mathrm{e}$. Electron capture is an alternative decay mode for unstable isotopes whenever the energy for decay by positron emission is insufficient. Positron emission requires at least 1.022 MeV binding energy difference between the mother and daughter isotope, corresponding to the rest mass of an electron-positron pair. If the energy difference is smaller, EC is the only way out for the decay. Therefore, decay by EC is sometimes also referred to as *inverse positron decay*.

2.4.4 Alpha decay

In heavy nuclides, strongly bond subunits of two protons and two neutrons forming an He-nucleus may often escape from their parent nucleus by a tunneling process through the Coulomb barrier of the nuclear potential. After successful tunneling, the alpha particles have a typical kinetic energy of about 5 MeV. Alpha particle decay changes the charge of the mother nucleus by $\Delta Z = -2$ and the mass number by $\Delta A = -4$. The decay has the general nuclear reaction form:

$$^{A}_{Z}X \rightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}\mathrm{He} + E_\mathrm{kin}\,.$$

In the nuclear chart in Fig. 2.4 α-decay always occurs along the diagonal or parallel to the valley of stable isotopes. An example is the decay of radium into radon:

$$^{226}_{88}\mathrm{Ra} \rightarrow {}^{222}_{86}\mathrm{Rn} + {}^{4}_{2}\mathrm{He} + E_\mathrm{kin}\,.$$

Alpha particle emission is the most common radioactive decay mode for the heaviest nuclides. Often an entire series of decays is observed, starting from uranium, thorium or neptunium until after a number of α-emissions a stable isotope is formed. For instance, ^{238}U requires eight α-emissions before reaching the stable isotope ^{206}Pb. These heavy nuclides were originally formed during supernova explosions and they can still be found in nature because of their very long lifetimes in the order of 5×10^9 a.

2.4.5 Decay schemes

After emission of α- or β-particles the daughter nucleus often remains for a short time in an excited high angular momentum state. The transition to the ground state occurs via emission of electromagnetic radiation, termed *γ-radiation* to distinguish it from characteristic x-ray radiation originating from the excitation of atomic shells. The energy of γ-photons ranges from keV to MeV. A typical example for the decay of the isotope ^{137}Cs is shown in Fig. 2.5. ^{137}Cs is a β^--emitter, i.e., after β^--emission ^{137}Cs transforms into the isotope ^{137}Ba with the same mass number. There are two possibilities for the decay, called *branching*. Either the ^{137}Cs isotopes emit a β^--electron with the maximum energy of 1.17 MeV, bringing ^{137}Ba immediately into the ground state; or ^{137}Cs emits a β^--electron with a smaller energy of 0.51 MeV, leaving ^{137}Ba* in an excited nuclear state that goes into a stable ground state after emission of a γ-photon of 0.66 MeV. The first transition has a probability of only 5 %, whereas the majority of transitions (95 %) occur via the intermediate excited ^{137}Ba* level into the ground state ^{137}Ba. The asterisk indicates a metastable state of the nucleus. The half-life of ^{137}Cs is about 30 years, independent of the branching, whereas the γ-emission is relatively prompt with a $T_{1/2}$ of 2.5 minutes. The hard γ-radiation of 0.66 MeV is used for radiation therapy and many other applications.

Another interesting example of branching are the isotopes $^{39-41}$K of potassium, already introduced. ^{39}K and ^{41}K are stable isotopes, only ^{40}K is radioactive with a very long half-life of $T_{1/2} = 1.3 \times 10^9$ a. This isotope may either decay by electron capture (EC) with a probability of 11 % or by electron emission with a probability of 89 %, transforming the radioisotope into either stable Ar or Ca. The decay scheme is shown in Fig. 2.6.

The natural abundance of ^{40}K is only 0.012 %. Nevertheless we can estimate that about 0.02 g of ^{40}K is in our body, decaying continuously with an activity of about

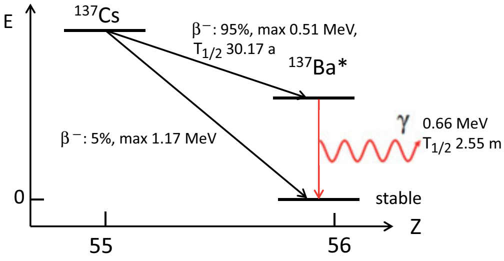


Fig. 2.5: Radioactive decay of ^{137}Cs and branching ratio.

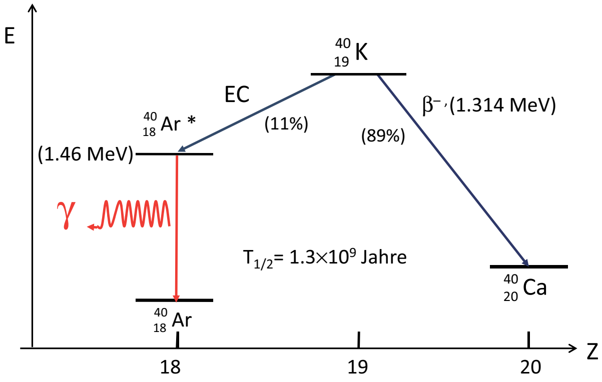


Fig. 2.6: Decay scheme of the isotope ^{40}K.

5000 events per second. As electrons do not have a long range, most of the electrons emitted through ^{40}K decay remain in our body. But some escape and can be counted with full-body scintillation counters.

One final example is technetium (Tc), which is the most used isotope for scintigraphy. As none of the Tc isotopes are long lived, Tc cannot be found in nature but must be produced artificially. The isotope ^{99}Tc is particularly useful for scintigraphy because of its short lifetime and emission of γ-radiation with a low energy of 140 keV corresponding to x-ray energies used for radiography. The starting isotope is ^{99}Mo, which is a by-product of ^{235}U nuclear fission in a nuclear research reactor. ^{99}Mo decays by β^--emission into an excited state of ^{99}Tc*; the metastable ^{99}Tc* state decays with a half-life time of 6 h by γ-radiation into the ground state: $^{99}_{43}\mathrm{Tc}^\star \rightarrow {}^{99}_{43}\mathrm{Tc} + \gamma$. The decay scheme is represented in Fig. 2.14. Finally ^{99}Tc decays by another β^--emission into the stable isotope ^{99}Ru.

For all radioactive decays the energy and momentum must be conserved as well as a couple of further conservation laws concerning the symmetry of space and time. However, by analyzing β-decay in detail, it was already noticed early that the energy and momentum conservation appear to be violated, unless another charge-neutral particle is postulated. This mysterious particle is the neutrino. Although there is no doubt that this elementary particle exists, the interaction with matter is so weak that until now the rest mass could not be determined unambiguously. Furthermore, during β-decay parity is violated, meaning that a small difference in the decay rate is noticeable whether the β-emission succeeds parallel or antiparallel to the nuclear spin. This violation can only be rectified by postulating a new force acting during the β-decay, which is called the electro-weak force. This force must be of extremely short range, shorter than the range of the strong nuclear interaction that binds the nucleons in a nucleus. This short interaction controls in nuclei the conversion of protons into neutrons and vice versa as well as the β^--decay of free neutrons.

2.5 Radioactivity

2.5.1 Exponential decay law

Radioactive decay is a stochastic process. In a large amount of identical radioisotopes it is not known when a particular nuclide will decay. However, it is possible to exactly state after which time half of the original radioisotopes will have decayed. This is called the half-life or the *half-life time*. Radioactive decay is described by the decay law: if N_0 is the number of radioisotopes existing at time $t = 0$, and the rate of decay is proportional to the radioisotopes still existing, multiplied by a proportionality constant λ:

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -\lambda N,$$

then after time $t > 0$ the number of nuclides $N(t)$ that have not decayed follows from integration over time t since the start $t = 0$ with the integration constant $N(t = 0) = N_0$:

$$N(t) = N_0 \mathrm{e}^{-\lambda t}.$$

This expression is known as the *exponential decay law*. Note that in this context N refers to the number of radioisotopes in a sample; the letter N may also designate the number of neutrons in a nucleus. From the context the difference should be clear. Here λ is the probability that a particular nuclide will decay in 1 second. λ is also called the *decay rate*. The inverse $1/\lambda$ is the *average lifetime* T characteristic of the decay, $T = 1/\lambda$. T is the average time a nucleus (nuclide) survives before it decays by any of the particle emissions including γ-radiation. The exponential decay law expresses the result of a statistical process. Therefore fluctuations of T are expected. However, due to the large number of isotopes in a typical sample, fluctuations from the average λ or T are very small.

After one lifetime T the number of nuclei which still exist has decreased by the factor e: $N(T) = N_0/e$. This means that after time T, 63 % of all nuclides in a sample have decayed and 37 % still remain. Often not the lifetime T but the half-life is quoted $T_{1/2}$. According to its name, the half-life or half-life time $T_{1/2}$ is the time after which half of the radioactive isotopes have survived and the other half has decayed (Fig. 2.7). Half-life and average lifetime are related by:

$$T_{1/2} = \frac{\ln 2}{\lambda} = T \ln 2 = 0.693 \times T.$$

Thus the half-life is 30 % shorter than the average lifetime.

When radioactive substances are in the body, radioactivity not only decreases over time by decay but also by excretion. Therefore the effective half-life of many isotopes residing in the body is substantially shorter than the physical half-life. The biological half-life $T^{\mathrm{b}}_{1/2}$ and the physical half-life $T^{\mathrm{p}}_{1/2}$ add reciprocally and yield the effective

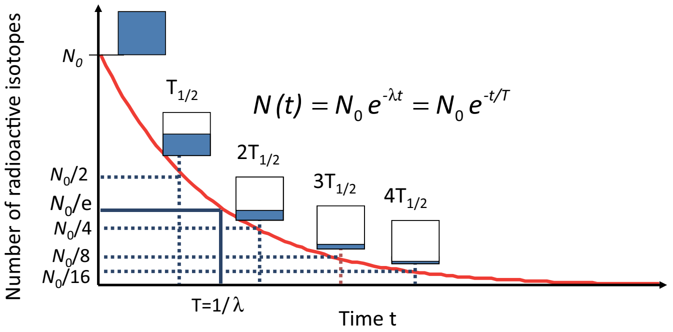


Fig. 2.7: Exponential decay law of radioactive isotopes. The blue colored area in the boxes symbolizes the remaining radioisotopes at the respective times. Number of isotopes after one lifetime is indicated by solid blue line, number of isotopes after successive half-life times are indicated by dashed lines.

half-life for radioisotopes in the body:

$$T_{\text{eff}} = \frac{T^{\text{b}}_{1/2} \cdot T^{\text{p}}_{1/2}}{T^{\text{b}}_{1/2} + T^{\text{p}}_{1/2}}\,.$$

As an example we discuss the radioisotope ^{131}I which is used for the scintigraphy of the thyroid gland. ^{131}I has a physical half-life of 8 days and a biological half-life of 80 days. The combined effective half-life is 7.27 days. The isotope ^{131}I is atypical as the biological half-life is usually shorter than the physical half-life. In Tab. 2.1 half-life times of some relevant radioisotopes in clinical practice are listed. The isotopes ^{90}Sr, ^{210}Pb, ^{210}Po, and ^{233}U are not used for therapy or diagnostics, but they can get into the body by contamination through radioactive waste. The radioisotope ^{14}C is constantly generated by cosmic radiation and naturally incorporated into the body through the food chain. Only after death does the exchange of ^{14}C with ^{12}C stop and the remaining ^{14}C isotopes decay according to the physical half-life. This can be used for dating archeological remains.

Tab. 2.1: Selected radioactive isotopes and their physical and biological half-life times.

Nuclide	Physical $T_{1/2}$	Biological $T_{1/2}$	Critical organs
^{90}Sr	28.1 a	11 a	Bones
^{210}Pb	22 a	730 d	Bones
^{210}Po	138 d	40 d	Spleen
^{233}U	1.63×10^5 a	300 d	Bones
^{131}I	8 d	80 d	Thyroid
^{14}C	5570 a	35 a	Fatty tissue
^{99m}Tc	6 h	24 h	Bones
^{201}Tl	3.04 d	9.8 d	Heart, brain

2.5.2 Nuclear activity

The exponential decay law is not really practical to handle in the laboratory since the number of radioisotopes in a particular sample is barely known. On the other hand, radioactivity is recognized by the emission of α-, β- or γ-radiation and can be detected with specialized counters. The number of emissions per time $\mathrm{d}N(t)/\mathrm{d}t$ is termed activity A and corresponds to the count rate in the detector:

$$A(t) = -\frac{\mathrm{d}N(t)}{\mathrm{d}t} = \lambda N(t) = \lambda N_0 \mathrm{e}^{-\lambda t} = A_0 \mathrm{e}^{-\lambda t}.$$

The activity is proportional to the decay rate λ. The larger λ is, the larger the activity and the count rate are, and the shorter the lifetime $T = 1/\lambda$ is. The activity has the same mathematical form as the exponential decay law, but with the advantage of expressing a measurable quantity: the emissions per time interval. The SI unit of activity is: $[A]$ = events/s = Becquerel (Bq).

1 Bq is one event per second, or one count per second if the efficiency of the counter is 100 %. An older unit is Curie (C), which corresponds to the activity of 1 g of radium that corresponds to 3.7×10^{10} Bq. The unit Curie is still sometimes used in clinics. While 1 C is a dangerously high activity, typical activity levels of samples handled in clinics are more on the μC level.

As an example for activity we discuss again the isotope ^{40}K. A typical human body contains about 160 g of potassium, of which 0.012 % or 0.02 g are radioactive. This corresponds to a number of $N_0 = (0.02/40) \times 1\,\mathrm{mol} = 3 \times 10^{20}$ radioactive ^{40}K isotopes in the body. These isotopes decay via β^--emission with a half-life of $T_{1/2} = 1.3 \times 10^9$ a = 4×10^{16} s. Then we determine the activity as:

$$A = \frac{\ln 2}{T_{1/2}} \cdot N = \frac{0.693}{4 \cdot 10^{16}} \cdot 3 \times 10^{20} \approx 5000/\mathrm{s} = 5000\,\mathrm{Bq},$$

which is the number quoted earlier. The specific activity a is the activity per mass unit. The SI unit is $[a]$ = Bq/kg. One kg of pure ^{40}K isotopes has an activity of 2.5×10^8 Bq/kg. More practical is the unit Bq/g. For instance, the specific activity of 1 g of radium is $a = 3.7 \times 10^{10}$ Bq/g. In brachytherapy discussed in Chapter 12 it is desirable to have high specific activities, i.e., high activity in small amounts of material.

The exponential law allows simple relationships between activities after incremental decay times t, $2t$, $3t$, etc. If we write

$$A_1 = A_0 \mathrm{e}^{-\lambda t} \Rightarrow \frac{A_1}{A_0} = \mathrm{e}^{-\lambda t},$$

then the activity after $2t$ is:

$$A_2 = A_0 \mathrm{e}^{-\lambda 2t} = A_1 \mathrm{e}^{-\lambda t} = A_1 \left(\frac{A_1}{A_0}\right),$$

and after nt we have the general expression:

$$A_n = A_{n-1} \left(\frac{A_1}{A_0}\right) = A_1 \left(\frac{A_1}{A_0}\right)^{n-1}.$$

These expressions are useful in clinics when handling short-lived isotopes. For instance, when the present activity A_1 is known and the activity from an hour ago was A_0, the activity in a few hours can be predicted from the ratio A_1/A_0.

2.5.3 Decay chains

After decay of a nucleus A (mother nucleus) with the decay rate λ_A the daughter nucleus B is often also radioactive with a decay rate λ_B. Then the daughter nucleus decays into C, which again decays with rate λ_C into D, etc. until a stable isotope is reached. The successive decay of mother nucleus into daughter nuclei B, C, etc. is known as radioactive chain. Four of those series occur naturally, for a discussion we refer to textbooks listed at the end of this chapter for further reading. It is important for us to recognize that the activities of mother and daughter nuclei are interdependent. The activity of the daughter nuclei depends on the production rate of the mother nuclei. As long as A has not decayed, B is not born and cannot decay. The activity of the daughter nucleus A_B is therefore a function of the production rate of the daughter nucleus, which is equal to the decay rate of the mother nucleus $\lambda_A N_A$ minus the decay rate of the daughter nuclei $\lambda_B N_B$:

$$A_B = \lambda_A N_A - \lambda_B N_B.$$

The equation can be integrated to yield the number of daughter nuclei $N_B(t)$ as a function of time [1]:

$$N_B(t) = \frac{\lambda_A}{\lambda_B - \lambda_A} N_A(0)(\mathrm{e}^{-\lambda_A t} - \mathrm{e}^{-\lambda_B t}).$$

Similar equations can be written for a sequence of decays from A to B to C, etc. The respective equations are called *Bateman equations* named after the mathematician who worked them out first.

An example is shown in Fig. 2.8 for the case of $\lambda_A \gg \lambda_B$, i.e., the lifetime of the daughter nuclei is considerably longer than the lifetime of mother nuclei. Then the number of daughter nuclei has a maximum at the time when most of the mother nuclei have decayed, and only at a later stage is the exponential decay of the daughter nuclei expressed. Finally the decay from B to C forms a stable isotope. At the other extreme, when the mother isotope is long lived but the daughter isotope decays much faster ($\lambda_B \ll \lambda_C$), then the number of daughter isotopes grows rapidly at the beginning and after some time an equilibrium is reached where A and B decay with the same rate governed by λ_A. This condition is indeed realized for the decay of ^{99}Mo/$^{99\mathrm{m}}$Tc shown in Fig. 6.3.

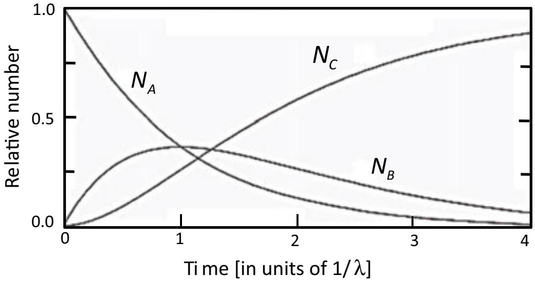


Fig. 2.8: Radioactive decay of successive mother (A) and daughter nuclides (B) with the same decay rates $\lambda_A = \lambda_B$. After decay of isotope B into C, isotope C is assumed to be stable.

2.6 Production of radioisotopes

All isotopes used in nuclear medicine are produced artificially by nuclear reactions. There are three main methods for isotope production, which are explicated in the following: neutron capture, nuclear fission, and charge particle irradiation. We start with a brief account of nuclear reactions and production rates in general terms that applies to all three methods and then turn to a presentation of the three isotope production methods.

2.6.1 Nuclear reactions

Nuclear reactions require a target containing the starting nucleus A and a projectile a that transforms isotope A into another isotope B by emission of particle b. The standard nomenclature for the reaction is:

$$a + A \rightarrow B + b$$

or equivalently:

$$A(a, b)B.$$

a and A are called the entrance channel, b and B are referred to as exit channel. Nuclear reactions either require or set free some energy, which is expressed as the Q value of the reaction:

$$Q = \Delta E_a + \Delta E_A - \Delta E_B - \Delta E_b,$$

where ΔE_x is the energy difference of particle x before and after reaction. If $Q > 0$, the reaction is exothermic, i.e., it releases energy, if $Q < 0$ the reaction is endothermic and requires energy. A more complete nomenclature for nuclear reactions is therefore:

$$A(a, b)B, Q[\text{MeV}].$$

Now we want to estimate how long it takes to produce radioactive material by irradiation. We assume a constant incoming flux J_a of particles a directed straight to the target, as indicated in Fig. 2.9. Then the flux or intensity is defined as number of particles N_a per area S and time t:

$$J_a = \frac{N_a}{St} = \frac{N_a}{Sl}\frac{l}{t} = n_a v_a .$$

The incident flux equals the product of particle density n_a and their average velocity v_a in an imaginary tube of length l.

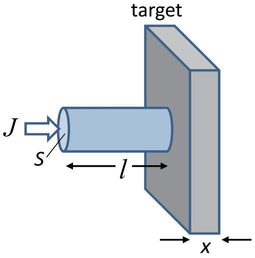


Fig. 2.9: Schematic for irradiation of a target.

Next we consider the *cross section* σ for isotope production. In the simplest form this is the ratio of successful events W per unit time and incident particle flux:

$$\sigma_a = \frac{W}{J_a} .$$

The unit of the cross section $[\sigma]$ is m^2. It is usual to quote the cross section in terms of barns, where 1 barn $= 10^{-28}\,\mathrm{m}^2 = 10^{-24}\,\mathrm{cm}^2$. Small cross sections are also quoted in terms of mb or μb.

If $\hat{N}_A$ is the total number of isotopes A in the target, and $\hat{N}_B$ are those that have been transformed to B in time t, then W can be expressed as:

$$W = \frac{\hat{N}_B}{\hat{N}_A t} = \frac{R_B}{\hat{N}_A} .$$

R_B is called the reaction rate for the production of isotope B. Re-assembling these equations, we find for the reaction rate:

$$R_B = \hat{N}_A \sigma_a J_a .$$

The last equation can be converted into practical and measureable units by taking into account the thickness of the target x, the target volume $V = Sx$ exposed to the incoming flux, and the number density of isotopes A in the target: $n_A = \hat{N}_A / V$:

$$R_B = n_A \sigma_a I_a x = \rho_A \sigma_a I_a \frac{N_{\mathrm{Av}}}{M_A} .$$

Here $I_a = J_a S = N_a/t$ is the flow of the incident particle beam, ρ_A is the mass density and M_A is the mass in atomic mass units of isotope A, and N_{Av} is the Avogadro number. Cross sections for particular isotopes and reactions can be found in nuclear tables.

Now we are ready to express the net rate for production of isotope B, which is the difference between the reaction rate R_B and the rate of decay of isotope B:

$$\frac{\mathrm{d}N_B}{\mathrm{d}t} = R_B - \lambda_B N_B = \hat{N}_A \sigma_a J - \lambda_B N_B .$$

Integration over time yields the number of isotopes B after time t plotted in Fig. 2.10:

$$N_B(t) = \frac{R_B}{\lambda_B}(1 - \mathrm{e}^{-\lambda_B t}).$$

Initially the increase in the number of isotopes B is controlled by the reaction rate R_B. But later the increase levels off and saturates at the value $N(t \to \infty) = R_B/\lambda_B$. The activity of isotope B follows from the last equation:

$$A_B = \lambda_B N_B(t) = R_B(1 - \mathrm{e}^{-\lambda_B t}).$$

The activity of isotope B has the same time dependence as the number of isotopes $N_B(t)$. Certainly, after switching off the incident particle beam, the activity of the isotopes B will decay as usual.

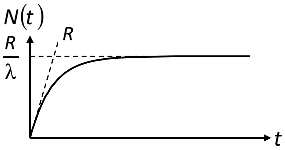


Fig. 2.10: Time dependence of the production of isotopes.

For isotope production it is important to come close to the saturation value R_B/λ_B. After one lifetime $t = T_B = 1/\lambda_B$ already 63 % of the saturation value has been reached, after two lifetimes 86 %, etc. Knowing the experimentally available target parameters $R_B = n_A \sigma_a I_a x$, the activity of the target after being exposed during an irradiation time t by a particle current I_a can easily be estimated.

2.6.2 Cyclotron generation of isotopes

In clinics radioisotopes are used for radiography and for radiotherapy. In some cases they have very short lifetimes, sometimes too short for delivery from an external laboratory to a clinic. Therefore specialized radiation clinics have their own cyclotron for the production of short-lived isotopes. A cyclotron is an accelerator for light, charged nucleons, such as protons, deuterium, or alpha particles. Production of radioisotopes by bombardement with charged particles is known as charge particle activation (CPA).

The physical principle of cyclotrons follows from the Lorentz force of charged particles in a magnetic field:

$$\vec{F}_\mathrm{L} = q(\vec{v} \times \vec{B}).$$

Here v is the particle velocity and B is the magnetic flux density, also known as magnetic induction. As v and B are oriented perpendicular to each other, we can omit the vector product and simply write: $F_\mathrm{L} = qvB$. On a circular trajectory the Lorentz force pulls the particles inwards and is balanced by the outward oriented centrifugal force: $F_\mathrm{C} = mv^2/r$. In equilibrium both forces balance:

$$qvB = \frac{mv^2}{r}$$

and we have a stable orbit with the orbital radius:

$$r = \frac{mv}{qB}.$$

Figure 2.11 shows the schematic of a cyclotron in top and side view. It consists of two D-shaped metal cans called *dees* with a gap in between and placed in a vacuum chamber. The proton source is positioned at the center of the cyclotron between the dees. The vacuum chamber is inserted between pole shoes of a large electromagnet forming a double C-shaped yoke. The coils are connected to a DC current source. Whenever a charged particle has completed a half circle in one of the dees and crosses the gap, it gets a kick by an RF generator connected with the dees, increasing the orbital velocity each half turn. Since the magnetic flux density B remains constant, the radius of the

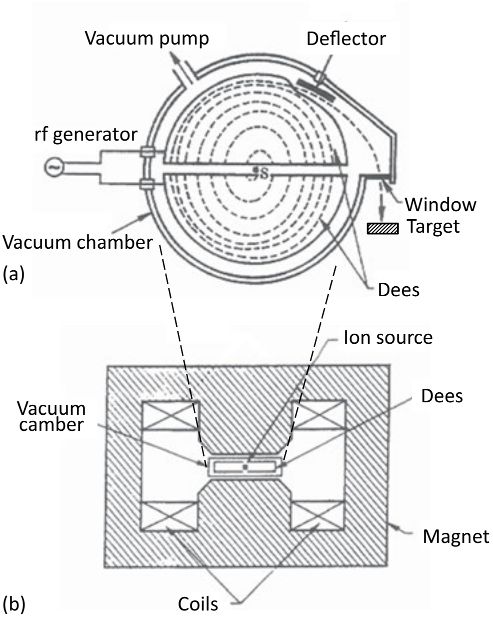


Fig. 2.11: Schematic of a cyclotron used for isotope production. (a) Top view of dees with ion source and target; (b) side view with magnet and coils. S is the source for protons being accelerated. The protons hit a target after penetrating a thin vacuum window.

orbit increases continuously forming an outward going spiral. While the circulation time T is independent of the radius and remains constant:

$$T = 2\pi \frac{m}{qB},$$

the kinetic energy increases and reaches a maximum when the particle orbit comes close to the radius R of the dees. Then the maximum kinetic energy is reached:

$$E_{\mathrm{kin,max}} = \frac{1}{2} \frac{(qBR)^2}{m}.$$

At the maximum energy the accelerated particles are deflected by an electric field in a condenser, then passing a vacuum window before hitting a target for isotope production via nuclear reaction.

This classic treatment of cyclotron kinematics should be corrected for relativistic mass increase when the particle velocity increases by more than 10 % of the speed of light. Then in all equations the mass m has to be replaced by $\gamma_0 m$, where $\gamma_0 = 1/\sqrt{1 - (v/c)^2}$. With RF frequencies between 10–30 MHz, kinetic energies of about 25 MeV are reached for protons and deuterons, and about 50 MeV for alpha particles.

With cyclotrons mainly γ-emitters for scintigraphy (Chapter 6) and short lived β^+-emitters for positron emission tomography (PET, Chapter 7) are produced. An example is the production of ^{18}F by proton bombardment according to the following nuclear reaction:

$$\mathrm{p} + {}^{18}\mathrm{O} \rightarrow {}^{18}\mathrm{F} + \mathrm{n}.$$

Protons hit an ^{18}O enriched target and convert the isotope ^{18}O by (p,n) exchange into ^{18}F. The cross section for this reaction has a maximum of 500 mb at about 5 MeV incident proton energy [1]. The natural abundance of the stable isotope ^{18}O is only 0.2 %. Therefore ^{18}O targets must be isotopically enriched. In most cases ^{18}O enriched water targets are utilized. After irradiation, fluorine is chemically extracted from the target and used for synthesizing fluoro-deoxy-glucose (^{18}F-FDG) required for PET scans. Using water targets has the additional benefit of thermalizing fast neutrons from the reaction in the target, which then can be captured by boron enriched protecting shields.

Table 2.2 lists some typical γ-emitting radioisotopes commonly used for scintigraphy in clinics. Isotopes produced by cyclotrons emitting positrons for PET and research are listed in Tab. 2.3. It has been suggested that $^{99\mathrm{m}}$Tc may also be produced by proton bombardment of an Mo target via the reaction: ^{100}Mo (p,2n)$^{99\mathrm{m}}$Tc as an alternative to the ^{235}U fission in a nuclear reactor. However, the $^{99\mathrm{m}}$Tc yield by cyclotron production is much smaller than by the fission process, such that for the foreseeable future the main source of $^{99\mathrm{m}}$Tc will still be reactor based.

Nowadays compact cyclotrons are available in clinics for radioisotope production by proton irradiation. The Eclipse Cyclotron shown in Fig. 2.12 from Siemens is an 11 MeV, negative ion, single particle accelerator designed for clinical and commercial production and distribution of ^{18}F radiotracers as well as other positron emitting radioisotopes such as ^{11}C, ^{13}N, ^{15}O, and ^{64}Cu.

Tab. 2.2: List of most common isotopes used for scintigraphy including energy of emitted γ-radiation and half-life time.

Isotope	Energy of γ-rad. [keV]	$T_{1/2}$ [h]	Production method
^{67}Ga	94, 184, 296	79	cyclotron p,n reaction
^{111}In	173, 247	67	cyclotron p,n reaction
^{123}I	159	13.1	cyclotron p,n reaction
^{201}Tl	x-rays after EC, 69–80	73	cyclotron p,n reaction

Tab. 2.3: Cyclotron based production of β^+-emitters for PET and clinical research.

Isotope	Reaction	Emitter	$T_{1/2}$ [min]
^{11}C	^{14}N(p,α)^{11}C	β^+	20.4
^{18}F	^{18}O(p,n)^{18}F	β^+	110
^{13}N	^{16}O(p,α)^{13}N	β^+	10
^{15}O	^{15}(p,n)^{15}O	β^+	2

Fig. 2.12: Compact cyclotron “Eclipse” for production of short-lived radioisotopes in radiation clinics (reproduced with permission from www.siemens.com/press).

2.6.3 Radioisotope production by fission

Nuclear reactors specialized for research and isotope production “burn” isotopically enriched ^{235}U. By capturing thermal neutrons the ^{235}U isotope is split in parts. Neutrons do not have to penetrate through a Coulomb barrier around the nucleon unlike protons. Therefore low energy neutrons are sufficient for neutron capture and subsequent fission reaction. The slower the neutrons are, the longer the reaction time is and the more likely ^{235}U nuclei become excited to a level of instability. The fission process splits the original nucleus into two unequal parts centered on the mass numbers 95 and 135, first realized by Otto Hahn and Fritz Strassmann in 1938, Nobel Prize 1944 for Chemistry. The fragments of the original uranium nuclei also release further neutrons. Parts of those neutrons are used for sustaining the fission reaction and parts are used for neutron scattering experiments, irradiation, and isotope production.

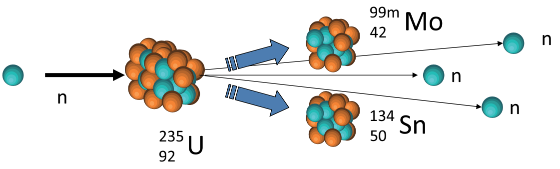


Fig. 2.13: Schematic illustration of the ^{235}U fission process.

There are several possibilities to produce $^{99}Mo/^{99m}Tc$ isotopes that are frequently used for scintigraphy: neutron capture, proton irradiation, and fission. The last one has by far the highest cross section and is the most efficient process [2]. During fission of ^{235}U the isotope ^{99}Mo is released as one of the fission fragments, sketched in Fig. 2.13. About 6.1 % of the fission products are ^{99}Mo. It has a half-life of 66 h and decays by β^--emission into the intermediate state ^{99m}Tc, which – in turn – has a half-life of 6 h and decays into ^{99}Tc by emission of 140 keV γ-radiation. This γ-radiation is utilized for scintigraphic imaging of the functioning of the brain, myocardium, thyroid, lungs, liver, gallbladder, kidneys, skeleton, blood, and last but not least for imaging tumors, discussed in more detail in Chapter 6. ^{99}Tc is also a β^--emitter, but with a very long half-life of 2.1×10^5 a, which can be regarded as stable on the time scale of a human life span. The biological lifetime is much shorter, in the order of one day via renal clearance. The decay scheme is shown in Fig. 2.14.

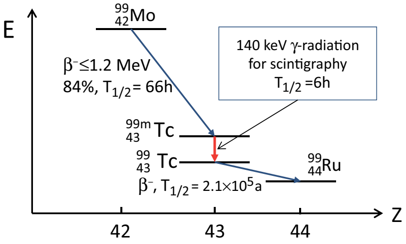


Fig. 2.14: Decay scheme of $^{99}Mo/^{99m}Tc$ into $^{99}Tc/^{99}Ru$.

Instead of extracting ^{99}Mo from nuclear fuel elements it is more practical to irradiate a metal plate of uranium containing about 20 % enriched ^{235}U. The uranium metal sheet is placed inside of a reactor moderator vessel to be exposed to thermal neutrons. Neutrons originating from fission of nuclear fuel at the core of the reactor are moderated to thermal energies by the surrounding water. After ^{235}U fission in the metal plate has approached saturation, typically after 6 days, the plate is removed and radiochemically processed. The radioactive ^{99}Mo isotopes are chemically separated in specialized isotope laboratories. The extracted ^{99}Mo/^{99m}Tc is split into smaller amounts of activity and safely deposited in generators and shipped to clinics for further elution of ^{99m}Tc (see Chapter 6 for further details). Worldwide there are only a few nuclear reactors available that specialize on ^{235}U irradiation and only a few laboratories have facilities for chemical extraction of ^{99}Mo/^{99m}Tc. This is in stark contrast to the heavy use of ^{99m}Tc for scintigraphy, estimated to be in the order of 30 million scintigrams taken annually worldwide. The heavy and increasing demand of this imaging modality would justify launching further centers for the production of ^{99}Mo/^{99m}Tc generators. One is soon to open at the research reactor FRM II in Munich, Germany.

2.6.4 Neutron activation

Stable isotopes can be converted into radioisotopes by neutron capture (NC). This process is called neutron activation and is characterized by an $A(\mathrm{n}, x)B$ reaction, where x stands for α-, β-, or γ-radiation. Neutron activation is used in many fields of science and technology, such as environmental science, geology, archeology, materials sciences and many others for detecting trace elements or for generating radioactive sources that can be used otherwise. In clinics, specific isotopes are used for radiography (scintigraphy and PET) and for radiotherapy (brachytherapy).

The cross section for neutron capture scales with $1/v_\mathrm{n}$, where v_n is the neutron velocity. Thermal neutrons with energies of about 25 meV are predestinated for such reactions. Neutron activation is rather easy and efficient. It only requires a “rabbit” system where specimens are enclosed in carriers fed into a tube that goes into a reactor vessel. After reaching the specified activity, the carrier is removed from the reactor vessel and further processed or analyzed. The relative ease of producing radioisotopes by neutron capture has also its downside. Neutron activation is not particularly specific. Most of the heavier isotopes can be activated. For clinical applications this constitutes a serious problem because of potential toxic contaminations. Therefore the utmost care is mandatory to activate only isotopically clean samples.

Concerning clinical applications of neutron activation we have to distinguish two variants: in vivo activation and in vitro activation. For in vivo activation, neutron-capturing but nonradioactive isotopes are first administered to the patient. The isotopes are then exposed to fast neutrons, which become moderated in the watery environment of the body and trigger (n, x) reactions locally in the tissue. The best known and successful application of this method is the boron neutron capture therapy (BNCT) via (n,α) reaction presented in Chapter 11. Here we focus on in vitro activation of samples in a thermal neutron environment.

Activation by thermal neutron capture can be used for a variety of nuclear decays, exemplified by the following reactions:

(a) (n,γ) reaction, example:

$$^{98}_{42}\mathrm{Mo} + \mathrm{n} \rightarrow {}^{99}_{42}\mathrm{Mo} + \gamma \quad (\sigma = 110\,\mathrm{mbarn})$$

(b) (n,γ) + β^- reaction, example:

$$^{130}_{52}\mathrm{Te} + \mathrm{n} \rightarrow {}^{131}_{52}\mathrm{Te}^{\star} + \gamma;$$
$$^{131}_{52}\mathrm{Te}^{\star} \rightarrow \beta^- + {}^{131}_{53}\mathrm{I} \quad (\sigma = 67\,\mathrm{barn})$$

(c) (n,α) reaction, example:

$$^{10}_{5}\mathrm{B} + \mathrm{n} \rightarrow {}^{7}_{3}\mathrm{Li} + \alpha \quad (\sigma = 767\,\mathrm{barn})$$

The first example refers to the already presented ^{99}Mo/$^{99\mathrm{m}}$Tc production. The cross section for this reaction is, though, rather small and therefore considered inefficient. Production via fission is preferred over thermal neutron capture.

The second example is a sequential transition via a short lived Te* radioisotope that decays further by beta emission to the radioisotope ^{131}I. This isotope is also a beta emitter with a half-life of 8 days, which is sometimes used for radiotherapy of the thyroid and less so for scintigraphy.

The third example shows the reaction already quoted for in vivo BNCT in conjunction with fast neutrons. This reaction is also used for detecting thermal neutrons by BF_3 gas-filled proportional counters or Geiger–Müller counters. The design of these counters is presented in Section 4.4.

Further examples of isotope production of clinical interest generated either by fission or by neutron capture are listed in Tab. 2.4 together with their main field of application.

Tab. 2.4: Radioisotopes generated either by nuclear fission or by neutron capture used in brachytherapy or scintigraphy.

Radioisotope	Reaction	Production	Application
^{137}Cs	^{235}U(n,γ)^{137}Cs → ^{137}Ba	fission	Brachytherapy
^{90}Sr	^{235}U(n,γ)^{90}Sr → ^{90}Y	fission	Brachytherapy
$^{99\mathrm{m}}$Tc	^{235}U(n,γ)^{99}Mo → $^{99\mathrm{m}}$Tc	fission	Scintigraphy
^{133}Xe	^{235}U (n,γ)^{133}Xe	fission	Scintigraphy
^{60}Co	^{59}Co(n,γ)^{60}Co → ^{60}Ni	neutron capture	Brachytherapy
^{192}Ir	^{191}Ir(n,γ)^{192}Ir → ^{192}Pt	neutron capture	Brachytherapy
^{125}I	^{124}Xe(n,e)^{125}I	neutron capture	Brachytherapy
^{103}Pd	^{102}Pd(n,γ)^{103}Pd	neutron capture	Brachytherapy
^{103}Pd	^{103}Ag(n,γ)^{103}Pd	neutron capture	Brachytherapy

2.7 Summary

1. Nucleons are composed of protons and neutrons.
2. Atomic number *A* is the sum of protons *Z* and neutrons *N*, the nomenclature is ^{A}X, *X* stand for a chemical symbol.
3. For light nucleons, the number of protons and neutrons is about equal. With increasing atomic number, nucleons tend to have more neutrons than protons.
4. Isotopes are atoms with the same number of protons but different number of neutrons.
5. Isotopes that have either too many or too few neutrons are unstable.
6. All isotopes with atomic mass beyond lead (^{208}Pb) are unstable.
7. Radioisotopes are unstable isotopes. They decay by emission of either α-, β^+-, β^--, γ-radiation, or by electron capture (EC) from the K-shell, converting radioisotopes into stable isotopes.
8. All radioisotopes decay according the exponential decay law.
9. Long-lived isotopes with half-life times in the range of the age of the earth can still be found in nature.
10. In decay chains the production of daughter nuclei depends on the decay rate of mother nuclei.
11. Radioisotopes can artificially be produced by proton irradiation in accelerators, by fission of heavy nucleons, and by thermal neutron capture.
12. ^{99m}Tc for scintigraphy is extracted from fission products.
13. Short-lived positron emitters for PET application are produced by proton bombardment in a cyclotron.

References

[1] Hess E, Takacs S, Scholten B, Tarkanyi F, Coenen HH, Qaim SM. Excitation function of the ^{18}O(p,n)^{18}F nuclear reaction from threshold up to 30 MeV. Radiochim Acta. 2001; 89: 357–362.

[2] Organisation for Economic Co-operation and Development. The supply of medical radioisotopes. The path to reliability. Nuclear Development. Paris: OECD Publishing; 2011.

Further reading

Krane KS. Introductory nuclear physics. John Wiley & Sons; 1988.

Lilley J. Nuclear physics, principles and applications. Wiley; 2013.

Cherry SR, Sorenson JA, Phelps ME. Nuclear medicine. 4th edition. Elsevier; 2012.

Dance DR, Christofides S, Maidment ADA, McLean ID, Ng KH, editors. Diagnostic radiology physics – A handbook for teachers and students. IAEA; 2014.

International Atomic Energy Agency (IAEA). Cyclotron produced radionuclides: Principles and practice. Technical Reports Series No. 465. Vienna: International Atomic Energy Agency (IAEA); 2008.

3 Interaction of radiation with matter

3.1 Introduction

In the last two chapters we discussed various forms of radiation produced either by accelerators, by nuclear reactors, or via radioactive decay: x-rays, electrons, positrons, α-particles, protons, neutrons. Table 3.1 summarizes these particles and their generation. Next it is essential to understand how these particles interact with matter. We need this information for several reasons. First, radiation is dangerous. Therefore appropriate radiation protection is mandatory. Second, some of the radiation is used for imaging. Imaging requires contrast and contrast is based on the various ways that radiation interacts with matter. Third, radiation is used for destroying cancerous cells. Therefore the biological effectiveness of radiation damage in healthy versus cancerous tissue requires a detailed analysis.

Tab. 3.1: Particles and their generation.

Particles	**Generation**
x-rays (up to 150 keV)	x-ray tube, synchrotron
x-rays (1–25 MeV)	Linear accelerator
Protons	Cyclotron
Neutrons	Nuclear reactor
γ-radiation	Radioactive decay
β^- and β^+	Linear accelerator Radioactive decay Pair production Neutron capture
α-particles	Radioactive decay Neutron capture

When a particle beam hits a target, three processes, sketched in Fig. 3.1, always take place independent of the type of radiation:

1. transmission
2. absorption
3. scattering

DOI 10.1515/9783110553116-005

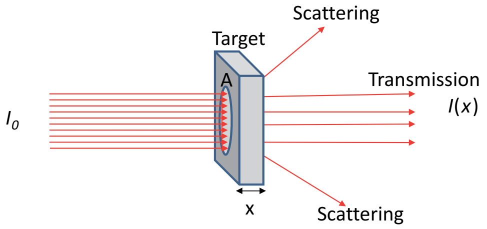


Fig. 3.1: Incident particle beam becomes attenuated in a target by absorption and scattering.

The primary beam intensity I_0 – also called flux – is defined as the number of particles N that impinge with perpendicular incidence on a slab of target material per cross-sectional area A and per time t:

$$I_0 = \frac{N}{\mathrm{At}} = \frac{\Phi}{t}.$$

The ratio $\Phi = N/A$ is called particle *fluence*. Intensity (flux) is therefore defined as the fluence rate: $I = \dot{\Phi}$, i.e., the first derivative of the fluence with respect to time. The SI unit of intensity is $[I]$ = particle number × $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The definition of fluence and intensity applies to all types of radiation listed in Tab. 3.1. The transmitted intensity $I(x)$ after passing a target of thickness x depends on particle type, particle energy, and target material. A general statement that would apply to all types of particles is not possible. The main distinction is between charged particles (electrons, positrons, protons, α-particles) and electromagnetic radiation (x-rays and γ-radiation). The interaction of neutrons with matter is again different and requires special consideration. In the following we first consider the interaction of electromagnetic (EM) radiation with matter, followed by similar considerations for charged particles and neutrons.

3.2 Interaction of EM radiation with matter

3.2.1 Mass attenuation coefficient

The energy distribution of a photon beam may be as sharp as a γ-emission line from radioactive decay or as wide as a bremsstrahlung spectrum. When passing through a solid, liquid, or gaseous target the primary beam intensity becomes attenuated, i.e., the intensity decreases exponentially with the thickness x of the target. The transmitted beam becomes attenuated by two processes: absorption and scattering. All those particles are transmitted that are neither absorbed nor scattered out of the area defined by the receiving detector. They form the net amount, described by the Lambert–Beer law, already introduced for the transmission of light in the visible range (see Section 8.6.5/Vol. 1):

$$I(E) = I_0(E)\exp(-\mu x).$$

Here μ is the linear attenuation coefficient, SI unit $[\mu] = \mathrm{m}^{-1}$. The physics of absorption is contained in the linear attenuation coefficient μ. As the total attenuation depends on the density of the target material, μ is customary normalized by the density for tabulation: μ/ρ (SI unit $[\mu/\rho] = \mathrm{m}^2/\mathrm{kg}$), which is known as the *mass attenuation coefficient*. In fact, μ/ρ is the sum of four contributions:

$$\frac{\mu}{\rho} = \frac{\mu_\mathrm{p}}{\rho} + \frac{\mu_\mathrm{c}}{\rho} + \frac{\mu_\pi}{\rho} + \frac{\mu_\mathrm{s}}{\rho}.$$

The attenuation for EM radiation is then written:

$$I(E) = I_0(E)\exp\left(-\left(\frac{\mu_\mathrm{p}}{\rho} + \frac{\mu_\mathrm{c}}{\rho} + \frac{\mu_\pi}{\rho} + \frac{\mu_\mathrm{s}}{\rho}\right)\rho x\right).$$

μ_p/ρ stands for the photoelectric effect, i.e., the complete absorption of x-ray photons and conversion into potential and kinetic energy of electrons.

μ_c/ρ is the coefficient for Compton scattering; the Compton effect refers to inelastic scattering of x-ray photons at bound electrons, where momentum and energy of the photon is partially transferred to electrons. Compton scattering has to be distinguished from elastic scattering (Thomson scattering) of x-rays used for structural analysis.

μ_π/ρ is the coefficient for pair production; high energy x-ray photons with an energy greater than the combined rest mass of electrons and positrons ($> 1.022\,\mathrm{MeV}$) can be converted into electron-positron pairs.

μ_s/ρ is the coefficient for coherent x-ray scattering. Coherent scattering is an elastic x-ray scattering process that does not change the energy of photons, but their direction.

The photoelectric effect and pair production both completely remove x-ray photons from the primary beam and convert their energy to another particle. This is what we call absorption. In the case of the Compton effect the x-ray photon energy and direction is changed by inelastic scattering but not removed. Nevertheless the Compton scattered photon is missed in the transmitted beam, since it is scattered into a different direction. Similarly, elastic scattering also steers the beam into a different direction, but does not change its energy. Thus the transmitted beam is attenuated by two absorption processes and two scattering processes. In the following we discuss these four processes in more detail.

3.2.2 Coherent scattering

Coherent scattering is an elastic x-ray scattering process, i.e., a process that does not transfer energy to other particles. Elastic scattering has different names depending on the target: scattering of EM waves at single bound electrons is referred to as *Thomson scattering*; scattering at single atoms is referred to as *Rayleigh scattering*; and scattering at nano- to microsized particles is referred to as *small angle scattering*. Thomson

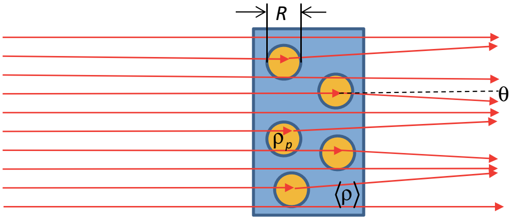


Fig. 3.2: Small angle x-ray scattering at particles of radius R that have a different electron density than the surrounding.

scattering determines the cross section for elastic x-ray scattering in general. Rayleigh scattering is a coherent scattering effect where the EM wave interferes with itself when scattered from a homogeneous cloud of charge representing the total number Z of electrons in an atom. Rayleigh scattering is directed in the forward direction, whereas no intensity is scattered backwards due to destructive interference. In the field of x-ray scattering this type of scattering is known to yield the so called atomic form factor, which modulates any type of x-ray scattering including small angle scattering. Small angle x-ray scattering (SAXS) is an elastic and coherent scattering process of EM waves at density fluctuations in the target. The scattering takes place at particles that have a different electron density ρ_p than the average electron density $\langle\rho\rangle$ of the surrounding medium in which these particles are embedded, illustrated in Fig. 3.2. For instance, if the target is homogenous water, no small angle scattering occurs; but if water contains erythrocytes, SAXS can be observed. The scattering intensity $I(\theta)$ as a function of scattering angle θ scales with the number density of particles n_p in the matrix, the squared difference in electron density between particles (erythrocytes) and surrounding medium (water) $(\langle\rho\rangle - \rho_\mathrm{p})^2$, and the 4th power of the ratio: wavelength to gyration radius R of the particle $(\lambda/R)^4$. The SAXS intensity then reads:

$$I(\theta) \propto n_\mathrm{p}(\langle\rho\rangle - \rho_\mathrm{p})^2 \left(\frac{\lambda}{R\theta}\right)^4 .$$

In all cases relevant for x-ray radiography the x-ray wavelengths range from 0.1 nm (10 keV) to 0.01 nm (100 keV), whereas the particle radius of cells is on the order of micrometer. The (λ/R) ratio is therefore in the order of 10^{-4}–10^{-5}, indicating that SAXS drops off quickly with increasing x-ray energy (decreasing x-ray wavelength) and takes place at very small angles of about 5/100°. Nevertheless, SAXS can cause blurring of high resolution x-ray images when all other geometric effects have been taken care of. SAXS leads to apparent attenuation of the primary incident beam intensity if the recording detector can discriminate between parallel beams and those which make a small angle θ with respect to the parallel orientation.

3.2.3 Photoelectric effect

The photoelectric (PE) effect is a two-step process shown schematically in Fig. 3.3. First an x-ray photon is absorbed by an atom and the total x-ray energy is transferred to a core electron, which may subsequently go to an unoccupied energy state in an outer shell, or may leave the atom with a kinetic energy of $E_e = E_1 - E_b$, where E_1 is the primary x-ray photon energy and E_b is the binding energy of the photoelectron, which in this example is the binding energy of the K-shell electron: $E_b = E_K$. In the second step the electron hole left in the inner shell is filled by an electron from the next higher shell, as we have already discussed for the production of characteristic x-ray radiation via electron impact in Chapter 1. The x-ray fluorescence radiation with energy $E_2 = E_K - E_L$ always has lower energy in comparison to the primary incoming beam: $E_2 < E_1$.

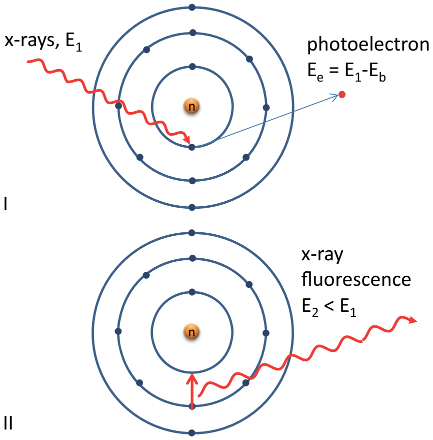


Fig. 3.3: Photoelectric effect takes place in two steps. First step: x-ray absorption and transfer of energy to an electron. Second step: filling of core hole by an electron from the outer shell and emission of x-ray fluorescence radiation. n at the center stands for nucleus.

The linear x-ray mass attenuation coefficient for the PE effect $\mu_p(E, Z)/\rho$ depends strongly on the energy of the incident x-rays and on the atomic number Z. The mass attenuation coefficient of Fe is plotted in Fig. 3.4 for illustration. Two aspects can immediately be recognized: (1) with increasing energy the mass absorption coefficient decreases; (2) when the x-ray energy equals the binding energy of electrons in the K-, L-, or M-shell, the absorption coefficient shows a sharp increase. These absorption edges are important for contrast enhanced x-ray radiography. In general the dependence of $\mu_p(E, Z)/\rho$ on energy and atomic number of the target material is expressed in the following relation:

$$\frac{\mu_p}{\rho} \approx \lambda^3 Z^3 .$$

Because of the strong Z dependence of the mass absorption coefficient, lead with a high Z ($Z = 82$) is used for protection against x-rays, barium ($Z = 56$) is used for contrast enhancement of x-ray radiographic images of the gastrointestinal tract, whereas low Z beryllium ($Z = 4$) or aluminimum ($Z = 13$) are often used as x-ray “transparent” windows.

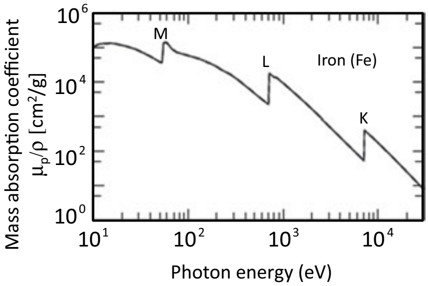


Fig. 3.4: Mass absorption coefficient of Fe as a function of energy.

Considering radiation damage effects by PE we have to realize that the entire energy of an x-ray photon is transferred to an electron in one step. The kinetic energy of the photoelectron is then: $E_e = E_1 - E_b$ as quoted before. In soft tissues composed mainly of H, C, N, O, P, S, and a few light metals, the binding energy of about 50–500 eV can be neglected in comparison to the high energy x-ray photons. After PE conversion high energy photoelectrons with kinetic energies of some 100 keV and more will scatter strongly due to Coulomb interaction with ions in the material. Scattered photoelectrons will emit synchrotron radiation and excite other electrons in atoms from inner shells, generating new x-ray photons, etc. until after a cascade of events including scattering, excitation, emission, reabsorption the photoelectron comes to a stop. This scenario is identical to the one β-particles experience in matter, which is discussed below and indicated in Fig. 3.5. Because of this straggling process, high energy photoelectrons in tissue have a range of a few tenths of millimeters. The biological damage done by x-ray absorption is entirely due to these high energy photoelectrons and determines the biological quality factor Q_R for radiation discussed in Chapter 4.

It should be noted that PE absorption in matter is a stochastic process that diminishes the x-ray intensity continuously and exponentially, but does not change the energy of those photons that have not been absorbed and which are transmitted. This

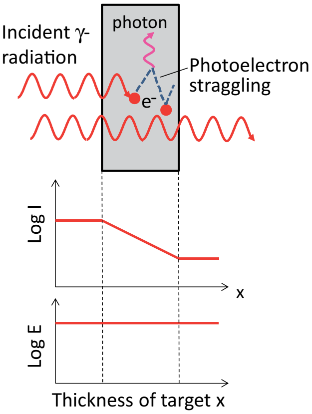


Fig. 3.5: Absorption of photons in matter via the photoelectric process. The photon intensity decreases exponentially with thickness *x* of the target, whereas the photon energy does not change. At the same time, high energy photoelectrons are generated, which have a certain straggling distance in the material before coming to rest.

situation is illustrated in Fig. 3.5. Therefore, x-rays going through matter do not have a range in contrast to charged particles, but they have a half-thickness $x_{1/2}$ where half of the intensity has diminished:

$$x_{1/2} = \frac{\ln 2}{(\mu/\rho)\rho}.$$

3.2.4 Compton effect

The *Compton effect* is an inelastic and incoherent scattering process of x-rays or γ-rays at single bound electrons. The Compton effect is to be distinguished from coherent elastic x-ray scattering discussed in Section 3.2.2. In Compton scattering the incident photon is not absorbed as in the PE effect but rather scattered in a different direction. Nevertheless, it leads to an attenuation of the primary beam. Due to inelastic scattering, some energy is transferred from the scattered photons to the electrons (Fig. 3.6). Using conservation laws for energy and momentum the wavelength change $\Delta\lambda$ of the photons can be calculated as a function of photon scattering angle θ:

$$\lambda_\mathrm{f} - \lambda_\mathrm{i} = \Delta\lambda = \frac{h}{m_0 c}(1 - \cos\theta) = \lambda_\mathrm{C}(1 - \cos\theta) \geq 0.$$

Here m_0 is the rest mass of electrons and the subscripts i, f designate initial and final wavelengths. The prefactor $\lambda_\mathrm{C} = h/(m_0 c) = 0.00243\,\mathrm{nm}$ is known as the Compton wavelength for electrons, which is a natural constant. The maximum energy loss of photons occurs at a scattering angle of $\theta = 180°$ and is $\Delta\lambda = 2\lambda_\mathrm{c}$. Compton scattering becomes increasingly important with increasing photon energy, producing high energy electrons in a forward scattering cone of opening angle 2ϕ. The cone narrows with increasing photon energy. When exposed to biological tissue much radiation damage can be caused within this cone of hot electrons. Biological effects of radiation are discussed in more detail in Chapters 4 and 8.

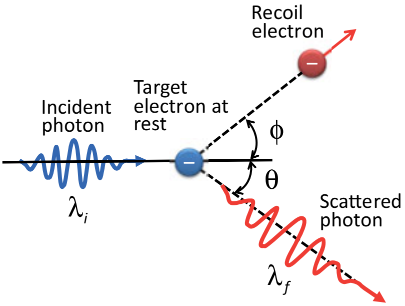


Fig. 3.6: Schematic of the Compton scattering process (graph adapted from http://hyperphysics.phy-astr.gsu.edu/hbase/hph.html).

3.2.5 Pair production

Pair production is a relativistic process. If the photon energy E_γ is larger than twice the rest mass of an electron, then in the field of a nucleus the photon energy can be converted into a pair of particle and antiparticle, i.e., in an electron-positron pair. Electrons and positrons are equal in all physical aspect, but they have opposite charge and consequently opposite magnetic moments. The onset energy for pair production is

$$E_\gamma \geq 2m_0c^2 = 1.022\,\text{MeV}.$$

Any photon energy in excess of the threshold energy is diverted in equal amounts to the electron-positron pair, which then fly apart at an angle of 180° for reasons of momentum conservation, as shown schematically in Fig. 3.7. Pair production is the inverse process to pair annihilation.

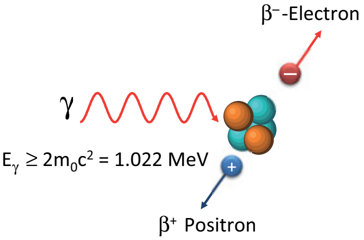


Fig. 3.7: In the field of a nucleus a high energy photon may be converted into an electron-positron pair. The threshold energy for pair production is 1.022 MeV.

3.2.6 Comparison of photon-electron interactions

Finally we discuss the relative importance of the mass attenuation coefficients for coherent scattering, photoelectric absorption, Compton scattering, and pair production. For photon energies up to about 0.5–1 MeV, the photoelectric effect is the dominating process, always accompanied by Compton scattering. In the high energy region beyond 1 MeV, pair production takes over. In the intermediate region inelastic Compton scattering prevails. Plotting all four mass attenuation coefficients together on a logarithmic scale in Fig. 3.8 for the case of Fe, the sum shows a minimum at about 1 MeV and is higher for lower and higher energies.

The important point to realize is the fact that x-ray attenuation by any of the discussed means does not simply reduce the intensity. Attenuation implies that radiation is converted into other types of energy and often produces hot electrons that may have high biological effectiveness. For instance, when x-ray radiotherapy is performed with photons of energy 20 MeV and higher, the photons are converted in equal amounts to hot Compton electrons and hot electron-positron pairs.

In summary, the far most important process in clinical practice is the photoelectric effect and Compton scattering for x-ray radiography with photon energies in the

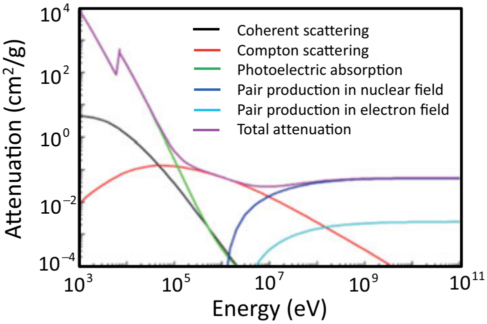


Fig. 3.8: Mass attenuation coefficients plotted as a function of energy. The sharp absorption edge at 7.1 keV refers to the K-edge of Fe (adapted from www.wikiwand.com, © Creative Commons).

range of 50–150 keV. For x-ray radiotherapy with energies beyond 4 MeV, the photoelectric effect plays no role. At these high photon energies only Compton electrons and electron-positron pairs contribute to the therapeutic effect.

3.3 Interaction of charged particles with matter

3.3.1 Alpha particles

α-particles originate from nuclear decay of heavy isotopes. After emission from radioisotopes, these charged particles have a very high kinetic energy of about 5 MeV. When penetrating matter, α-particles have a well-defined range in contrast to photons. Figure 3.9 impressively shows the range of α-particles in air emitted from a polonium source at the center. The picture was recorded with a cloud chamber, tracing the straight trajectories of the α-particles until they come to a sudden stop within a very short distance. The total path length is typically 5 cm in air and the straggling at the end is not more than 5 % of the total path length.

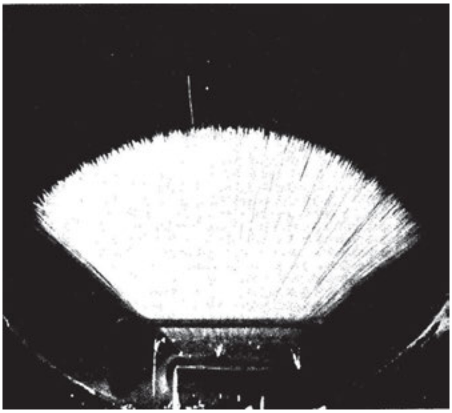


Fig. 3.9: Alpha particle emission during decay of ^{212}Po in ^{208}Pb visualized with a cloud chamber (reproduced from W. Finkelnburg, Introduction to Atomic Physics, Springer [1]).

First we consider the energy loss of α-particles when penetrating matter. Because of their charge the main interaction with matter is of Coulomb type that results in inelastic scattering with electrons. Since the mass ratio of α-particles to electrons is very large (about 8000) each collision with an electron carries off only little energy from the α-particle, in the order of 30 eV per collision. While electrons are scattered over a wide range of angles, the path of α-particles remains more or less straight. This situation is sketched in Fig. 3.10 and is fundamentally different from attenuation via the photoelectric effect.

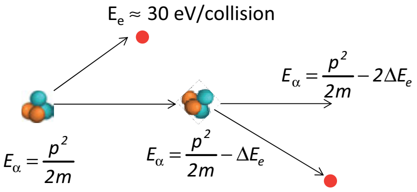


Fig. 3.10: α-particles collide with electrons and lose energy little by little until they are finally stopped after traversing a distance *R*.

In Fig. 3.11 the number of charged particles *N* in a particle beam is plotted from the start at $x = 0$ to the final stop at $x = \langle R \rangle$. Also plotted in the same graph is the average energy loss $\langle \mathrm{d}E/\mathrm{d}x \rangle$ per distance *x* traveled. The ratio $\langle \mathrm{d}E/\mathrm{d}x \rangle$ is called *Linear Energy Transfer* (LET). LET is the rate at which energy is deposited by a particular type of radiation when passing through matter. The unit of LET is usually keV/μm.

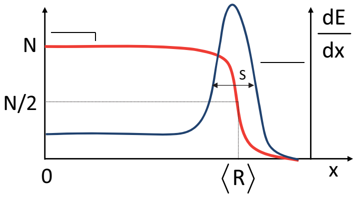


Fig. 3.11: Range of charged particles in matter. *N* is the number of particles in the beam and d*E*/d*x* is the energy loss per distance traveled, known as linear energy transfer (LET). ⟨*R*⟩ is the range of charged particles and the dashed line indicates the distance at which the original number of particles has dropped to half.

We notice that $N(x)$ stays relatively constant over a long distance and then suddenly drops to zero. The distance where the number of particles drops to half the original value is called the average *range* of charged particles $\langle R \rangle$. We notice that the LET $\langle \mathrm{d}E/\mathrm{d}x \rangle$ is rather small and constant in the beginning, but then increases dramatically where the number of particles starts to decrease, followed by a sharp drop. The peak in $\langle \mathrm{d}E/\mathrm{d}x \rangle$ at $\langle R \rangle$ is called Bragg peak (not to be confused with the Bragg peak in coherent x-ray diffraction, but the discoverer is the same William Henry Bragg, who together with his son William Lawrence Bragg received the Nobel Prize in Physics in 1915 for their x-ray diffraction work). The full width of the peak at half maximum is the *straggling* distance S over which the particles get finally stopped. A typical range for 5 MeV α-particles in air at 1 bar is 5 cm, the straggling distance is about 2.5 mm.

The energy loss per distance traveled of charged particles by passing through matter with the atomic number Z is described approximately by the Bethe–Bloch equation and assuming that energy loss takes place by ionization of the target material this equation reads [2]:

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = 4\pi N_\mathrm{A} r_0^2 m_0 c^2 \rho \frac{Ze^2}{A\beta^2} \ln\left(\frac{W}{E_\mathrm{ion}} - \beta^2\right).$$

Here N_A is the Avogadro number, r_0^2 is the classical electron radius and m_0c^2 is the electron rest mass, $\beta = v/c$ and v is the projectile velocity, ρ is the density of the target material, Ze and A are respectively the ion charge and atomic mass of the target material, W the kinetic energy of the projectile, and E_ion is the mean ionization energy of light atoms in the target. The total energy loss is found from integration over the path up to rest at $\langle R\rangle$.

$$E_\mathrm{tot} = \int_0^R \frac{\mathrm{d}E}{\mathrm{d}x}\,\mathrm{d}x.$$

The total energy is equivalent to the initial energy of the charged particle immediately before penetrating some slab of material with atomic number Z. Integration yields the range R. For the range of α-particles in air at standard temperature, pressure, and humidity conditions, an empirical equation is quoted as follows [3]:

$$R_{\alpha,\mathrm{air}} = 0.325E^{3/2}.$$

Here the energy is in MeV and the range in cm. For scaling from air to another material, the following conversion, known as Bragg–Kleeman rule, is used:

$$R_{\alpha,1} = \frac{\sqrt{A_1}}{\rho_1}\frac{\rho_\mathrm{air}}{\sqrt{A_\mathrm{air}}} R_{\alpha,\mathrm{air}}.$$

Here A and ρ are the respective atomic numbers and mass densities for air and material 1. For air the effective numbers are: $A_\mathrm{air} = 14.6$ and $\rho_\mathrm{air} = 0.001293$ g/cm^3, and for tissue: $A_\mathrm{tissue} = 9$ and $\rho_\mathrm{tissue} = 1$ g/cm^3. The range equation can also be converted to other charged particles, like protons or carbon ions. For conversion to protons we find [2]:

$$R_\mathrm{p} = \frac{m_\mathrm{p}}{m_\alpha}\left(\frac{Z_\alpha}{Z_\mathrm{p}}\right)^2 R_\alpha - k = 1.0072R_\alpha - k,$$

where k (≈ 20 cm) is a constant accounting for differences in ionization energy of electrons between protons and α-particles. These empirical equations depend on the particle energy and have to be rescaled for specific energy ranges. In any case, the important point to recognize from this section is the fact that heavy charged particles have a well-defined range $\langle R \rangle$ in materials of density ρ, which scales with the initial particle kinetic energy E:

$$\langle R \rangle \approx E^{\delta},$$

where δ is an appropriate exponent ranging between 1.5 and 1.8. This relation implies firstly that the range of charged particles in tissue can be tuned by the initial kinetic particle energy and secondly that within this range most of the particle energy is deposited. The possibility to tune the position of the Bragg peak in the target area by the incident particle energy is one of the major advantages of particle beam therapy over x-rays, which do not feature a range. This is discussed in more detail in Chapter 10 on proton therapy.

3.3.2 Beta particles

β-particles are also charged. But their behavior is quite different from α-particles and protons due to their much smaller mass. They may collide with other electrons in the material elastically or inelastically, i.e., with or without energy transfer. In either case scattering implies a large angular change of the trajectory. Deflection of the beam from the original path entails centripetal acceleration, which – in turn – generates synchrotron radiation. Fast β-particles quickly lose their energy through emission of photons, either via Cherenkov radiation or synchrotron radiation, or by ionization of the target material. Cherenkov radiation (named after the Russian physicist and Nobel laureate of physics in 1958 Pavel Alekseyevich Cherenkov) occurs whenever the speed of charged particles v_p in a medium with refractive index n is higher than the phase velocity of light in the medium: $v_p > c/n$. Cherenkov radiation is the light analogue to supersonic sound propagation. Like for sound, the radiation is emitted in a cone behind the traveling particle with an opening angle $\cos\theta_c = 1/n\beta$, where $\beta = v_p/c$. The energy loss through emission of synchrotron and Chernenkov radiation is proportional to:

$$-\frac{dE_\beta}{dx} \propto Z^2 E_\beta,$$

whereas energy loss through ionization has the form:

$$-\frac{dE_\beta}{dx} \propto Z \ln(E_\beta).$$

This shows that the energy loss through radiation is more dominant, in particular for high Z target materials.

Similar to α-radiation, β-particles have a maximum range $\langle R \rangle$. But in contrast to α-particles the range is much longer and straggling is more extended. The range of β-particles depends on their primary energy and the atomic number of the target. The greater straggling is in part also due to the larger energy distribution of β-particles to begin with since during β-decay the energy is shared with a neutrino. At the end of their range, β^+-particles annihilate together with an electron from the target, emitting their combined rest mass in the form of two high energy photons, traveling in opposite directions. Positron annihilation and γ-emission tomography (PET) make use of the annihilation process for imaging of tumors, which is the topic of Chapter 7.

Summarizing, for the range of α-, β-, and γ-particles we can state: α-particles have a well-defined range; β-particles have a much longer range for the same energy and their straggling is more extended; γ-particles and x-rays have no range, but a well-defined half-intensity thickness. Typical ranges are 5 cm in air for α-particles versus 5 m for β-particles, but little attenuation of γ-radiation in air. These different ranges are sketched in Fig. 3.12 and tabulated in Tab. 3.2.

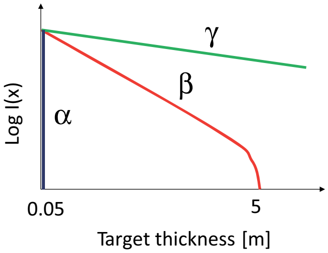


Fig. 3.12: Qualitative overview of the range of different particles in air.

3.4 Interaction of neutrons with matter

Neutrons are charge-neutral particles as the name indicates. Therefore neutrons lack Coulomb interaction and cannot lose energy by ionization of matter. From a radiation point of view, free neutrons have two possibilities to interact with matter: collisions with other nuclei and absorption. Which interaction dominates depends on the neutron energy. At high energies collisions prevail, at low energies resonant absorption in some nuclei becomes more likely. We distinguish between thermal neutrons, epithermal neutrons, and fast neutrons. Thermal neutrons have energy in the order of room temperature or about 1/40 eV = 25 meV. Epithermal neutrons are all those with energy between 0.1 eV and 1 MeV, fast neutrons have energies beyond 1 MeV.

When fast neutrons with initial energy E_0 penetrate a target they collide with other nuclei in the target and scatter. Assuming billiard-type elastic collision, the energy transfer solely depends on the scattering angle θ and mass number A of the target atom (see Fig. 3.13). In the forward direction ($\theta = 0$) no energy is transferred and the neutron energy remains maximal, whereas by backscattering ($\theta = \pi$) the energy transfer is largest and the neutron energy is minimal:

$$\theta = 0: E_{\max} = E_0; \quad \theta = \pi: E_{\min} = E_0 \left(\frac{A-1}{A+1} \right)^2 .$$

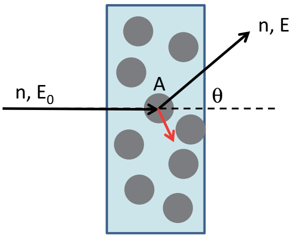


Fig. 3.13: Scattering geometry for neutrons hitting a target consisting of nuclei of mass number A. θ is the scattering angle.

Obviously the energy transfer to the target material is most effective for $A = 1$ (protons) and for atoms with small mass numbers A. After the first collision with the target atoms the energy of the scattered neutrons is equally distributed between the maximal and the minimal value, i.e., no scattering angle is preferred. However, after further collisions the neutron energy decreases step by step reaching a value that corresponds to the thermal energy of the target atoms. In the end the neutrons have a Maxwellian energy distribution centered at the thermal energy $3k_\mathrm{B}T/2$. If the target consist of pure hydrogen, only 18 collisions are required before reaching thermal energy. Neutrons with energy of 2 MeV, typical for fission neutrons in nuclear reactors, need about 80 collisions in water and a range of about 40 mm before being thermalized. The process of slowing down fast neutrons by elastic collisions is called moderation. Obviously, hydrogen-rich materials such as water or paraffin are very effective neutron moderators. This is also the case when neutrons penetrate tissue. Radiation therapy with fast neutrons utilizes neutron moderation in tissue, a topic discussed further in Chapter 11.

As fast neutrons reach low energies by moderation, often they are captured by isotopes. The absorption cross section of neutrons as a function of neutron energy usually has three characteristic regions, as illustrated schematically in Fig. 3.14 for a generic isotope: for $E < 1$ eV the absorption cross section decreases steadily with increasing neutron energy. In this region the absorption cross section is inversely proportional to the neutron velocity v_n. This region is therefore referred to as the '$1/v_\mathrm{n}$' region. The epithermal energy range beyond 1 eV is called the 'resonance region'. In this region the cross section raises sharply to high values in narrow energy bands. These sharp peaks in the absorption cross section are termed resonance peaks. They occur whenever the affinity of a nucleus for neutrons closely matches discrete quantum energy levels, i.e., when the combined binding energy plus kinetic energy of a neutron equals the energy difference between the ground state and an excited state. Beyond the resonance regime, absorption becomes negligible and elastic scattering of fast neutrons prevails, as already discussed above. Neutron activation of isotopes by thermal neutrons is presented in Chapter 2 and plays an important role for brachytherapy, discussed in Chapter 12.

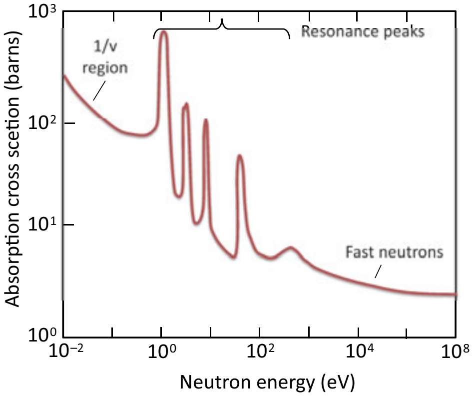


Fig. 3.14: Absorption cross section for neutrons. Three regions can be distinguished: $1/v$ region, resonance region, and fast neutron region.

Summarizing the various interactions of radiation with matter in an easy to remember comparison goes as follows: α-particles are stopped by a shirt, β-particles are slowed down within the thickness of the skin, fast neutrons are moderated to thermal energies within the top 4 cm, and γ-radiation will penetrate the body, as illustrated in Fig. 3.15.

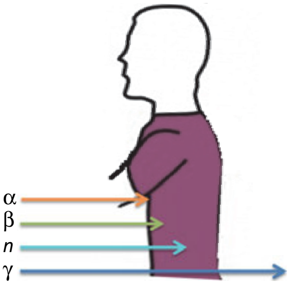


Fig. 3.15: Penetration of radiation in the body.

Tab. 3.2: Typical ranges for particles in water and lead. For γ-rays the thicknesses quoted and marked by an asterisk are for the half-value thickness.

Particle	Energy	Range in water or tissue	Range in lead
α	4 MeV	5 μm	5 μm
β	1 MeV	5 mm	0.5 mm
γ	100 keV	5 cm*	0.1 mm*
γ	1 MeV	10 cm*	1 cm*
Neutrons	2 MeV	40 mm	≈ m

3.5 Summary

1. Particle beams that hit a target are characterized by attenuation of the primary beam intensity, transmission, and scattering.
2. The attenuation of photon beams is governed by four effects of different importance in different energy regimes: small angle scattering, photoelectric effect, Compton scattering, pair production.
3. Small angle x-ray scattering is the scattering of electromagnetic waves at particles that have a different electron density than their surroundings.
4. For x-rays in the energy region up to 0.5 MeV the photoelectric effect is most important.
5. The photoelectric effect consists of two steps: absorption of photon and emission of fluorescence radiation.
6. The mass attenuation coefficient for the photoelectric effect has pronounced edges at the binding energy of electrons in the K-, L-, or M-shells.
7. The photoelectric effect of hard x-rays generates high energy photoelectrons.
8. Photoelectric absorption decreases the intensity of the primary x-ray beam but does not change the energy of the x-ray photons.
9. Photoelectric absorption causes an exponential decrease of intensity as a function of thickness penetrated.
10. x-ray or γ-ray photons do not have a range but a half-thickness, where half of the intensity has diminished.
11. X-ray and γ-rays produce hot photoelectrons whose range in tissue is similar to β-particles.
12. The Compton effect involves the inelastic scattering of photons at a bound electron.
13. Compton scattering transfers energy from photons to electrons.
14. Compton scattering changes the direction of the incident photon.
15. Compton scattering becomes increasingly important with increasing photon energy.
16. Pair production involves the materialization of photon energy into a pair of electron and positron.
17. Pair production has a threshold energy of 1.022 MeV photon energy.
18. α-particles have a defined range, which depends on the initial energy of the particle, the atomic number of the target material and its density.
19. The energy loss of α-particles is continuous. At the end of the range, the energy loss has a peak, called Bragg peak.
20. The linear energy transfer LET is defined as the energy loss per distance traveled.
21. LET for alpha particles is high at the end of their range.
22. Electrons and positrons also have a range which is much longer than that of alpha particles.
23. Positrons annihilate together with an electron at the end of their range.
24. β-particles lose energy mainly by Cherenkov and synchrotron radiation.
25. The range of β-particles is greater than for α-particles. The straggling range is also more extended.
26. Fast neutrons are slowed down by elastic scattering.
27. Water is the most effective moderator for fast neutrons.
28. The neutron absorption cross section scales with $1/v_n$.

References

[1] Finkelnburg W. Introduction to atomic physics. Springer Verlag; 1958.
[2] Livingston M, Bethe HA. Nuclear physics. Reviews of Modern Physics. 1937; 9: 245–390.
[3] Lapp RE, Andrews HL. Nuclear radiation physics. Prentice-Hall; 1972.

Further reading

Khan FM, Gibbons JP. The physics of radiation therapy. 5th edition. Wolters Kluwer; 2014.
Lilley J. Nuclear physics, principles and applications. John Wiley & Sons; 2013.
Kaplan I. Nuclear physics. 2nd edition. Addison Wesley; 1963.
Tavernier S. Experimental techniques in nuclear and particle physics. Berlin, Heidelberg: Springer-Verlag; 2010.

Useful website

Hyperphysics Webpage on Compton scattering provides a calculator for the incident and final photon and electron energy: http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/compton.html#c1

4 Radiation detection and protection

4.1 Introduction

Dosimetry is a combination of precise radiation measurements, regulatory work, and surveillance. Dosimetry quantifies dose levels from any radiation source that produces ionizing radiation, such as x-ray generators, radioisotopes, and accelerators. Furthermore, dosimetry sets boundaries for radiation exposure to living organisms that is considered safe versus unsafe in the sense of causing radiation sickness. Finally, dosimetry determines adequate dose levels for radiography and radiotherapy. Specifically, dosimetry has to fulfill the following tasks in civilian areas and an additional one for radiology clinics:

1. quantification of dose emitted by radiation sources;
2. designation of radiation control areas with increased dose above levels considered safe;
3. protection, safeguarding, and surveillance of people working in control areas;
4. guidance for the safe handling of radioisotopes and other radiation sources;
5. develop codes of practice for radiography and/or radiotherapy.

International and legally binding conventions have been established concerning these topics. The *International Atomic Energy Agency* (IAEA) and the *International Commission for Radiological Protection* (ICRP) frequently publish recommendations and guidelines on radiation protection and safeguarding.

All numbers provided in terms of dose have a statistical meaning based on a population rather than on individuals. For instance, if it is stated that an accumulated dose may lead to death within a certain number of years or may shorten the lifespan by so many years, this may or may not apply to an individual; even for a large population the numbers for low dose exposure have a large variance. However, the accumulated dose when serious radiation sickness is expected and the lethal dose are rather well defined. Since radiation experiments cannot be performed on people, much of the experience with long-term radiation exposure comes from nuclear bombs (Hiroshima, Nagasaki), nuclear tests, and major catastrophes (Chernobyl, Fukushima).

4.2 Definitions of dose

The dose of radiation D_R is defined as the mean energy $\langle E\rangle$ of ionizing radiation imparted to a volume that contains a mass m of material, or, in short, the mean energy per mass:

$$D_\mathrm{R} = \frac{\langle E\rangle}{m}.$$

DOI 10.1515/9783110553116-006

This definition is independent of exposure time and independent of the type of radiation, which can be from any radioactive material, nuclear reactor, or particle accelerator. The unit of dose is $[D_R]$ = Joule/kg. 1 J/1 kg = 1 Gray (Gy). The definition of dose is illustrated in Fig. 4.1. 1 J is the energy that is required to increase the temperature of 1 l of water by 2.4×10^{-4} K. This does not sound like much, but in terms of radiation 1 Gy is a dose that already causes severe radiation sickness.

To be more specific, the mean energy imparted to the volume with the mass unit m is the difference of radiant energy going into the volume, minus the radiant energy that has not been absorbed by the mass, plus the sum of all changes of rest energies of nuclei and elementary particles that may occur within the volume:

$$\langle E\rangle = \langle E\rangle_{\text{in}} - \langle E\rangle_{\text{out}} + \sum_i E_{\text{r},i}.$$

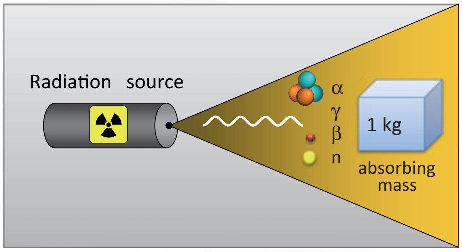


Fig. 4.1: Definition of dose is independent of the type of radiation and relies solely on the energy absorbed by some mass.

Obviously the definition of dose has some shortcomings. In particular, the different ionization probabilities of photons and charged particles are not accounted for by this definition. Therefore the dose by itself cannot express how dangerous a particular radiation is for biological tissues. To rectify this missing information, a quality factor Q_R for radiation is introduced characterizing different types of radiation with respect to their biological effectiveness. Q_R is also called *relative biological effectiveness* (RBE). A table of quality factors is given in Tab. 4.1. The different quality factors are mainly an expression of how likely ion bonds are broken by radiation, given the same energy of particles from a radioactive source. Obviously α-particles are particularly dangerous as they are 20 times more efficient at breaking ionic bonds and forming radicals on their track than γ-rays (through photoelectrons) or β-particles. Although the RBEs listed in Tab. 4.1 appear to be constant factors, this is actually not the case. The RBE depends on the *linear energy transfer* (LET) of particles, as will be discussed in Section 8.3.

Taking into account the quality factor Q_R for a certain type of radiation and the dose D_R of the same radiation, the equivalent dose for some tissue H_T is defined as:

$$H_T = D_R \times Q_R.$$

Tab. 4.1: Quality factors for the dose of ionizing radiation. RBE = relative biological effectiveness.

Radiation source	Quality factor Q (RBE)
x-rays, γ-rays, β	1
Thermal neutrons	2.3
Fast neutron	10
Protons	2
α-particles	20

The unit of the equivalent dose is $[H_\mathrm{T}]$ = Sievert, 1 Sv = 1 J/kg. Equivalent dose H_T is a measure of potential health effects on the human body due to low levels of ionizing radiation.

If the tissue is exposed to several different types of radiation, the sum has to be taken over all radiation sources:

$$H_\mathrm{T} = \sum_i D_{\mathrm{R},i} \times Q_{\mathrm{R},i}.$$

Different organs in the body have different levels of sensitivity to radiation. To take these into account, weighting factors w_j have been introduced for different body parts when evaluating a full-body equivalent dose. The tissue-weighted sum of equivalent doses in all tissues and organs of the body is given by the expression:

$$D_\mathrm{eff} = \sum_j w_j \times H_{\mathrm{T},j}; \quad \sum_j w_j = 1.$$

D_eff is called the *effective dose*. According to the recommendations of the ICRP the weighting factors are highest for breast, colon, lung, stomach, bone marrow (0.12) and lowest for bone surface, brain, salivary gland, and skin (0.01) [1]. All weighting factors should add up to 1.0 for full-body exposure.

Next we define the dose rate, which is the dose per time of exposure. The dose rate is either defined as $D_\mathrm{R}/\Delta t$ and is measured in Gray/a or defined as $H_\mathrm{T}/\Delta t$ and is measured in Sv/a. Here "a" stands for annum or year. Any appropriate time unit can be used in these definitions, but mSv/a is a most frequently used unit.

The units for activity (Becquerel, defined in Section 2.5.2), dose (Gray, Sievert), and dose rate (Gray/a, Sievert/h, etc.) are SI units. They are the only ones used in this text. The older unit rad (radiation absorbed dose) is replaced by Gray, rem (röntgen equivalent man) is replaced by Sievert, and Curie is replaced by Becquerel. The conversion factors are as follows:

$$1\ \mathrm{Curie} = 3.7 \times 10^{10}\ \mathrm{Bq}$$

$$1\ \mathrm{rad} = 10^{-2}\ \mathrm{Gy}$$

$$1\ \mathrm{rem} = 10^{-2}\ \mathrm{Sv}$$

4.3 Energy fluence and kerma

Fluence Φ is defined as the number of particles N accumulated over time per unit area: $\Phi = N/A = I \times t$, where I is the intensity of particles. At constant intensity I the fluence increases linearly in time. In the following we consider only neutral or uncharged particles such as x-ray photons with a fluence Φ_{ph}. For estimating the radiation dose it is essential to take into account the energy of the photon beam. Therefore, another quantity is defined, the *energy fluence*. For a monoenergetic beam of photons, the energy fluence Ψ_{ph} is the product of the photon fluence Φ_{ph} and the energy E per photon:

$$\Psi_{\mathrm{ph}} = \Phi_{\mathrm{ph}} \times E_{\mathrm{ph}}.$$

The SI unit of Ψ is $\mathrm{J\,m^{-2}}$. When taking an x-ray radiograph it is important to know the dose per exposure. This can be estimated from the information provided above. As the x-ray beam passes through matter, it deposits energy in a two-step process. First, photons transfer energy to charged particles either via the photoelectric process, Compton scattering, or pair production. Second, the charged particles generated lose energy in matter by ionization and excitation of atoms and molecules over some straggling distance before coming to rest. The energy transferred from photons to charged particles is called *kerma* (K), which stands for *kinetic energy released in matter*. Kerma is determined by the product of energy fluence and the *mass energy transfer coefficient* to be defined next.

The mass energy transfer coefficient μ_{tr}/ρ considers only those photon absorption processes that result in kinetic energy of charged particles. Elastic scattering of photons attenuates the beam intensity, but does not generate highly energetic charged particles. Compton scattering, in contrast, is the main source of dose in x-ray radiography. For pair production, relevant for x-ray radiotherapy, only the excess energy beyond the threshold energy of 1.022 MeV is converted into kinetic energy of electrons and positrons. Thus the mass energy transfer coefficient μ_{tr}/ρ is always smaller than the mass attenuation coefficient. The ratio depends on the primary energy and ranges from 0.7 for 20 keV x-rays to 0.2 for 60 keV x-rays. This is due to the fact that at higher energies Compton scattering becomes more important in comparison to photoelectric effect (see Fig. 3.8), contributing less to hot electrons. But this picture changes again when going to very high energy photons of up to 25 MeV for radiation treatment of cancer where Compton scattering and pair production dominate.

Kerma is defined as the photon energy fluence of an x-ray beam converted into kinetic energy of charged particles. Quantitatively Kerma is given by:

$$K = \Psi_{\mathrm{ph}} \left(\frac{\mu_{\mathrm{tr}}}{\rho} \right).$$

This definition can be rephrased as follows: kerma is the sum of all initial charged particle energies $\mathrm{d}E = \sum_i \mathrm{d}E_i$ that have been produced by photons (more generally: by uncharged particles) per unit of mass dm:

$$K = \frac{\mathrm{d}E}{\mathrm{d}m}.$$

The SI unit of K is J kg^{-1}.

As we have seen before, the *dose* (D) is defined as energy deposited by ionizing radiation per unit mass m of irradiated material:

$$D_\mathrm{R} = \frac{E}{m}.$$

Dose has the same unit as kerma $[D] = \mathrm{J\,kg^{-1}}$, where $1\,\mathrm{J\,kg^{-1}} = 1$ Gray (Gy). But the definition of dose is broader. Dose takes into account all kinds of ionizing radiation of charged and uncharged particles: photons, protons, neutrons, kaons, etc. In contrast, kerma is limited to the fraction of charged particles that has been produced by uncharged particles. Accordingly the dose can be calculated as:

$$D_\mathrm{R} = \Psi\left(\frac{\mu}{\rho}\right),$$

where Ψ is the energy fluence of any kind of ionizing radiation and μ/ρ is the corresponding mass absorption coefficient.

In radiography it is important to know the dose for a single image, for a complete CT, a scintigraphic body scan, or the dose administered to a cancerous volume. The dose for an x-ray radiograph can be calculated relatively easily because all parameters are well defined and known: x-ray intensity determined by the power of the x-ray tube, exposure time, average mass absorption coefficient for tissues and bones, and quality factor for x-rays.

As an example we assume the following values: x-ray intensity $I = 10^{10}$ photons/cm^2 s, mean x-ray photon energy of 60 keV for a bremsstrahlung spectrum with a kVp = 100 keV, exposure time $t = 50$ ms, average mass absorption coefficient for tissues and bones at 60 keV radiation is about 5 cm^2/g (see Fig. 3.8), RBE for x-rays is 1, which yields a dose according to:

$$D_\text{x-ray} = I \times \Delta t \times E \times \mathrm{RBE} \times \left(\frac{\mu}{\rho}\right) = \Phi \times E \times \mathrm{RBE} \times \left(\frac{\mu}{\rho}\right)$$

of 24 mSv. It is more difficult to estimate the dose for scintigraphic imaging because of many unknowns. The radioisotopes administered with a known activity spread out over the body and accumulate at hot spots such that the fluence is ill-defined. The time of exposure can only roughly be estimated by the effective lifetime of the radioisotopes used. Nevertheless, the following effective dose values have been estimated for scintigraphy:

- 6 μSv for a 3 MBq ^{51}Cr measurement of glomerular filtration rate of the kidneys;
- 37 mSv for a 150 MBq ^{201}Tl cardiac stress test;
- 3.5 mSv for a common bone scan with 600 MBq of $^{99\mathrm{m}}$Tc.

Tab. 4.2: Dose for different medical radiation procedures with x-rays.

Dose [mSv]	Medical procedure
0.1–0.3	x-ray projection radiograph of the thorax
20	Full-body x-ray CT
0.006	Renal filtration with γ-rays
3.5	Bone scintigraphy with γ-rays
37	Cardiac stress test with γ-rays

In the last example the activity is much higher than in the other two cases. But the activity is spread over a much larger area, such that the dose is not exceedingly high. The quoted dose values for some medical procedures are listed in Tab. 4.2.

For radiation treatment with charged particles (cp) a slightly different equation is used:

$$D_{\mathrm{cp}} = \Phi \times \mathrm{LET} \times \mathrm{RBE} \times \left(\frac{1}{\rho}\right).$$

Here LET is the energy loss per distance within the Bragg peak and ρ is the density of the target material.

4.4 Dosimeters

For measuring the dose and dose rates of ionizing radiation a number of detectors (dosimeters) are available. Dosimeters can be classified as personal dosimeters carried by individuals while they are in a radiation control area, and those which are used for monitoring radiation levels in control areas. Personal dosimeters are important for estimating the radiation dose deposited over some time on a person carrying the dosimeter. In fact, wearing personal dosimeters is mandatory in all control areas with an average dose level higher than 5 mSv/a. Figure 4.2 shows in panels (a) and (b) personal film badge dosimeters and in panel (c) a pocket dosimeter. All these dosimeters are used to monitor a cumulative dose over time until they are read out. Film dosimeters are usually read out once per month, while pocket dosimeters allow the wearer an immediate read out after exposure to x-rays or γ-rays.

Film badge dosimeters indicate the dose by the blackening of an x-ray film, which shows up after the film is developed. With different absorbers in front of the film the type of radiation and its energy can be roughly estimated. In panel (a) of Fig. 4.2 a film badge dosimeter is opened to display various compartments for monitoring different radiation types and levels. Finger dosimeters (Fig. 4.2 (b)) are used by personnel that need to work with their fingers close to an open radiation source.

Direct reading pocket ionization dosimeters resemble fountain pens in shape and size. The dosimeters contain a small ionization chamber filled with gas. In the center of the ionization chamber is a split anode wire connected at the base and the other

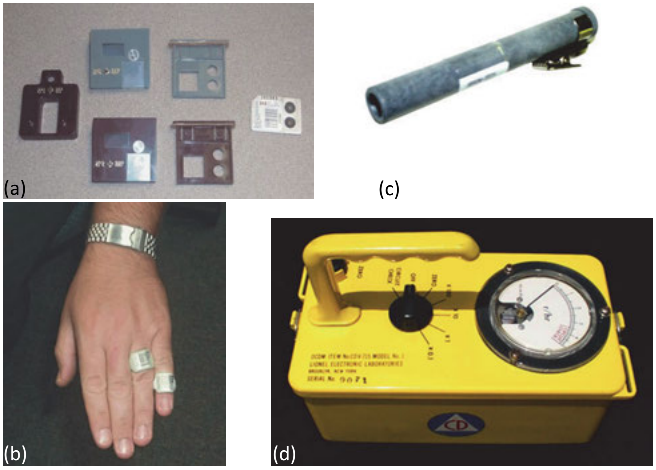


Fig. 4.2: (a) Film dosimeter; (b) finger dosimeter; (c) pocket ionization dosimeter; (d) handheld radiation monitor.

end is free. When charged to positive potential, electrostatic repulsion deflects these wires apart. Radiation causes ionization of the gas in the chamber; the electrons are attracted to the positive potential and at the same time discharge the wire. This in turn reduces the deflection, which can be read out with the help of a lens. Pocket dosimeters are being replaced more and more by digital electronic dosimeters. The output of a radiation detector is collected and amplified. When a predetermined charge from radiation has been reached, the collected charge is discharged by simultaneously triggering an electronic counter. The counter then displays the accumulated exposure and dose rate in digital form.

For monitoring the radiation level of an area, many types of handheld radiation monitors are available. An example is displayed in Fig. 4.2 (d). They are all based on discharge tubes with a central anode wire, schematically outlined in Fig. 4.3. Depending on the voltage applied, they can be categorized as ionization chambers, proportional counters, or Geiger–Mueller detectors. When ionizing radiation enters through a thin window, the gas in the chamber becomes ionized and the free charges are separated in a voltage gradient: negative charges drift to the anode wire and positive charges go to the grounded cathode chamber wall. The processes that occur after ion pairs have formed depend on the voltage applied. The charge versus voltage is plotted in Fig. 4.4.

At low voltages, all charges that have been produced by incoming ionizing particles are collected without amplification. In this first plateau region the operation of the discharge tube is referred to as *ionization chamber*. No special requirement of the gas is specified and even air can be used, in which case no window is needed. Since the charge from a single particle is too small, ionization chambers are not useful as

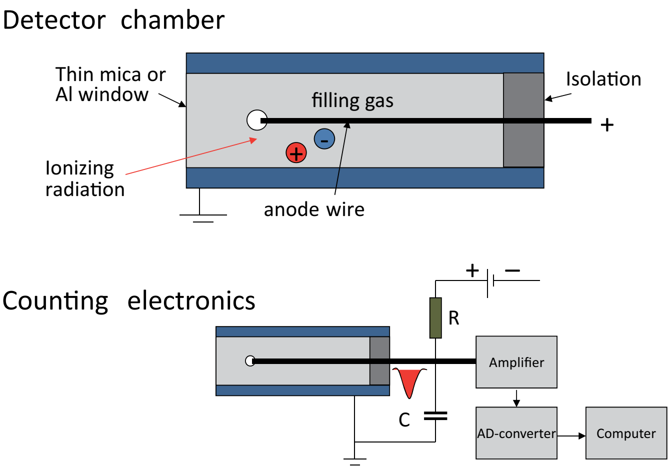


Fig. 4.3: The detector chamber for the ionization detector, for the proportional detector, and for the Geiger–Mueller detector holds detecting gas, an anode wire at positive potential, and a thin window for radiation to enter. The counting electronics is shown in the lower part of the figure.

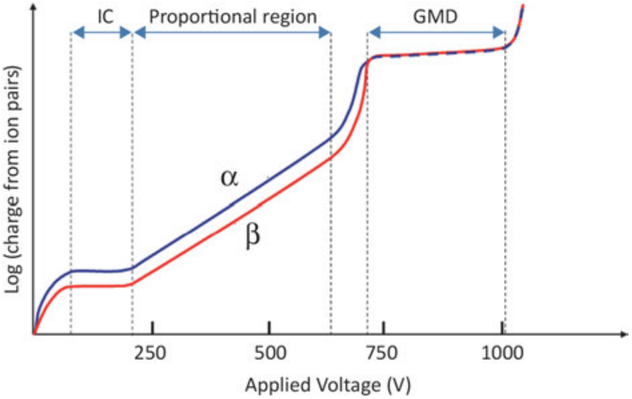


Fig. 4.4: Charge versus voltage characteristics for the different regions of discharge tube operation. The ionization chamber (IC) operates at low voltages in the plateau region. The proportional counter operates in the region where the charge collected is proportional to the voltage applied. The Geiger–Mueller detector (GMD) operates again in the plateau region, but at much higher voltage. The blue curve is the charge–voltage characteristics for α-particles, the red curve for β-particles.

particle counters. Instead they are used in current mode, where the detected current is proportional to the count rate. In this mode, ionization chambers are very robust and can be used for detection of very high activities without dead time problems.

Proportional counters contain gases that can easily be ionized by charged particles. With the higher voltage gradient in proportional counters, the ion pairs become accelerated on their way to the electrodes, thereby strongly multiplying the charge collected. Therefore the proportional counter can be used as a single event counter after electronic amplification of the charge signal. The total charge collected per particle is proportional to the applied voltage and to the energy of the detected particle. This latter property makes the proportional counter interesting for spectroscopic investigations, where not only the activity is of interest but also the energy of particles.

Geiger–Mueller detectors operate at even higher voltage in the plateau region of the charge–voltage curve of Fig. 4.4, all other properties being equal to the proportional counter. At these higher voltages the amplification of the charge can be considered as an avalanche that is independent of the exact voltage and independent of the particle energy. Therefore Geiger–Mueller detectors are pure event counters.

In both proportional counters and Geiger–Mueller detectors the discharge would sustain unless it is quenched quickly by a special additive to the gas so that the chamber is ready for detecting the next particle. Discharge and quenching requires some time during which no other particle can be detected. This time is called *dead time*. For low count rates the dead time is not important but eventually limits the count rate. If the dead time is about 10 µs, count rates up to 10^4 counts/s should be possible without much correction. For higher count rates, dead time corrections must be applied. The corrected count rate N_{corr} of a number of recorded events N over a measurement time T is:

$$N_{\text{corr}} = \frac{N}{N - N\frac{\tau}{T}},$$

where τ is the dead time of the counter.

All three detectors have difficulties detecting x-rays or γ-rays. This can be partially improved by adding high Z noble gases to the gas mixture of the detector, such as Ar or Xe. Much more appropriate are scintillation counters for detecting high energy photons and they are used for scintigraphic imaging, described in Chapter 6.

4.5 Radiation exposure

People are exposed to many types of radiation, voluntarily or not. In general we distinguish between three different sources of radiation:

(a) cosmic
(b) terrestrial
(c) medical/industrial

Isotopes produced by cosmic radiation are mainly due to the activity of our sun or other solar systems. They eject mainly π-mesons, protons, and α-particles, which interact with the earth's atmosphere and produce isotopes of carbon (^{14}C), tritium (^{3}H), and to a lesser extend ^{10}Be, ^{26}Al and ^{36}Cl. ^{14}C finds its way into our body through the food chain. At sea level the radiation received per person due to cosmic radiation is on average 0.3 mSv/a. This dose increases dramatically at higher altitudes. In normal

Tab. 4.3: Typical dose per person and dose limitations in various countries. For dose limitations in specific countries or specific facilities, local regulations should be consulted.

Dose and dose rate	Radiation source
0.01 mSv/a	Nuclear power plants
0.3 mSv/a	Cosmic sources
0.1–1 mSv/a	Cosmic ray exposure during flights
1 mSv/a	Terrestrial sources
1 mSv/a	Medical/industrial sources
3.5 mSv/a	Average radiation exposure
<1 mSv/a or < 0.1 μSv/h	Maximum tolerable radiation exposure without further radiation control
1–6 mSv/a	Survey area
> 6 mSv/a	Control area
>3 mSv/h	Off-limit area
20 mSv/a	Occupational dose limit
250 mSv	Maximum lifetime occupational dose
1000 mSv	Radiation sickness
3000–4000 mSv	50 % death toll within 3–6 weeks
5000–8000 mSv	Lethal

flight altitude of 10 km the dose rate is already 10 μSv/h and with one transatlantic flight we accumulate about 0.1 mSv. Dose and dose rates are listed in Tab. 4.3.

Terrestrial radioactivity is due to long-lived isotopes, mainly from ^{40}K, ^{232}Th, ^{235}U, ^{238}U and their decay products. One important radioisotope is the noble gas radon (^{222}Rn), which decays with a half-life of 10.5 a into ^{218}Po. Since Rn as a noble gas atom is not bound in molecules, it emanates from uranium-rich rocks and accumulates in closed rooms, such as poorly vented basements and when inhaled Rn reaches our lungs. As an α-emitter its biological effectiveness is particularly high. After decay to ^{218}Po, the daughter isotope is not only radioactive but also toxic. The dose rate from terrestrial sources fluctuates strongly and depends on geological conditions. Another important contribution to terrestrial radiation is the decay of the natural isotope ^{40}K in our body, causing about 5000 Bq from 1.3 MeV β-radiation. This internal source of radiation accumulates to 0.3 mSv/a. Altogether, on average one can assume that the dose rate from terrestrial sources is about 1 mSv/a.

Industrial activity and medical treatment are further sources of radiation exposure. Coal burning power plants set free ^{14}C that would otherwise be bound. Nuclear power plants contribute a negligible amount to environmental radioactivity, unless there is a major accident, such as in Chernobyl in 1986 and Fukushima in 2011. Much of the radiation from the Chernobyl disaster came from ^{137}Cs fallout. This is a strong β^-- and γ-emitter with a half-life of 30.2 a. Moreover, in addition to the high level of radioactivity, Cs when incorporated with food replaces K and disturbs cell functioning. The radiation exposure from normal x-ray imaging without tumor treatment is estimated to be about 1 mSv annually. In total, the dose rate from industrial/medical applications per person is estimated to be about 1.2 mSv/a.

Summing up the three main sources of radioactivity: cosmic, terrestrial, and medical, we estimate a total dose rate of about 3.5 mSv/a any person receives on average. This is below the radiation control level, which starts at 5 mSv/a according to Tab. 4.3. At the other extreme, a lethal dose of 5–8 Sv is applied locally during radiotherapy for destroying cancer cells, as discussed in Chapter 8.

4.6 Radiation protection

The basic guideline for radiation safety and radioprotection is stated as follows: "As Low As Reasonably Achievable", which is quoted as the *ALARA* principle.

The three most important measures for radiation protection are as follows:

1. reduction of exposure *time*
2. gaining *distance* from the radiation source
3. choice of proper and sufficient *shielding*

The first point is obvious and does not need further explanation.

The easiest and most effective measure against radiation is distance. For a radiation source that emits in all directions (4π), the inverse square law $1/r^2$ holds, meaning that at double distance from the source the intensity drops by a factor of four because the surface area exposed has quadrupled. This is illustrated in Fig. 4.5. The quadratic distance law does, however, not hold in case of directed beams, such as for x-ray tubes and radiation from accelerators. Furthermore, any incorporation of radioactive material via food or via inspiration through the respiratory tract should be strictly avoided.

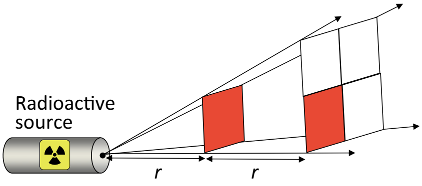


Fig. 4.5: Illustration of the quadratic distance law. The intensity of a radioactive source drops by a factor of $(1/2)^2 = 1/4$ in double distance $2r$ from the source.

Appropriate shielding against radiation can be an art. The range of particles in different environments gives a first hint to the best choice of shielding. Some of them are listed in Tab. 3.2. X-rays and γ-radiation can be shielded most effectively by high Z materials such as lead sheets or lead bricks. The required thickness can be determined as follows. First the half-value layer (HVL) thickness follows from the exponential attenuation law according to: $\mathrm{HVL} = 0.693/\mu$ (see Section 3.2). Depending on the acceptable radiation level, the total thickness of the required shielding is then given by $(1/2)^n$, where n is the number of HVL to be used. For instance, for a reduction in intensity to 1 %, $n = 6$ is necessary. α-particles are easy to stop within 5 cm of air or by a sheet of paper. But any skin contact with an α-source should be strictly avoided. β-particles can be shielded with a thin aluminum metal sheet. It is considerably more difficult to shield fast neutrons. First, fast neutrons need to be slowed down to thermal energies by a sequence of collisions in water or paraffin. Then thermal neutrons can be captured by highly neutron absorbing materials, such as boron, cadmium, or gadolinium. But thermal neutron capture does not come for free. For instance, the absorption reaction ${}^{10}_{5}\mathrm{B}(\mathrm{n},\alpha){}^{7}_{3}\mathrm{Li}$ releases α-particles, and the other reactions release β- and γ-radiation. Therefore additional shielding is required for capturing secondary particles.

4.7 Summary

1. Dose *D* is a measure of radiation absorbed by matter. The dose of 1 Gray equals the absorption of 1 Joule of radiant energy in 1 kg of matter.
2. The unit Sievert considers the relative biological effectiveness (RBE) of radiation.
3. Equivalent dose is defined as the dose multiplied by a quality factor *Q* specific for each type of radiation.
4. At the same energy α-particles are 20 times more effective in destroying biological tissue than γ- or β-particles.
5. Fluence is the number of particles impinging on an area.
6. Energy fluence is the fluence of photons multiplied by the energy per photon.
7. The mass energy transfer coefficient considers only those absorption processes that result in kinetic energy of ionizing charged particles.
8. Kerma is defined as the energy fluence of an x-ray beam converted into kinetic energy of charge particles
9. The annual exposure of people to radiation from cosmic, terrestrial, or medical sources amounts on average to 3–3.5 mSv.
10. For monitoring radiation in controlled areas personal dosimeters and various detectors are available and must be carried.
11. The three most important measures against radiation are time, distance, and shielding.
12. Dosimetry quantifies dose levels and sets boundaries for radiation exposure considered safe versus hazardous to people.

References

[1] International Commission on Radiological Protection. The 2007 recommendations of the International Commission on Radiological Protection. ICRP Publication 203. ICRP; 2007.

Further reading

IAEA. Dosimetry in diagnostic radiology: An international code of practice. Technical Reports Series No. 457. IAEA; 2007.

Khan FM, Gibbons JP. The physics of radiation therapy. 5th edition. Wolters Kluwer; 2014.

Dance DR, Christofides S, Maidment ADA, McLean ID, Ng KH, editors. Diagnostic radiology physics – A handbook for teachers and students. IAEA; 2014.

5 X-ray radiography

5.1 Introduction

X-ray radiography is the most frequent imaging modality for noninvasively visualizing inner parts of the body. It is also the oldest method dating back to 1895 with the famous x-ray image that W. C. Röntgen took of his wife's hand. In the early days the radiation hazards associated with radiation exposure were not appreciated. Up until the 1970s it was common practice to check the fitting of shoes with an x-ray machine in the shoe store, called shoe-fitting fluoroscope!

X-rays penetrate opaque materials since the energy (frequency) of x-ray photons is much higher than any energy band gap of materials or any plasma frequency of electrons in materials. On their way through gases, liquids, or solids x-rays are attenuated but not refracted. Therefore x-ray lenses in the common sense do not exist. Only attenuation offers an imaging modality by shadow projection of partially transparent and inhomogeneous media.

In Chapter 3 we learned about x-ray attenuation under the assumption of well-defined conditions typical for physics laboratory experiments: the incident x-ray beam is assumed to be monochromatic, the absorbing material has a well-defined thickness, and the density is taken to be constant. In medical practice the situation is much more complex. Radiographs are taken with a broadband x-ray source (modified bremsstrahlung spectrum), the absorbing tissue is highly inhomogeneous, and the thickness varies with the shape of organs. Nevertheless with x-ray radiographs we try to gain information about inner organs from transmitted x-ray beams that expose an appropriate x-ray detector. The magic word is contrast. A homogeneous sample would not offer any contrast. In this chapter we learn about three possibilities of contrast formation: attenuation contrast, phase contrast, and scattering contrast. We start with attenuation contrast and discuss the more specialized contrast modalities in subsequent sections.

5.2 Image formation

5.2.1 Attenuation contrast

We consider an x-ray beam of intensity I_0 impinging at perpendicular incidence on a medium with varying lateral mass attenuation coefficient and thickness, as indicated in Fig. 5.1. After transmission the intensity is recorded by a suitable detection system, such as a flat-panel detector or an x-ray film. For didactical reasons we sometimes show exposures of x-ray films, where the gray scale is logarithmically proportional to the exposed intensity. Obviously the thinner material with lower mass absorption

DOI 10.1515/9783110553116-007

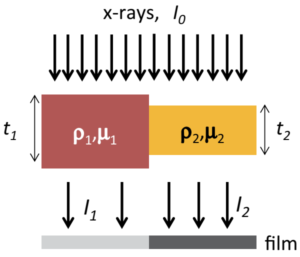


Fig. 5.1: X-ray attenuation in a laterally inhomogeneous medium recorded by an x-ray film forming a gray scale contrast.

coefficient attenuates the x-ray beam less and therefore the blackening of the x-ray film is stronger than in neighboring regions where the absorbing thickness is larger and the mass attenuation coefficient is higher. Ideally we will observe a sharp boundary between high and low absorbing material for a parallel incident beam. The difference in gray scale on the recording film is called contrast, i.e., the degree to which bright and dark areas of an image differ in brightness or in optical density.

Quantitatively contrast C is defined as the difference in transmitted intensity in neighboring regions normalized by the sum:

$$C = \frac{I_1 - I_2}{I_1 + I_2} = \frac{I_0 e^{-\frac{\mu_1}{\rho_1}\rho_1 t_1} - I_0 e^{-\frac{\mu_2}{\rho_2}\rho_2 t_2}}{I_0 e^{-\frac{\mu_1}{\rho_1}\rho_1 t_1} + I_0 e^{-\frac{\mu_2}{\rho_2}\rho_2 t_2}} .$$

In the simple example of Fig. 5.1 we have taken the mass attenuation coefficients μ/ρ_1 and μ/ρ_2 to be constant within their respective thicknesses t_1 and t_2 corresponding to one specific x-ray energy E. To be more realistic, we need to consider a more general situation depicted in Fig. 5.2. Again an x-ray beam is incident parallel to the x direction and homogeneously extended in the yz plane. The x-ray beam has a wide spectral distribution typical for a bremsstrahlung spectrum. An object with a cross section A and with a spatially varying attenuation coefficient $\mu(x, y, z, E)$ intersects the beam.

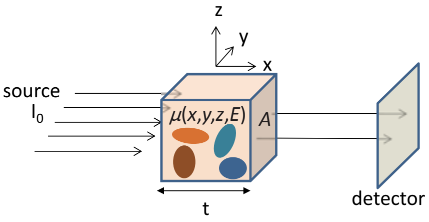


Fig. 5.2: Geometry of x-ray absorption via an object of thickness *t* and cross-sectional area *A* in the *yz* plane.

For such an inhomogeneous object and for a wide energy band x-ray source, the Lambert–Beer equation (Chapter 3) has to be recast into the more general integral form:

$$I_t(y, z, E) = I_0(y, z, E) \exp\left(- \int_0^t \mu(x, y, z, E)\, dx \right),$$

where the line integral is taken over a varying absorption coefficient along the x direction. The line integral is repeated for all points within the cross-sectional area of the object $A = y \cdot z$. Taking into account the quantum efficiency $\eta(E)$ of the detector on the one hand and the fact that the detector is not energy dispersive on the other hand, the intensity recorded by the detector is:

$$I_{\text{det}}(y,z) = \int_0^{E_{\max}} I_0(y,z,E)\eta(E)\exp\left(-\int_0^t \mu(x,y,z,E)\,\mathrm{d}x\right)\mathrm{d}E,$$

where the energy integral is taken over the bremsstrahlung spectrum. This equation is called *"signal equation"* for x-ray imaging via *projection radiography*. $I_{\text{det}}(y,z)$ is called a *projection image* of $\mu(x,y,z,E)$. Larger line integrals cause lower output intensity at the detector $I_{\text{det}}(y,z,E)$ and vice versa. Actually, x-ray detectors as well as x-ray films collect the transmitted intensity over the exposure time T. The number of photons per area integrated over the exposure time T is then:

$$N_{\text{film}}(y,z) = \int_0^T \mathrm{d}T \int_0^{E_{\max}} I_0(y,z,E)\eta(E)\exp\left(-\int_0^t \mu(x,y,z,E)\,\mathrm{d}x\right)\mathrm{d}\lambda.$$

Figure 5.3 illustrates an attenuation profile as detected by an x-ray film.

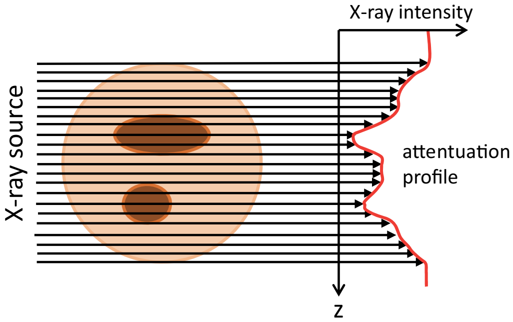


Fig. 5.3: Attenuation profile as recorded by an x-ray film or by a flat-panel detector. The length of the arrows symbolizes the local x-ray intensity behind the object. Many line scans repeated in the y direction combine to a two-dimensional picture in the yz plane.

X-rays are attenuated in bones more strongly than in the surrounding tissue because of the high calcium content (atomic number 20) of bones. When exposed to x-rays for taking a radiograph, bones form an x-ray shadow, shown schematically in Fig. 5.4. Radiographs by x-ray shadow formation are called *projection radiography*, because they convert a three-dimensional object into a two-dimensional shadow image. On the recording x-ray film bones appear in light gray (low x-ray exposure of the film), while the surrounding tissue appears in dark gray (high x-ray exposure of the film).

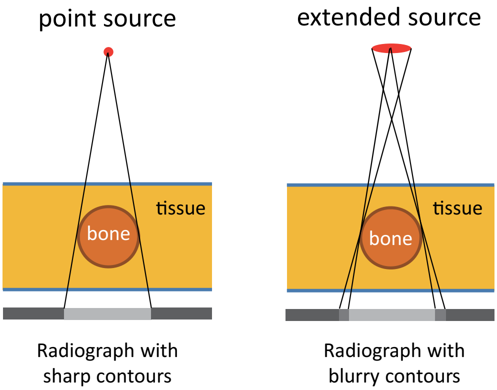


Fig. 5.4: Shadow from absorbing objects (bones) taken with a point source and an extended source.

When the radiograph is taken with a point-like x-ray source, the rays are not parallel, as assumed in Figs. 5.1–5.3 but diverge as in Fig. 5.4. Nevertheless, boundaries between light and dark regions appear sharp, i.e., contours of shadows appear sharp forming a high contrast image. Note that the shadow area on the film is enlarged as compared to the object size by simple geometrical means. If, however, the x-ray source is extended, a half-tone, called *penumbra*, separates the bright and dark regions and the radiograph appears blurry.

Another potential source of fuzziness is inelastic Compton scattering and elastic small angle x-ray scattering (Fig. 5.5). Small angle scattering (in medical literature also referred to as coherent scattering) occurs from any object with a different density than the surrounding matrix. Small angle scattering is very useful for analyzing nano- to microsized objects and it is extensively applied in the investigation of soft matter. However, the relevant scattering angles are rather small when using hard x-rays (in the order of $< 0.05°$). Therefore small angle scattering can safely be neglected as a source of fuzziness in x-ray projection radiographs. On the other hand, Compton scattering becomes increasingly important with increasing x-ray energies (see cross sections in Fig. 3.8) and contributes noticeably to fuzziness. To remedy the Compton scattering problem, a lead containing grid with parallel lamellas is placed in front of the film that allows only the straight-through x-rays to reach the recording detector. This also solves the problem of using extended x-ray tubes, depicted in Fig. 5.4. But the grid causes another problem: it forms an obscuring shadow on the radiograph. To overcome this obstacle, in practice the grid can be oscillated in front of the detector.

Figure 5.6 compares the three main factors that affect the contrast of radiographs: (a) half-tone from an extended source; (b) small angle scattering and Compton scattering; (c) insufficient contrast due to similar attenuation coefficients of object and surrounding matrix. Figure 5.7 compares radiographs from the pelvic space taken with and without an oscillating grid in front of the x-ray film. The difference in contrast is impressive. There are additional reasons for blurry images that lie in the x-ray detection method to be discussed later.

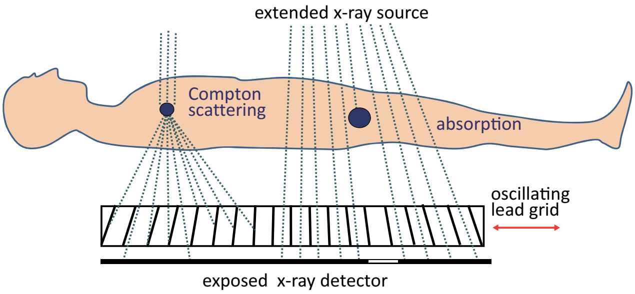


Fig. 5.5: A rigid or moving lead grid in front of the exposed x-ray detector suppresses stray x-rays and improves the contrast.

Fig. 5.6: Three types of contrast affected by an extended source: (a) penumbra; (b) Compton scattering; (c) insufficient difference of attenuation coefficients for object and surrounding.

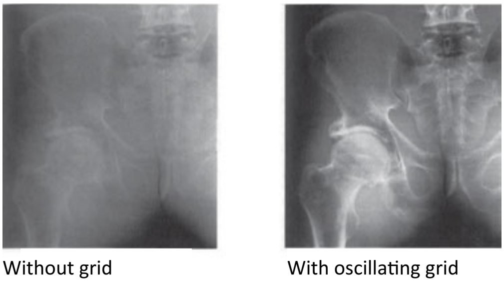


Fig. 5.7: Radiograph of the pelvic space taken with and without an oscillating grid in front of an x-ray film.

5.2.2 Beam hardening

Radiographs taken with a standard medical x-ray tube like the one shown in Fig. 1.8 usually use the entire bremsstrahlung spectrum for imaging at operating voltages between 80–140 kVp. However, the low energy “soft” part of the x-ray spectrum is highly absorbed in the skin, causing biological damage without contributing to image contrast. Therefore, for safety reasons the spectrum is modified by passing the x-rays first through an Al absorber. This eliminates low energy x-rays, while higher energy x-rays are hardly attenuated by the Al foil. A schematic x-ray spectrum before and after filtering with an Al foil is reproduced in Fig. 1.3 and in Fig. 5.8. Thus, radiographs are usually taken with a filtered broad x-ray bremsstrahlung spectrum. In general, by employing various types of filters (Al, Cu, etc.) the average bremsstrahlung spectrum is shifted to higher energies. This often serves to vary or enhance contrast and is referred to as x-ray *beam hardening*. Thus filters have two effects: they reduce the primary flux preferentially in the soft x-ray regime, and they shift the average energy of the bremsstrahlung spectrum to higher energies. Filters for dual energy contrast enhancement are discussed in Section 5.3.3.

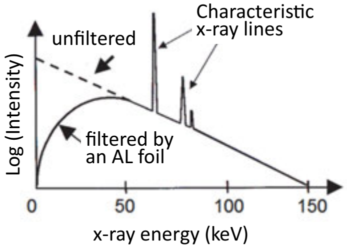


Fig. 5.8: Bremsstrahlung spectrum and characteristic spectrum of an x-ray tube before and after filtering the radiation with a thin Al foil.

5.2.3 X-ray films

As already alluded to, x-ray radiographs were traditionally recorded with x-ray films. As film recording has nowadays been entirely replaced by digital recording (see next section), film recording remains only of historical interest. Therefore a brief summary of some physical properties of x-ray films is justified but may be skipped if not incidentally required. Figure 5.9 (a) shows the layering in a standard x-ray film. In the emulsion layer transparent AgCl/AgBr particles are embedded in a water soluble gelatin like in a traditional camera film. Upon x-ray exposure the silver halogenides become reduced and are converted to metallic Ag. If at any specific spot at least five or more Ag^+ ions are reduced, they cluster up into a metal particle storing a latent image. Upon developing in a watery solution containing a reducing agent, the silver particles grow in size to amplify the latent image. Subsequently, in an oxidizing solution called fixer,

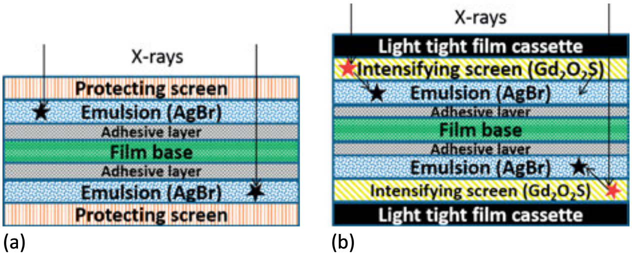


Fig. 5.9: Left: layer arrangement of an x-ray film; right: layering in an x-ray film with intensifying fluorescence screen.

the unexposed halogenides and all other residuals are washed out. After drying, the film is inspected with the help of a viewing light box. When viewed against the light source, optically dark areas correspond to a high density of grainy silver particles, which are due to a high x-ray exposure implying low x-ray attenuation in the exposed body part. Vice versa, lighter areas on the film are less exposed to x-rays, meaning a higher attenuation in body parts. The x-ray radiograph is a negative image: higher x-ray exposure of the film produces darker images. The lateral resolution of x-ray films depends on the density of AgCl/AgBr particles in the emulsion and the grain size of silver particles after reduction. Both can be controlled to a certain extent. The resolution is usually quoted as 0.1 to 0.06 mm, which is the “pixel” size of radiographic films.

The quantum efficiency of the described film process is rather low. Only 5 % of x-ray intensity is absorbed in the emulsion layer and thus contributes to darkening the film. In more advanced versions an additional, intensifying screen is added on either side of the film containing rare earth metal ions, in most cases gadolinium oxysulfide (Gd_2O_2S), as sketched in Fig. 5.8 (b). X-rays are more effectively absorbed by the high Z material of Gd_2O_2S and sequentially emit fluorescence light in the green visible region. Thus x-rays are effectively converted into visible light, which, in turn, exposes the silver halide emulsion layer. The resulting conversion efficiency is increased by a factor of three to about 15 %. The screen is placed inside a light-tight cassette and tightly pressed against the film during exposure to avoid any blurring.

Although the conversion of x-ray intensity into visible light intensity is linear, the conversion of visible light into darkening of the x-ray film is not linear but logarithmic. The optical density (OD) of the x-ray film is defined as the negative logarithm to the base 10 of the film transparency T, when viewed against a visible light source:

$$\mathrm{OD} = -\log_{10} T; \qquad T = \frac{I}{I_0}.$$

An optical density of 1 corresponds to a transparency of 10 %, requiring an x-ray exposure of 7 µGy. 20 % transparency is reached at an exposure of 10 µGy. Optical density as a function of exposure is plotted in Fig. 5.10. The lower toe part is affected by noise, for

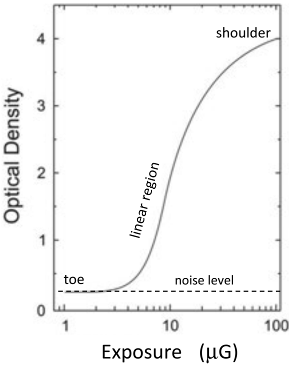


Fig. 5.10: Optical density of an x-ray film plotted versus the x-ray exposure. Note the logarithmic scale of the *x* and *y* axes.

high exposure the film goes into saturation. The relevant linear region in this double logarithmic plot is in between the toe and the shoulder.

5.2.4 Digital recording

There have been many efforts to use digital sensors for x-ray imaging. However, because of the large field of view, CCD chips are too small, and stacking them up into an array is too costly. The easiest way to get a digital image is by laser scanning an x-ray film. While this was often practiced in the past, it was not a step forward in the direction of digital image processing and live x-ray videos that completely digital recording (DR) promises. Significant progress in DR has been achieved in recent years [1]. Two types of detectors are distinguished: direct and indirect, their principal designs are shown in Fig. 5.11. By direct detectors we understand the conversion of x-ray photons into electron-hole pairs, the charge being collected in a high voltage potential at a suitable pixel electrode. Indirect detectors convert the x-ray energy first into visible photons, which are then detected by photodiodes. Light sensitive scintillators consist either of Gd_2O_2S or CsI. Gd_2O_2S has a higher conversion efficiency, whereas the columnar growth of CsI crystals has a better photon guiding ability to the photodiodes underneath. This means that the generated visible photons travel straight to the photon detector without being scattered in the scintillator material nor detected by a neighboring diode. Suppression of scattering in scintillator materials is important to avoid blurring and to increase contrast. Amorphous silicon (a-Si) on glass substrates is commonly used as photodiode. The thin film transistors (TFT) have a typical size of $0.2 \times 0.2\,\mathrm{mm}^2$ or less, each acting as a photodiode and representing a pixel in a large detector array. The electrical signal generated by the photodiode is then amplified and encoded by electronics into a gray or color scale, generating sensitive and accurate x-ray images.

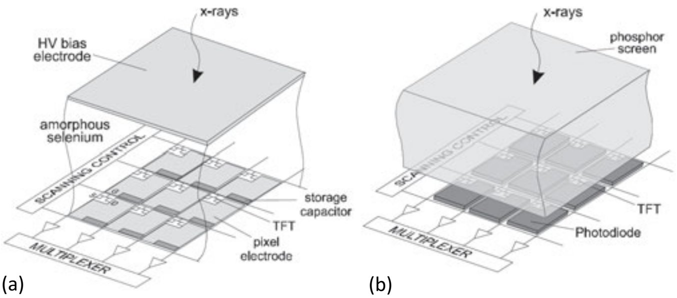


Fig. 5.11: (a) Direct and (b) indirect flat-panel x-ray detectors (reproduced from http://hepwww.rl.ac.uk/Vertex03/Talks/Row/Rowlands.pdf).

Flat-panel x-ray detectors are costly. On the other hand they are more sensitive than x-ray films, i.e., they have a higher quantum efficiency, allowing a lower dose of radiation for any picture taken. Furthermore, detection is much faster such that *real time radiography* (RTR) becomes feasible with 1 to 20 frames per second depending on the image processing required. This is indeed an important step forward that facilitates RTR assistance during surgical interventions. Presently commercial TFT panels have a size of 41 × 41 cm, with an array of 2048 × 2048 diodes. This yields a spatial resolution of 0.2 mm, considering only the physical size of the diodes. Comparing with traditional x-ray films, the spatial resolution of TFT panels is lower by a factor of 2–3.

5.2.5 System integration

Figure 5.12 depicts all parts required for taking an x-ray projection radiograph: x-ray source, filter, support table for body and parts to be imaged, oscillating grid, and detector.

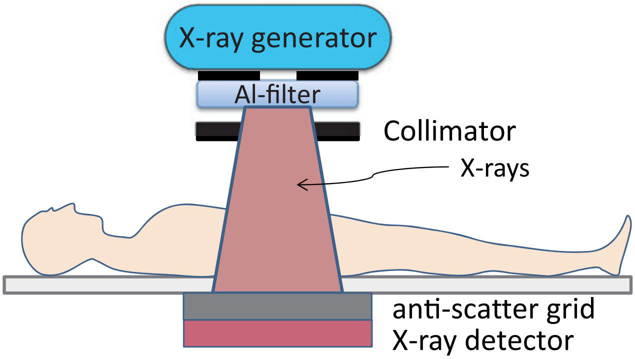


Fig. 5.12: Schematic of the arrangement for taking x-ray projection radiography pictures.

Depending on which part of the body is to be imaged, three main types of x-ray machines are used for projection radiography (Fig. 5.13): vertical panel, horizontal table with tilt option, and C-frame. The C-frame has the smallest detector field but the largest flexibility and is therefore often used in operating theaters for monitoring. In recent years with the use of flat-panel detectors sensitivity has dramatically increased, such that RTR with high time and spatial resolution can be recorded during surgical procedures. An example for monitoring the proper placement of stents during cardiac surgery is shown in Fig. 5.14.

A special kind of vertical panel radiography is the mammography of the female breast for tumor recognition. To secure high sensitivity to any changes in the soft tissue, the voltage of the x-ray generator is reduced to about 25 kVp and often the Mo-Kα

Fig. 5.13: X-ray machines for projection radiography. (a) Vertical panel for chest radiographs (http://medicaloutfitter.net); (b) horizontal table with tilt option (www.medicalexpo.com); and (c) with a C-arm for real time imaging during surgery (reproduced from http://usa.healthcare.siemens.com).

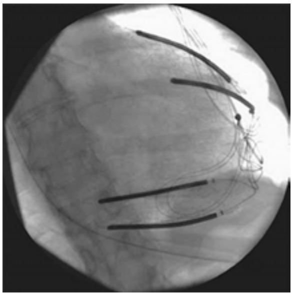


Fig. 5.14: X-ray image recorded with a C-arm mobile imaging machine during cardiac surgery and placement of stents (from https://www.healthcare.siemens.de/surgical-c-arms-and-navigation/mobile-c-arms/arcadis-avantic# publication Arcadis Avantis).

characteristic radiation from a molybdenum target is used as x-ray source. Often the breast is squeezed against the panel containing the film in order to produce a homogeneous absorption thickness (Fig. 5.15). Traditionally, highly sensitive black and white films were used for mammography. But recently *digital mammography*, also called full-field digital mammography (FFDM), has replaced x-ray films. The advantage is higher sensitivity and therefore lower radiation dose per screening. A variant of the FFDM is the digital *breast tomosynthesis* (DBT), which is a kind of three-dimensional computer tomography (CT) imaging of the breast. CT is discussed in Section 5.4.

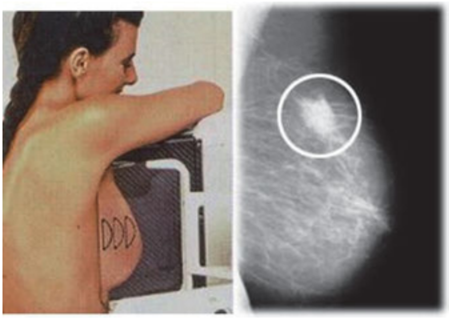


Fig. 5.15: Mammography of a female breast for tumor recognition. In the encircled area a higher absorption is observed, which could indicate the presence of tumorous tissue.

5.3 Contrast enhancement

When imaging the upper chest, the gastrointestinal tract, or blood vessels often the contrast to the surrounding tissue is not sufficient to recognize fine details, meaning that the attenuation profile is too flat. In this situation it is desirable to enhance the contrast artificially. There are several means for achieving higher attenuation contrast:
(a) injection of contrast agents
(b) digital subtraction angiography (DSA)
(c) dual x-ray energy (DXE) imaging
(d) phase contrast imaging (PCI)
(e) phase contrast tomographs (PCT)

5.3.1 Contrast agents

Any contrast enhancing agent (CA) should fulfill the following conditions: it should attenuate the x-ray beam more strongly than the surroundings; be easily taken up by the parts of the body to be imaged without altering their shape or function; have a short biological lifetime; and be pharmacologically and physiologically inert.

For imaging the gastrointestinal (GI) tract, barium with a high atomic number ($Z = 56$) is used. Recalling from Chapter 1 that the absorption coefficient scales with the third power of the atomic number $\mu \approx Z^3$, Ba is certainly a good candidate for high x-ray absorption. For administering, barium is bound in barium sulfate ($BaSO_4$), which is a dry, white, chalky powder that is mixed with water to make a barium sulfate liquid. About 30 minutes after filling the liquid into the GI tract, it coats the inside walls of the esophagus, stomach, and intestines. Hence the inside wall coatings, size, shape, and contour become visible on x-ray radiographs by a white (highly absorbing) contrast. An example is shown in Fig. 5.16. The $BaSO_4$ slurry is used solely for diagnostic studies of the GI tract.

Fig. 5.16: Contrast enhancement of the gastrointestinal tract with the help of barium sulfate powder in solution (reproduced from https://openi.nlm.nih.gov/, © Creative Commons).

Imaging of blood vessels with projection radiography is called *angiography*. The name comes from the Greek word angeion for "vessels". For contrast enhancement a contrast agent is injected into the bloodstream. The contrast agent or liquid "dye" contains iodine, which is also a high Z element ($Z = 53$). Iodine may be bound either in an organic (nonionic) or ionic compound. An example from the family of nonionic water-soluble iodinated CAs, named iopromide, is shown in Fig. 5.17. It is one of the most frequently used CAs for x-ray radiography including CT imaging of the brain. Since these small molecules in the bloodstream are cleared out quickly by the kidneys, renal tolerance has to be considered before administering. If tolerance is not an issue, the renal clearance of iopromide can also be used for intravenous urograms, i.e., the radiography of kidneys, ureters, and bladder. Nevertheless, there is considerable interest in replacing iodine contrast agents by nanoparticle contrast media with less nephrotoxicity, which is a topic to be discussed further in Chapter 14 on nanomedical applications.

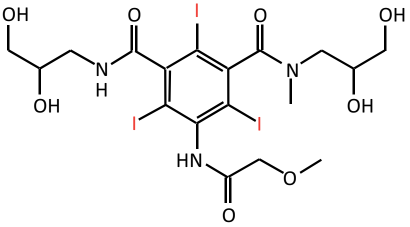


Fig. 5.17: Chemical structure of the iodinated contrast agent iopromide. Three iodine atoms are attached per molecule.

5.3.2 Digital subtraction angiography (DSA)

Often angiography is combined with digital subtraction radiography, also known as *digital subtraction angiography* (DSA). For this method two digitally recorded radiographs are taken before and after administering the dye. Obviously the object to be imaged should not move in between. The DSA at a position (x, y) is then an image of the absorption coefficient difference $\Delta\mu$ before (mask) and after injection:

$$\mathrm{DSA}(\Delta\mu) = \ln(I_{\mathrm{dye}}(x, y)) - \ln(I_{\mathrm{mask}}(x, y)).$$

The image taken before injection is called the mask, as it masks off the field of interest. A DSA radiograph of blood vessels in a hand is reproduced in Fig. 5.18. The fine branching of the blood vessels would not be visible in a normal radiograph. A sequence of images is taken of the same area with a rate of 1–6 frames per second. Then it can be decided during image processing how many of them need to be subtracted from the mask in order to gain an optimized contrast.

Fig. 5.18: Angiography of the blood vessels in the hand (reproduced from Michel Royon, Wikimedia, © Creative Commons).

5.3.3 Dual x-ray energy imaging (DXE)

Even without administering any contrast agent, and without using absorption edges, contrast between bones and tissue can be enhanced through the use of two different x-ray energies. This allows visualizing structures which are otherwise obstructed, such as tissue behind bones. Using proper energies, either bones or tissue can be emphasized. This is particularly important for taking images of the chest, where one would like to get images of the lung behind the ribs. This can be achieved by referring to the mass attenuation coefficient for bones in tissue plotted in Fig. 5.19. At low x-ray energies the absorption in bones is higher than in tissue, while at high energies the attenuation is overall weaker and similar in bones and tissue. As an example let us assume that we take radiographs with energies indicated by the green and yellow areas in Fig. 5.19. In the green area the attenuation coefficients are 1 and 0.5 for bones and tissue, respectively, whereas in the yellow region they are 0.33 and 0.2 (arbitrary

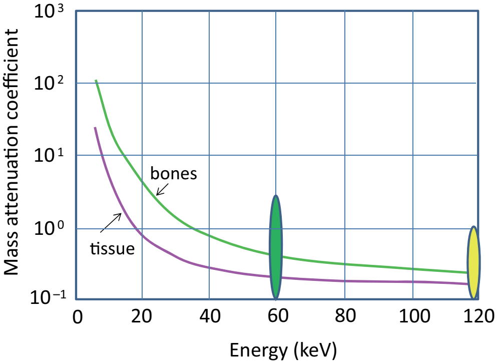


Fig. 5.19: Mass attenuation coefficient of bones and tissues as a function of x-ray photon energy.

units). We now apply a scaling factor of 3 to the high energy image and a scaling factor of 1 to the low energy image before taking the difference. This yields $3 \times 0.33 - 1 \times 1 = 0$ for the bones, but $3 \times 0.5 - 3 \times 0.2 = 0.9$ for the tissue. Next we apply a scaling factor of 1 to the high energy image and 2.5 to the low energy images. Then taking the difference we obtain $1 \times 1 - 0.33 \times 2.5 = 0.205$ for the bones and $1 \times 0.5 - 0.2 \times 2.5 = 0$ for the tissue. This way we can emphasize either bones or tissue.

How can two different x-ray energies be selected with a standard medical x-ray tube? There are two possibilities. Either one may change the accelerating voltage and take a picture at lower peak voltage (60 kVp) and another one at higher peak voltage (120 kVp). Although the spectrum is broad, the mean x-ray energy is shifted from low to high by increasing the voltage. Or one may use different absorber sheets, such as Al and Cu. A thin Cu sheet in front of the x-ray tube removes the soft energy spectrum and hardens the radiation compared to Al filtered radiation. Therefore, taking a picture with an Al filter an another one with a Cu filter provides the weighted change of the x-ray spectrum and a proper contrast. Another question is whether the two images should be taken simultaneously or sequentially. There are advantages and disadvantages for both methods. When taking radiographs simultaneously, two film cassettes are stacked on top of each other sandwiching a Cu absorber sheet (or any other metal sheet) in between. Then both recordings are spatially congruent, but the lower cassette receives much less intensity than the top one. If taken sequentially, the intensities are similar but the patient should not move between exposures to avoid a blurry difference image. An example for a dual energy imaging is shown in Fig. 5.20. The original radiographs taken at two different average x-ray energies are shown in the top panels, and the weighted difference images in the lower panels emphasizing either tissue (left) or bones (right).

5.3.4 Phase contrast imaging

Conventional x-ray imaging uses contrast via attenuation of x-rays. But x-ray contrast can also be generated by exploiting subtle effects of x-ray refraction that change the phase of x-ray radiation. This method is termed *phase contrast imaging* (PCI). PCI permits visualization of soft tissue structures that would not be detectable by conventional x-ray radiography. In addition, PCI requires a much smaller radiation dose to the body for a comparable image quality. Thus PCI has great potential for special applications; its general clinical use is, however, still hampered by specific requirements that can easily be fulfilled in laboratories but less so in clinical environments.

The physics of x-ray contrast lies in the refractive index for x-rays:

$$n = 1 - \delta + i\beta .$$

The real part of the refractive index is slightly smaller than 1 by about one part in 10^{-6}. The consequence of this is an x-ray wave with a phase velocity and a wavelength that

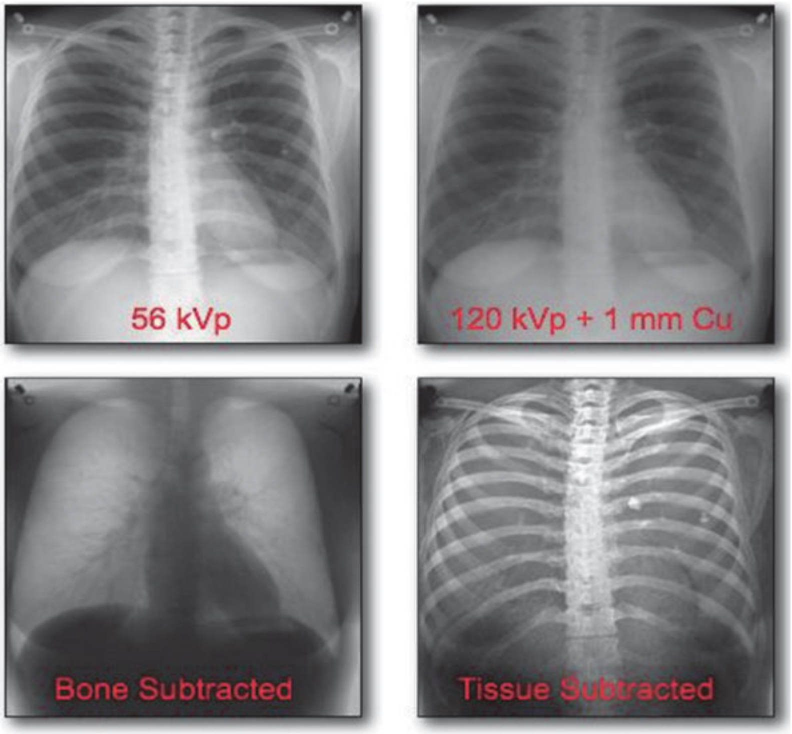


Fig. 5.20: Dual energy imaging of the original radiographs and after taking weighted differences to visualize either soft tissue (bones subtracted) or bones (tissue subtracted) of the chest (reproduced from http://en.wikibooks.org/wiki/Basic_Physics_of_Nuclear_Medicine/Dual-Energy_Absorptiometry, © Creative Commons).

is slightly bigger in matter than in vacuum, in contrast to visible light. The imaginary part is related to the attenuation coefficient via the following relation:

$$\mu = \frac{4\pi\beta}{\lambda\rho_a} .$$

Here λ is the x-ray wavelength and ρ_a is the atomic number density.

Upon traveling a distance t through matter, the corresponding phase change is:

$$\Phi = \frac{2\pi}{\lambda} \int \delta(t)\,\mathrm{d}t$$

and the attenuation becomes:

$$-\log\left(\frac{I}{I_0}\right) = \frac{4\pi}{\lambda\rho_a} \int \beta(t)\,\mathrm{d}t .$$

It turns out that δ for tissues and for diagnostic x-ray energies can be three orders of magnitude larger than β. This means that over the same distance t the phase shift of x-rays can be much larger than the intensity change by attenuation. While this is a convincing argument for PCI, it turns out that in practice measuring phase shifts is much harder than measuring intensities. A number of methods have been proposed in the past, only one of them is briefly presented, others are discussed in [2].

The simplest method of phase contrast imaging is the so called *in-line holographic imaging* approach [3], illustrated in Fig. 5.21 (a). This method requires the wavefront of the incoming beam to be parallel but not necessarily monochromatic. When the beam strikes an object, the wavefront becomes distorted due to different phase velocities of the x-ray beam in and outside of the object and due to refraction at surfaces that are not perpendicular to the wavefront. All together this leads to edge effects, i.e., edges are emphasized. To allow the effect to develop, a certain distance between object and film is required (air gap), which is just the opposite of conventional x-ray radiography where all unnecessary air gaps should be avoided. The phase contrast depends on the distance of the x-ray source to the object and from the object to the detector. If the source is too close, the incident beam will not be sufficiently parallel. If the detector is too close, distorted and undistorted rays will overlap and the contrast is lost. In a more advanced stage one or two diffraction gratings G can be placed between object and detector to enhance the contrast, as shown in Fig. 5.21 (b). True holographic imaging requires an undistorted reference beam, which interferes with the distorted beam behind the object and before the screen. The in-line holographic imaging methods is a simplified version of PCI, but rather difficult to analyze quantitatively. An example of a phase contrast enhanced x-ray image is shown in Fig. 5.22. An early review of PCXI is given in [4].

Phase contrast imaging using additional diffraction gratings has produced remarkably contrast rich images, as shown in Fig. 5.23 from [5]. All three images from the same object (mouse) are intrinsically perfectly registered as they are extracted from the same data recorded with a grating interferometer. Examples of regions of enhanced contrast as compared to a conventional attenuation image in (a) are

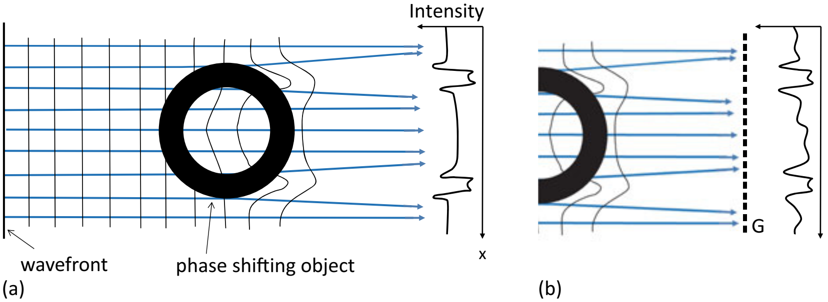


Fig. 5.21: Distortion of a planar wavefront due to refraction effects in an object without additional grating (a) and with an enhancement diffraction grating G (b) (adapted from [3] by permission of IOP Publishing).

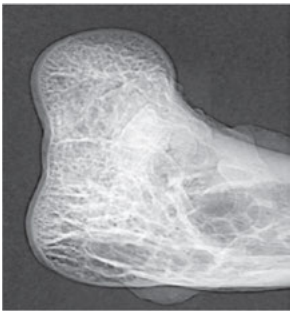


Fig. 5.22: In-line x-ray holography of the femur of a duck, using a 51 keV monochromatic x-ray beam at the synchrotron Spring 8, Japan. The cartilage of the bone is clearly visible (reproduced from [3] by permission of IOP Publishing).

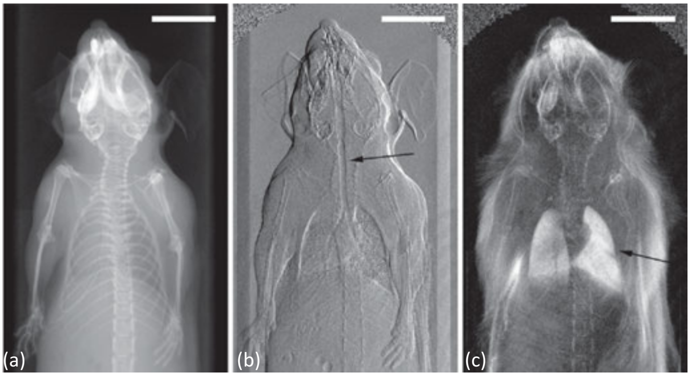


Fig. 5.23: In vivo multicontrast x-ray images of a mouse. (a) Conventional x-ray image based on attenuation. (b) Differential phase contrast image based on x-ray refraction. (c) Dark-field image based on x-ray scattering. The white bars correspond to 1 cm (courtesy of F. Pfeiffer and reproduced from [4] by permission of Nature Publishing Group).

marked with arrows, showing the refraction of the trachea (b) and the scattering of the lungs (c). Phase contrast imaging can provide three-dimensional tomographic images [6], and even the fourth dimension, time, has recently been explored [7].

5.4 Computed tomography (CT)

X-ray projection radiography based on x-ray attenuation provide two-dimensionally resolved pictures of body parts projected into the yz plane (coronal plane) with x-rays incident along the x direction, the axes are defined in Fig. 5.2. Projection radiography is highly valuable for a survey of large body parts, uncovering bone fractures, foreign objects, infections, indications of cancer and much more. Any spatial variation of the absorption coefficient in the x direction is integrated over and remains unresolved. Dual energy x-ray imaging is a crude attempt to gain depth resolution in the z direction. *Computed tomography* (CT) adds the missing spatial resolution in the z direction. The term *tomography* is derived from the Greek word *tomos*, meaning slice or section, and *grapho* meaning to write. Tomography is used in many fields of science and technology, including radiology. By CT we understand a diagnostic tool that combines usual x-ray attenuation contrast with computation to generate cross-sectional radiographic images of the body in the *xy* plane (horizontal or axial plane). For CT imaging the x-ray tube together with a detector bank on opposite sides of a circle rotate about a stationary patient. During rotation many frames are taken by digital recording. Upon one full rotation one slice of the body is completely projected in all directions, and with the help of a fast computer a picture of the cross section is generated. Figure 5.24 shows the general layout of a CT scanner.

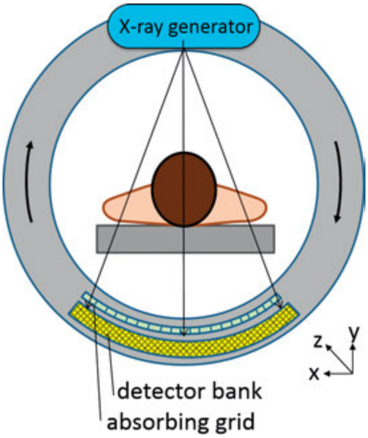


Fig. 5.24: CT geometry: x-ray generator and detector bank rotate simultaneous about a stationary patient.

The basic idea of CT scans is illustrated in Fig. 5.25. If radiographs from an inhomogeneous body containing various objects with different mass absorption coefficients are taken from different orientations, then the attenuation profile will always be different. From these different profiles it should be possible to reconstruct the spatial location and orientation of the objects inside the body. The procedure is as follows. First a cross-sectional slice is subdivided into voxels. Voxels are three-dimensional volume elements filling up the space of an object, in analogy to pixels, which fill up a two-dimensional area. Voxels are well known through 3D printing.

Each voxel at position (x_i, y_i, z_i) is assumed to have on average a specific mass absorption coefficient $\mu_i(x_i, y_i, z_i)$. The coefficient is projected onto the *xy* plane and each pixel of a voxel is given a CT number representative for the x-ray absorption in the voxel.

The assignment of CT numbers to each pixel follows according to the *Hounsfield scale*. In the Hounsfield scale the x-ray attenuation of water is taken as reference. Anything which has a higher linear attenuation than water is assigned a positive number, in the opposite case a negative number. CT numbers on the Hounsfield scale are there-

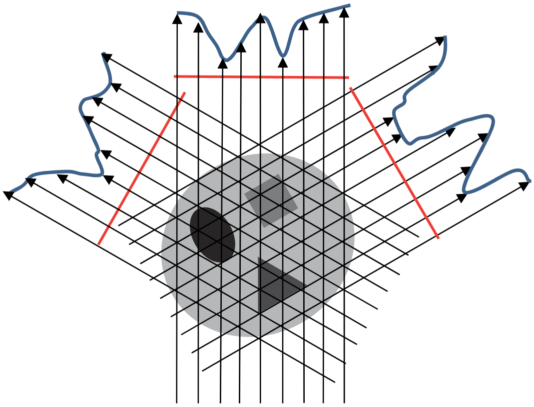


Fig. 5.25: Attenuation profiles from three different orientations.

fore defined as:

$$\mathrm{CT}(x_i, y_i, z_i) = 1000 \times \frac{\mu_{\text{object}}(x_i, y_i, z_i) - \mu_{\text{water}}}{\mu_{\text{water}}}.$$

Although the result is just a number, it is given an artificial unit, the *Hounsfield unit* (HU), in recognition of Sir Godfrey Hounsfield, who pioneered CT scanners and received the Nobel Prize for Medicine in 1979.

According to the definition, water with a linear attenuation of $0.19\,\mathrm{cm}^{-1}$ has 0 HU. For hard x-rays air has an absorption coefficient which is nearly zero, thus the CT number of air is −1000 HU. Bones with a linear attenuation coefficient of roughly $0.38\,\mathrm{cm}^{-1}$ have a CT number of +1000 HU. Each CT number is associated with a gray scale. Bones and air are the two extremes of the Hounsfield scale spanning 2000 shades from white for bone to black for air. Since the eye cannot distinguish between 2000 shades of gray, usually a Hounsfield window is set: a range of ±500 HU for a rough overview, and a range of ±100 HU for a detailed inspection. The Hounsfield scale is graphically shown in Fig. 5.26.

In CT terminology, tissues are named hypodense if they have a negative HU; vice versa, tissues are hyperdense if they have a positive HU. Two pixels with the same HU are isodense.

We will now illustrate by a simple example how to localize a voxel with higher attenuation than its surrounding. In Fig. 5.27 a slice made up of nine voxels is shown. Four projection radiographs are taken from different orientations ($\theta = 0°, 45°, 90°, 135°$) and the intensities are recorded. The total linear attenuation coefficient is calculated from the intensities, knowing the thickness of the object:

$$-\frac{1}{t(\theta)} \log \frac{I(\theta)}{I_0} = \mu(\theta).$$

The task is to find the location of a voxel with the highest attenuation in the slice. For each projection, the attenuation coefficients are recorded in a matrix according to the

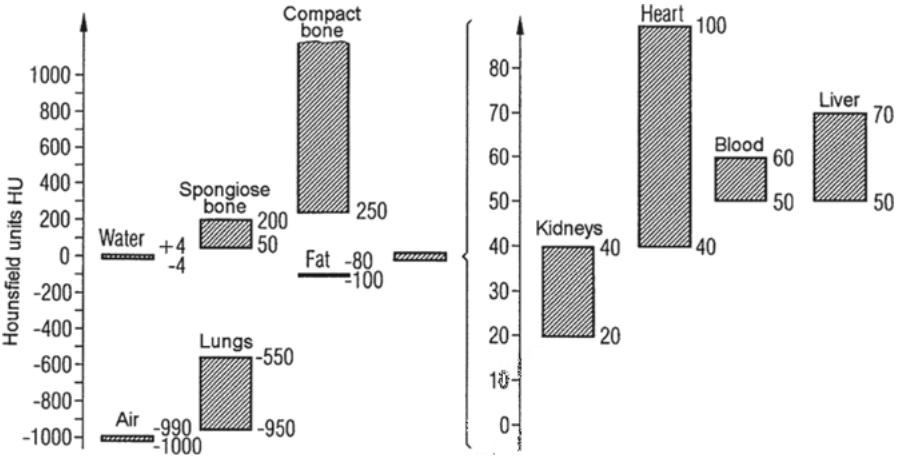
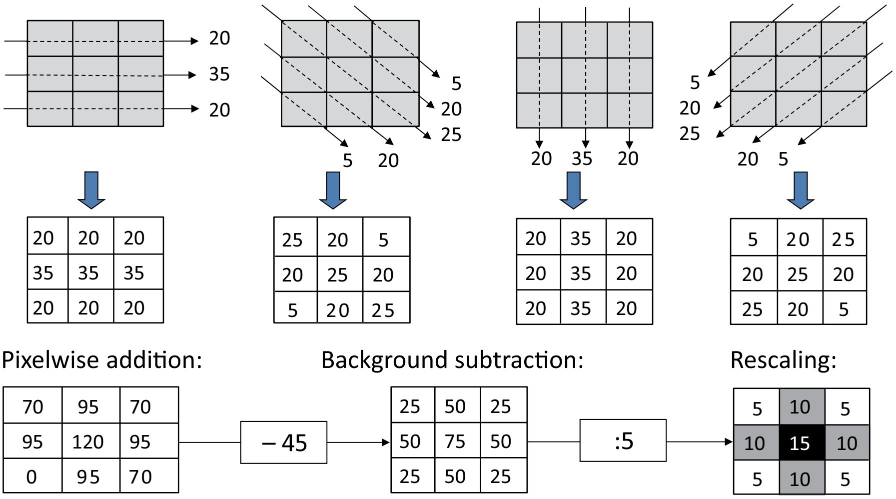


Fig. 5.26: Hounsfield scale for x-ray absorbing tissues, bones, and organs.



20	20	20
35	35	35
20	20	20

25	20	5
20	25	20
5	20	25

20	35	20
20	35	20
20	35	20

5	20	25
20	25	20
25	20	5

Pixelwise addition:

70	95	70
95	120	95
0	95	70

– 45

Background subtraction:

25	50	25
50	75	50
25	50	25

:5

Rescaling:

5	10	5
10	15	10
5	10	5

Fig. 5.27: Procedure for finding the voxel with the highest attenuation.

projections taken. These numbers are added up in a sum matrix on the lower left side. After background subtraction and rescaling it becomes obvious that the voxel with the highest attenuation is at the center.

In practice, a number of technical issues need to be considered for engineering CT equipment and for the proper analysis of projections taken. First of all the beam from the x-ray tube diverges as indicated in Fig. 5.24, instead of being parallel as suggested in Figs. 5.25 and 5.27. This is not a problem as long as x-rays do not cross before detection. To avoid ray crossing and also to suppress detection of scattered x-rays, grids are used in front of the detector bank just as in the case of flat projection radiography. X-ray generator and detector bank are mounted rigidly on opposite sides of a rotating gantry and this gantry rotates about a stationary patient in the center. The rotation and recording speed can be as high as 200–300 rpm in order to take quasi-static pictures of moving parts like the heart. During rotation the supporting table can be translated in the z direction to record continuous slices of the body forming a spiral as illustrated in Fig. 5.28.

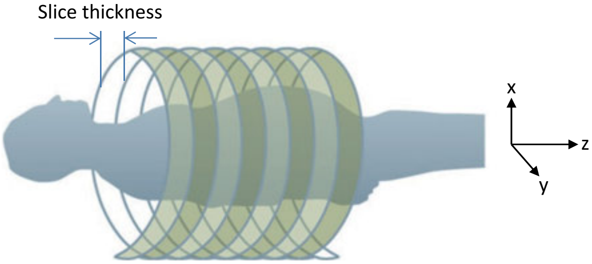


Fig. 5.28: Geometry for taking slices by rotating source and detectors while simultaneously translating the patient in the *z* direction (reproduced from www.siemens.com/press).

Detector sampling rates are typically 2 kHz. If the gantry rotates with a speed of 120 rpm, i.e., 500 ms per rotation, then 1000 projections can be taken during one full rotation. The detector bank sits on a circle with a radius of about 500 mm, which amounts to 3140 mm for the complete circumference. If 1000 projections are taken with a single detector, then the sampling distance between frames is about 3 mm. However, the detector bank covers a sector containing many detectors, increasing the resolution to about 0.3 mm. The detector sector in front view subdivided into a number of individual detectors is schematically seen in the left panel of Fig. 5.29.

Furthermore, the detector bank contains a multislice arrangement of detectors, seen by the side view in the middle panel of Fig. 5.29. The thickness of a slice is determined by the size of the detectors. For high resolution the slice should be as thin as 2 mm. The divergence of the x-ray beam in the z direction has to be confined by collimators to the detector size. If an organ spanning 200 mm needs to be scanned, 100 rotations are required for full coverage. The recording time of each rotation depends on the rotational speed which we assume to be 0.5 s. Thus a full scan takes about 50 s. This long scan time can be drastically reduced by using a multidetector with a multislice array. Presently, solid state detectors with a width of 0.5 mm are available. In detector banks up to 256 detectors are mounted in parallel to each other in the z direction, covering a total area of up to 120 mm. For exploiting the multidetector array over its total width of 120 mm, the x-ray beam divergence has to be relaxed accordingly. Now only two rotations are needed for scanning 200 mm requiring only about 1 s. Obviously, the radiation exposure of the patient is much reduced. The raw data taken by

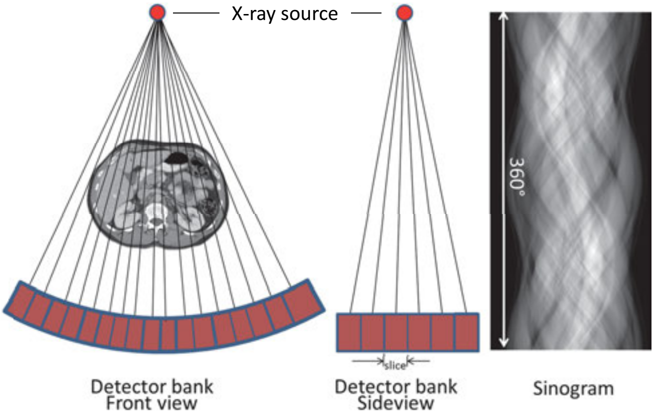


Fig. 5.29: Schematic view of a sector detector bank seen from the front and from the side. Only six detectors are shown in parallel defining $M = 6$ slices; up to $M = 256$ detectors in parallel are used nowadays. The sinogram in the right panel is a sequence of pictures taken during one 360° rotation by all M detectors simultaneously. Reconstruction of the sinogram yields the image of a slice shown in the left panel (sinogram courtesy of Dr. Siegel, University of Maryland, School of Medicine).

the detectors during one full rotation, called sinogram, is shown in the right panel of Fig. 5.29. The sinogram has to be reconstructed with appropriate algorithms to obtain real space pictures of slices like the one shown in the left panel of Fig. 5.29 [8]. This is usually done by a 2D Radon transform, named after the Austrian mathematician Johann Radon (1887–1956). The measured intensity values are represented in a coordinate system of the variables: rotation angle θ and time t, representing a sinogram. The Radon back transformation, similar to a Fourier transformation, results in an approximate reconstruction of the object [9].

CT scans can be combined with contrast enhancement procedures as discussed before and with angiography of blood vessels. The x-ray generator can be tuned to higher or lower accelerating voltages as needed, and the x-ray beam can be hardened with the help of different absorbers. Most recently dual energy CT scanners have been developed for further contrast enhancement. Present state of the art is a CT voxel size of about $0.5 \times 0.5 \times 0.5\,\text{mm}^3$ and a gantry rotation of 214 rpm or one revolution in only 280 ms. For special applications such as the inner ear the voxel size can be reduced to $0.25 \times 0.25 \times 0.25\,\text{mm}^3$. Furthermore, ECG synchronization with a time resolution of better than 75 ms has been reached, and a dose of less than 2.8 mSv is required for full exposure. A modern version of a CT scanner is shown in Fig. 5.30 with open and closed gantry.

Finally, in Fig. 5.31, we want to show an example of a modern CT radiographic image, here of the head and the upper chest. To obtain contrast in the blood vessels, a contrast enhancing agent was injected. The dose delivered to the person was as low as 1.1 mSv for taking a high contrast radiograph of such an amazing resolution and richness of detail.

Fig. 5.30: CT unit without cover and with cover. X-ray generator and detector bank is housed inside behind the cover (www.healthcare.siemens.com/computed-tomography/).

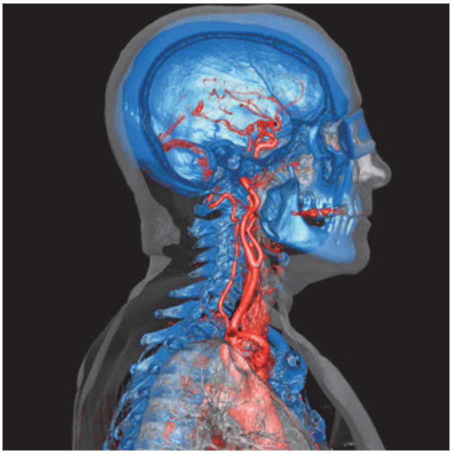


Fig. 5.31: Three-dimensional CT radiograph of the head and upper chest including part of the vascular system (from www.medicalradiation.com/ provided by www.healthcare.siemens.com).

5.5 Risks and comparisons

State-of-the-art x-ray radiographs not only provide an unprecedented imaging quality with respect to spatial resolution and contrast. The radiation exposure time has also decreased dramatically due to the use of highly sensitive detectors, fast scanning, and enormous computer power for image processing. Reduced exposure time translates into less x-ray radiation dose. Some average values are given in Tab. 5.1 for standard radiographic examinations. The higher dose for mammography compared to a chest radiograph is due to the soft x-rays used, which are more strongly absorbed.

Tab. 5.1: Dose for various x-ray radiographs.

Exposure	**Dose**
Average radiation exposure	3.5 ± 1 mSv/a
Chest projection radiograph	0.1 mSv
GI tract with $BaSO_4$ contrast	15 mSv
Mammography	3 mSv
CT skull	1.5–2.3 mSv
CT chest	5.8 mSv
CT cardio-angiogram	7–13 mSv

These low dose levels still compete with zero dose when compared to MRI. Therefore it is appropriate to make a first comparison between these two advanced imaging modalities that have reached a very high technological standard. What are the advantages and disadvantages? When should one or the other method be used for imaging? A comparison is given in Tab. 5.2. In general, one can state that imaging the lung is a challenge for both methods; but given the choice, x-rays still show more contrast. In most cases these two methods are complementary and images taken by one method can be confirmed by the other method.

Tab. 5.2: Comparison of x-ray and MRI scanning techniques for various body parts and diagnostics.

Body part	**X-ray radiography**	**MRI**
Tendons and ligaments	Not seen in CT scans	Shows up in MRI
Fracture of bones and vertebrae	Better seen in CT	
Brain tumor		Better recognized by MRI
Tears and organ injuries, bleeding in brain	CT scans are quicker	
Spinal cord injuries		MRI is preferred
Pneumonia	CT is preferred	
Early cancer recognition	CT is preferred	

5.6 Summary

1. X-ray attenuation contrast is defined as the difference of transmitted intensity in neighboring regions normalized by the sum.
2. Contrast on x-ray films or flat panels is achieved by attenuation of the penetrating radiation in tissues of different density and thickness.
3. The transmitted intensity in projection radiography follows from the Lambert–Beer equation where in the exponent a line integral is taken over the attenuation path with spatially varying attenuation coefficient.

4. Extended x-ray sources and Compton scattering cause a penumbra at the border of shadow projections.
5. Oscillating grids in front of the recording detector improve contrast by suppressing stray x-rays.
6. Beam hardening is required to eliminate soft x-rays that cause high dose without contributing to image quality.
7. Most x-ray radiographs are recorded digitally with a flat-panel detector containing thin film transistors.
8. X-ray machines for projection radiography are either vertical, horizontal, or have a C-frame.
9. Mammography is taken with lower x-ray kVp (35 kV) and often with Mo-Kα characteristic radiation in contrast to normal radiographs taken with 80–140 kVp.
10. Contrast can be enhanced by injection of contrast agents, by digital subtraction angiography, by dual x-ray energy imaging, or by phase contrast imaging.
11. CT is a diagnostic tool that combines usual x-ray attenuation with computation to generate cross-sectional radiographic images of the body in the xy-plane.
12. The gray scale of CT images is determined by the Hounsfield scale.
13. The Hounsfield scale compares the attenuation of water with attenuation of objects in the body.
14. High speed CT scans of the thorax take only 3 seconds and the dose is in the order of 6 mSv.
15. Radiography and CT compares favorably to MRI for scanning fractures, lung, and for early cancer recognition.

References

[1] Lanc L, Silva A. Digital imaging systems for plain radiography. New York: Springer Science+Business Media; 2013.
[2] http://en.wikipedia.org/wiki/Phase-contrast_X-ray_imaging
[3] Lewis RA. Medical phase contrast x-ray imaging: current status and future prospects. Phys Med Biol. 2004; 49: 3573–3583.
[4] Zhou SA, Brahme A. Development of phase-contrast X-ray imaging techniques and potential medical applications. Physica Medica. 2008; 24: 129e148.
[5] Bech M, Tapfer A, Velroyen A, Yaroshenko A, Pauwels B, Hostens J, Bruyndonckx P, Sasov A, Pfeiffer F. In-vivo dark-field and phase-contrast x-ray imaging. Nature Scientific Reports. 2013; 3: 3209.
[6] Bech M, Jensen TH, Bunk O, Donath T, David C, Weitkamp T, Le Duc G, Bravin A, Cloetens P, Pfeiffer F. Advanced contrast modalities for X-ray radiology: Phase contrast and dark-field imaging using a grating interferometer. Z Med Phys. 2010; 20: 7–16.
[7] Hoshino M, Uesugi K, Yagi N. 4D x-ray phase contrast tomography for repeatable motion of biological samples. Rev Sci Instruments. 2016; 87: 093705.
[8] Kevin L, La Rivière PJ. Sinogram restoration in computed tomography with an edge-preserving penalty. J Med Phys. 2015; 42: 1307–1320.
[9] Gabor HT. Fundamentals of computerized tomography: image reconstruction from projections. 2nd edition. Springer Verlag; 2009.

Further reading

Bushberg JT, Seibert JA, Leidholdt EM Jr, Boone JM. The essential physics of medical imaging. 3rd edition. Lippincott Williams & Wilkins, Wolters Kluwer; 2012.

Smith NB, Webb A. Introduction to medical imaging: physics, engineering and clinical applications. Cambridge Texts in Biomedical Engineering; 2010.

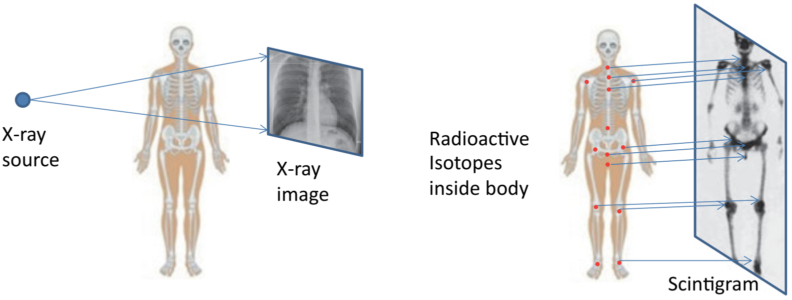
Useful website

www.upstate.edu/radiology/education/rsna/radiography/index.php

6 Scintigraphy

6.1 Introduction

In x-ray radiography and tomography the source of radiation (x-ray tube) is outside the body. Contrast on x-ray films or flat panel detectors is achieved by attenuation of the penetrating radiation in tissues of different density and thickness. In scintigraphy the source of radiation is inside the body. It is administered to the body by injecting radioactive isotopes into the blood circulation. The radioactive isotopes emit γ-radiation, which again is detected by x-ray films or by scintillation detectors. The difference between these two imaging modalities is highlighted in Fig. 6.1. Scintigraphy is not as general an imaging technique as x-ray radiography, the latter being used for inspection of healthy and diseased body parts alike. In contrast, scintigraphy is used for a number of specialized tests. Those include scans for tumor recognition, investigations of the metabolism of bones, of thyroid disorders, kidney clearance, lung ventilation, and of cardiac stress tests. These different scintigraphic applications will be presented in the following.



(a) X-ray radiography

(b) Scintigraphy

Fig. 6.1: Comparison of x-ray radiography and scintigraphy. Note that the source of radiation is outside the body in the case of x-ray radiography, whereas it is inside the body in the case of scintigraphy.

In short, scintigraphy is a nuclear imaging method using radioactively marked molecules that take part in the metabolism of the body. Therefore this method is only applicable to body parts that show perfusion, which excludes nails, teeth or hair. Some basic information on nuclear physics for understanding scintigraphic methods is provided in Chapter 2 and radiation dose including safety issues are discussed in Chapter 4.

DOI 10.1515/9783110553116-008

6.2 Isotopes for scintigraphy

All radioisotopes used for scintigraphy are produced artificially. The most common ones applied in nuclear medicine for scintigraphy are listed in Tab. 6.1. The ideal radioisotope is a high energy pure γ-emitter akin to the metabolism of the body. Unfortunately, isotopes of the light elements that are frequent in organic molecules, such as H, C, N, O, and P do not feature γ-emission lines in the proper energy range of 80–300 keV. Only iodine with the isotopes ^{123}I and ^{131}I is part of the thyroid metabolism and can be used for diagnostics (^{123}I) and for therapy (^{131}I). Other useful γ-emitters for scintigraphy, like ^{99}Tc, ^{111}In, and ^{211}Tl, need to be bound to ligands that promise to show similar and compatible biochemical behavior.

The isotope ^{99m}Tc is the most common γ-emitting radioisotope for scintigraphy. It is used for imaging of tumors of the skeleton, in the brain, kidneys, liver, gallbladder, lungs, etc. and for perfusion studies of blood in the heart and liquids in the kidneys. The main applications of ^{99m}Tc and other γ-emitting isotopes for scintigraphy are listed in Tab. 6.1. Before administering radioisotopes they need to be tagged onto radiopharmaceuticals with specific functions that transport them to the targeted destination. Some frequently used radiopharmaceuticals for tagging ^{99m}Tc and their targets are summarized as follows:

– ^{99m}Tc chemically tagged to an organic molecule $C_{13}H_{25}N_4O_3Tc$ known as exametazime (HMPAO), is able to cross the blood-brain barrier. HMPAO is therefore used for cerebral blood flow (CBF) imaging.

Tab. 6.1: Radioisotopes and their main use in scintigraphy (data are reproduced from www.nucmedtutorials.com/dwclinical/in111.html).

Isotope	Energy of γ-photons [keV]	Physical lifetime	Application
^{99m}Tc	140	6.02 h	Most widely applied radioisotope for diagnostics. Different radiopharmaceuticals are used for brain, bone, liver, spleen, kidney imaging, and for blood flow studies.
^{67}Ga	94, 184, 296	79.2 h	Gallium citrate used for demonstrating the presence of several types of malignancies.
^{111}In	173, 247	67 h	Used for tumor recognition
^{123}I	159	13.3 h	Widely used to diagnose thyroid disorders
^{201}Tl	135, 167	73 h	Used in cardiac stress tests
^{133}Xe	81	5.3 d	Used for studying lung ventilation
^{51}Cr	321	28 d	Used for radiolabeling of red blood cells for specialized in vivo/in vitro studies, not for imaging

- For bone imaging, a radiopharmaceutical known as hydroxymethylene (HMDP) or methylene diphosphonate (MDP) is used. The phosphonate concentrates in the mineral phase of the bone by chemisorption. Accumulation in the bone depends on the amount of blood flow, which is higher in cancerous areas than in healthy parts.
- Imaging of renal glomerular filtration is done by tagging ^{99m}Tc to mercaptoacetyl triglycine, which is known as an MAG$_3$ scan.
- ^{99m}Tc tagged to methoxyisobutylisonitrile (MIBI) ligands is used for myocardial perfusion imaging. Scans performed with the use of MIBI are commonly referred to as “MIBI scans”.

6.3 Isotope generators

Isotope generators provide short-lived radioisotopes in clinics for diagnostics and for therapy. Short-lived radioisotopes are extracted from longer-lived radioisotopes by *elution*. In equilibrium both nuclides, mother and daughter, appear to decay with the same half-life.

The best example for the elution process is the ^{99}Mo/^{99m}Tc generator. After ^{99}Mo is produced by fission reaction of ^{235}U (see Section 2.6.3), it is chemically purified and passed on an ion exchange column composed of alumina (Al_2O_3). The doubly negative charged and water-soluble molybdate $^{99}MoO_4^{2-}$ ion is firmly bound to the alumina substrate, whereas the single charged pertechnetate $^{99}TcO_4^-$ ion is less bound to the substrate. Whenever ^{99}Mo decays to ^{99m}Tc, technetium goes into solution and can be washed out. The column is placed in a lead container with tubes attached to allow column elution. The column is schematically shown in Fig. 6.2.

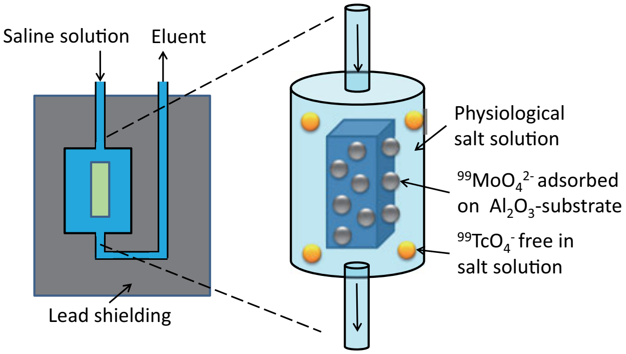


Fig. 6.2: Generator system for elution of ^{99m}Tc from a column containing $^{99}MoO_4^{2-}$ ions bound to alumina substrate.

The activity of ^{99m}Tc reaches a maximum after 23 h and then decays with the same effective half-life as ^{99}Mo, illustrated in Fig. 6.3 (a). Elution at the peak maximum is optimal and usually practiced in clinics. After elution the activity drops dramatically but is recovered after a few hours. Therefore several elutions are possible during one week, as depicted in Fig. 6.3 (b). This process is sometimes also referred to as “milking” the generator. The elute-containing salt solution and ^{99m}Tc radioisotopes are safely placed in a vial, from which a patient’s dose is withdrawn and the dose is calibrated before administering.

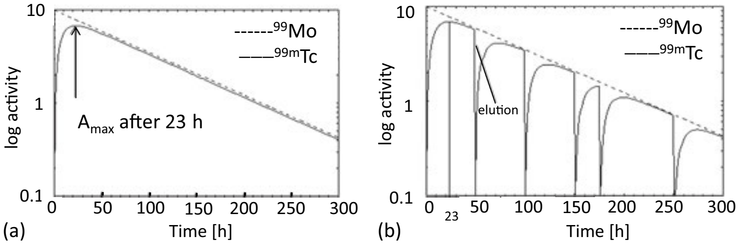


Fig. 6.3: (a) Build-up of ^{99m}Tc in a ^{99}Mo/^{99m}Tc generator. Notice the maximum ^{99m}Tc activity after 23 h. (b) Multiple elutions of ^{99m}Tc are possible during one week’s time.

6.4 Recording of scintigrams

Traditionally scintigraphy was recorded with x-ray films, just as already described in Chapter 5 for x-ray radiography. Today x-ray films have been completely replaced by two-dimensional scintillation detectors. 2D images provide information on the location of the γ-emitting radioisotope in the body and record the temporal evolution of the radioisotope distribution. As the radionuclide emits γ-radiation into a solid angle of 4π, collimation in front of the detector is absolutely necessary for enhancing contrast and spatial resolution. The contrast is usually high, since only the radioisotopes but not the rest of the body contribute to the image, in distinction to attenuation radiography via x-rays. However, the spatial resolution is much lower (≈ cm) compared to x-ray radiography (< mm). Figure 6.4 shows a full body scan of the ventral and dorsal part of the body after three hours of ^{99m}Tc injection. Another impressive sequence of scintigrams is depicted in Fig. 6.5. The perfusion of liquids through the renal system from the kidneys to the bladder was recorded over a time span of 12 minutes, probing the tubular filtration by using a ^{99m}Tc-MAG$_3$ scan.

For a full body bone scintigraphy scan using ^{99m}Tc the average activity injected intravenously is in the order of 500 MBq, sufficient for recording a full body γ-radiation image and low enough to keep the patient radiation level at a reasonable level. The corresponding dose is about 0.1–0.2 µSv. Im comparison, for a full body x-ray CT scan, a dose of more than 5 mSv is deposited, depending on the scanning speed.

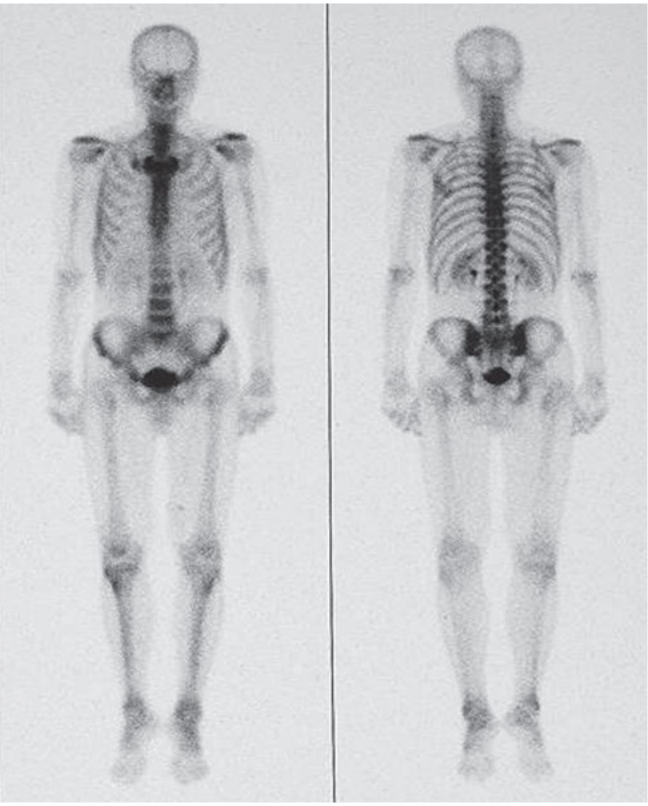


Fig. 6.4: Ventral and dorsal scintigram of the skeleton (adapted from https://en.wikipedia.org/wiki/Bone_scintigraphy#/, © Creative Commons).

In x-ray radiography, digital recording is performed with 2D flat-panel TFT detectors, which provide high resolution images. In scintigraphy, because of the intrinsically lower resolution, usually lower cost 2D scintillation detectors are used instead. Scintigrams recorded with scintillation counters are referred to as *single photon emission* (SPE) scintigrams.

The schematic of a scintillation detector is shown in Fig. 6.6. A scintillation crystal in front of a photon counter converts γ-rays into visible light. One γ-quant produces in the order of 5000 photons in an NaI crystal drifted with Tl ions. High atomic number Tl ($Z = 81$) is added to NaI to increase the quantum efficiency of x-ray detections. The visible photons hit a photocathode, where they produce about 400 photoelectrons per γ-quant. These electrons – in turn – are then accelerated in a cascade of dynodes, called photomultiplier, each time producing more electrons, until they arrive at the anode and deliver a sizeable current pulse. The pulse height at the anode is proportional to the energy of the incident γ-ray. The current pulse is then converted into a voltage pulse, which can further be amplified by electronics. Because of the

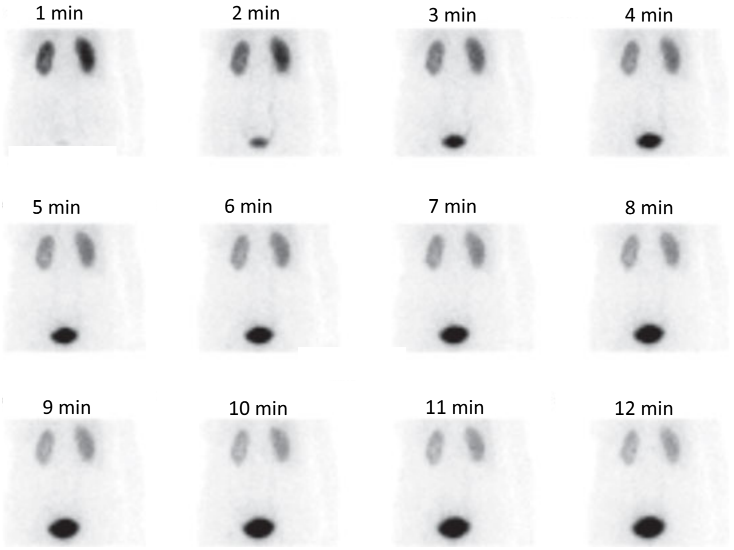


Fig. 6.5: Time sequence of the perfusion of liquids through kidneys and bladder (adapted from www.people.vcu.edu/mhcrosthwait/).

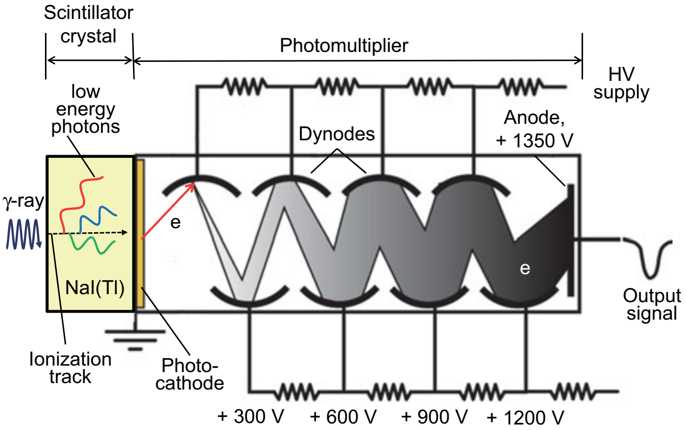


Fig. 6.6: Schematic of scintillation counter including a scintillation crystal, photocathode, and photomultiplier.

proportionality between energy of the incident γ-ray and the pulse height, scintillation counters are also used for γ-spectroscopy. These detectors are much faster than ionization chambers or Geiger–Mueller detectors discussed in Section 4.4, and their respective dead time is much shorter.

The scintillation detectors are arranged in an array covered by a scintillation crystal, as illustrated in Fig. 6.7. A collimator in front reduces the divergence and con-

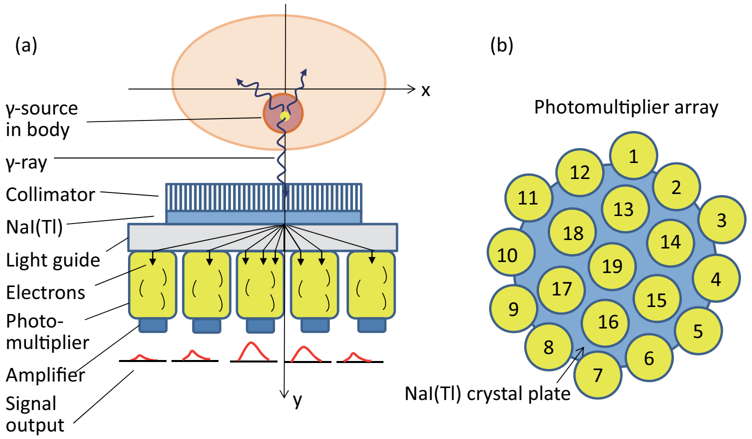


Fig. 6.7: (a) Detector system for scintigraphy consisting of collimator, scintillation crystal, and an array of photomultipliers. (b) Top view of photomultiplier array. Numbers help identifying the distribution of intensity.

tributes to the contrast. The collimator and the diameter of the scintillation counter determine the spatial resolution. Long and narrow scintillation tubes increase the resolution but lower the count rate. A light-guiding glass material channels the photons to the photocathode of the scintillation counter, where the photons are converted into electrons and amplified by a cascade of dynodes, as already described.

Planar scintigraphy or projection scintigraphy using a single detector head has the disadvantage of providing just one projection, similar to an x-ray radiograph that lacks depth information. Another disadvantage is the high dose patients are exposed to while the spatial resolution is rather limited. Projection scintigraphy with just one detector array should only be applied if there is a justified suspicion for a tumor, inflammation, etc. requiring confirmation. Another justified application is to test the functionality of certain organs like the renal system or the respiratory system. For scanning the thyroid there is in fact no alternative to planar scintigraphy, a typical examination setting is reproduced in Fig. 6.8. In all other cases single photon emission computed scintigraphy (SPECT) should be used as discussed in the next section.

6.5 Single photon emission computed tomography (SPECT)

The shortcomings of flat-panel projection scintigraphy can be overcome by taking scans simultaneously under different angles. One detector can be rotated around the patient as done in x-ray CT scans, or several planar detector arrays may be mounted on a circle recording simultaneously. One particular detector arrangement is shown in Fig. 6.9. Two detector panels may enclose an angle of 90° or 180° as seen in the

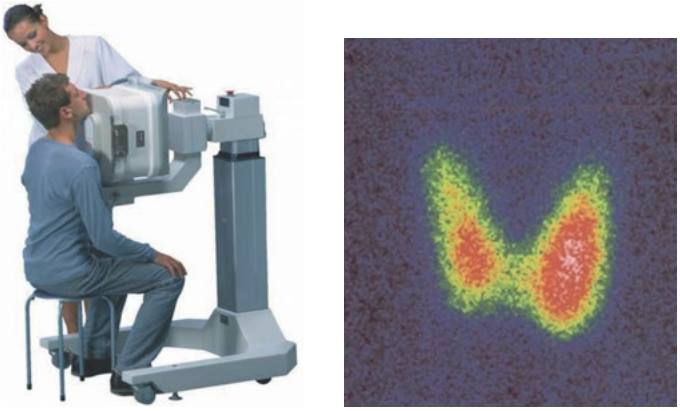


Fig. 6.8: Left: SPE scintigraphy of the thyroid with a flat-panel camera. Right: Tc-99m scintigraphy of the thyroid gland (reproduced from https://openi.nlm.nih.gov/, © Creative Commons).

Fig. 6.9: SPECT Siemens Symbia SW06 with two opposing detector heads (reproduced from www.siemens.com/press).

illustration, and three detector panels may be arranged at 60° angles. Each planar detector array is similar to the one used for SPE. In *single photon emission computed tomography* (SPECT), however, rotatability of the detector arrays and computer power are added for higher contrast and depth resolution of the scintigrams taken.

The radiation that the camera receives depends on the depth of the γ-emitter within the body. Through algorithms taking into account the absorption effect at different projection angles, the depth can be determined. An example is given in Fig. 6.10 (a). The γ-source inside the body is symbolized by a red star. It emits in all directions equally, but we consider for now only the vertical or z direction. The intensity on the recording panel is reduced by absorption in the tissue with different absorption coefficients μ_i. At the detector position the intensity, after passing 4 different voxels in this example, is then:

$$I(z) = I_0 \exp[-(\mu_1 + \mu_2 + \mu_3 + \mu_4)\Delta z].$$

In panel (b) two columns are outlined with different voxel sequences having different attenuation coefficients. However, the sum is identical and therefore also the detected intensity is the same: $I_1 = I_2 = I_0 \exp[-10\Delta z]$. Obviously, this unidirectional projection does not provide any contrast between left and right column, although the sequences are completely different.

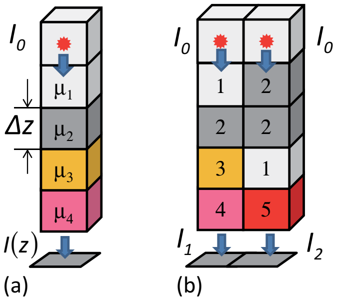


Fig. 6.10: Intensity at the detector after absorption in material stacks with different absorption coefficients.

Depth information can, however, be gained by adding projections from different directions as sketched in Fig. 6.11. Each voxel has a different unknown attenuation coefficient. The attenuated intensity from each of the 4 projections is listed in matrices P_1–P_4. All numbers are then added up in a sum matrix, attenuation due to background is subtracted, and the result is normalized to the smallest common denominator. The final matrix yields the spatially resolved attenuation coefficients of all individual voxels. With this method the contrast can be enhanced and different organs can be identified.

Similar to x-ray CT, SPECT images are taken at a sequence of angles with the detectors spiraling around the patient. After recording and data manipulation a complete three-dimensional image can be generated. However, to keep the dose level low, the

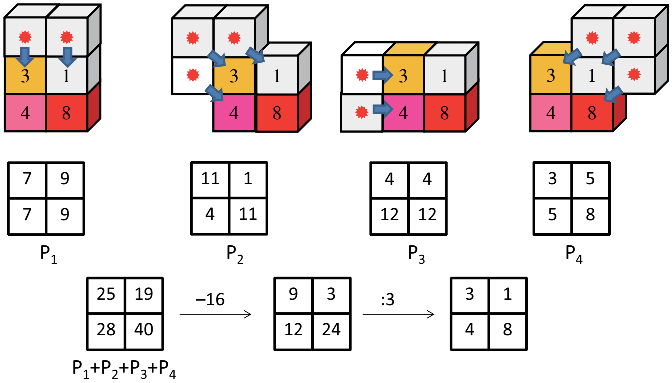


Fig. 6.11: Four different projections from the γ-source to flat panel detector heads.

count rate is also low and counting over 30 s per image or more is required, adding up to about a half hour for a full body scan, during which the patient ought not to move. In recent years SPECT detection has been greatly improved to reduce the acquisition time. But in most cases only specific parts are scanned, such as the brain or the heart, which takes less time.

SPECT is mainly applied for the following diagnostic procedures:

- Myocardial perfusion SPECT is used for inspecting the vitality of the myocardium. The sequence of the pictures in Fig. 6.12 shows the perfusion of blood through the heart chambers during different phases of the action potential. Any damage that might have been caused by a heart attack can be evaluated. Figure 6.12 compares the preoperative and the postoperative myocardial perfusion after setting a bypass.
- Bone SPECT is used to localize regions with different and conspicuous metabolism.
- Brain perfusion SPECT is used for diagnosis of normal functionality versus Parkinson's disease, Alzheimer's disease, and epilepsy.

An overview on recent advances in SPECT imaging can be found in [1, 2].

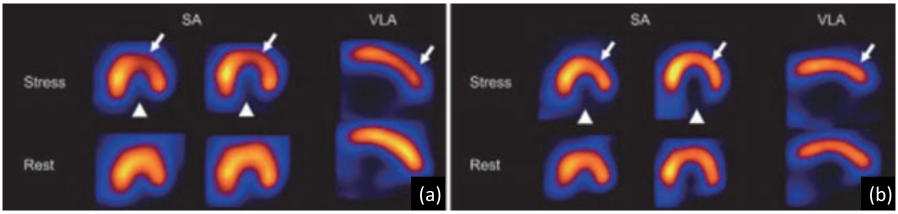


Fig. 6.12: (a) Preoperative and (b) postoperative myocardial SPECT. The patient underwent a left carotid-to-subclavian artery bypass grafting. Postoperative myocardial SPECT shows improved stress perfusion in the anterior and anterolateral walls when compared with the preoperative SPECT (arrows). SA = short axis; VLA = vertical long axis (reproduced from https://openi.nlm.nih.gov/, © Creative Commons).

6.6 Summary

1. Scintigraphy records γ-emission of radioisotopes that are injected into the body.
2. ^{99m}Tc is the most frequently used radioisotope in scintigraphy.
3. ^{99m}Tc is a decay product of ^{99}Mo, which in turn is a fission product of ^{235}U after slow neutron capture.
4. ^{99m}Tc is extracted in the clinic with help of a generator, which separates mother (^{99}Mo) and daughter isotopes (^{99m}Tc) after the activity of the daughter isotope has reached the first maximum.

5. ^{99m}Tc is used for brain, bone, liver, spleen, kidney imaging, and for blood flow studies.
6. Scintigrams are digitally recorded projection images with an array of scintillation counters.
7. Scintigraphy using scintillation counters is known as single photon emission (SPE).
8. SPE scintigraphy with a single flat-panel detector provides two-dimensional projection scintigrams.
9. 2D projection SPE scintigrams are used for full body scanning, for thyroid inspection, and perfusion studies of the myocardium and the renal system.
10. Arrangement of two or three flat-panel detectors allows depth resolution (SPECT), extending perfusion studies to the brain, and myocardial investigations under various stress conditions.

References

[1] Madsen MT. Recent advances in SPECT imaging. Journal of Nuclear Medicine. 2007; 48: 661–673.

[2] Slomka PJ, Pan T, Berman DS, Germano G. Advances in SPECT and PET hardware. Progress in Cardiovascular Diseases. 2015; 57: 566–578.

Further reading

Mettler FA, Guiberteau MJ. Essentials of nuclear medicine imaging. 6th edition. Elsevier-Saunders; 2012.

Bushberg JT, Seibert JA, Leidholdt EM Jr, Boone JM. The essential physics of medical imaging. 3rd edition. Lippincott Williams & Wilkins, Wolters Kluwer; 2012.

Smith NB, Webb A. Introduction to medical imaging: physics, engineering and clinical applications. Cambridge Texts in Biomedical Engineering; 2010.

Wernick MN, Aarsvold JN, editors. Emission tomography: the fundamentals of PET and SPECT. Elsevier; 2004.

Useful websites

https://nucmedtutorials.com/
www.people.vcu.edu/mhcrosthwait/

7 Positron emission tomography

7.1 Introduction

Positron emission tomography (PET) is a nuclear imaging modality used in clinics for cardiologic, neurologic, and oncologic studies. The PET method is based on the annihilation of positrons and electrons via converting their rest mass into two γ-photons flying in opposite directions. These two γ-photons are detected in a fashion similar to SPE or SPECT. Positron annihilation spectroscopy (PAS) is also used in condensed matter physics for determining the density and diffusivity of defects in solids. PAS is the opposite effect to pair production, which occurs when photons interact with nuclei at photon energies beyond 1 MeV. Pair production is important for cancer treatment with very hard x-rays, discussed in Chapter 9. However in this chapter we consider PET as an analytic tool.

In order to record PET, first positron (β^+) emitters have to be taken internally via radiopharmaceuticals. After annihilation with an electron and emission of two γ-photons the radiation is recorded externally with an array of scintillation counters, and tomographic images are formed with the help of computers.

For PET, low atomic number radioisotopes with β^+-emission are required. These are all produced artificially by proton bombardment of light atoms. Upon proton uptake, these elements have a deficiency of neutrons and therefore decay over time by converting protons back into neutrons, thereby emitting positrons. The general reaction path is written as ${}^{A}_{Z}X + \mathrm{p} \rightarrow {}_{Z-1}^{\;\;A}Y + \beta^+$ or in short notation ${}^{A}X(\mathrm{p,n}){}^{A}Y$. In most cases, the β^+-emitters have very short lifetimes ranging between 2 minutes and 110 minutes. Hence an isotope production line via accelerator or generator must be available either in-clinic or nearby. Nuclear reactions by proton bombardment are presented in Chapter 2 and Tabs. 2.3 and 7.1 provide a list of isotopes used for PET.

7.2 Basic principle of PET

After production of β^+-emitters with a proton accelerator, radiochemistry is required to synthesize PET tracers for intravenous or oral incorporation. These tracers perfuse or diffuse to the target organ where they preferentially accumulate. Once emitted, the high energy β^+-particles are deaccelerated through collisions with other electrons and by emitting synchrotron and Cherenkov radiation. As soon as the β^+-particles are slowed down and come to rest, particle (β^-) and antiparticle (β^+) annihilate together. Their joint rest mass is converted instantaneously into two γ-photons. Because of momentum conservation, these two γ-photons fly in opposite directions (180°), cancelling their total momentum. The rest mass and therefore the γ-radiation energy is $2m_\mathrm{e}c^2 = 2 \times 0.511\,\mathrm{MeV}$. Figure 7.1 schematically shows the stopping of a positron

DOI 10.1515/9783110553116-009

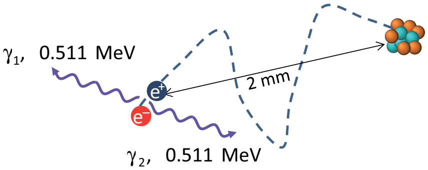


Fig. 7.1: Electron–positron annihilation via emission of two γ-photons propagating in opposite directions. After emission from a radioisotope, positrons travel for about 2 mm before slowing down and annihilating together with electrons in the body.

and the prompt emission of γ-radiation. The average distance between β^+-decay and γ-emission in soft tissue is about 1–2 mm (see Tab. 7.1 for specific isotopes).

Both γ-photons emitted simultaneously in opposite directions are used for imaging. They are detected by a coincidence detection scheme determining the angle of emission and the time of arrival. The schematic of a PET coincidence detector is shown in Fig. 7.2. Scintillation detectors are fixed around a ring with the patient in the center. If detector D1 registers a γ-photon, detector CD1 on the opposite side of the ring should also detect a γ-photon within a set time shorter than a nanosecond. If these conditions are fulfilled, the event is registered positive and counted, otherwise the event in D1 is rejected. This scheme ascertains that photons detected are due to β^+-β^--annihilation and not due to any other unrelated γ-emission. Detection based on this principle is called *annihilation coincidence detection* (ACD).

The double conditional counting with respect to angle and to time provides very clean, essentially background-free spectra. Because of the coincidence counting, proximity effects between counters can be neglected and collimators in front of the detectors are not required. In spite of the “clean” detection scheme, some limitations should be mentioned. Obviously the angular resolution increases with the number of

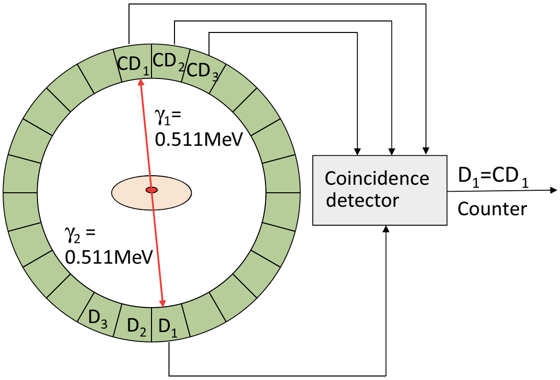


Fig. 7.2: Schematic for PET coincidence counting. The red point marks the origin of two photon γ-emission. Scintillation detectors are arranged on a circle around the source in the center.

detectors mounted on the ring. However, the physical size of the scintillation detectors sets an upper limit. Furthermore, coincidence may not only occur by a true event, starting from the center of the detection ring, as illustrated in Fig. 7.3 (a). Another possibility is an event which originates off-center and makes it into the counter by scattering, case (b) in Fig. 7.3. This process has a longer optical path length and could be suppressed by tightening up the time discriminator. One nanosecond corresponds to light traveling over a distance of 30 cm. With time coincidence set to ≤ 1 ns, events of type (a) and (b) cannot be distinguished, better time resolution is required. A third kind of coincidence, shown in panel (c), is the one of the rare event that two unrelated annihilation processes occur at the same time and one photon of each pair arrives at detectors 180° apart. This false coincidence cannot be avoided, but should be very rare.

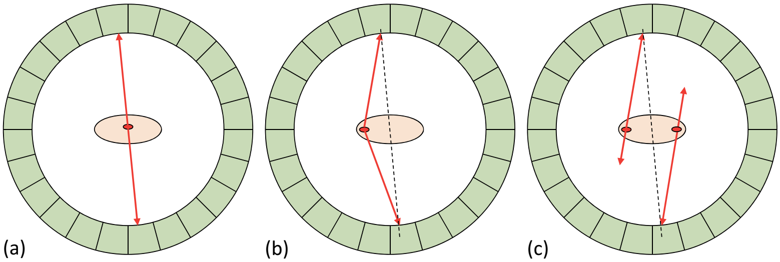


Fig. 7.3: Three different coincidence events: (a) true coincidence; (b) coincidence by scattering; (c) random coincidence.

The 2γ-emission is an incoherent process, meaning that the emission goes into 4π. However, the detector ring covers only a 2π slice out of all events, the other photons remain undetected. The detection efficiency can be enhanced by adding more detector rings, as indicated in Fig. 7.4. Coincidence counts from neighboring rings can be bunched together to enhance the counting statistics of the center ring. But in order not to lose spatial resolution, coincidence detectors on different rings have to correlate crosswise, i.e., detector D_{i+1} correlates with detector CD_{i-1}, etc. Having several de-

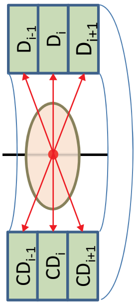


Fig. 7.4: Side view of PET scanner containing three rings of detectors.

tector rings, one may also consider extending the two-dimensional detection scheme into a three-dimensional detection arrangement. Some systems offer the option of performing either 2D or 3D data acquisition. In 2D data acquisition the recordings of each ring is kept separate, but a few neighboring rings may be binned together. In 3D recording coincidence events between all detector rings are properly recorded and used for a 3D tomographic reconstruction.

Another consideration is the quantum efficiency of the detector system for very hard 0.511 MeV γ-photons. Let us assume an efficiency of 0.7, i.e., 70 % of all γ-photons arriving in the detector are converted into photons of the visible spectrum. Then the efficiency to detect two photons is only 49 %. It is evident that the quantum efficiency of the scintillation detector should be as high as possible. For PET, the usual NaI(Tl) scintillation detectors are replaced by bismuth germinate (BGO = $Bi_4Ge_3O_{12}$) detectors. Bismuth has a slightly higher atomic number (83) than thallium (81) and therefore higher quantum efficiency for hard γ-photons. However, the drawback is a significantly lower light output and longer light decay time than for NaI(Tl), increasing the dead time of the detector such that time coincidence detection below 30 ns becomes unrealistic. Because of these shortcomings further detectors have been tested and are still under development. Promising materials are lutetium oxyorthosilicate (LSO), lutetium-yttrium oxyorthosilicate (LYSO), and gadolinium oxyorthosilicate (GSO). Detectors with these materials combine high density, high Z values, and better timing resolution reduced to about 5–6 ns.

Overall, PET detection systems feature a higher spatial resolution of about 3–6 mm compared to SPECT, which has a spatial resolution of about 2–4 cm. The resolution of PET is only limited by the number of detectors on the ring, by the straggling length of the positrons before they come to rest and annihilate together with electrons (2–3 mm), and by Compton scattering of γ-photons in tissue, causing a deviation from 180° angular correlation. In practice there are also limiting factors related to patient movement mainly during respiration. All other factors set aside, the intrinsic straggling length of positrons, depending on their initial kinetic energy after decay, is the most severe limitation with respect to spatial resolution.

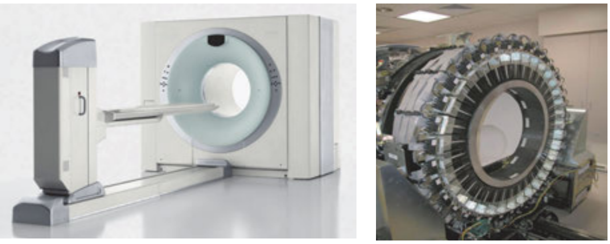

(a) (b)

Fig. 7.5: PET set-up. (a) Full set-up from Siemens (TruePoint PET·CT eco); (b) cover removed from the detector ring (reproduced from www.siemens.com/press).

Figure 7.5 (a) shows a complete PET machine including detector ring and stretcher for the patient, and in (b) the cover of the detector bank is removed. During a scan of a particular organ either the detector ring or the patient is translated across the region of interest. Admittedly from the outside a PET unit looks very similar to an x-ray CT unit. However, the inside ring has a completely different design. The x-ray CT unit contains an x-ray generator and an opposing detector bank, which both together rotate rapidly around the patient. In the case of a PET machine all detectors are stationary. The source of radiation is in the body and the detectors inside of the ring cover 360° for coincidence detection.

In recent years imaging systems have been combined which complement each other. For instance, if a cancerous region is identified with PET CT or SPECT, then this region can be scanned with x-ray CT, providing much higher spatial resolution. System combinations of SPECT/x-ray CT, of PET/CT and of PET/MRI have been discussed and constructed for improving diagnostic accuracy. An overview on recent advancements of scanners with improved image contrast, image resolution, and reduced image noise can be found in [1].

7.3 Isotopes for PET

As already mentioned, PET requires positron emitters that are artificially produced by high energy proton bombardment in an accelerator. Fortunately the light and biocompatible elements C, N, O, and F can be transmuted to positron emitting radioisotopes. This is in contrast to SPECT, where only heavy γ-emitting radioisotopes are available whose biocompatibility is a delicate issue. For PET, authentic radiopharmaceuticals are administered to the patient, i.e., pharmaceuticals with identical properties to those without radioisotopes, whereas for SPECT analogous but not identical radiopharmaceuticals are used. This is a big advantage of PET over SPECT. After production of β^+-emitting radioisotopes, radiochemistry is required to synthesize PET tracers bound in molecules similar to glucose. The synthesized tracer is then administered either by intravenous injection or by inhalation. Subsequently, the PET scanner detects gamma rays and registers coincidence events. After coincidence recording, the data are processed to exhibit locations of high gamma count rates originating from pair annihilation. Those are the areas where PET tracers have accumulated. The sequence of processes from isotope production to image processing is illustrated in Fig. 7.6 by respective icons.

Returning to the radioisotopes of the elements C, N, O, F, listed in Tab. 7.1. These isotopes allow true nuclear biochemistry. The replacement of a stable isotope in a biomolecule by a radioisotope such as ^{11}C, ^{13}N, or ^{15}O does not change the biochemistry of the tracers and permits an undisturbed imaging of the metabolic process. The

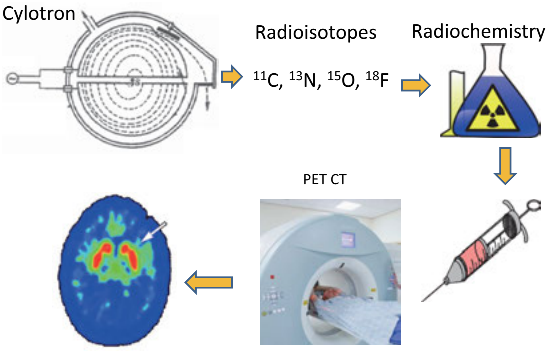


Fig. 7.6: Procedures required for PET scans with positron emitting radioisotopes. First, suitable radioisotopes are produced with a cyclotron by proton bombardment of light elements. Usually the cyclotron is located in the radiation clinic because of the short life time of the β^+-emitting radioisotopes. Then the isotopes are bound into authentic radiopharmaceuticals for administering to the patient. PET scans including data acquisition and fast data processing provide cross-sectional slices that map areas with enhanced activity indicative for changes of metabolism symptomatic for diseases.

Tab. 7.1: Positron emitting radioisotopes used for PET CT. Gen = Generator (Ge-68 and Sr-82).

Isotope	$T_{1/2}$ [min]	Production	E_{max} [MeV]	Range [mm]
C-11	20.5	$^{14}N(p,\alpha)^{11}C$	0.96	1.1
N-13	10.0	$^{16}O(p,\alpha)^{13}N$	1.19	1.4
O-15	2.1	$^{14}N(d,n)^{15}O$	1.72	1.5
F-18	110	$^{18}O(p,n)^{18}F$	0.9	1.0
Ga-68	68	Ge-68 Gen	1.9	1.8
Rb-82	1.27	Sr-82 Gen	3.15	2.1

exemption is ^{18}F as fluorine is not part of the body biochemistry. Nevertheless, PET CT with F-18 is particularly important, and therefore this procedure is discussed in more detail in the next section.

7.4 Use of F-18 in PET scans

Fluorine has only one stable isotope: $^{19}_{9}F$. When oxygen $^{18}_{8}O$ is bombarded with protons from a cyclotron, the reaction $^{18}_{8}O(p,n)^{18}_{9}F$ takes place, producing the unstable isotope $^{18}_{9}F$, which decays by β^+-emission back into $^{18}_{8}O$. The decay scheme is shown in Fig. 7.7. Usually pure or enriched ^{18}O-water is used as target and the proton beam energy is typically 18 MeV, although the cross section has a maximum at 5 MeV, (see

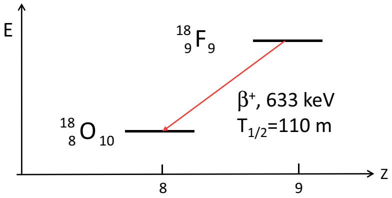


Fig. 7.7: Decay scheme of F-18 via β^+-decay.

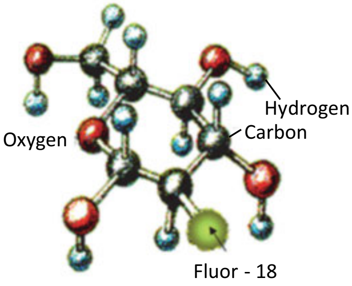


Fig. 7.8: ^{18}F-FDG molecule where ^{18}F is substituted for a hydroxyl group.

Section 2.6.2 for further information). The significance of ^{18}F β^+-emitting radioisotope in nuclear medicine is due to both a short half-life of only 110 min and a unique metabolism.

In the radiochemistry laboratory, F-18 is primarily synthesized into fluoro-deoxy-glucose (^{18}F-FDG) for use in PET scans. ^{18}F-FDG is a glucose analog where the positron emitting ^{18}F isotope substitutes for a hydroxyl group (see Fig. 7.8). ^{18}F-FDG is then used as a tracer in PET scans. In the body, ^{18}F-FDG follows an in vivo metabolic pathway similar to glucose, except that it is not metabolized to CO_2 and water, but remains trapped within the tissue. This makes it well suited to use as a glucose uptake tracer and is of particular importance in oncology because proliferating cancer cells have a higher than average rate of glucose metabolism.

In standard metabolism, glucose reacts with ATP to form glucose-6-phosphate and finally fructose-6-phospate by isomerization. The first part of FDG metabolism is identical to that of glucose. However, in the case of FDG metabolism stops after phosphorylation to FDG-6-phosphate and FDG becomes enriched by metabolic trapping. This is advantageous for early recognition of cancerous tumors, as they show an enhanced glucose metabolism. Thus enhanced γ-radiation indicates enrichment of radioactive ^{18}F-FDG revealing tumor cells in the PET images [2]. An example of a bronchial carcinoma is shown in Fig. 7.9 in comparison with other imaging modalities. On the CT image in the left panel the carcinoma is barely visible, but stands out in the PET image of the middle panel. The right panel shows a combination of CT and PET with impressive quality and a clear message of how valuable PET is for cancer recognition.

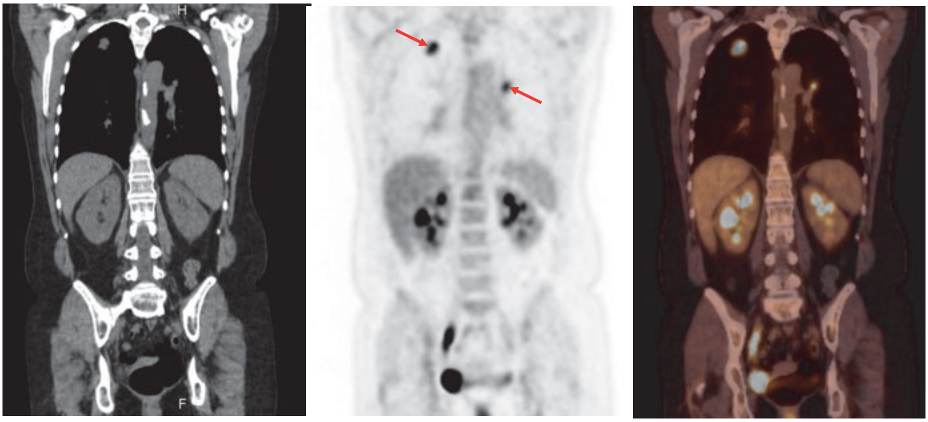


Fig. 7.9: Images of a bronchial carcinoma recorded with different techniques: left panel CT; middle panel ^{18}F-FDG PET scan demonstrating intense uptake in lung cancer of right upper lobe (arrow) and in lymph nodes on the lower left lobe (arrows); right panel shows a combined CT/PET image (reproduced with permission of Dr. Margot Jonas, Clinic Knappschaft, Bochum, Germany).

While ^{18}F-FDG is particularly useful for monitoring glucose metabolism, other tracers are used for scanning different metabolic processes as listed in Tab. 7.2. Before the introduction of functional MRI, PET was the first method to image the brain and to distinguish between normal activity and various abnormal conditions, such as Parkinson's disease, Alzheimer's disease, or Huntington's disease. While f-MRI has replaced part of the brain PET scans, it still remains useful for the detection of early stages of certain dementias, when the damage is too diffuse to be distinguished with MRI.

Tab. 7.2: Positron tracers and their applications in nuclear medicine.

Tracer	Indicator	Application
^{18}F-FDG	Glucose metabolism	Oncology cardiology, neurology
^{68}Ga-Dotatoc	Somatostatin receptors	Oncology
Na^{18}F		Bone metastasis
^{11}C-Acetate	Cholesterol metabolism	Cardiology
^{11}C-Methionin	Amino acid metabolism	Oncology
^{13}N-NH$_3$	Perfusion	Cardiology
$H_2{}^{15}O$	Perfusion	Cardiology, neurology

Concluding, we want to compare both methods PET and SPECT that use radioisotopes for detecting various diseases. A noncomprehensive comparison is given in Tab. 7.3. PET has the big advantage of being able to use light isotopes in connection with authentic radiopharmaceuticals. However, this advantage has a high prize, as the production of the isotopes is very expensive and because of very short half-lives the isotopes have to be produced by a proton accelerator (cyclotron) in the clinic or in a nearby facility.

Tab. 7.3: Comparison of PET and SPECT.

PET	SPECT
Emits positrons, detects gamma radiation	Emits gamma radiation, detects gamma radiation
Higher spatial resolution	Lower spatial resolution
Limited half-life of radiopharmaceuticals	Longer lived radioisotopes
Requires on-site cyclotron	Radioisotopes from external lab
Use of authentic radiopharmaceuticals	Use of analog radiopharmaceuticals
Costlier scanner	Less capital intensive scanner

7.5 Summary

1. PET is based on coincidence detection with high sensitivity and high homogeneous spatial resolution of 3–6 mm.
2. The spatial resolution of PET is limited by the number and size of the detectors, noncollinearities of the 2γ-emission process, and by the straggling length of the positrons.
3. Coincidence measurements do not require collimator blades in front of the detectors.
4. Fast scintillation counters with high light output and low dead time are essential for high sensitivity.
5. PET is based on positrons emitted from light isotopes with short half-life.
6. Application of PET requires an isotope laboratory in close reach.
7. The radioisotopes are integrated into biological active and authentic radiopharmaceuticals.
8. 18-F FDG is the most frequently used tracer for PET.
9. 18-F FDG takes part in the metabolism of glucose, but stops after phosphorylation. The trapped 18-FDG reveals areas with enhanced metabolism.
10. Among the imaging methods available, PET is the most expensive because of the combination of accelerator, radiochemistry, and scanning unit.

References

[1] Slomka PJ, Pan T, Berman DS, Germano G. Advances in SPECT and PET hardware. Progress in Cardiovascular Diseases. 2015; 57: 566–578.
[2] Acker MR, Burrell SC. Utility of 18F-FDG PET in evaluating cancers of lung. J Nucl Med Technol. 2005; 33: 69–74.

Further reading

Saha GP. Basics of PET imaging: physics, chemistry, and regulations. 3rd edition. Springer Verlag; 2016.
Santiago JFY. Positron emission tomography with computed tomography (PET/CT). Springer Verlag; 2015.
Wernick MN, Aarsvold JN, editors. Emission tomography: the fundamentals of PET and SPECT. Elsevier; 2004.

Useful website

www.people.vcu.edu/mhcrosthwait/PETW/index.html

Part B: **Radiotherapy**

8 Cell cycle and cancer

8.1 Introduction

The present chapter is intended to provide a basic understanding of the cell's life cycle and the difference between normal and cancerous cells. This information is a prerequisite for the rationale behind cancer treatment plans using radiotherapy discussed in Chapters 9–12. For more specific and detailed information on cell cycle the reader is referred to standard biology [1] or physiology textbooks listed under "Further reading". Here we focus the discussion on the relation between dose and cell survival rate that is contextual information for the following chapters on radiotherapy.

Genes and a number of messenger proteins determine when and how fast cells grow, when they divide (mitosis) and when they die (apoptosis). Some cells live for only a very short time, such as the cells in the epithelial layer of the intestines (1.5 days), some have a decently long lifespan such as the cells of the skin epidermis (20 days) or the red blood cells (up to 60 days), some cells live for a long time, such as the osteocytes (25–30 years), and some cells never die, such as the nerve cells. Those cells which die over time need to be replaced. Those that never die cannot be easily replaced, even when injured, which is the problem with spinal cord injuries. It may happen by still unknown and uncontrolled reasons that a normal cell suddenly turns into a malfunctioning cell that overrides all rules of a 'normal' cell's life and death, i.e., the cell becomes cancerous. Radiotherapy is presently one method out of three to fight cancerous cells; the other two are chemotherapy and surgery.

8.2 Life cycle of cells

The normal life cycle of a cell can be subdivided into two main phases: interphase and mitotic phase (M), see Fig. 8.1. Interphase and mitotic phase constitute one cell cycle. During the interphase the normal cell activity takes place including growth and DNA replication. 90 % of a cell cycle is occupied by the interphase. The remaining 10 % is filled with the mitotic phase, where the cell divides itself into two identical copies. After cell division is completed, both cells have the exact same genetic information and the interphase starts over again simultaneously in both new cells.

The cell's life cycle starts with the growth phase G_1 during which the cell enlarges and synthesizes new proteins. Close to the end of this phase there is a checkpoint that controls stop or go to the next phase. In the case of stop, the cell goes into a rest phase G_0 and remains there until called upon, for instance for repairing injuries. Most cells of the body are in the rest phase G_0 and fulfill their specified duties. If the checkpoint gives a green light, the cell will complete the cycle and divide. The checkpoint

DOI 10.1515/9783110553116-010

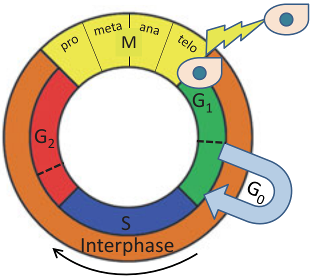


Fig. 8.1: Life cycle of a cell. G_1 = growth phase, G_0 = rest phase, S = replication phase, G_2 = second growth phase, M = Mitosis. The mitotic or cell division phase is subdivided into four subphases: prophase, metaphase, anaphase, and telophase. Dashed lines indicate checkpoints. At the end of the telophase two identical new daughter cells are formed.

is very important for cell regulation. If the stop signal is overwritten by some means, uncontrolled replication of the cell may occur, as we will see later on.

The total genetic information is contained in any cell and is encoded in double stranded DNA molecules. The genetic information of humans – the genome – is written into 46 *chromosomes*, each one containing one long DNA molecule. These 46 chromosomes come in 23 homologous pairs, one of each pair is inherited from mother and father. Both chromosomes in a pair contain essentially the same genetic information governing the basic function of organs. This is true for the first 22 chromosomes. However, the 23rd pair is different for female (XX) and male (XY), determining the sex. Three pairs of chromosomes in their G_1 phase are sketched in Fig. 8.2 (a) representative for all others.

During the S phase cells replicate their genetic material, ensuring that each daughter cell will receive an exact copy later during the mitotic phase. During DNA replication the two strands are copied in opposite directions (leading and lagging strands) using different transcription methods. After replication is completed, the DNA condenses into chromosomes, which is a heavily coiled and folded up version of long DNA strands. Only in this condensed phase can chromosomes be recognized by a microscope. Each duplicated chromosome has two *sister chromatids* (joined copies of the original chromosome), which separate during cell division. The sister chromatids are attached to each other by a *centromere*, see Fig. 8.2 (b). The *centromere* is a narrow “waist” of the duplicated chromosome, where two chromatids are most closely attached. Centromeres are important for organizing the associated chromatids in the cell and for their mechanical separation in the mitotic phase. Once separated by cell division, the chromatids are called *chromosomes*.

The complex topological copying process of the DNA offers many opportunities for miscoding. Therefore, at the end of phase S and in the G_2 phase there is another checkpoint that confirms the proper replication. Mistakes are repaired, but if the DNA damage is too great, the cell’s development is prevented from proceeding to the next stage, the mitotic phase. Instead, the cell prepares itself for an apoptosis, i.e., a programmed cell death. All fragments of the cell are disassembled and disposed of to avoid any further harm.

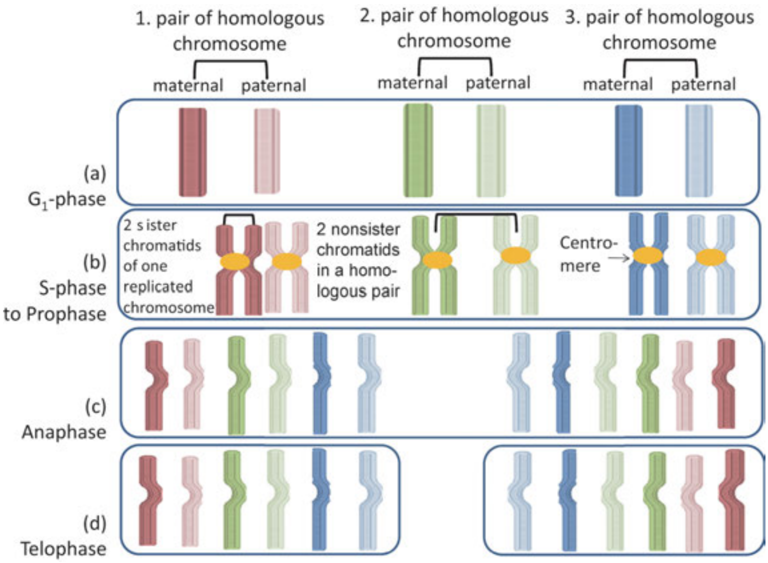


Fig. 8.2: Cell division starting from the G_1 phase to the end of the M phase. (a) In the G_1 phase all 46 chromosomes appear in homologous pairs, only three representative ones are sketched. (b) In the S phase DNA replication and synthesis takes place, which condenses to sister chromatids in the prophase at the beginning of the M phase. (c) In the anaphase the sister chromatids are pulled apart to opposite halves of the cell. (d) In the final telophase a complete separation and generation of two new cells occurs.

If the checkpoint in phase G_2 is passed, the cell will further grow before entering the mitotic phase (M). The M phase is subdivided into prophase, metaphase, anaphase, and telophase. In the prophase spindle fibers attach to the centromeres on one end and to anchor points on opposite ends of the cell, called centrosomes. In the metaphase the chromatids align in the center of the cell, in the anaphase the chromatids are pulled apart by the spindle fibers (Fig. 8.2 (c)), each sister moving to the opposite end of the cell, and finally in the telophase the cell divides by cleavage, forming two new daughter cells (Fig. 8.2 (d)). The original cell content is completely taken over by the two daughter cells. One life cycle of a cell is completed and two new life cycles can now begin.

Once the G_1/S checkpoint has been passed, the remaining typical cell cycle time is about 10–40 hours. Completion of the G_1 phase takes about 30 % of the cell cycle time, the S phase 50 %, G_2 phase 15 %, and the M phase only 5 %, i.e., between 0.5 and 2 hours.

At each end of a chromosome there is a wound-up protein called *telomere*. This is a countdown clock that shortens with each cell division. After 50 divisions, the shortened length of telomeres causes mitosis to stop. Once this point is reached, the cell is commanded to commit apoptosis (suicide).

Normal cells contain an extensive feedback system inside and outside for keeping a dynamical balance between mitosis and apoptosis. Cells are also able to sense their surroundings and respond to changes. For instance, if a cell senses that it is surrounded on all sides by other cells, it will stop dividing. In this way, cells will grow when needed but stop when in contact with other cells. In the case of a gap due to injury, cells fill in the gap left by a lesion but stop dividing when the gap has been sealed. Cancer cells do not exhibit contact inhibition. They grow even when they are surrounded by other cells, causing a mass of cells to form.

8.3 Cancerous cells

The World Health Organization (WHO) has classified abnormal masses of tissue called *neo-* (new) *plasms* (formation or creation) into four main groups: *benign neoplasms*, in situ neoplasms, *malignant neoplasms*, and neoplasms of uncertain or unknown development. *Malignant neoplasms*, simply called *cancers*, exhibit cell growth and cell division which exceeds and is uncoordinated with that of normal tissue. Growth and division persist in the same excessive manner even after eliminating the stimuli which invoked the original change, and the new growth has virulent or adverse properties in the body [2]. The paradigm in cancer research is to identify a close connection between *carcinogenesis*, the process that initiates and develops cancer, and alterations or mutations of the genome.

Whatever the cause is for damaged or mutated DNA, if such alterations are not stopped in time, the damaged DNA may replicate and act as an oncogene, i.e., a gene that causes cancer. These oncogenes have the ability to avoid apoptosis, to evade surveillance by the immune system, and to override telomerase, which otherwise would stop cell replication after 50 divisions.

In normal cells there are at least two important known checkpoints, located at G_1/S and G_2/M involving tumor suppressors. The G_1/S checkpoint calls for a stop when the DNA is damaged or when the surroundings signal that cell growth and cell division is not required. After replication and before division, checkpoint G_2/M controls again any irreparable DNA damage, which may have occurred during the S phase. During carcinogenesis, these checkpoints fail to prevent abnormal cells from going through the cell cycle. Abnormal cells that are not eliminated by the checkpoints pile up within otherwise normal tissue. If those remain only at the original site, the lump is called a *benign tumor* or *primary tumor*. If the abnormal cells invade surrounding tissues and *metastasize* by exporting cancer cells to other parts of the body, where they may initiate additional tumors, they are called *malignant tumors* or *secondary tumors*.

The checkpoint G_1/S is controlled by a protein named p53 acting as a tumor suppressor. Because of its control function, p53 has been described as “guardian of the genome”, providing stability and preventing mutations of the genome. Accordingly, the gene TP53 responsible for coding the protein p53 has been named tumor suppressor gene. By binding to the DNA the protein p53 has many additional control functions. It can activate DNA repair in case of damage, or alternatively command apoptosis if the DNA damage turns out to be irreparable. In cancerous cells it was found that the gene TP53 contains numerous mutations. If the TP53 gene is damaged, tumor suppression via p53 control is severely inhibited. Cancer research pays much attention to the functionality and the frequency of mutations in the tumor suppressor gene and how this affects p53 with respect to various types of cancer; an overview is given in [3].

Cancerous cells do not need to divide faster than normal cells. But as soon as the balance between cell division (mitosis) and cell death (apoptosis) is disturbed, such that cells can further divide and grow, cancer is initiated. Thus, cancer is defined as uncontrolled cell division and unlimited cell life.

While mutations are observed in most cancer cells, mutation accumulation is not the only cause for carcinogenesis. Certain precancerous lesions were found that were not induced by genetic alterations but by cell lineages driven by opportunistic competition [4]. Cancer may be the interplay of genetic and environmental effects, such as asbestos, air pollution, ionizing radiation, inadequate nutrition, virus infection (i.e., cervical cancer), or psychic stress weakening the immune system. Heredity associated with cancer might also include those mutations directly linked to neoplastic transformation of cells and to general health conditions such as the immune system.

There have been attempts to understand cancer as one of multiple intrinsic robust states of the endogenous network, i.e., communication network between cells [5]. The dynamical interaction among endogenous agents like hormones can generate many locally stable states for long periods of time with normal biological functionality, such as cell growth, cell division, and apoptosis. The statistical nature may cause a sudden transition from one stable state to another. If the switched local state is not optimized to the interest of the entire organism, the organism becomes “sick”. It is hypothesized that tumor cells belong to one of those robust states that have suddenly switched [5]. Such new perspectives might lead to additional treatments of cancer beyond the current three methods, radiotherapy, chemotherapy and surgery.

Typically tumors develop in the following six stages illustrated in Fig. 8.3: (1) mutation; (2) extension; (3) attack; (4) penetration; (5) angiogenesis; (6) invasion and distribution. The starting point is a cell containing DNA material that has transformed a normal gene into an oncogene overriding all checkpoints. The indicated red cell in panel (a) undergoes cell division, although at this location and at this time there is no call for replacement of an existing cell. In the second step the cancerous cell replicates itself without control (b). After about a million cell divisions the tumor has the size of a pin head, too small to be detected by any of the imaging modalities discussed in Chapter 13–15/Vol. 1 and Chapter 5–7 of this volume. In the course of further mutations and replications the tumor cells start attacking the connective tissue (c). Up to this point the tumor is still locally confined. This stage is sometimes also called *primary tumor*. A primary tumor may be present in the body for months or years before clinical symptoms develop. Mortality during primary tumor is minimal, with the exception of brain

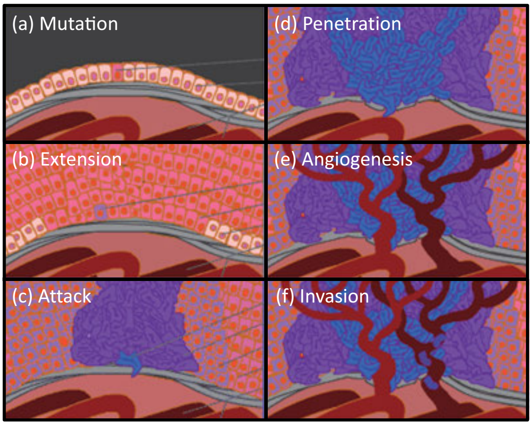


Fig. 8.3: Different steps of cancer development. For explanations see text (adapted from www.worldofteaching.com).

tumors. Further attack in step 4 leads to penetration of the connective tissue (d). From this point on the cancerous cell mass is no longer confined, but still too small to be detected. As the uncontrolled cell growth needs oxygen and nutrients for further growth, the cancer cells send out messages that initiate the advancement and distribution of blood vessels into the cancerous region (e). This step is called *angiogenesis* and marks the transition from a *benign* to a *malign tumor*. With oxygen supply the tumor can now grow rapidly. With a billion of cells the *malignant tissue* or *malign tumor* has reached the size of a grape and can be detected by imaging methods. In the last step (f) cancer cells invade the circulation and from there other organs of the body are affected. This process is called *metastasis* and is responsible for the formation of *secondary tumor* cells at other locations. Metastases more commonly cause death. Alternately, tumor cells can also spread via the lymph system.

Since cell division is a binary process, the number of cancer cells is, naively speaking, expected to grow exponentially. Thus the tumor volume $V(t)$ should double at constant time intervals Δt_d. For small tumors indeed a linear relationship $\log(V(t)) \sim t$ has been observed with an average doubling time Δt_d of about 2–3 months. As the tumor grows larger, the logarithmic volume dependence levels off, presumably because of insufficient oxygen and nutrition supply via penetrating blood vessels.

There are many possibilities for the immune system to attack cancer cells once they have been recognized. Nonetheless, some cancer cells make it to distant locations, settle in and transform normal cells into cancer cells. Unfortunately, the biochemical processes behind these disastrous stages of cancer development – although sought after by many laboratories around the world – are still unknown. As long as the genetic origin of cancer development is not completely understood and drugs for impeding cancer cell division are not available yet, we still need to apply rather crude methods of fighting cancer via radiotherapy, chemotherapy, and surgery. Hopefully with the help of fast sequencing techniques the key mutations can be identified and targeted drugs may be synthesized that stop uncontrolled cell division. At present the leaky and porous blood vessels (vasculature) in the tumor volume are the main target points for pharmaceuticals and nanoparticles. Specific and targeted nanoparticles can be excited by magnetic or optical means to produce local heating causing hyperthermia that kills cancer cells. These promising procedures will be further discussed in Chapter 14 on nanomedical applications.

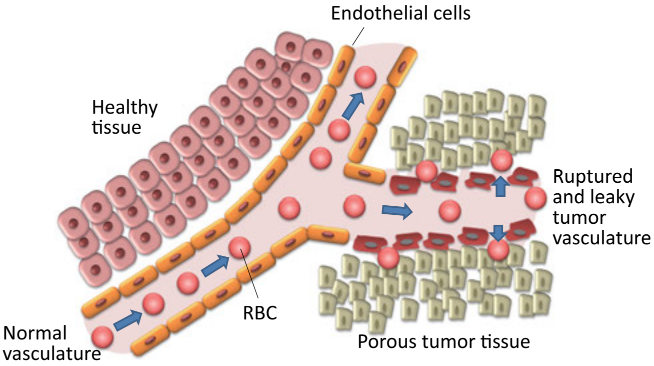


Fig. 8.4: Normal vasculature with blood-tight vessels supply healthy tissue with oxygen and nutrients. Ruptured and leaky vasculature penetrates porous tumor tissue.

Healthy versus tumorous vasculature is shown schematically in Fig. 8.4. Normally endothelial cells line the entire circulatory system including heart and brain up to the smallest capillaries. This monolayer of cells has an important function as a barrier between vessel lumen and surrounding tissue and for controlling the passage of gases, liquids, particles, and chemicals across the boundary. In the brain the endothelium is particularly tight, allowing only the passage of water, some gases (O_2, CO_2), hydrophobic lipid-soluble molecules, and a small amount of molecules such as glucose and amino acids that are essential for neural activity. It does not allow the diffusion of microscopic objects, such as bacteria or nanoparticles, or larger hydrophilic molecules like some contrast agents to enter the cerebrospinal fluid (CSF). This tight endothelium in the brain is known as *blood-brain barrier* (BBB). In tumorous tissue the vasculature is ruptured and leaky. Ruptured blood vessels enable contrast agents and even nontargeted nanoparticles to percolate across the wall where they help identify tumor volumes and assist in treatment procedures.

8.4 Radiation response of cells

Radiotherapy is used for cancer treatment either to control tumor growth or to destroy tumor cells. The aim of radiation treatment is to cause either mitotic or apoptotic death. Alternatively, radiation is supposed to alter molecular signaling pathways and in particular to block the G_1/S and G_2/M checkpoints to inhibit further cell division.

Before applying any radiation for cancer treatment a number of issues need to be considered first. What kind of damage will be caused by radiation? Which radiation is best for the treatment of specific tumors, low LET radiation (x-rays, γ, β, protons) or high LET radiation (α, neutrons), for a definition of LET, see Chapter 3? What is the appropriate dose? What is more effective: single lethal dose or fractionated dose? How can normal tissue be protected? These issues are discussed in the following.

8.4.1 Cell response to radiation

We start our discussion with the first question. Normal cells and cancer cells react in the same way to radiation. Both types of cell have the ability to self-repair if the dose is not lethal. But normal cells require a higher dose of radiation for complete apoptosis than cancer cells. We come back to this important point further below. The sensitivity to radiation varies during the life cycle of a cell. In Fig. 8.5 the single cell survival rate as a function of dose for x-rays applied during different phases of the cell life cycle is plotted. The semilog plot shows that the radiosensitivity is highest during the M phase, whereas the late S phase is the most radiation resistant phase; the G_1 phase takes an intermediate position. Note that the initial radiosensitivity of the S phase to low dose radiation is very low, expressed by an extended shoulder before the linear term (in the semilog plot) takes over. The shape of the survival curve S (not to be confused with the S phase) as a function of dose D is usually described by the following expression:

$$S(D) = \mathrm{e}^{-\alpha D-\beta D^2} .$$

The first exponential term describes the linear dependence in the semilog plot, whereas the second term defines the shoulder, signaling delayed dose response. The shoulder is typical for low LET radiation, whereas with high LET radiation only a linear dependence is observed. $\alpha = 1/D_0$ characterizes the slope of the survival curve, which not only distinguishes different phases of the cell cycle, but also different types of cancer cells and their oxygen content (oxicity), as we will recognize in the following.

The open uncoiled structure of the DNA and lack of compactness may explain the radioresistance in the S phase to low LET radiation, whereas the lack of repair mechanisms may contribute to the high radiosensitivity in the M phase. The difference is as great as a factor of 2.5 according to the plot in Fig. 8.5.

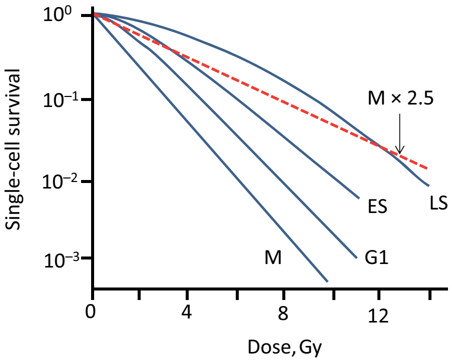


Fig. 8.5: Single cell survival fraction versus dose upon high energy x-ray radiation. ES (LS) = early (late) stage of S phase. The dotted red line indicates that the radiosensitivity of the M phase is higher than for the LS phase by a factor of 2.5.

8.4.2 Fractionation

Since the M phase covers only a short time span in the cell's life cycle, the probability to hit the M phase is higher if the total dose is broken down into smaller portions administered in a sequence of time intervals. This is called *fractionation*. Each fractionated dose is not sufficient to kill tumor cells, but the accumulated total dose will finally succeed. Furthermore, after first radiation exposure cells readjust their cell clock with respect to their cell cycle to be more in phase with one another. However, during the interval time they also have a chance to self-repair the damage. Therefore the total dose has to be increased compared to single dose exposure, as indicated in Fig. 8.6. The survival fraction of cells versus dose is plotted, which is typically 2 Gy per fraction.

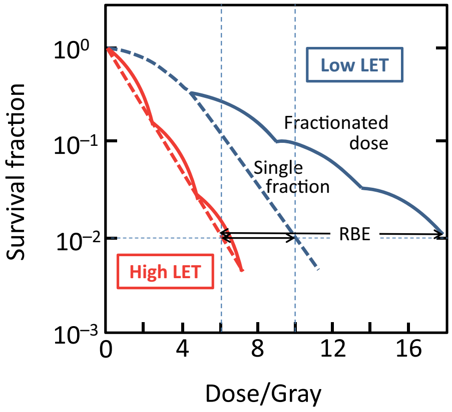


Fig. 8.6: Survival fraction as a function of dose for single fraction (dashed lines) and fractionated dose application (solid lines) for both low LET and high LET radiation.

The example shows x-ray radiation with low LET compared to high LET radiation. The comparison defines the *relative biological effectiveness* (RBE), discussed in Section 8.4.4. The survival rate per fraction is an important number, because the fractions are repeated 30–40 times during a complete treatment. If the survival rate per fraction is 0.8, after 30 fractions the survival rate is $(0.8)^{30} = 10^{-3}$; however, if the survival rate for a single fraction is 0.7, the final survival is reduced by two orders of magnitude to 2×10^{-5}. So the dose has to be finely tuned in order to have the best effect.

8.4.3 Biological effect of radiation

The sensitivity of cells to radiation is due to cutting chemical bonds of chromosomes. We distinguish between direct and *indirect action* [6]. In the case of *direct action*, photons or secondary photoelectrons break bonds in strands of the DNA, as sketched in Fig. 8.7. Furthermore, radiation generates free water radicals such as H^+, OH^-, or H_2O^+ in cellular water, which increases the toxicity in a cell and can also break bonds in the DNA, known as *indirect action*.

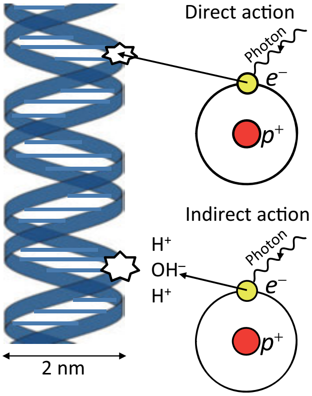


Fig. 8.7: Comparison of direct action on the DNA through the photoelectric effect and indirect action through formation of free radicals (adapted from Hall and Giaccia, [6]).

Most effective direct radiation damage occurs when not only one strand breaks (*single strand break*, SSB) but when both strands of DNA (*double strand break*, DSB) are hit simultaneously. Then the proper repair of the DNA is impeded and the survival chance of the cell is drastically reduced. The strands may still be partially repaired, but lose their ability to reproduce. Often these cells still pass from the G_2 to the M phase, but experience a mitotic catastrophe in the M phase. Cancer cells are considered dead if they have lost their ability to reproduce. This may not occur immediately, but may require a few more cell divisions. Malfunctioning cancer cells can be recognized in histological cuts by changes in the cell morphology, in particular by shrinkage and bubble formation (blebbing) of cell membranes.

Cancer cells have less ability than healthy cells for DNA repair between fractionated doses. With each subsequent mitosis, the cumulative effects of unrepaired DNA result in an apoptosis of these tumor cells. In this case the original cell together with fragments of DNA breaks into smaller parts, while the cell membrane and organelles inside remain intact. These apoptotic cell fragments are finally removed by phagocytosis [7]. Phagocytosis is a process by which scavenger cells, such as white blood cells, ingest debris of apoptotic cell material for destruction and removal. More on phagocytosis can be found in Section 14.2.1.

8.4.4 Relative biological effectiveness (RBE)

In the case of x-ray radiation, direct or indirect actions occur more frequently with increasing photon energy. According to Fig. 3.8 the absorption cross section via photoelectric effect decreases with increasing photon energy. Beyond 100 keV the photoelectric effect is irrelevant, and beyond 4 MeV the photon energy in matter is completely converted to Compton electrons and pair production. This results in an increasing chance for DSB, reaching a maximum probability at an energy transfer of about 100 keV/µm (100 eV/nm). At this energy the reaction distance between two hits corresponds to the diameter of a DNA molecule or about 2 nm (Fig. 8.8). For lower energies the reaction distance is larger, causing mainly SSB, and for higher energies the short reaction distance reduces the effectiveness of radiation damage. However, the most effective energy loss per distance of 100 eV/nm cannot be reached with x-rays or photoelectrons but requires heavy particles like protons, neutrons or α-particles that feature higher LETs. An overview on the LET for different particles and energies is given in Tab. 8.1.

All these facts call for a reconsideration of the relation between linear energy transfer (LET) of radiation $\langle dE/dx \rangle$ and relative biological effectiveness (RBE). In Chapter 3 we discussed the interaction of radiation with matter and the range of different types of radiation. In general, radiation with high LET (α-particles, neutrons) has a short range, that with low LET (x-rays, γ, β, p) has a large range. As can be recognized from the plot in Fig. 8.9, the RBE first increases with increasing LET, then goes

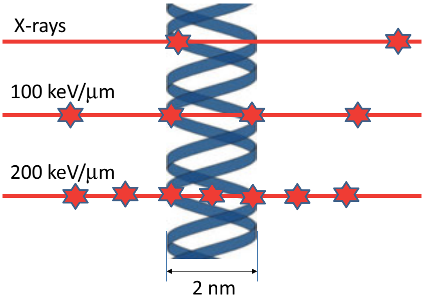


Fig. 8.8: Primary damage to the DNA for radiation with different LET.

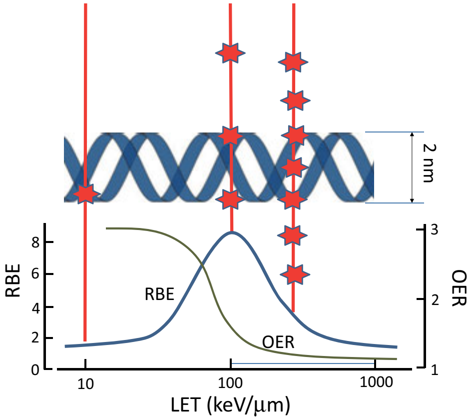


Fig. 8.9: RBE versus LET goes through a maximum. At the maximum the probability is high for double strand breaks in the DNA. The stars symbolize interactions of radiation with DNA strands along their track. The maximum RBE can only be reached with high LET particles, but not with x-rays or γ-rays. Over the same range OER drops from 3 for low LET to 1 for high LET radiation.

Tab. 8.1: Linear energy transfer for different types of radiation according to Hall and Giaccia [6].

Radiation	LET [keV/μm]
1.33 MeV ^{60}Co-γ-rays	0.2
250 kVp x-rays	2.0
10 MeV protons	4.7
150 MeV protons	0.5
14 MeV neutrons	12
2.5 MeV α-particles	166

through a maximum at 100 keV/μm, and decreases again at higher LET. At 200 keV/μm the RBE has dropped again and at 1000 keV/μm the RBE is back to 1. The maximum can be reached with a few MeV α-particles, but not with x-rays or γ-radiation. The RBE of the latter radiation always stays on the left side of the peak.

The RBE of radiation can directly be determined from Fig. 8.6 comparing survival fractions as a function of dose for different radiation sources. RBE is defined as the ratio:

$$\mathrm{RBE} = \frac{\text{Dose (x-ray)}}{\text{Dose (particle)}}.$$

By this definition the RBE of x-rays is 1. Now we compare the dose required in a single fraction for a survival probability of 1 %. The required x-ray dose is 10 Gy, whereas the high LET particle dose is 6 Gy. This yields an RBE of 1.7. However, when comparing fractionated low LET dose with fractioned high LET dose, for the same survival probability, according to Fig. 8.6, we obtain a ratio of 18/6 = 3. From this comparison we conclude that RBE depends on the way the radiation is fractionated. For low LET, fractionation makes a big difference, whereas for high LET radiation the difference is negligible. Nevertheless, fractionation is essential for both types of radiation because of the recovery that healthy tissue needs between fractions.

In Fig. 8.9 the oxygen enhancement ratio (OER) versus LET is also plotted. OER is defined as the ratio:

$$\mathrm{OER} = \frac{\text{Dose to produce effect without oxygen}}{\text{Dose to produce same effect with oxygen}}\,.$$

We notice that for low LET radiation the OER is almost 3, whereas for high LET radiation the OER drops to one. This implies that for low LET radiation the oxygenation of cancer cells is a very important parameter to obtain a positive treatment result, whereas for high LET radiation the oxygenation level of cancer cells has no effect on the result. Hypoxic and even anoxic cancer cells can be treated with high LET radiation without additional oxygen conditioning. For low LET radiation oxygen is important to increase the indirect action effect of radiation. The effect of oxygen is further discussed in Section 8.4.6.

8.4.5 Radiation treatment plan

Now we discuss typical treatment plans for radiotherapies of tumors in clinics. Usually a dose of 2 Gy per session is administered 5 times per week for a period of 6–8 weeks. Thus the total dose is 60 to 80 Gy. Between fractions cancerous cells have a chance of self-***r**epairing* DNA damage, ***r**eoxygenation* of the tumor cells, ***r**edistribution* within the cell cycle, and ***r**epopulation* of cells by cell replication. These are the four “R’s” in radiation biology and oncology that need to be considered. But most important is the fact that normal cells circumstantially affected by the radiation also have time for repair between fractions. At the low dose of a single session radiotherapy has the effect of *growth delay* but not of permanent tumor recession or cure as seen in Fig. 8.10.

The growth delay increases with increasing dose. Furthermore, cancer cells lose their ability to survive. Figure 8.11 (a) shows the tumor control fraction as a function of dose. At a fraction of 50 %, half of the cancer cells have died. This point is reached

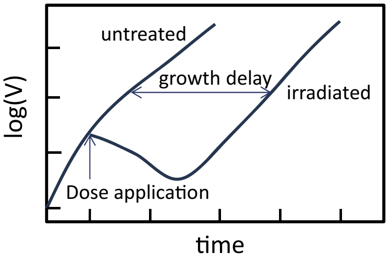


Fig. 8.10: Tumor volume *V* on a log scale is plotted versus time. At the onset of irradiation, tumor volume recesses but can grow again after some delay time. The growth delay time depends on the dose.

after an accumulated dose of about 50 to 60 Gy. Below this dose cancer cells have a chance of self-repair and repopulation as already mentioned. With increasing dose the chances are high to hit the stem cancer cells in particular, as those are responsible for metastasizing normal cells.

Normal cells have a higher radiation resistance than cancer cells. Permanent damage of normal cells occurs at a higher dose, as indicated by the red survival curve in panel (a) of Fig. 8.11. Under optimal conditions of a radiation treatment plan, on average more than 90 % of cancer cells are killed with a 5 % hit rate of normal cells causing severe side effects.

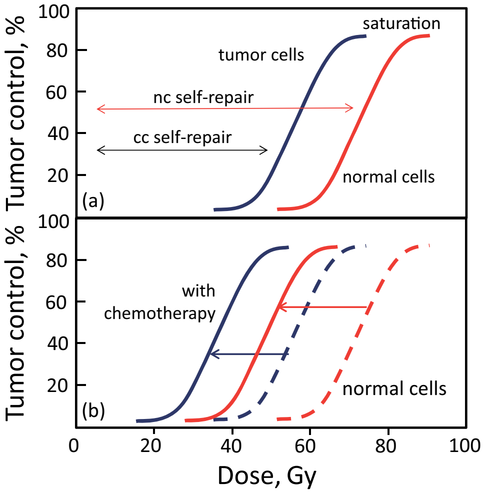


Fig. 8.11: (a) Tumor control versus dose for normal cells (nc) and cancer cells (cc). (b) With chemotherapy the radiation sensitivity increases and the tumor control shifts to lower dose.

8.4.6 Hypoxia and chemotherapy

Solid lumps of cancer cells are percolated by blood vessels for nutrition and oxygen supply. However, the penetration of blood vessels during angiogenesis is not a well-organized process. The supply of oxygen is intermittent and the periphery remains oxygen deficient as illustrated in Fig. 8.12. Those oxygen deficient hypoxic or even anoxic cancer cells turn out to be more radiation resistant than cells with sufficient oxygen supply. Since the early recognition of this effect, there has been speculation about its cause. Meanwhile research has affirmed that the presence of oxygen during irradiation increases the concentration of free and toxic radicals via the indirect action of radiation (see Fig. 8.7). For instance, free photoelectrons can combine with O_2 to yield O_2^-, which is a strong oxidizing agent with high reactivity towards DNA. The combined effect of SSB at lower LET by direct radiation action and oxygen reactivity

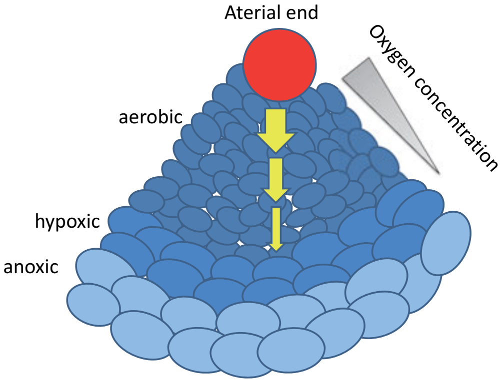


Fig. 8.12: Oxygen distribution in a lump of cancer tissue. The oxygen supply decreases with distance from blood vessels.

via indirect radiation action reduces the tumor survival fraction. The difference is appreciable: a survival fraction of 0.01 requires 10 Gy in oxic cells, but 28 Gy in hypoxic cells, which is almost a factor of 3. During fractionated radiation exposure some hypoxic cells may turn into oxic cells, such that in the end all cells are killed. This is an important point to consider, which temporarily changes the shape of the survival fraction versus dose and determines the treatment result with radiotherapy. Before radiotherapy is applied, cancer tissue can be exposed to higher oxygen pressures and additional radiation sensitizer can be administered. One of them is chemotherapy which enhances the radiation sensitivity of the hypoxic cells and shifts the tumor control to lower dose, as indicated in Fig. 8.11 (b). This shows that for low LET radiation the OER is high, in contrast to high LET radiation with no visible enhancement effect via chemotherapy and therefore an OER = 1. An overview on hypoxia and radiation therapy is provided in [8].

The combination of radiotherapy and chemotherapy shows a number of further synergetic effects. Repopulation is suppressed, i.e., the recovery from a sublethal dose is reduced. Also noticed is an accumulation of tumor cells in the more radiation sensitive G_2 and M phase. Chemotherapy not only improves blood circulation, but also perfusion of other chemosensitizers and therefore improves the overall radiation sensitivity. This allows a broader therapeutic bandwidth of fractionated and total dose.

Radiation therapy is like the fire service that arrives after the house is already burning. Present and future developments concentrate on an early diagnosis of cancer on a molecular and cellular level.

8.5 Summary

1. The life cycle of a cell can be subdivided into interphase and mitosis.
2. Cell division takes place during mitosis.
3. Mitosis is subdivided into prophase, metaphase, anaphase, and telophase.

4. Tumors have six stages of development: (1) mutation, (2) extension, (3) attack, (4) penetration, (5) angiogenesis, (6) invasion and distribution.
5. Cancer is defined as uncontrolled cell division and unlimited cell life.
6. Tumor is an abnormal mass of tissue that arises from cells of pre-existent tissue, and serves no useful purpose.
7. Normal cells require a higher dose of radiation for complete apoptosis than cancer cells.
8. Radiosensitivity is highest during the M phase, whereas the late S phase is the most radiation resistant phase. The survival probabilities are different by a factor of 2.5.
9. Fractionation is a radiation treatment plan with the total dose broken down into smaller portions administered in a sequence of time intervals.
10. Cells are affected by radiation through direct action (bond breaking) and indirect action (increased toxicity).
11. The most effective LET is reached with heavy particles like protons, neutrons or α-particles.
12. The four “R’s” in radiation biology are: Repairing, Reoxygenation, Redistribution, and Repopulation.
13. Oxygen deficient hypoxic and anoxic cancer cells are more radiation resistant than cells with sufficient oxygen supply.
14. Chemotherapy enhances the radiation sensitivity of hypoxic cells and shifts the tumor control to lower doses.
15. Chemotherapy allows a broader therapeutic bandwidth of fractionated and total dose.
16. The oxygen enhancement ratio (OER) is defined as the radiation dose required for a certain survival fraction without oxygen and the dose required for the same survival fraction with oxygen.
17. Radiation treatment with high LET radiation is independent of the oxygen level of cancer cells. For high LET radiation the OER is one.

References

[1] Campbell NA, Reece JB. Biology. 9th edition. Benjamin Cummings; 2009.
[2] Kumar V, Robbins S, Zhai Q, Chen J. Textbook of pathology. Peking University Medical Press; 2009.
[3] Edlund K, Larsson O, Ameur A, Bunikis I, Gyllensten U, Leroy B, Sundström M, Micke P, Botling J, Soussi T. Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultradeep sequencing of human tumors. 2012; 109: 9551–9556.
[4] Wang G, Zhu X, Hood L, Ao P. From phage lambda to human cancer: endogenous molecular-cellular network hypothesis. Quantitative Biology. 2013; 1: 32–49.
[5] Ao P, Galas D, Hood L, Zhu XM. Cancer as robust intrinsic state of endogenous molecular-cellular network shaped by evolution. Medical Hypotheses. 2008; 70: 678–684.
[6] Hall EJ, Giaccia AJ. Radiobiology for the radiologist. 7th edition. Lippincott Williams & Wilkins, Wolters Kluwer; 2012.
[7] Poon IKH, Hulett MD, Parish CR. Molecular mechanisms of late apoptotic/necrotic cell clearance. Cell Death and Differentiation. 2010; 17: 381–397.
[8] Rockwell S, Dobrucki IT, Kim EY, Marrison ST, Vu VT. Hypoxia and radiation therapy: Past history, ongoing research, and future promise. Curr Mol Med. 2009; 9: 442–458.

Further reading

Alberts BE, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, Walter P. Molecular biology of the cell. 6th edition. Garland Science; 2014.

International Atomic Energy Agency (IAEA). Radiation biology: A handbook for teachers and students. Training Course Series No. 42. Vienna: International Atomic Energy Agency (IAEA); 2010.

Weinberg RA. The biology of cancer. 2nd edition, Garland Science, Taylor & Francis Group; 2013.

Kerr DJ, Haller DG, van de Velde CJH, Baumann M, editors. Oxford textbook of oncology. 3rd edition. Oxford University Press; 2016

Podgorsak EB, technical editor. Radiation oncology physics: a handbook for teachers and students. Vienna: International Atomic Energy Agency; 2005.

Useful websites

www.cancerquest.org/cancer-cell-division.html

http://wiki.cancer.org.au/oncologyformedicalstudents/Clinical_Oncology_for_Medical_Students#_ga=1.264882507.554164266.1468659718

http://aboutcancer.com/introduction_page.htm

9 X-ray radiotherapy

9.1 Introduction

In this and in the following chapters we discuss a variety of cancer treatment procedures by physical methods using different kinds of radiation summed up under the term *radiation therapy* or in short *radiotherapy* (RT). Radiotherapy is one branch of radiology, the other one is *radiography dealing with* diagnostics. The treatment of tumors is always performed by a multimodality management plan, which includes radiotherapy, medication treatment (chemotherapy) and possibly surgery before or after radiation exposure. Moreover, the management plan has to consider the specific organ and its functionality that is to be preserved as far as possible, while surrounding organs and in particular the spinal cord have to be protected from radiation. These medical and oncological aspects are not discussed here in any further detail, we concentrate our discussion on radiotherapeutic aspects. Radiotherapy is intended to deposit a lethal dose of radiation to a confined volume of cancerous tissue in the body, while keeping radiation damage to surrounding healthy tissue as minimal as possible. How this is achieved is the central topic of this chapter.

X-ray radiotherapy (XRT) has a long history. After the first excitement about the imaging capabilities of penetrating x-rays discovered by Röntgen in 1895, it soon became clear that radiation can also heal by causing damage. Radiation was used as early as 1899 to successfully treat squamous cell carcinoma of nose and cheek. But radiotherapy as a cancer treatment modality has taken off only during the last 50 years. The early applications of XRT were performed with x-ray tubes and with radioactive sources. Linear accelerators (linac) were developed for high energy particle physics in the 1930s and became available for XRT in the 1960s to reduce complications that occurred with the use of ^{60}Co γ-sources. A modern version of such a linac-based cancer treatment gantry, boosting the x-ray energies into the MeV region, is shown schematically in Fig. 9.1.

Starting in the 1970s, lead blocks were used to define a treatment window and to reduce the dose to normal tissue. In the 1980s, multileaf collimators were invented leading to a 3D conformal therapy. Later in the 1990s CT imaging was combined with increasing computer power for establishing 3D treatment plans. This includes computerized intensity modulated radiation therapy (IMRT), an RT that delivers variable beam intensity throughout the irradiated volume from multiple beam angles. Concurrently new imaging technologies (MRI, CT, US) improved localization and definition of tumor volumes. More recent developments of XRT advance to smaller subfields of the tumor volume and to higher resolution IMRT. Of particular interest is the combination of robotics, newest accelerator technologies, and patient tracking systems for high precision RT with reduced side effects, advancements that are presented in Section 9.7.

DOI 10.1515/9783110553116-011

Fig. 9.1: Modern medical linac used for cancer treatment. The gantry housing a linac, collimators, and shutters circles around the patient at the center for irradiating tumors with high energy x-ray photons incident from different angles (courtesy Varian Medical Systems International AG, Switzerland; all rights reserved).

9.2 Radiation treatment planning

Any cancer treatment requires careful and precise planning beforehand. This is not specific to XRT but applies to any RT. In the first step a multidisciplinary team consisting of oncologists, surgeons, and radiotherapists work out a treatment protocol after the tumor has been localized and characterized. This also includes an assessment of the cancer stage and whether the treatment is intended to be curative or palliative. Also a decision has to be made whether RT should be applied before surgery or after surgery. Then the patient is called again for precise localization of the tumor in the body, and determining its volume, shape, and electron density by one or several imaging modalities: MRI, CT, PET, or US. CT is the preferred imaging modality as it provides accurate pictures with a gray scale that depends on the electron density, which is the required input data for dosimetry calculations. MRI is more sensitive to soft tissue than CT, and US provides additional information on superficial tumor masses not detected by the other imaging methods. PET is a powerful but expensive imaging modality that is particularly sensitive to malignant cells. If combined with CT, tumors can be precisely identified with respect to location and shape, thus improving the definition of the tumor delineation, i.e., the gross tumor volume (GTV) as seen for instance in Fig. 7.9. Input data from all imaging modalities are integrated into a computer simulation program and used for calculating a 3D image of the tumor in the coordinate system of the body.

The information from 3D imaging of the cancer volume is then used for defining a radiation plan by considering the depth of the tumor in the body and the dose to be delivered. Once the dose to volume has been accurately calculated, the patient goes to a simulator to determine the location of the tumor with respect to the labo-

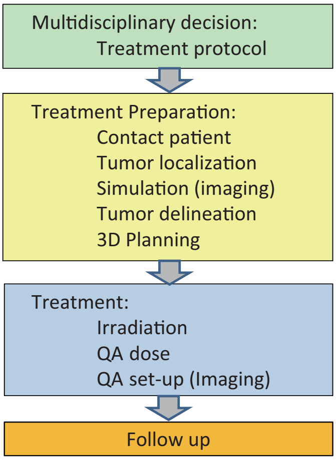


Fig. 9.2: Flow chart for radiation treatment of cancer.

ratory frame. The simulator is a setup identical in dimensions to the one used for XRT but furnished only with x-ray imaging equipment. The coordinates are determined by taking a series of x-ray radiographs to make sure that the tumor is in the correct position relative to treatment couch and gantry after the patient has been immobilized. The third step is the actual radiation treatment according to the radiation plan. Quality assurance (QA) during irradiation is required to avoid systematic errors, such as tumor movement. After irradiation the tumor size is checked again with different imaging modalities. In case of local recurrence, a radiation boost is administered within a decreased target volume. Further follow-up checks conclude the treatment. A flow chart of the different steps is shown in Fig. 9.2. Note that this describes only solid types of tumor. Hematological cancer types affecting blood, bone marrow, and lymph nodes require different treatment plans. As alredy mentioned, treatment planning applies not only to XRT but also to the other RTs presented in Chapters 10–12.

9.3 Absorbed dose by high energy photon beam

Before continuing the discussion of cancer treatment via XRT we first need to develop an understanding of the principal physical processes that occur when high energy photons hit a soft matter target. More precisely, for determining the dose to a volume of cancerous tissue we need to know in detail the x-ray interaction with matter as well as the dose deposited as a function of depth. Part of this has already been discussed for lower energy x-ray radiography in Chapters 1, 3, and 4. Here we focus on very hard MeV x-rays and γ-radiation used for radiotherapy.

9.3.1 Depth-dose profiles

Table 9.1 lists the three contributions of photon interaction with matter as a function of energy: photoelectric effect (PE), Compton scattering (CS), and pair production (PP). For a discussion of these effects we refer to Fig. 3.8 and adjoining text. At high photon energies beyond 4 MeV, which are relevant for XRT, Compton scattering, and pair production are the dominant interaction processes. The energy transferred by inelastic scattering of photons to CS electrons and to PP electrons is the dose-defining radiation that causes damage of tumor cells. It is important to realize that it is not photons that are responsible for LET and RBE of tumor cells, but hot electrons.

Tab. 9.1: Relative weight of interactions for photons penetrating water from low energies up to high energy incident photon energies. PE = photoelectric, CS = Compton scattering, PP = pair production (data reproduced from [1]).

Photon energy [MeV]	Relative weight of interactions [%]		
	PE	CS	PP
0.01	95	5	0
0.025	50	50	0
0.06	7	93	0
0.150	0	100	0
4	0	94	6
10	0	77	23
24	0	50	50

At 10 MeV incident photon energy, 77 % of the scattering events are due to CS. Assuming that the photon scattering angle is in the order of 10°, the energy transfer to Compton electrons is about 23 % or 2.3 MeV. According to Fig. 9.3 these electrons have a range of about 1 cm in water. For a photon energy of 24 MeV the energy transfer to Compton electrons under the same condition increases to 50 %, producing 10.6 MeV electrons with a range of about 5 cm in water.

Range and dose of electrons can be measured with experiments that use an electron accelerator with tunable energy and water as a phantom target [2, 3]. Inside the water tank, detectors can be easily positioned and shifted. The setup is schematically shown in Fig. 9.3 (a). The body has an electron density similar to water. Therefore simulations with water reflect conditions of soft tissue rather well. The exit slit of the accelerator is placed at a standard 100 cm source-to-surface distance (SSD), which is a typical distance between x-ray source and patient used in clinics. Nevertheless, simulation of electron depth-dose with this experimental arrangement is different from the real situation, where hot Compton electrons are generated within the tissue and without crossing an air gap. On the other hand, the stopping power of air is rather low compared to water, so that the 100 cm air gap causes little error in the estimate of Compton electron range in tissue. According to Fig. 9.4 the dose-depth profile of electrons in water features a rapid increase below the surface, an extended plateau region and a rapid drop-off, defining a range. Plateau and range depend approximately linearly on the incident electron energy. See also Chapter 3 for a discussion of the range of charged particles.

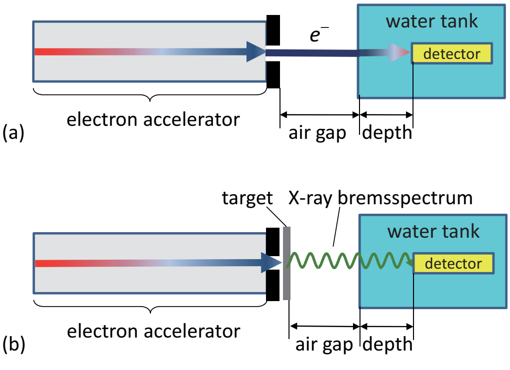


Fig. 9.3: Experimental arrangement for determining the dose versus depth for electrons in water (a) and for x-ray photons in water (b). Under standard source-to-surface distance (SSD) conditions the air gap is 100 cm.

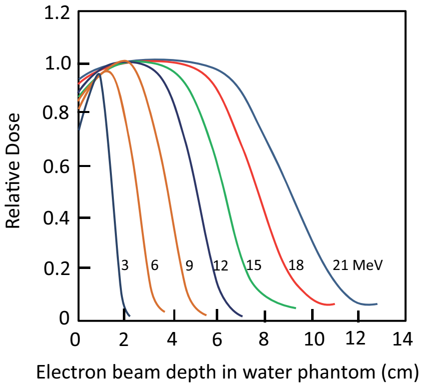


Fig. 9.4: Relative dose as a function of penetration depth of high energy electron beam in water at a standard 100 cm source-to-surface distance (SSD) (adapted from [2] and [3]).

Next we consider the dose-depth relationship for a photon beam. The experimental condition for determining the dose-depth profile of x-rays is similar to that for electrons and is schematically shown in Fig. 9.3 (b). Electrons are first accelerated to high energies as in the previous case. Then the electrons are converted to x-rays by bombarding a target that generates a bremsstrahlung spectrum. Since the bremsstrahlung spectrum has a broad photon energy distribution, it is not the photon energy that is quoted in respective graphs but the accelerating potential which determines the cutoff energy of the photon spectrum. In Fig. 9.5 (a) the *percent photon dose* (PPD) versus depth is plotted for two bremsstrahlung spectra generated at 6 MV and 15 MV. PPD is a relative dose normalized by the maximum dose at a certain distance below the sur-

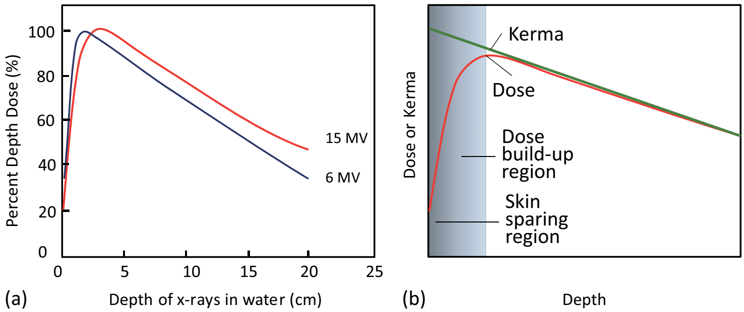


Fig. 9.5: (a) Percent depth-dose curve for photon beams in water under standard conditions of 100 cm source-to-surface distance. Note that photon energy is given in terms of accelerating potential (MV), which determines the cut-off energy of the bremsstrahlung spectrum. (b) The region of dose build-up.

face. The curves are characterized by a steep build-up of dose below the surface, a maximum depending only weakly on the energy, and a slow almost linear drop-off beyond the maximum.

From Fig. 9.5 (a) we recognize that the maximum dose is reached at a depth of 0.5 to 3 cm below the surface, depending on the x-ray energy. Near the surface the dose is very low, which is known as *skin sparing effect*. Beyond the maximum the dose drops off almost linearly. The initial build-up of dose below the surface can be explained as follows. When the x-ray beam enters the phantom or the body, it starts generating Compton electrons and electrons/positrons from pair production. Those particles travel a certain distance in the forward direction before being stopped and depositing their energy by producing further secondary photoelectrons. As an example we consider a 15 MV x-ray bremsstrahlung spectrum. The average photon energy is 2/3 of the maximum energy or 10 MeV. We have already learned that 10 MeV photons produce 2.3 MeV Compton electrons at a scattering angle of 10°, which are stopped within a range of 1 cm. So it is natural that the maximum dose occurs at the respective depth below the surface. Beyond the maximum the dose drops off because the intensity of the x-ray beam decreases and consequently also the number of Compton electrons produced per distance traveled. Dose build-up and skin sparing region are shown again on an enlarged scale in panel (b) of Fig. 9.5.

In Fig. 9.5 (b) kerma as a function of depth is plotted. Kerma is defined as the energy fluence of an x-ray beam converted into kinetic energy of charged particles (see Chapter 4). According to this definition, kerma starts right away as soon as x-ray photons interact with matter and photon energy is converted to Compton electrons and electrons-positrons pairs. Those hot electrons/positrons travel a certain distance before depositing their energy and contributing to the dose, as explained before. Hence, kerma is high from the very beginning without build-up region, followed by a continuous drop-off with diminishing x-ray intensity.

9.3.2 Dose to isocenter

When calculating the dose to a tumor volume two different methods are commonly used as illustrated in Fig. 9.6 [3, 4]. The first one, known as *constant source-to-surface distance (SSD)* assumes a constant distance from the source to the surface of the patient. The second method is the isocentric setup with a constant *source-axis distance* (SAD). The center of the target volume is placed at the machine's isocenter such that the distance to the target point is kept constant for all beams at different angles. For radiotherapy, the SAD technique is easier to implement as no readjustment of the patient in relation to the radiation source is required. On the other hand, under SSD conditions the dose to volume is easier to simulate for establishing a treatment plan. SSD implies, however, that the isocenter is on the skin and therefore the treatment couch together with the patient has to be moved laterally and vertically with respect to the source in order to meet this condition.

In the following we work out a simplified example for the calculation of dose to the isocenter assuming SAD conditions. The dose provided by linacs is calibrated and quoted in Gy per monitor unit (Gy/MU), where 1 MU = 1 cGy. We assume a three field exposure from anterior and posterior with 100 MU from either side using a 10 MeV photon beam. According to Fig. 9.5 (a) the PDD at a distance of 10 cm from skin to isocenter is 73 % for the anterior field. As there is no further obstruction in this field, the weighing factor is taken as 1. After exposure of 100 MU from anterior the dose at the isocenter is 73 cGy. For the posterior field the distance from skin to isocenter is longer and the PDD is consequently lower, assumed to be 47 %. An additional weighing factor of 0.7 takes into account the smaller posterior field, widening of the beam, and bones in the path with a higher absorption coefficient than soft tissue. The dose delivered to the isocenter from posterior is then $2 \times 0.47 \times 0.7 \times 100$ cGy $= 2 \times 33$ cGy $= 66$ cGy. This is in fact lower than expected. If the treatment plan requires that 70 % of the anterior side should be delivered from either posterior side, then the dose should be $2 \times 51.1 = 102.2$ cGy instead of 66 cGy. In order to reach this level, the MU must be boosted for each of the posterior beams by a factor $51.1/33 = 1.55$, i.e., 155 cGy. Then the total dose delivered to the isopoint is 73 cGy + 102.2 cGy $= 175.2$ cGy. Assuming that the prescribed dose is 200 cGy at the isocenter, all three field exposures have to be increased by a factor $200/175 = 1.143$, which is 114 MU from anterior and 2×177 MU from posterior. All numbers are presented in Tab. 9.2.

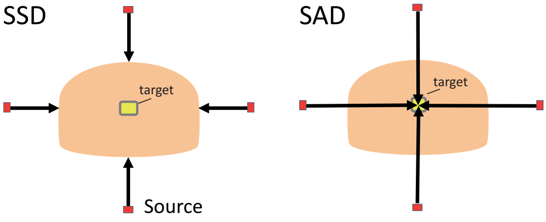


Fig. 9.6: Two different methods for planning radiotherapy. Left: constant source-surface distance (SSD). Right: constant source-axis distance (SAD).

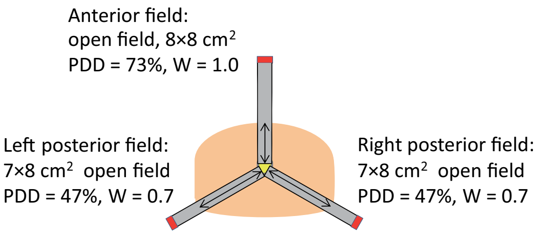


Fig. 9.7: Three field exposure from anterior and posterior for dose calculation in SAD geometry.

Tab. 9.2: Conversion of dose to monitor units for reaching defined total percent photon dose (PDD) at the isocenter (IC).

Field	Starting MU [cGy]	PDD at IC [cGy]	Weight factor	Weighted PDD at IC [cGy]	Boosted PDD at IC [cGy]	Rescaled MU [cGy]	Rescaled MU for 200 cGy at IC [cGy]
Anterior	100	73	1.0	73	73	100	114
Left post.	100	47	0.7	33	51.1	155	177
Right post.	100	47	0.7	33	51.1	155	177

9.4 Target volume

The first task in radiation treatment planning is the precise localization of the tumor volume to be irradiated. According to the recommendations of the International Commission on Radiation Units and Measurements (ICRU), Reports 50, 62, and 71 [5] the following distinctions should be made, which are also illustrated in Fig. 9.8:

1. Gross Tumor Volume (GTV) is the macroscopic tumor volume, as defined by imaging methods.
2. Clinical Target Volume (CTV) is a volume of tissue that contains the GTV and/or subclinical malignant disease with a certain chance to become relevant for therapy. CTV is an anatomical and clinical judgement, not a physical measurement. Clinical judgment tries to answer the question whether beyond the clearly defined tumor mass any additional cancerous tissue might be present, how far microscopic branches may reach, what the chances are for local recurrence, and what the growth rate of histological known tumors is. If the tumor has been surgically removed prior to RT, than no GTV can be defined and CTV is the volume of subclinical disease.
3. Planning Target Volume (PTV) is a geometrical concept, introduced for treatment planning and evaluation. It is the recommended volume for determining beam size and dose distribution ensuring an adequate dose is delivered to all parts of the CTV. PTV also includes possible tumor movement due to organ movement inside the body, patient movement on the treatment coach, and machine tolerance. PTV is defined in the reference frame of the treatment room and assuming that it is located in the isocenter of the gantry. Usually the PTV is taken as: PTV = CTV plus a shell of 1 cm thickness.

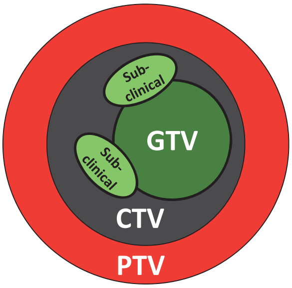


Fig. 9.8: Defining volumes according to the ICRU Report 62. GTV = Gross Tumor Volume; CTV = Clinical Target Volume; PTV = Planning Target Volume.

With the help of image guided radiotherapy (IGRT), the position of the tumor volume during RT can be checked continuously and eventually corrected. Although IGRT adds extra dose to the patient, the overall improvement of the geometric accuracy justifies its use during XRT and may decrease the volume of tissue being irradiated. Figure 9.9 compares a CTV with and without the use of IGRT.

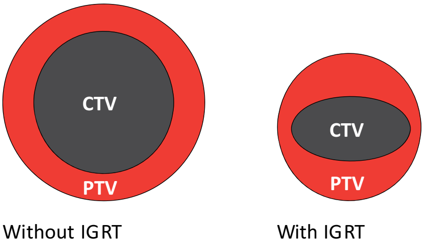


Fig. 9.9: Image guided radiotherapy increases the accuracy of localizing the clinical target volume and thereby helps reduce the planned target volume.

A positional accuracy of 3.5 % is presently the accepted tolerance. But the required accuracy depends on the specific situation. High-dose small-volume RT such as in brain irradiation requires more accuracy and precision than a large or total body irradiation. Here we use the terms *precision* and *accuracy*. High precision implies hitting the same spot with minimal scatter, but the narrow Gaussian distribution may lie outside of the target area. Accuracy implies hitting the target, but eventually with a wide Gaussian distribution about the target's center. High accuracy and high precision implies hitting the center of the target with minimal scatter. To reach this goal in brain tumor irradiation the head is completely immobilized by a stereotactic head frame, as presented later in Section 9.8.

Having defined the aim of RT and volumes to be treated, the next step is the definition of dose to PTV per fraction, number of fractions, and total dose to PTV. This is the topic of the next sections.

9.5 Multileaf collimator and intensity modulated radiotherapy

In the past, beam size and treated volume (TV) were defined by simple rectangular collimator blades between radiation source and target. The collimator consists of two adjustable slits arranged perpendicular to each other. According to the ICRU, this selected TV area should at least receive the absorbed dose determined as the minimum dose to the PTV. According to the schematic in Fig. 9.10, TV is naturally larger than PTV and normal tissue is also exposed to high dose. With such a rectangular slit system, single, double, or multifold exposures can be performed, as shown in Fig. 9.11.

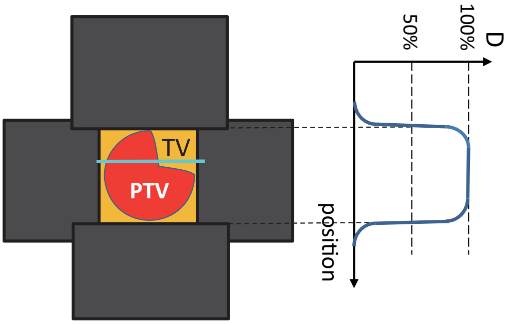


Fig. 9.10: (a) Simple rectangular double slit collimator defining the treated volume (TV) containing the PTV to be exposed; right panel shows a dose profile for the slice indicated by the horizontal line. 50 % dose level defines the beam width.

Note that with a single field exposure the dose on the proximal side of the tumor needs to be more than 100 % of the calculated dose at the tumor position due to the attenuation of the beam in the tissue, see Fig. 9.11 (a). With two exposures from anterior and posterior the dose at the target center needs to be only 50 % in order to add up to 100 % from both sides. Hence the dose on the proximal side of the tumor can be lowered to about 60 % (Fig. 9.11 (b)). Similarly, with a four field exposure the single dose from either side is only 25 % to add up to 100 % at the target center (Fig. 9.11 (c)).

The example in Fig. 9.11 shows that by employing an increasing number of exposures from different sides the dose outside of the target center decreases, while at the tumor position the isodose contour becomes better defined. This is an important aspect for sparing the healthy tissue from radiation dose and side effects. It shows that even with a linear dependence of dose versus depth high dose can be accumulated in a specific location at a certain depth if the exposure is administered from many different angles. In any case, exposures using single or double slit collimators are not sufficiently precise for a high quality RT. Therefore wide-open slit collimators are used only for palliative treatments.

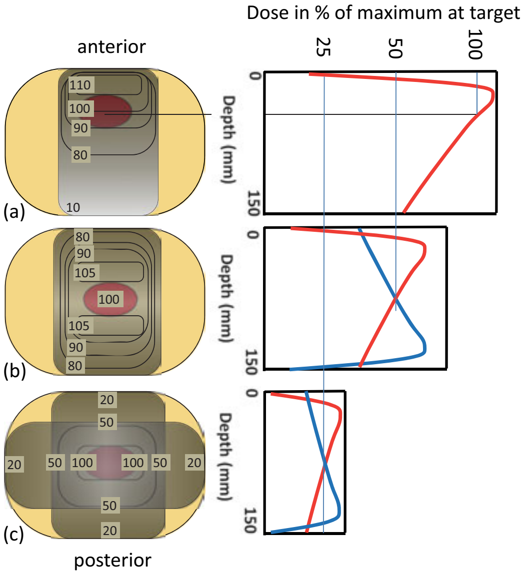


Fig. 9.11: Radiation exposure using a rectangular collimator (shown in Fig. 9.10). (a) Single field exposure from anterior; (b) double field exposure from anterior and posterior; (c) four field exposure. Lines indicate isodose in percent. In the case of four field exposure (anterior, posterior, left, and right) the total dose at the center of the tumor adds up to 100 % with some sparing of healthy tissue at a dose level of 25 %.

To overcome the problems that occur with a double slit, first lead blocks were cast to conform to the individual shape of the PTV. However, multiple lead blocks were required to adapt to the individual PTV seen from different angles. This was an expensive and time consuming procedure. In the 1980s the concept of multileaf collimators (MLC) was introduced consisting of two opposing banks of moveable tungsten sheets (Fig. 9.12) [4]. The higher the number of sheets, the more precisely the conformity can be adjusted. These slits are pneumatically driven and computer controlled. While the gantry moves around the patient, the beam is stepwise opened at specific angular positions and the sliding leaves are adapted to the projected profile of the PTV. This conformal MLC RT spares the surrounding normal tissue and much reduces radiation side effects. An example for MLC applied to the prostate is shown in Fig. 9.13. The MLC is adapted to the shape of the prostate and radiation treatment sparse exposure of colon and rectum.

Having fast reacting, high speed moveable collimator sheets it was only a small step to introduce *intensity modulated radiation therapy* (IMRT, sometimes abbreviated as IMXT). IMRT combines MLCs for defining the PTV and intensity modulation to deposit the required dose at each individual location. The principle is explained in Fig. 9.14. Let us first consider one fixed gantry angle. The exposure time at one fixed angle is subdivided in time intervals Δt_i. During each time interval the collimator slits

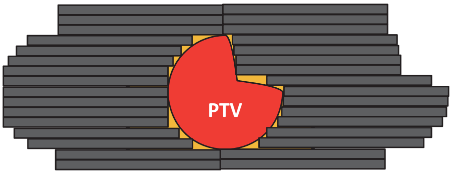


Fig. 9.12: Multileaf collimator that conforms to the actual projected profile of the PTV.

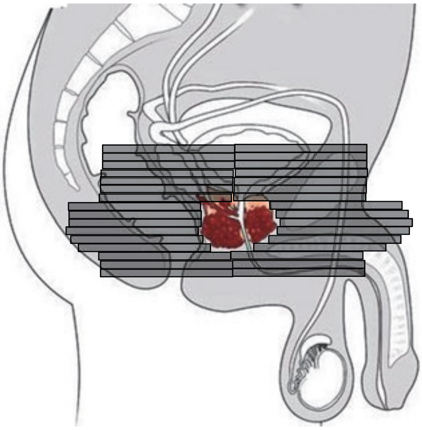


Fig. 9.13: MLC application for prostate radiotherapy with conformal adaptation that spare the colon and rectum.

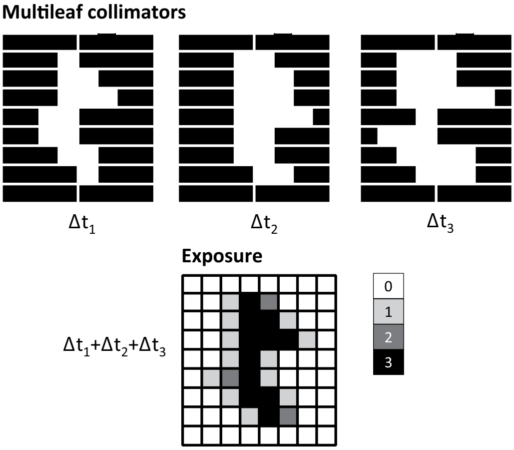


Fig. 9.14: During exposure at a fixed angle of the gantry, MLC open and close during time intervals Δt_i to yield a modulated dose distribution in the PTV. In the upper panel white fields correspond to open slits, black fields to closed slits. In the lower panel white fields correspond to zero exposure, black fields to highest exposure for open slits during all time intervals.

are moved under computer control to yield a laterally modulated intensity distribution. The voxel that is open all the time receives the highest dose, a neighboring voxel that is only open for part of the time gets less radiation, etc. Thus the time modulation of the collimator slits translates into lateral dose modulation, as indicated in Fig. 9.14.

The intensity modulation procedure is repeated at all angular positions of the gantry, as exemplified in Fig. 9.15. The accumulated dose to the tumor volume then corresponds exactly to the simulated radiation treatment plan. Clearly, QA during the whole procedure has to confirm the agreement between anticipated dose and actually delivered dose. With IGRT the proper positioning of the patient can be controlled and eventually corrected. The latter procedure requires two beams: one at lower energies for imaging and another one at higher energies for XRT.

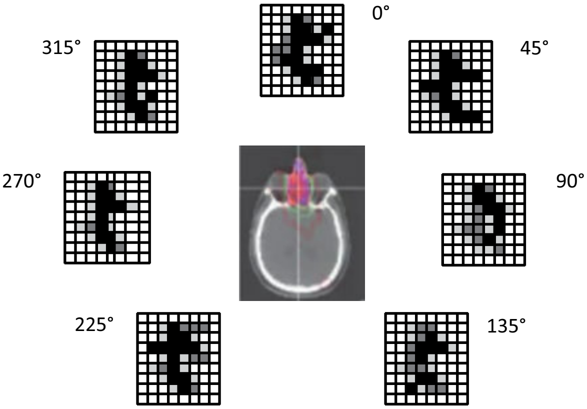


Fig. 9.15: Intensity modulated fields for all gantry angles add up to a precisely defined dose to volume distribution of the tumor volume.

The overall goal of the treatment plan and its execution is the delivery of a homogeneous isodose within the PTV with a sharp contour at the edges and a rapid drop-off of dose beyond. With continuously improving treatment plans supported by advanced imaging methods and 3D simulation programs this goal is achieved nowadays better than ever. Nevertheless, we should keep in mind that x-ray beams precisely delivered to the PTV locally create Compton electrons and electrons from pair production. These have their range, causing radiation damage beyond the well-defined borders of the PTV. This effect is known as *penumbra*. Penumbra can occur for *geometric* reasons due to an extended source size and transmission through slits that do not completely block the beam, see Fig. 5.4. Lateral spreading of dose due to the scattering of the x-ray beam and straggling range of electrons is known as *physical* penumbra. Geometric penumbra can be reduced by well-designed collimators. Physical penumbra is intrinsic. Because of a more precisely defined range, proton radiation therapy discussed in the next chapter features a better defined physical penumbra. Further information about IMRT can be found in [3, 6, 7].

9.6 Linear accelerators for x-ray generation

It should be recalled that x-rays for cancer treatment are much harder (higher energies) than x-rays for imaging. For imaging, electron acceleration energies up to 150 kVp are used. For cancer treatment a typical value in the past was 250 kVp. This was the upper limit of what could be achieved with standard x-ray tubes and static electrical fields. For higher energies, γ-radiation such as a ^{60}Co source with an average energy of 1.3 MeV was used in the past and is still applied for the so called γ-knife discussed in Section 9.8. Radiation therapy has greatly benefitted from developments that took place in high energy physics. Elementary particles (electrons and protons) are being accelerated to very high energies (MeV to TeV) with linear accelerators (linac) using microwave cavity technology. A spin-off of these developments are linacs for generating very hard x-rays in the MeV range for cancer treatment. The advantage of using these very high energy x-rays is a much enhanced LET, therefore higher RBE and shorter exposure times. For more details on linacs, we refer to Section 1.3.2 and [8].

For clinical applications linacs that have a physical length of 0.5 m to 1.5 m producing electron energies between 4 MeV and 25 MeV are used. According to the schematic in Fig. 9.16, electron bunches are generated in a klystron and fed into evacuated cavities where they become accelerated by high power microwaves with frequencies of usually 3 GHz. At the end of these cavities, the electron bunches are guided by a bending magnet to an exit slit, where a 1.5–2 mm wide electron beam hits a tungsten target. Like in a standard x-ray tube, the high energy electrons produce a bremsstrahlung

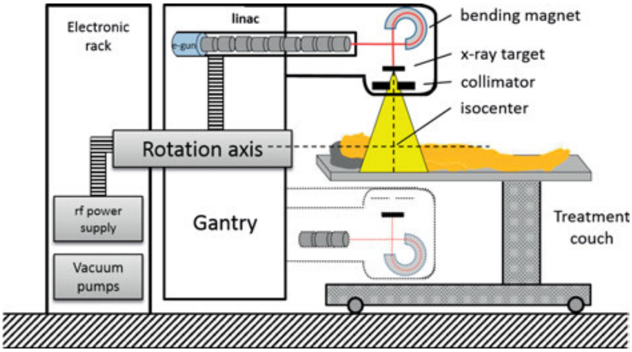


Fig. 9.16: Medical linac used for cancer treatment. The gantry houses a linac, bending magnet for beam guidance, collimators, and shutters. It circles around the isocenter where the patient's tumor is positioned by adjusting the treatment couch.

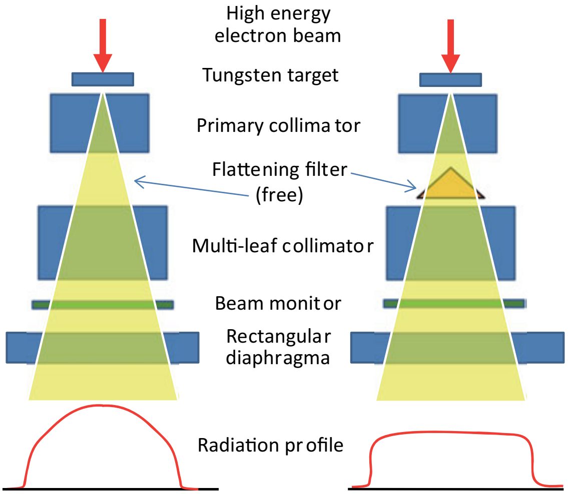


Fig. 9.17: X-ray radiation profile with and without flattening filter.

spectrum of hard x-ray radiation with cut-off energy according to the maximum energy of the electrons.

The exit x-ray beam is roughly confined by tungsten “jaws” and further refined by MLCs, as described in the previous paragraph. Electron gun, accelerator, target and collimators are housed in a gantry (see also Fig. 9.1). Since the gantry is rather bulky, rotation is only possible in a coplanar plane around one horizontal axis and all beams meet at the isocenter. The patient on the treatment couch is moved such that the cancer target coincides with the isocenter. Two positions for radiation treatment, anterior and posterior, are indicated in Fig. 9.16.

The energy distribution of the x-ray beam is very broad, which is typical for a bremsstrahlung spectrum. Furthermore, the intensity is often inhomogeneously distributed in the lateral direction due to x-ray scattering processes in the tungsten target. To improve beam homogeneity, a flattening filter (FF) is inserted after the tungsten target (Fig. 9.17). If the flattening filter is not inserted, the treatment is called flattening filter free (FFF) exposure. Furthermore, the x-ray intensity is monitored by a dual ionization chamber before the beam exits through the MLC to the target. The monitor is calibrated with a phantom of similar density to body tissue. The dose is then provided in monitor units (1 MU = 0.01 Gy). Alternatively, the tungsten x-ray target can be removed and the electron beam will directly be funneled to the target for treatment. For this purpose the 1.5–2 mm wide electron beam is widened by inserting a scattering foil instead of the tungsten target. Note from the discussion of Fig. 9.4 that the range of electrons is much shorter than for x-rays, there is no skin sparing effect, and the buildup of dose is minimal. Therefore treatment with an electron beam is only justified for near surface exposure, for instance for skin cancer. However, skin cancer is usually treated either by brachytherapy (Chapter 12) or by laser exposure (Chapter 13).

In recent years it has been shown that the high energy photons generated in clinical linear accelerators also produce photon induced neutrons by (n,γ)-reactions. Whenever photons with energies higher than the binding energy of neutrons hit a target, neutrons can be release. As the binding energy of neutrons in nuclei is in the order of 4 to 8 MeV, this is a phenomenon encountered by medical linacs. The neutrons generated either penetrate the head of the gantry, or they are captured by other materials. In the latter case, neuron activation occurs, which is usually followed by β^--de-excitation and prompt γ-emission. For linacs operating at energies above 10 MeV, a significant neutron production rate has been observed and after turning off linacs still an afterglow of radiation is detected consisting of fast neutron background and gamma spectral lines. Therefore it is necessary to take extra radiation protection measures to shield patients and clinical staff from exposure to this radiation.

9.7 Cyberknife technology

In recent years a novel and revolutionary invention in clinical radiation therapy has taken place, which is known as cyberknife (CK). It overcomes a number of problems of more conventional linacs for RT by combining three major technology advancements. These are: (1) precision robotics, (2) image guiding system, and (3) breathing compensation. The high precision of the CK allows treatment with much higher submillimeter definition of tumor areas in the brain, spine, lung, liver, kidneys, and prostate, which reduces the PTV and decreases the risk from exposure of healthy tissue.

CK technology is based on the same principle of high energy x-ray photon production via a linac as we have discussed in the previous section. However, the linac is operated at higher frequencies of 9.3 GHz instead of the 3 GHz. This produces a much steeper field gradient, allowing shortening the microwave cavities. Thus the CK linac producing 4–6 MeV electrons is shorter (0.5 m to 1 m) and more lightweight (≈ 150 kg) than standard linacs for RT. At the same time, the lower photon energy reduces the risk of photon-neutron production. The short linac is placed on a robotic arm with six degrees of freedom, as is common for industry robots, with a positional accuracy of 0.2 mm. A second arm carries the treatment couch. The positioning of the accelerator and angle of exposure is much more flexible than in a standard gantry, which has only one coplanar plane of rotation. Instead the CK can deliver beams from thousands of non-coplanar, isocentric or non-isocentric angles.

The image guiding system is a computer controlled position tracking system, using an infrared camera in combination with x-ray imaging. Both together monitor patient movement and compensate for small deviations from the original plan. Therefore the patient does not need to be immobilized as is necessary for standard XRT equipment. The synchrony camera (Fig. 9.18) is an infrared camera that follows LED markers placed on the body and compares their position with gold markers inside the body, which are tracked by two x-ray sources. X-rays from both tubes are oriented at

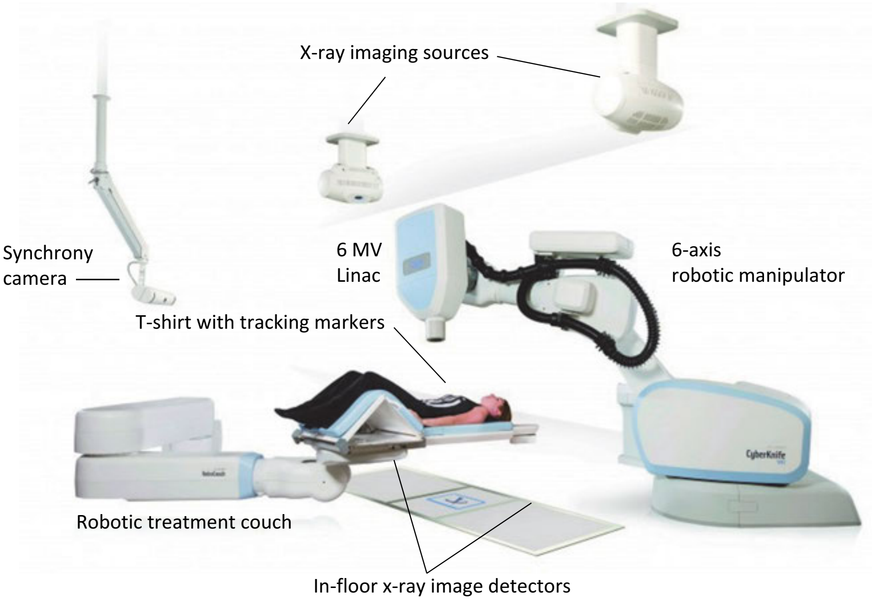


Fig. 9.18: Cyberknife treatment facility (adapted from ACCURAY www.accuray.com/ and Wikipedia, © Creative Ccommons).

right angles and meet at the center of the target. X-ray images therefore deliver a stereoscopic picture which is matched with the CT images taken during the planning and simulation phase. Reference points are compared and any deviations are corrected for by robot repositioning. With this tracking system the spatial accuracy of target location is within 0.6 mm.

Moreover, technical innovations allow treatment of tumors that move while breathing, such as tumors of the lung. Also liver and kidneys move while breathing up to several millimeters, which constitutes a big problem in standard RT. Due to the movement, the PTV has to be much larger than the GTV. However, CK technology overcomes this problem by an intelligent image tracking system as described above. The external breathing motion is recorded by an infrared camera and balanced with the internal tumor position as seen by x-ray imaging using pre-implanted gold markers. The robot arm follows the breathing pattern and delivers the radiation precisely to the target. Raytracing procedures and Monte Carlo simulations complement the optimized radiation treatment, which typically takes 30–120 minutes. Thus CK treatment is a form of highly focused and position tracking radiotherapy also named *stereotactic radiotherapy* and also known as stereotactic radiosurgery (SRS). Obviously the radiosurgery here is noninvasive.

Treatments with CK have excellent tumor coverage, steep dose gradients and tight dose conformity independent of the target shape. The system tracking capability eliminates the need for gating techniques otherwise required for following periodic patient motion like breathing or heartbeat. It also makes breath holding techniques and restrictive headframes obsolete, providing greater comfort for the patient during XRT. CK treatment facilities are now becoming available at various clinics and they will likely replace traditional XRT facilities in the long run. In [9] cyberknife technology for prostate cancer treatment is reviewed.

9.8 Gamma knife

Gamma knife is an RT using ^{60}Co isotopes emitting γ-radiation of about 1.3 MeV photon energy. ^{60}Co isotopes are produced by neutron capture of ^{59}Co via the reaction ^{59}Co(n,γ)^{60}Co with a lifetime of 5.26 years. The decay via β^- and prompt γ-radiation consists of two γ-lines emitting at 1.173 MeV and 1.33 MeV. More information concerning ^{60}Co isotopes can be found in Section 12.2.3. Gamma knife is specifically designed for RT of tumors in the brain. It cannot be used for other purposes. Similar to XRT, γ-RT does not have a well-defined range, but the exposure from many different angles deposits a highest dose at the target center. Outside of the target edge, the radiation rapidly drops off with an insignificant dose to adjacent normal tissue. For precisely delivering the dose to the target volume, the head of the patient has to be fixed in a stereotactic frame. This ensures that the head position is reproducible between treatment planning and during exposure. After the head is fixed, it is covered with a stainless steel helmet containing a large number of boreholes. The holes are plugged up with cylindrical tubes accepting the ^{60}Co isotopes. Figure 9.19 shows schematically a stereotactic frame with helmet and gamma radiation emanating from different sides to a target at the isocenter. The number of plugs filled determines the dose at the target position. The patient is moved on a hydraulic couch into a radiation unit and irradiation can start. γ-knife is the only RT modality where γ-rays from many different sides all simultaneously hit the target. All other irradiation procedures have to move the source sequentially from one position to the other. This saves some treatment time, which is beneficial for the patient constrained in an uncomfortable stereotactic head frame, but is insignificant with respect to cell cycle time.

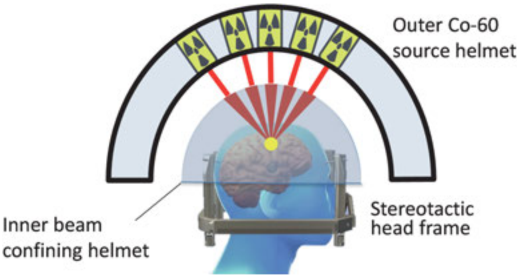


Fig. 9.19: Stereotactic headframe for gamma knife treatment of brain tumors.

Summarizing, γ-knife RT is used for treatment of tumors in the brain. In the past this method was justified by the fact that hard x-rays of about 250 kVp do not have sufficient intensity after penetrating the skull and not sufficient LET to damage tumors. However, this problem has been overcome with the generation of very hard x-rays via linacs providing x-rays beyond the γ-energy of ^{60}Co. The second requirement of precisely positioning the head and suppression of any movement is fulfilled by the target tracking possibility using cyberknife technology. Furthermore, larger tumors can rather be treated with the cyberknife than with the γ-knife since the focus can be moved, which is not possible with the stereotactic frame of a γ-knife. For these reasons γ-knife RT may no longer be necessary in future brain radiation therapy. This would also circumvent handling of radioactive isotopes, starting from production of isotopes via neutron irradiation, safeguarding highly radioactive material during use, and finally depositing low radioactive nuclear waste. An overview on the use of γ-knife treatment in neurosurgery is given in [10].

9.9 Summary

1. Any cancer treatment requires careful and precise planning beforehand.
2. Quality assurance (QA) during irradiation is required to avoid systematic errors.
3. At 10 MeV incident photon energy about 80 % of the scattering events are due to Compton scattering, the rest is due to pair production.
4. The dose-depth profiles of electrons and photons in soft matter are very different.
5. High energy electrons feature a plateau and a range with respect to particle intensity.
6. Photon beams do not have a range, but their intensity diminishes with penetration depth due to Compton scattering and pair production.
7. For calculating dose to a tumor volume two different methods are commonly used: *constant source-to-surface distance (SSD)* and a constant *source-axis distance* (SAD).
8. For defining the target volume, three different volumes are distinguished: Gross Tumor Volume (GTV), Clinical Target Volume (CTV), and Planning Target Volume (PTV).
9. Multileaf collimators (MLC) consist of two opposing banks of moveable tungsten sheets that can be adapted to the shape of the tumor volume.
10. *Intensity Modulated Radiation Therapy* (IMRT) is a combination of MLC and intensity modulation for each individual angle of radiation exposure.
11. Penumbra can occur for *geometric* reasons due to an extended source size and transmission through slits, and for *intrinsic* physical reasons due to the range and straggling of particles in tissue.
12. Linacs are used to boost the energy of electrons into the MeV range, which is not possible with standard x-ray tubes.
13. High energy electrons accelerated in a linac hit a tungsten target and produce very hard bremsstrahlung x-ray photons.
14. The advantage of using very high energy x-rays is a much enhanced LET, resulting in higher RBE and shorter exposure times.
15. To improve beam homogeneity, a flattening filter (FF) is inserted after the tungsten target.

16. Cyberknife combines three major advancements in technology and overcomes a number of problems of more conventional linacs for radiation therapy: (1) precision robotics, (2) image guiding system, and (3) breathing compensation.
17. Gamma knife is a radiation therapy using γ-radiation from a ^{60}Co isotope source, emitting 1.3 MeV photons.
18. Gamma knife is specifically designed for RT of neurosurgical disorders in the brain.

References

[1] Johns HE, Cunningham JR. The physics of radiology. 4th edition. Springfield, IL: Charles Thomas; 1983.

[2] Sardari D, Maleki R, Samavat H, Esmaeeli A. Measurement of depth-dose of linear accelerator and simulation by use of Geant4 computer code. Reports of Practical Oncology and Radiotherapy. 2010; 15: 64–68.

[3] Khan FM, Gibbons JP. The physics of radiation therapy. 5th edition. Wolters Kluwer; 2014.

[4] Price P, Sikora K, Illidge T, editors. Treatment of cancer. 5th edition. Hodder Arnold Publication; 2008.

[5] International Commission on Radiation Units and Measurements Prescribing. Recording and Reporting Photon Beam Therapy. ICRU Reports 50, 62, 71.

[6] Mayles P, Nahum A, Rosenwald JC, editors. Handbook of radiation therapy physics, theory and practice. CRC Press, Taylor & Francis Group; 2007.

[7] Podgorsak EB, editor. Radiation oncology physics: a handbook for teachers and students. Vienna: International Atomic Energy Agency; 2005. Available at http://www-pub.iaea.org/mtcd/publications/pdf/pub1196_web.pdf

[8] Anderson R, Lamey M, MacPherson M, Carlone M. Simulation of a medical linear accelerator for teaching purposes. Journal of Applied Clinical Medical Physics. 2015; 16(3): 1–17.

[9] Tan TJ, Siva S, Foroudi F, Gill S. Stereotactic body radiotherapy for primary prostate cancer: a systematic review. J Med Imaging Radiat Oncol. 2014; 58: 601–611.

[10] Ganz JC. Gamma knife neurosurgery. Springer Verlag; 2011.

Further reading

Williams JR, Thwaites D, editors. Radiotherapy physics in practice. 2nd edition. Oxford University Press; 2000.

Hall EJ, Giaccia AJ. Radiology for the radiologist. 7th edition. Wolters Kluwer, Lippincott Williams & Wilkins; 2012.

Useful websites on cyberknife technology

www.accuray.com/solutions/treatment-delivery/tomotherapy-treatment-delivery
www.accuray.com/keywords/cyberknife

10 Charged particle radiotherapy

10.1 Introduction

Charged particle radiotherapy is an external beam radiation treatment of tumors with light charged particles (electrons, positrons) or heavy charged particles (protons, α-particles, carbon ions, argon ions). Electron beam radiation is either a by-product of Compton scattering and pair production or applied directly when removing the tungsten target in an XRT gantry. This has been discussed in the context of XRT in Chapter 9. Positrons are used for diagnostics in PET, but not for treatment, see also Chapter 12 on brachytherapy. In this chapter we consider only heavy charged particles, more specifically protons and carbon ions.

Unlike x-rays, proton beams can be steered to a predefined target where they deposit their energy and stop. Thus a proton beam can be more finely tuned to deliver a high dose of radiation to a controlled tumor volume with less radiation risk to healthy tissue on the anterior and posterior sides of the tumor. This is particularly important when tumors are located close to critical organs such as the central nervous system (CNS) or when tumors cannot be removed by surgical means like in the brain. Proton beam therapy goes back to a paper published by Robert R. Wilson in 1946 where he suggested the application of newly available particle beams from cyclotrons for medical treatment and in particular for cancer irradiation [1]. According to the Mayo Clinic in Minnesota, *proton beam therapy* (PBT) has by now been applied with great success to many kinds of tumors, including brain, breast, esophageal, eye, gastrointestinal, gynecological, head and neck, liver, lung, lymphoma, prostate, soft tissue, spine, and many pediatric cancers. Children with cancer benefit most from PBT since their organs are still developing. For adults, the benefits of PBT compared to more conventional and less costly radiation treatments remains a subject of discussion as long-term comparative studies are still missing. The same arguments hold for heavy ion beam radiation like carbon ion beams.

The higher the energy of charged particle beams, the higher is the production cost of these machines and the treatment cost for patients. XRT requires a 0.5–2 m long electron linac producing 4–25 MeV electrons at a cost of roughly $ 1–5 M. PBT entails an accelerator of about 250 MeV at a cost of $ 200 M and more. This high investment capital and subsequently also high treatment cost necessitate a careful justification of why PBT is needed and for which types of cancer it is most beneficial. In comparison to XRT there are only a few operating PBT centers in the world (≈ 50) and far fewer centers for carbon ion beam therapy (≈ 10). In [2] an up-to-date list of operating and planned facilities can be found. As the demand for new facilities grows and advanced smaller accelerators are being developed, the investment capital may eventually drop, making PBT more viable. In this chapter we will first introduce the main characteristics

DOI 10.1515/9783110553116-012

of charged particle beams and their interaction with tissue and continue later with a discussion of facilities and practicalities.

10.2 Range of protons in tissue

The interaction of photons with matter is a rather simple affair. Only two types of cross sections are of any importance for photon energies beyond 1 MeV: Compton scattering and pair production. In comparison, the interaction of protons with matter is complex [3]. The most relevant interaction channels are sketched in Fig. 10.1. These can be subdivided in electronic interactions (a) and nuclear interactions (b–d). Electronic interaction consists of ionization of atoms and transfer of kinetic energy to electrons, similar to the Compton effect. The nuclear interactions can be either Coulomb scattering at the nucleus, also known as Rutherford scattering (b); small angle elastic recoil scattering of nuclei (c); or large angle inelastic recoil including nuclear reactions and excitations, comprising a number of reaction products, such as electrons, neutrons, γ-radiation, and eventually further hadrons and mesons (d). Proton radiotherapy requires proton energies of 70 MeV to about 250 MeV. In this energy range the processes (a–c) dominate and (d) can be excluded.

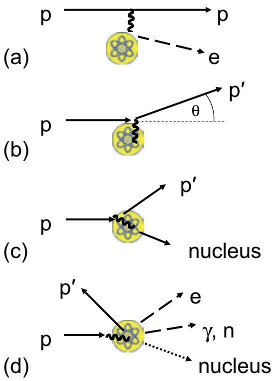


Fig. 10.1: Interaction channels of high energy protons with nuclei in the target. (a) Ionization interaction; (b) Rutherford scattering; (c) elastic recoil scattering; (d) nuclear reactions and excitations.

At the start of the path through the body, energy transfer of protons is small and consists mainly of ionization according to interaction channel (a). In water and tissue mainly the K-shell electrons are excited. The scattering range of knocked out electrons is in the order of 1 mm, while the proton path remains straight. In this region the *linear energy transfer* (LET) is low and therefore also the dose to the body. After successive scattering events the protons will slow down, the *linear energy transfer* (LET) increases dramatically and reaches a maximum. In the region of the maximum the dose varies almost inversely with the remaining proton energy (see Fig. 10.3 (a)). After some more straggling the protons come to a complete stop. The peak in LET as a function of depth is known as the *Bragg peak* (see also Chapter 3 and Fig. 3.11). The average range $\langle R_{\text{tissue}} \rangle$ of protons in tissue can be estimated with a semi-empirical equation [1, 4]:

$$\langle R_{\text{tissue}}(\text{cm}) \rangle = 2 \times 10^{-3} (E_{\text{p}})^{1.8},$$

which covers the energy range of PBT. Here E_{p} is the kinetic energy of protons just before entering the tissue in units of MeV and R is in units of cm. This function is plotted in Fig. 10.2 and gives a good impression of initial proton energies needed for reaching deeper parts of the body at a depth of not more than 25–30 cm.

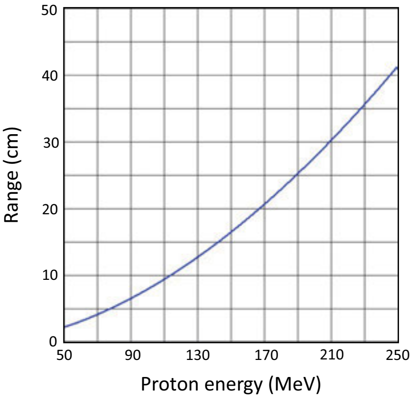


Fig. 10.2: Range of protons in tissue as a function of initial proton kinetic energy at the point of entry into the body.

The characteristic Bragg peak defining a range for protons in matter, in contrast to x-ray radiation, is the main argument for *proton beam therapy* (PBT). Within the region of the Bragg peak the energy loss per distance travelled is highest and so the dose as well as the *relative biological effectiveness* (RBE) for killing cancerous tissue. In this Bragg peak both electrons and protons contribute to the ionization of the target tissue and thus to the RBE. In Fig. 10.3 (a) the range of protons and of x-rays is compared. The x-ray dose piles up below the surface, goes through a maximum and then drops off without a defined range, as already discussed in Chapter 9, Fig. 9.5. In contrast, with PBT the Bragg peak of highest dose can be placed right into the region of the cancer volume by tuning the initial energy of the proton beam. By varying the incident proton energy and beam intensity a Bragg peak plateau is formed that covers the entire cancer volume (Fig. 10.3 (b)). The plateau region is known as *Spread-out Bragg peak* (SOBP).

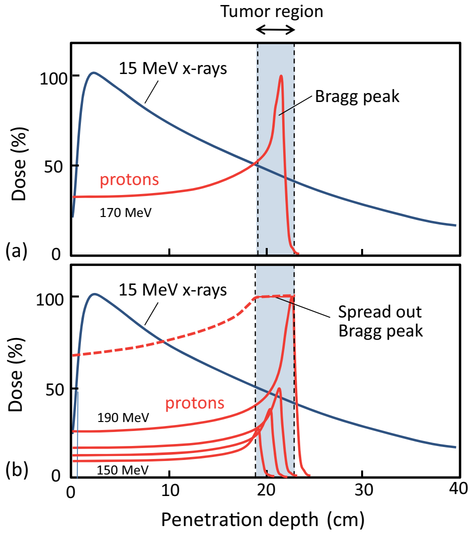


Fig. 10.3: (a) Comparison of dose versus depth for x-rays and protons. Plots are normalized to maximum dose. X-rays do not have a well-defined range, in contrast to protons, which stop a little beyond the Bragg peak. The Bragg peak with the highest LET and RBE can be placed at the location of the tumor volume by tuning the incident energy. (b) By varying the incident energy and intensity of the proton beam a plateau region of the size of the tumor is exposed by proton beam radiation.

10.3 Dose profiles of charged particles

The dose deposited in the tumor region leads to direct and indirect bond breaking in DNA (see Fig. 8.6), where double strand breaks are most effective in reducing the survival chances of cancer cells. The survival fraction of cancer cells as a function of dose D is plotted in Fig. 10.4 for photons, protons, and carbon ions. According to the definition, the RBE for photons is 1. At a survival fraction of 0.1 the RBE for carbon ions is $\mathrm{RBE}_{\mathrm{carbon}} = D_{\mathrm{photon}}/D_{\mathrm{carbon}} = 2.4$, whereas for a survival fraction of 0.01 the $\mathrm{RBE}_{\mathrm{carbon}} = 2$. In all cases the RBE of carbon ions is significantly higher than for photons. For protons, an RBE of 1.1 ± 0.1 has been determined independent of the dose, which is only 10 % higher than for photons. The OER for protons was determined to be about 2.7 ± 0.1 [5]. The higher RBE of protons and carbon ions entails a lower total dose administered for a complete treatment. In the case of hard x-rays, usually a total dose of 60 to 80 Gy is administered in fractions of about 2 Gy per session. According to the higher RBE of protons and carbon ions, the total dose and the number of sessions can be reduced while the dose per session is kept constant.

Another advantage of PBT is easily recognized in Fig. 10.5 showing a sketch of the dose profile in tissue. The targeted cancer volume is assumed to be at the center. With the well-defined range of proton beams and tunability of their position, SOBP can cover the total tumor volume just with irradiation from one side. The back side (distal)

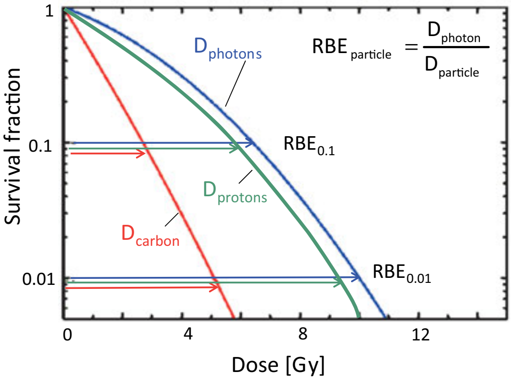


Fig. 10.4: Comparison of survival fraction of cancer cells as a function of dose for photons, protons, and carbon ions.

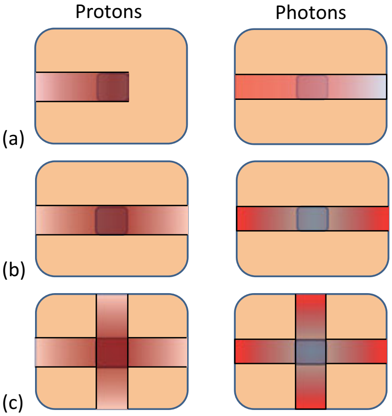


Fig. 10.5: Comparison of proton and photon RT for single field exposure (a), double opposing field exposure (b), and quadruple field exposure (c) of a cancer region at the center. The color gradients indicate dose gradients.

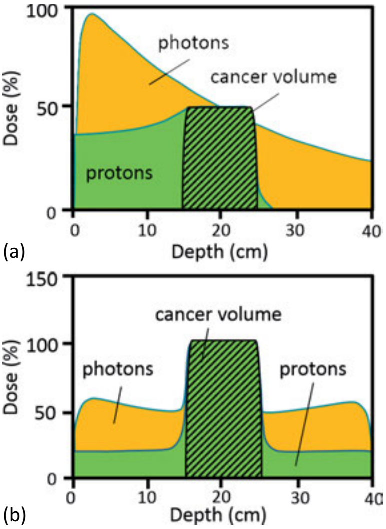


Fig. 10.6: Comparison of photon and proton irradiation of a cancer volume by a single field exposure from the left side (a) and for a four field exposure with the same respective dose of x-rays and protons to reach a 100 % dose level at the cancer volume (b).

is not exposed at all, in contrast to XRT. This is crucial if sensitive organs on the distal side are to be spared from radiation such as the spinal cord. On the other hand, if the treatment plan requires exposure from several sides, then proton therapy still has the advantage of a much lower dose even on the front (proximal) side. Figure 10.5 schematically compares exposures with protons and photons from different sides, the respective dose profiles are plotted in Fig. 10.6.

For a single field exposure shown in the top panel of Fig. 10.5 (a) the x-ray dose reaches a maximum dose of 100 % some 20 millimeters below the skin, dropping to 50 % in the target area. The same 50 % dose level in the target area is reached with a proton beam, but at a much lower accumulated dose of about 40 % on the proximal side. If we now aim for 100 % dose in the target area via a four field exposure as shown in Fig. 10.5 (c) and 10.6 (b), then for XRT we need a maximum dose of 50 % on the proximal side of all four directions, dropping to 4 × 25 % = 100 % in the target area. In contrast, for PBT the same dose is reached at the center while the dose on each proximal side is lowered to 20 %. This example clearly shows that in all cases – single, double, or quadruple exposure fields – PBT is more effective than XRT in sparing the proximal part of tissue. This helps to reduce acute side effects in healthy tissue such as inflammation and nausea. However, in the case of multiple field exposure proton beams do not protect the skin as much as photons do, where the build-up of dose reaches a maximum after penetrating about 2 cm below the skin (Fig. 10.3).

In PBT the absence of dose on the distal side and the lower dose on the proximal side of cancer regions results in an overall lower *nontarget dose* (NTD) compared to XRT. This provides a greater flexibility for selecting number and directions of proton beams and for distributing NTD over healthy tissue. The reduction of NTD can be significant and has been estimated to be about 50 % for prostate cancer. It is obvious that proton radiation has a distinct advantage over photon beam irradiation as concerns NTD. However, this advantage is credited to a considerably more complex and expensive facility, a subject to be discussed further in Section 10.6.

To be more quantitative we want to calculate the particle fluence Φ and beam current that is required for a single field and single session exposure with an accumulated dose of 2 Gy. The particle fluence can be calculated using the dose equation quoted in Chapter 4 for charged particles:

$$D_{\text{charged particle}} = \left(\frac{1}{\rho}\right) \sum_i (\Phi_i \times \text{LET}_i \times \text{RBE}_i).$$

Here the symbols have their usual meaning and the sum goes over all charged particles that are in the particle track and which are produced by the primary particle, here assumed to be protons. ρ is the density of the targeted tissue. The LET of protons depends strongly on their kinetic energy. This is demonstrated again in Fig. 10.7, where the LET of protons is plotted versus residual depth in water before coming to rest. The proton beam enters from the right side with an initial energy of 180 MeV yielding a range of 24 cm according to Fig. 10.2. At the beginning the LET is small, but raises sharply close to the end of their track. This is the essence of the Bragg peak discussed previously. Assuming an average stopping power of LET = 40 MeV/cm across the Bragg peak at a density of 10^3 kg/m^3 for water, and an RBE = 1.1, we find for the required fluence Φ:

$$\Phi = \frac{\rho D}{\mathrm{RBE} \times \mathrm{LET}} = \frac{10^3\,\mathrm{kg/m^3} \times 2\,\mathrm{J/kg}}{1.1 \times 40 \times 10^6 \times 1.6 \times 10^{-19}\,\mathrm{J}/10^{-2}\,\mathrm{m}} = 2.84 \times 10^{12}\,\frac{\mathrm{protons}}{\mathrm{m^2}}.$$

Multiplying the fluence with the elementary charge yields a beam current density of 4.5×10^{-7} A/m^2 if delivered within one second, a current that is easy to handle by proton accelerators.

A word of warning is warranted here. Calculations of required fluence like the one performed here are only good for rule-of-thumb estimates and should never be used for determining the fluence to a patient. For patient treatment, certified radiation treatment plans have to be used.

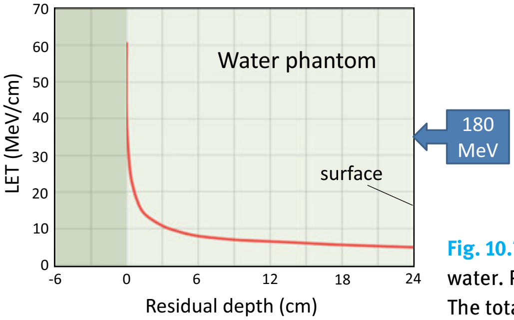


Fig. 10.7: Linear energy transfer (LET) of protons in water. Protons enter from the right in a water tank. The total range for 180 MeV protons is 24 cm.

10.4 Examples for proton beam therapy

In Chapter 9 on x-ray and γ-RT we have seen how the cancer volume is defined by distinguishing gross tumor volume (GTV), clinical target volume (CTV), and planning target volume (PTV). These considerations apply equally to PBT. Furthermore, target volume conformity by using multileaf collimators (MLC), intensity modulated radiation therapy (IMRT), and image guided radiation therapy (IGRT), already discussed in Chapter 9 for XRT, apply similarly to PBT and is called in analogy *intensity modulated proton therapy* (IMPT). The details will not be repeated here. But there is also an important difference to be noted between IMRT and IMPT. By virtue of the Bragg peak, a real three-dimensional (3D) intensity modulation can be achieved with IMPT in contrast to 2D IMRT using photons. An example is shown in Fig. 10.8 where a comparison is made of dose distributions for treatment of the posterior fossa with protons, photons, and IMRT. Posterior fossa is a brain tumor located near the bottom of the skull causing medulloblastoma in patients of young age. Radiation treatment covers the brain and also critical organs such as the cochlea. A dose of 100 % of the cochlea with standard

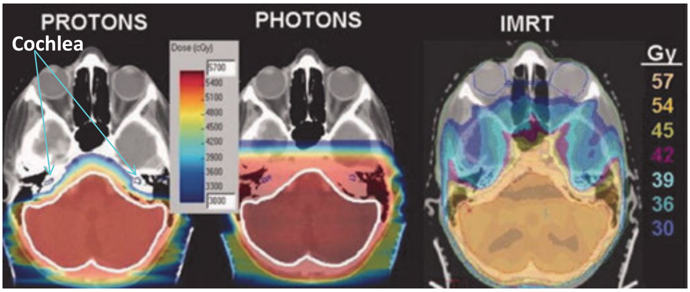


Fig. 10.8: A comparison of dose distributions for treatment of the posterior fossa in a patient with medulloblastoma, using protons, standard photons, and IMRT. Substantial sparing of the cochlea, outlined in blue and indicated by cyan colored arrows, is only possible with protons. Note the different color coding for PBT, XRT, and for IMRT (reproduced from [6] by permission of John Wiley and Sons Inc.).

slit XRT can be reduced to 90 % with the use of MLC XRT, to 33 % with IMRT, and to merely 2 % using protons [6].

Because of the exceptionally high local control and precision, PBT has successfully been applied for treating uveal melanoma [7, 8]. Uveal melanoma is the most common ocular tumor occurring on the retina close to the optic nerve. Surgery is critical and almost impossible. IMRT does not provide the required local control nor does it spare the nearby optic nerve. In contrast, PBT is the treatment modality of choice for this type of cancer. Proton therapy of uveal melanoma requires a ≈ 70 MeV proton beam for the depth of the eye (< 25 mm), which is moderate compared to PBT requirements of other cancer types. A schematic outline of the treatment setup is depicted in Fig. 10.9. Similar to the γ-knife therapy, the patient's head has to be fixed during irradiation. Ta-markers implanted beforehand provide individual coordinates of the patient's eye, which are required for the treatment plan and irradiation simulation. Eye movement is monitored by a fixation light and by video recording. Individually shaped collimators are adapted to the tumor size and location. A plastic range monitor serves for additional range control.

A new development in PBT is the use of a pencil beam [9, 10]. Scanning beam PBT with a pencil is illustrated in Fig. 10.10. A narrow, only millimeter wide proton beam penetrates from one side and the energy is tuned such that the Bragg peak lies within the PTV. The proton beam is then steered precisely across the tumor volume by adjusting magnets and the depth is controlled by the incident proton beam energy. With such a pencil proton beam the entire PTV is covered point by point, essentially "dose painting" the radiation into the PTV. Pencil beam PBT is very effective in treating the most complex tumors with curved contours, like those in the prostate, brain, and eye, while leaving healthy tissue and other critical areas intact. The pencil beam PBT is a very powerful tool that could eventually supersede MLC and replace IMPT techniques.

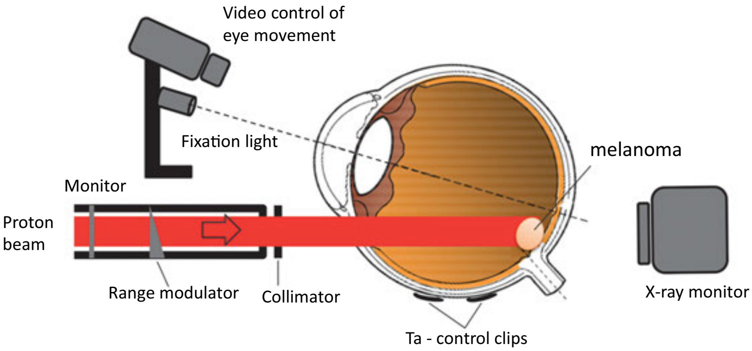


Fig. 10.9: Proton therapy of ulvea melanoma. During RT the head is fixed and the eye should not move. Ta clips on the back side of the eyeball define individual eye coordinates.

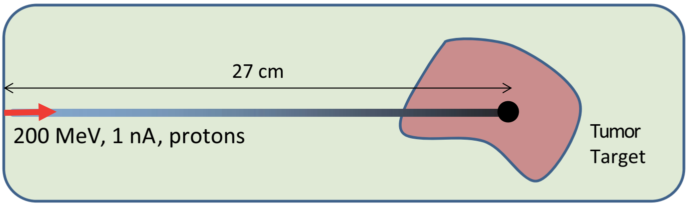


Fig. 10.10: A pencil proton beam arriving from the left penetrates through tissue and stops in the tumor target. The beam can be steered in the lateral direction and energy tuning covers the depth until the entire tumor volume is completely covered.

For all PBT treatments it is necessary to fix the patient just like for conventional XRT. Stereotactic fixation of the head is particularly important for PBT of tumors in the brain. Furthermore, irradiation synchronized with the respiration cycle of patients is used to compensate for organ movement. This requires a variable cycle operation of the synchrotron with injection, acceleration and deacceleration of proton bunches on the time scale of seconds. Compensation for patient or organ movement on a shorter time scale of milliseconds, which is the hallmark of cyberknife technology (see Section 9.7), may also become available for PBT with the next generation of proton synchrotrons (see Section 10.6). Future PBT operating practice will include tumor tracking systems using markers monitored by x-rays and a standby control of the proton beam instead of decelerating the beam immediately after irradiation [11]. This will allow a much faster monitoring of and response to organ movement.

10.5 Carbon ion beam therapy

Protons have a superior range property compared to x-rays. But the RBE of 1.1 is not much higher than for x-rays. PBT is therefore characterized as low LET radiation. Irradiation with much heavier carbon ions features similar advantages as proton irradiation including control of Bragg peak position by incident beam energy, well-defined range, and low radiation beyond the Bragg region. Moreover, because of the heavier mass of carbon ions, multiple scattering and range straggling is approximately three times less than for protons, resulting in a sharper edge in the lateral direction (perpendicular to the beam) and, with some limitations to be discussed below, in the longitudinal direction (parallel to the beam). Edge sharpness is an important factor for deep-seated tumors, where beam widening due to scattering may blur the PTV edge, known as penumbra effect.

The RBE of carbon ions in the Bragg peak is two times higher than that of protons (see Fig. 10.4). Carbon ions are therefore classified as high LET radiation, featuring enhanced therapeutic benefits in treating radiation-resistant tumors. Furthermore, they have a low oxygen enhancement ratio (OER), meaning that they are effective for treating hypoxic cells, which otherwise resist XRT with high OER. And they are less dependent on the cell cycle, implying that cell death occurs (necrosis) independent of the cell phase [12]. For these reasons carbon ion radiation therapy (CIRT) has been proposed as an alternative to PBT and XRT. The dose–depth profile for all three types of radiation is compared in Fig. 10.11. Although the dose on the proximal side is nearly the same for protons and for carbon ions, there is a notable difference on the distal side, where the tail in the dose distribution is assigned to a fragmentation of carbon ions. The fragments consist of H, He, Li, Be, and B ions, as has been identified in studies using a water phantom [13]. The lighter fragments have their own and longer range and therefore straggle in the forward direction. This annoying tail is, however, considered tolerable compared to the other benefits that CIRT offers. On the other hand, if a crucial organ that cannot tolerate the dose is located in the region of the tail, CIRT is not appropriate.

CIRT requires much higher beam energy to reach the same point in the tissue as PBT. For instance, a 200 MeV proton beam has a Bragg peak at 25 cm in water. To reach the same depth a carbon energy of 4800 MeV is required, which is a factor of 24, twice more than the mass difference would suggest. This makes application of CIRT even more complex and expensive than PBT. While it is much more difficult to steer proton beams around patients than an electron beam, the heavy machinery required for CIRT appears to make beam steering almost impossible. Nevertheless, one facility for CIRT with a rotatable gantry exists, located at the University of Heidelberg, Germany. A picture of the gigantic facility is shown in Fig. 10.12. For further discussions on the present status of carbon beam therapy we refer to the review in [14].

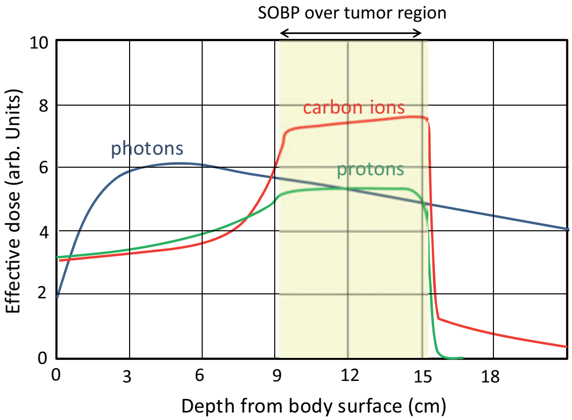


Fig. 10.11: Dose distribution as a function of depth from the body surface for photons, protons, and carbon ions. The dose is given in effective units and scaled such that photon and proton dose have the same value at the tumor center. The difference between proton dose and carbon ion dose within the tumor region is attributed to their different RBE. The carbon ion dose tails off because of fragmentation of carbon ions before coming to rest. SOBP = spread-out Bragg peak (SOBP).

Fig. 10.12: Gantry for carbon ion radiation treatment at the Heidelberg Ion Beam Therapy Center is a 360° rotating beam delivery system. It weighs 670 tons, is 25 m long, and 13 m in diameter spanning three stories (photo reproduced from Heidelberg University Hospital https://www.klinikum.uni-heidelberg.de/Gantry,image-2012.129328.0.html).

10.6 Accelerators and gantries for proton beam therapy

Proton radiation therapy requires large-scale facilities which are beyond the scope of local hospitals. Usually such facilities are part of a regional or national healthcare center. A general layout of a PBT facility is sketched in Fig. 10.13 (a). It consists of four main parts: (1) ion source and linear accelerator for ramping up the energy of proton bunches to 200–250 MeV; (2) synchrotron to store the protons on a circular track guided by bending magnets and to maintain their energy until they are needed; (3) beam delivery system to the gantries; (4) gantries for proton delivery to patients in treatment rooms. The number of gantries determines the number of patients that can be treated simultaneously.

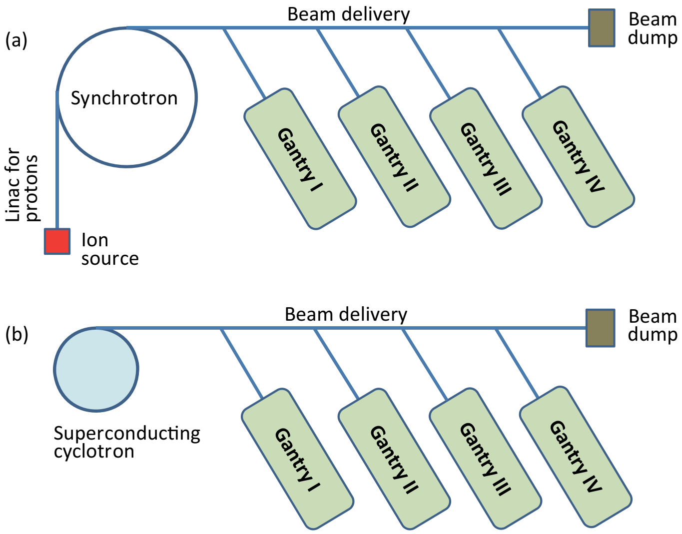


Fig. 10.13: (a) Basic layout of a proton radiation facility for clinical use consisting of linac, synchrotron storage ring, beam delivery lines, and a set of gantries for patient treatment. (b) Alternative layout replacing linac and synchrotron by a cyclotron.

Alternative concepts for the layout of proton radiation facilities replace proton linacs and synchrotrons by cyclotrons or by synchrocyclotrons, the schematic is shown in Fig. 10.13 (b). Synchrocyclotrons have a design similar to conventional cyclotrons, but with a varying RF field to compensate for relativistic effects as protons approach the speed of light. Using superconducting coils, synchrocyclotrons can be much reduced in size. Recently systems have been developed that are considerably smaller and lighter than accelerators used so far, allowing alternative strategies for beam delivery [11, 15]. In one concept a 20 ton synchrocyclotron is designed to rotate around the patient for PBT similar to an x-ray gantry. Another concept proposes the use of a dielectric wall, which is a 2 m long accelerator for reaching 200 MeV protons. The dielectric wall uses a high voltage gradient insulator to handle high electric field stresses without being short circuited, a concept which is promoted by the Lawrence Livermore Laboratory in the USA. Although this technology is not yet ripe for application in PBT, there is a clear challenge for finding cost efficient and space saving alternatives to presently available facilities. Development of novel accelerator concepts can be expected in the coming years.

Heavy ion treatment, such as with carbon ions, requires even bigger facilities, greater investment cost for the contractor and higher treatment cost for the patient. These developments are very risky as new accelerator technologies may soon make bulky and costly machinery obsolete. In addition, more effective nanomedical cancer treatment methods may replace RT altogether by using specific and targeted nanoparticles. Nanoparticles for cancer treatment are discussed in Chapter 14.

Although proton therapy has clear advantages over other radiotherapies as outlined in this chapter, there is not yet sufficient clinical evidence from long-term studies on the benefits of PBT over more conventional radiotherapies. For a status report, see [16]. In the case of ocular melanoma PBT has a clear edge [6]. Also for treating children with malignant central nervous system (CNS) tumors PBT is preferable because it lessens the chance of harming healthy, developing tissue [17]. Children may receive proton therapy for cancers of the brain, the spinal cord, and the eye. For other types of cancer the main advantage is a better sparing of surrounding tissues thereby reducing side effects. The essential question, whether or when survival or cure can be improved by applying PBT, can presently not be answered due to lack of long-term experience, but it is a topic of current medical investigations [18]. As more proton radiation facilities come online around the world, experience will considerably increase in the near future.

10.7 Summary

1. Interaction of protons with matter is dominated by electronic interactions, Coulomb interaction with the nucleus, and small angle recoil of nuclei.
2. For proton beam therapy (PBT) protons of energies 200–250 MeV are required.
3. Protons have a well-defined range in matter.
4. The dose per depth shows a peak at the end of the proton track known as the Bragg peak.
5. The position of the Bragg peak can be tuned by varying the incident energy of the proton beam.
6. RBE of protons is 10 % higher than for photons and is highest in the Bragg peak.
7. LET of a proton beam is low in the beginning and increases sharply in the Bragg peak.
8. Proton radiation therapy is low in LET on the proximal side.
9. The proximal dose is lower than for XRT and is zero on the distal side of the target volume.
10. PBT has better conformity to cancer volume than other radiation treatment methods.
11. Pencil type PBT has the ability to dose paint the entire volume of cancer tissue.

12. PBT is useful for local, well-defined tumors and for tumors near sensitive parts of the body, such as near the optic nerve, cochlea, brain, and spinal cord.
13. PBT is more effective than XRT in sparing proximal tissue.
14. PBT does not spare the skin as much as x-ray.
15. Carbon ions have an RBE of 2.2–2.4 and are therefore classified as high LET radiation.
16. Carbon ion radiation treatment has a low OER.

References

[1] Wilson RR. Radiological use of fast protons. Radiology. 1946; 47: 487–491.
[2] Particle Therapy Co-Operative Group, www.ptcog.ch/
[3] Lilley J. Nuclear physics, principles and applications. John Wiley & Sons; 2013.
[4] Wilson RR. Range, straggling, and multiple scattering of fast protons. Phys Rev. 1947; 71: 385.
[5] Kanemoto A, Hirayama R, Moritake T, Furusawa Y, Sun L, Sakae T, Kuno A, Terunuma T, Yasuoka K, Mori Y, Tsuboi K, Sakurai H. RBE and OER within the spread-out Bragg peak for proton beam therapy: in vitro study at the Proton Medical Research Center at the University of Tsukuba. J Radiat Res. 2014; 55: 1028–1032.
[6] Greco C, Wolden S. Current status of radiotherapy with proton and light ion beams. CANCER. 2007; 109: 1227–1238.
[7] Damato B, Kacperek A, Errington D, Heimann H. Proton beam radiotherapy of uveal melanoma. Saudi Journal of Ophthalmology. 2013; 27: 151–157.
[8] Groenewald C, Konstantinidis L, Damato B. Effects of radiotherapy on uveal melanomas and adjacent tissue. Eye. 2013; 27: 163–171.
[9] Ojerholm E, Kirk ML, Thompson RF, Zhai H, Metz JM, Both S, Ben-Josef E, Plastaras JP. Pencil-beam scanning proton therapy for anal cancer: a dosimetric comparison with intensity-modulated radiotherapy. Acta Oncologica. 2015; 54: 1209–1217.
[10] Wisenbaugh ES, Andrews PE, Ferrigni RG, Schild SE, Keole SR, Wong WW, Vora SA. Proton beam therapy for localized prostate cancer 101: basics, controversies, and facts. Reviews in Urology. 2014; 16: 67–75.
[11] Umezawa M, Ebina F, Fujii Y, Matsuda K, Hiramoto K, Umegaki K, Shirato H. Development of compact proton beam therapy system for moving organs. Hitachi Review. 2015; 64: 506–513.
[12] Hamada N, Imaoka T, Masunaga SI, Ogata T, Okayasu R, Takahashi A, Kato TA, Kobayashi Y, Ohnishi T, Ono K, Shimada Y, Teshima T. Recent advances in the biology of heavy-ion cancer therapy. J Radiat Res. 2010; 51: 365–383.
[13] Haettner E, Iwase H, Krämer M, Kraft G, Schardt D. Experimental study of nuclear fragmentation of 200 and 400 MeV/u 12C ions in water for applications in particle therapy. Phys Med Biol. 2013; 58: 8265–8279.
[14] Schlaff CD, Krauze A, Belard A, O’Connell JJ, Camphausen KA. Bringing the heavy: carbon ion therapy in the radiobiological and clinical. Radiation Oncology. 2014; 9: 88.
[15] Fukumoto S. Cyclotron versus synchrotron for proton beam therapy. Proceedings of the 14th International Conference on Cyclotrons and their Applications, Cape Town, South Africa, 2014.
[16] Ollendorf DA, Colby JA, Pearson SD. Proton beam therapy. Final Evidence Report, Health Technology Assessment Program. 2014.
[17] Main C, Dandapani M, Pritchard M, Dodds R, Stevens SP, Thorp N, Taylor RE, Wheatley K, Pizer B, Morrall M, Phillips R, English M, Kearns PK, Wilne S, Wilson JS. The effectiveness and safety of proton beam radiation therapy in children with malignant central nervous system (CNS) tumours: protocol for a systematic review. Systematic Reviews. 2016; 5: 124.

[18] Allen AM, Pawlicki T, Dong L, Fourkal E, Buyyounouski M, Cengel K, Plastaras J, Bucci MK, Yock TI, Bonilla L, Price R, Harris RR, Konski AA. An evidence based review of proton beam therapy: The report of ASTRO's emerging technology committee. Radiotherapy and Oncology. 2012; 103: 8–11.

Further reading

Yajnik S. Proton beam therapy: how protons are revolutionizing cancer treatment. Springer Verlag; 2012.
Goitein M. Radiation oncology: a physicist's-eye view. Springer Verlag; 2008.
Delaney TF, Kooy HF, editors. Proton and charged particle radiotherapy. Wolters Kluwer; 2008.
Podgorsak EB, Radiation physics for medical physicists. 2nd edition. Springer Verlag; 2014.
Bethge K, Kraft G, Kreisler P, Walter G. Medical applications of nuclear physics. Springer Verlag; 2004.
Chao AW, Chou W, editors. Reviews of accelerator science and technology. Vol. 1. Applications of accelerators. World Scientific Publishing; 2008.
Chao AW, Chou W, editors. Reviews of Accelerator science and technology. Vol. 2. Medical applications of accelerators. World Scientific Publishing; 2009.
Amaldi U. Particle accelerators: from big bang physics to hadron therapy. Springer Verlag; 2014.

Useful websites

Particle Therapy Statistics: http://theijpt.org/doi/pdf/10.14338/IJPT.14-editorial-2.1
https://en.wikipedia.org/wiki/Proton_therapy

11 Neutron radiotherapy

11.1 Introduction

Neutrons like protons are hadrons, however neutrons are uncharged. They are more difficult to produce, and because they are neutral, accelerators and magnets cannot be used for tuning their energy or steering the beam. Neutron irradiation requires a neutron source and a radiological shielding environment that are very different to those for charged particles. Among the various external radiation treatments of cancer, neutron radiation therapy is presently not often prescribed.

Why and when should neutron radiation therapy (NRT) be considered? Neutron treatment is of interest if extended (inoperable) and radiation resistant tumors are difficult or impossible to destroy with low linear energy transfer (LET) radiation, such as photons, electrons, or protons. Fast neutrons, in contrast, have high LET, high relative biological effectiveness (RBE) (see Tabs. 4.1 and 8.1), and the clinically most important Oxygen Enhancement Ratio (OER) parameter is close to 1 instead of 3 for low LET radiation. While with low LET radiation damage is done through single strand break (SSB) of DNA and via generation of toxic radials, in high LET neutron radiation the damage is achieved mainly by double strand breaks (DSB). If a tumor cell is damaged by low LET radiation, it has a certain chance to repair an SSB in the DNA and to stay active. Due to the very high ionization density along the neutron track, DSB can occur already with a single neutron hit illustrated in Fig. 11.1. The chance of self-repair after a DSB is low.

In general, fast neutrons can control very large tumors, because unlike low LET radiation, neutrons do not depend on the presence of oxygen to kill the cancer cells. Thus neutron radiation therapy is the method of choice if the cell division rate is slow and tumorous or malignant cells show signs of oxygen deficiency (hypoxia). In addi-

Fig. 11.1: Neutrons kill cancer cells by destroying both strands of DNA (reproduced from NIU, Institute for Neutron Therapy, Fermilab, USA).

DOI 10.1515/9783110553116-013

tion, the RBE of neutrons is not affected by the time or stage in the life cycle of cancer cells, as is the case for low LET radiation.

Because the RBE of fast neutrons is so high (≈ 10–12), the required tumor dose to kill cancer cells is about 1/3 the dose required with low LET radiation. A full course of neutron therapy is delivered in only 5 to 12 sessions, compared to 30–40 sessions needed for low LET radiation.

We distinguish between three main methods for neutron radiation therapy:

1. Irradiation with fast neutrons of about 2 MeV generated by nuclear fission.
2. Irradiation with high energy neutrons ($E \approx 5-15$ MeV) produced by nuclear reaction of a proton or deuterium beam (0.1–10 MeV) with an appropriate target.
3. Irradiation with epithermal neutrons (1–10^6 eV) that cause only minor cell damage, but after thermalization in the tissue can react with ^{10}B, releasing dose relevant α-particles, known as Boron Neutron Capture Therapy (BNCT).

There are only a few places worldwide that offer one or the other type of neutron radiation therapy: in Seattle (USA), Tomsk (RUS), iThemba LABS (South Africa), and Munich (Germany).

11.2 Neutrons and neutron production

11.2.1 Neutron energies

Some basic properties of neutrons are stated in Chapter 2. In contrast to protons, neutrons cannot be found in nature as free particles. They are always bound in a nucleus. However, neutrons can be produced artificially by nuclear reactions, as fission product of some actinide isotopes (^{232}Th, ^{233}U, ^{235}U, ^{238}U, ^{239}Pu, ^{241}Pu), by spallation of heavy nuclei upon proton impact, or by nuclear reactions of light nuclei with protons. The mass of neutrons $m_n = 1.675 \times 10^{-27}$ kg is slightly larger than the mass of protons ($m_p = 1.673 \times 10^{-27}$ kg) and therefore free and unbound neutrons decay into the protonic ground state via β^--emission:

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu}_\mathrm{e} + E_\mathrm{kin},$$

where n, p, e^-, $\bar{\nu}_\mathrm{e}$ and E_kin denote neutron, proton, electron, electron-antineutrino, and kinetic energy, respectively. The lifetime (1/e) of neutron decay is 881 ± 5 seconds or about 14.7 minutes.

When neutrons are produced via nuclear reaction, fission or spallation, they are usually very fast. They can be slowed down by passing them through a moderator of low atomic mass A, where they thermalize via collisions with the nuclei of the moderator. According to classical mechanics this is most efficiently done with resting or moving particles of identical mass and as central collision. Protons and neutrons have about the same mass; therefore light or heavy water are by far the most efficient moderators with a typical diffusion length of 4 cm needed to thermalize fast neutrons. Further details are provided in Section 3.4.

We distinguish neutrons according to their kinetic energy:

- cold neutrons: $E_{\text{kin}} < 0.020\,\text{eV}$
- thermal neutrons: $0.025\,\text{eV} < E_{\text{kin}} < 0.15\,\text{eV}$
- epithermal neutrons: $1\,\text{eV} < E_{\text{kin}} < 1\,\text{MeV}$
- fast neutrons: $0.5\,\text{MeV} < E_{\text{kin}} < 20\,\text{MeV}$
- high energy neutrons: $E_{\text{kin}} > 20\,\text{MeV}$

For neutron-based radiation therapy only epithermal, fast neutrons, and high energy neutrons are of interest. They are either used for direct irradiation of tissue or for neutron capture in boron that has been injected into tumorous tissue, discussed in Section 11.4.

11.2.2 Fast neutrons from fission

Nuclear research reactors usually burn isotopically enriched ^{235}U. Fission of heavy ^{235}U nuclei takes place by capturing thermal neutrons. The fission products consist of radioactive isotopes, two to three prompt or delayed fast neutrons, fast electrons and γ-radiation. For maintaining a fission chain reaction, fast neutrons must be slowed down to thermal energies by an adequate moderator, which in most cases consist of water because of its excellent moderation and the need to safely dissipate the heat generated by fission. After thermalizing, the fraction of neutrons diffusing back to the fuel assembly is used for continued and controlled ^{235}U fission; those neutrons diffusing after thermalization towards the outside of the fuel assembly can be guided out of the reactor vessel for neutron-based research, such as structure analysis via scattering, tomography, and irradiation. In contrast to thermal neutrons, fast neutrons cannot be guided out because of the inevitable moderation by the surrounding coolant and moderator environment. To extract fast or fission neutrons moderation has to be suppressed. A cross-sectional view of the reactor vessel of FRM II at the Technical University Munich with beam tubes for extracting thermal and fast neutrons is shown in Fig. 11.2. Beam tube SR10 is specialized on extracting fast neutrons for medical treatment. SR10 faces a sheet of enriched ^{235}U in a converter plate. Thermalized neutrons impinging onto this sheet induce fission reactions releasing fast neutrons. Some of those fast neutrons exit the reactor via beam tube SR 10. The heat load at the converter plate is much lower than in the fuel assembly in the reactor core. Therefore direct cooling and thereby moderation can be avoided.

For medical applications, beam tube SR10 guides fast neutrons to the patient. Between converter plate and patient a stack of filters stops background such as thermal and epithermal neutrons and 50 % of the γ-radiation generated during fission. Hence a clean fission neutron beam is generated with a mean energy around 1.9 MeV and

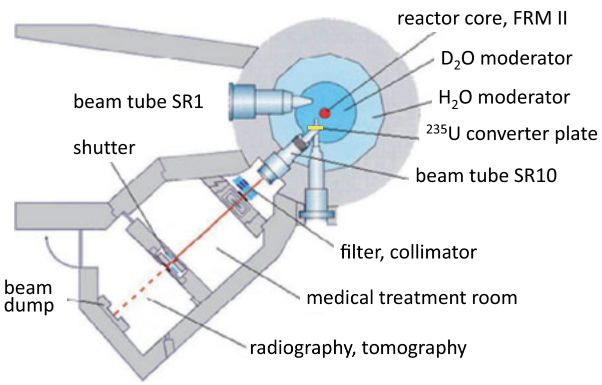


Fig. 11.2: Cross section of the nuclear research reactor FRM II at the Technical University in Munich, Germany. Thermal neutrons are guided out of the moderator (here heavy water) by beam tubes, for example beam tube SR1. SR10 is used for fast neutron radiation therapy. SR2–SR9 are omitted for clarity [1].

the flux is in the order of 10^8 to 10^9 n/(s cm^2). A multileaf collimator (MLC) is used for defining the area to be exposed. At this energy the half maximum depth of neutrons in human tissue is about 5 cm. The dose-depth profile is shown in Fig. 11.6. The dose is highest at the surface and then drops off with increasing depth and loss of energy. Since the RBE has a peak at about 2 MeV, fission neutrons are best suited for treating skin cancer or inoperable tumors close to the surface.

11.2.3 Accelerator-based neutron sources

Fast neutrons can also be generated with the help of accelerators. First, protons or deuterons are accelerated up to a few MeV energy with a linear accelerator (linac) or in a cyclotron; after reaching the specified energy the projectiles hit isotopically enriched targets to trigger nuclear reactions. The most frequent nuclear reactions used for medical neutron radiation therapy are shown in Fig. 11.3. The corresponding short notations for these charge exchange and fusion reactions are:

$$\begin{aligned} &{}^9\mathrm{Be(p,n)}{}^9\mathrm{B}, && Q = -1.85\,\mathrm{MeV} \\ &{}^9\mathrm{Be(d,n)}{}^{10}\mathrm{B}, && Q = +4.4\,\mathrm{MeV} \\ &{}^3\mathrm{H(d,n)}{}^4\mathrm{He}, && Q = +17.6\,\mathrm{MeV} \end{aligned}$$

The threshold energy for the reaction ^{9}Be(p,n)^{9}B is 2.29 MeV proton energy [3, 4]. ^{9}B is not stable but decays immediately into ^{8}Be by emission of a proton; ^{8}Be, in turn, splits up into two α-particles, such that the net reaction is: $^9\mathrm{Be(p,n)}^9\mathrm{B} \rightarrow 2 \times {}^4\mathrm{He} + \mathrm{p}$. With 5 MeV incident protons a wide energy spectrum of fast neutrons can be produced ranging from 0.25 MeV up to 3 MeV. Beam shape assemblies (BSA) are used to narrow down

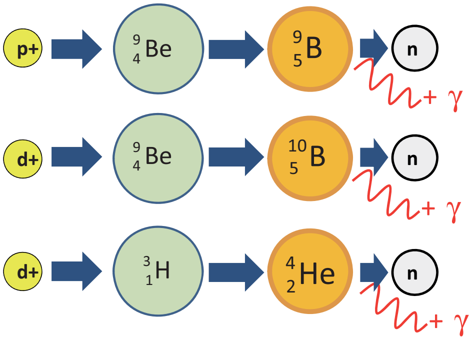


Fig. 11.3: Reaction schemes for the most important types of fast neutron production by proton and deuterium impact.

the neutron energy distribution for specific applications. Proton absorption cross sections for a wide energy range and corresponding neutron yields are plotted in Fig. 11.4 according to [4, 5].

For the ^{9}Be(d,n)^{10}B reaction there is no threshold energy, the neutron yield starts already at low deuteron energies and reaches the highest yield of almost 10^{10} n/s/μA at ^{2}H energy of 6 MeV. The exothermic reaction with $Q > 0$ yields neutron energies of 4.4 MeV at low energy ^{2}H particles hitting the target and increases continuously with increasing ^{2}H projectile energy. Therefore this reaction produces faster neutrons than the first reaction. The reaction product ^{10}B is a stable isotope, which itself is a strong absorber for thermal neutrons and is used for BNCT, to be discussed in Section 11.4.

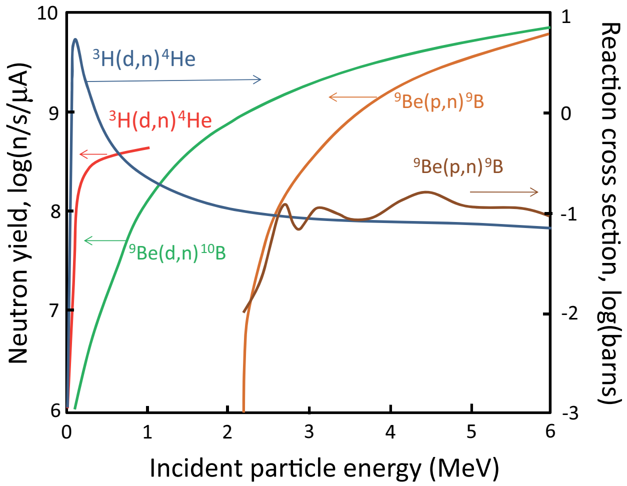


Fig. 11.4: Left axis reading: neutron yields for ^{9}Be(p,n)^{9}B, ^{9}Be(d,n)^{10}B and ^{3}H(d,n)^{4}He reactions as function of projectile energy. Right axis reading: reaction cross sections for ^{9}Be(p,n)^{9}B and ^{3}H(d,n)^{4}He in dependence on projectile energy (data from [4] and [5]).

The $^3H(d,n)^4He$ fusion reaction is important for clinical use, as it produces monoenergetic neutrons even for low energy ^2H projectiles. There is no threshold for the reaction, but a resonance-like cross section is located at 100 keV incident particle energy (see Fig. 11.4). The Q value of the reaction starts at 17.6 MeV, where the kinetic energy is partitioned between neutrons (14.05 MeV) and α-particles (3.5 MeV). With increasing projectile energy, the α-particles gain further energy, whereas the neutron kinetic energy stays almost constant. The neutron yield saturates already at projectile energies below 1 MeV. As this fusion reaction does not feature a threshold, it is of interest even for low energy linear accelerators or cyclotron facilities. The main problem is the handling of the radioactive tritium target. This problem is solved by using metal hydrides, such as Ti^3H_2 to trap gaseous $^3H_2$ in a “solid target”. Furthermore, the target is enclosed in an evacuated glass tube, similar to sealed off x-ray tubes for x-ray radiography, to prevent tritium from escaping into the environment.

Fast neutrons resulting from these three reactions can be used either for direct neutron irradiation of tumors, or as neutron source for BNCT treatment (see Section 11.4). In either case a neutron moderator is required to tune down the neutron energy to the appropriate level for the highest dose at the tumor location. The depth-dose relationship for fast neutrons is discussed in Section 11.3.

11.2.4 Accelerator facility

Charged particles ($^1H^+$, $^2H^+$) are usually accelerated in radio-frequency quadrupole (RFQ) linacs or in *drift tube linacs* (DTL). The linacs run in a pulse mode with a frequency of 10–20 Hz and the pulses are a few micro- to milliseconds long. A low pulse repeat frequency (PRF) is needed to provide sufficient travel time for all neutrons from the target to the patient, before the next pulse can be issued. Figure 11.5 shows the schematic of a setup for proton-neutron conversion presently being installed in Japan [6]. At the end of an RFQ-DTL accelerator a beryllium target is positioned for proton-neutron conversion, a neutron moderator assembly is used for reducing the neutron energy arriving from the target, and collimators for guiding neutrons to the tumor, in the illustration assumed to be in the brain. Heavy shielding is required for protection against other sources of radiation, such as γ-radiation and fast ions.

Accelerator-based neutron production has the advantage – in comparison to fission neutron sources – that a much smaller facility and less infrastructure are required for safe operation of a neutron beam. Such facilities could, in principle, be run by hospitals that already host an accelerator for isotope production.

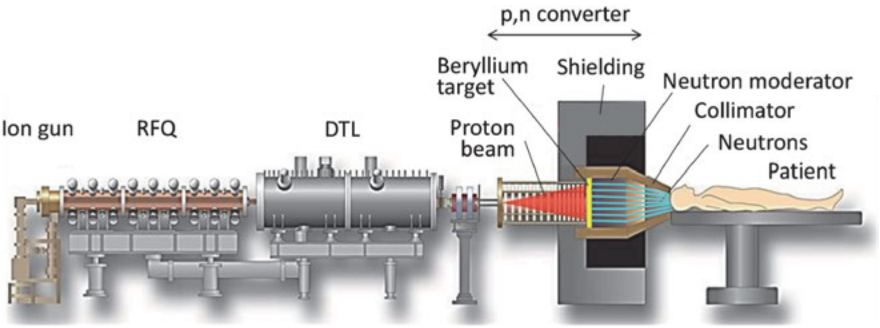


Fig. 11.5: Schematic of a linear proton accelerator irradiating targets to produce epithermal neutrons by a conversion reaction. RFQ = radio-frequency quadrupole linac, DTL = drift-tube linac (DTL) (adapted from [6]).

11.3 Linear energy transfer for fast neutrons

The *linear energy transfer* (LET) for high energy neutrons as a function of depth is plotted in Fig. 11.6. The shape of the dose-depth profile resembles that for high energy photons (Fig. 9.5). In both cases LET increases below the skin surface and reaches a maximum, beyond which it slowly tails off. Neutrons like photons have neither a range nor a Bragg peak at the end of their travel path, unlike protons. Although similar in shape, the reasons for a buildup of dose below the surface are different for photons and neutrons. In the case of photons, their energy is first converted to Compton electrons

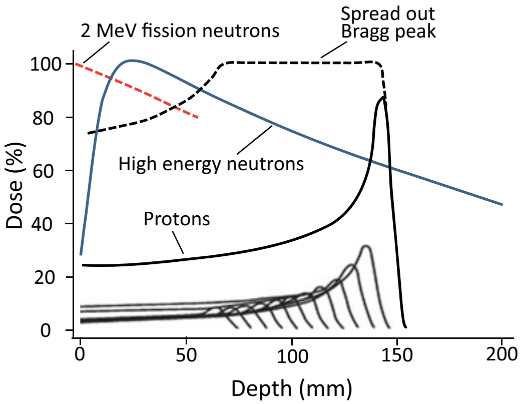


Fig. 11.6: Comparison of the dose-depth profile for protons and high energy neutrons (< 20 MeV). For fission neutrons, the maximum dose is at the skin and drops off continuously (red dashed line). The maximum range of these 2 MeV neutrons is only about 50 mm. Higher energy neutrons show buildup of dose below the skin. In the case of protons, many different energies are shown, which together define a range and a spread-out Bragg peak.

and electron-positron pairs before a relevant dose builds up within the range of fast electrons. In the neutron case, the maximum dose is reached after neutrons are slowed down to energies of about 2 MeV, where LET and RBE are highest. The maximum shifts to smaller depth for lower energy neutrons. Fission neutrons with energy of 2 MeV do not show a buildup of dose below the surface. Their dose is highest right at the skin and then drops off continuously, as shown by the red dashed line in Fig. 11.6. Once the neutrons are thermalized, LET is negligible.

Dose-depth measurements are carried out in phantoms, usually water, which behave similar to tissue but allow placing a dosimeter (ionization chamber) at various depths to the surface for measuring the energy deposited, as already discussed in Chapter 9. A conversion factor of 1.25 is used to compare water with tissue. For instance, 2 MeV fission neutrons are thermalized within a distance of 40 mm in water, which is assumed to be 50 mm in tissue.

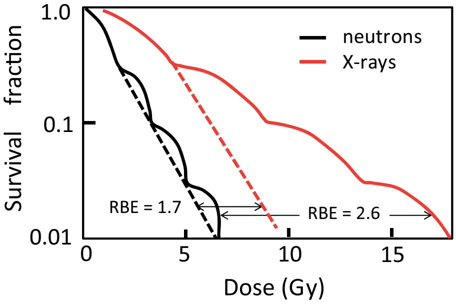


Fig. 11.7: Proportions of mammal cells surviving radiation as a function of dose. Dashed lines are for intense radiation, while solid lines are temporally fractionated radiation. Black lines represent neutron radiation and red lines x-ray radiation. Also shown is the RBE of neutrons for single exposure and multiple exposures (adapted from www.wikiwand.com/en/Cell_survival_curve).

The difference between XRT and NRT is accentuated by considering the survival curve of mammal cells upon radiation shown in Fig. 11.7. The survival curve for x-rays is characterized by a broad initial shoulder, whereas the survival curve for neutrons has almost no shoulder. This difference is typical for low LET versus high LET radiation. For low LET radiation the total dose increases when the dose is fractionated, i.e., delivered in a sequence of time intervals. Nevertheless, fractionation is important in order to let the healthy cells recover between fractions. In the case of NRT, the shoulder is much less pronounced and therefore the difference between single exposure and fractionated exposure is negligible. Anyhow, a fractionated dose is also applied in NRT for the same reasons as for low LET XRT. Now we compare the dose accumulated at survival fraction of 1 %. In the example given in Fig. 11.7 the x-ray dose required is 18 Gy in 4 fractions, whereas for neutron irradiation it is 7 Gy after the same number of fractions. This yields an RBE for neutrons of 2.6. However, in reality 18 Gy would be administered in 9 XRT sessions, whereas 7 Gy would require only 3 NRT sessions. This is one of the main advantages of high LET radiation: fewer exposures are required for the same dose. The other main advantage is the insensitivity to oxygen content in the tumor volume expressed in a low OER of about 1.

11.4 Boron neutron capture therapy

Boron neutron capture therapy (BNCT) is a very effective means of killing cancer cells. Here neutrons only assist the process; the damage is done by fast α-particles. BNCT proceeds according to the steps illustrated in Fig. 11.8. First, isotopically-enriched boron (^{10}B)-containing drugs are delivered to the tumorous tissue, preferentially by injection. Then by exposure to a neutron source, thermal neutrons (0.025 eV) are captured by ^{10}B. In response, ^{10}B breaks apart by ejecting energetic short-range α-particle (1.47 MeV) and ^{7}Li-ion (0.84 MeV). The short notation for this reaction is $^{10}\mathrm{B}(\mathrm{n},\alpha)^{7}\mathrm{Li}$. ^{10}B has a particularly large cross section for thermal neutron absorption: $\sigma_{\mathrm{abs}} = 3837$ barns. The energy of α-particles and ^{7}Li-ions is deposited mostly within the same cell containing the original ^{10}B atom, since the range of the particles is in

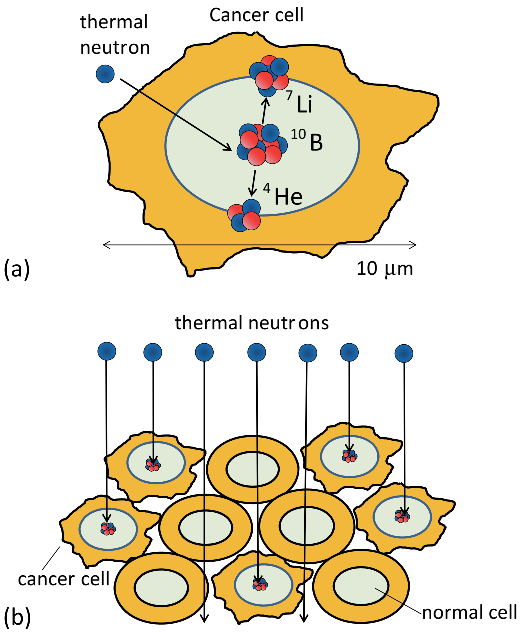


Fig. 11.8: Schematic of radiotherapy of cancer cells via capture of thermal neutrons in boron. (a) The reaction products ^{4}He and ^{7}Li are shown that have a range in the order of the cell's size. (b) Ideally only cancer cells contain ^{10}B, which are destroyed by thermal neutron capture, while normal cells are unaffected by neutrons.

the order of the cell size, or typically about 10 µm. In the past, BNCT was performed mainly with neutrons from a nuclear reactor. However, the more modern approach is the use of (p,n) or (d,n) converters by means of a linear proton (deuterium) accelerator as shown in Fig. 11.4 [4, 5]. Fast neutrons from such a converter can be moderated down to epithermal energies with an appropriate moderator. The final moderation to thermal energies is done within the top layers of the skin.

There are also other isotopes with high thermal neutron absorption cross sections, but ^{10}B is the best choice for several reasons:

- ^{10}B is not radioactive and chemically not toxic;
- ^{10}B is readily available (abundance about 20 % of naturally occurring boron);
- the emitted particles (^{4}He and ^{7}Li) have high LET;
- the chemistry of boron is well understood and allows it to be readily incorporated into a multitude of different chemical structures.

The radiation effect is local and limited to those cells that have taken up ^{10}B. Other cells are still irradiated by the thermal and epithermal neutrons, but the LET of thermal neutrons is rather low.

BNCT is mainly used for high-grade gliomas, cerebral metastases of melanoma, as well as head, neck and liver cancer. The standard boron delivery agents is borocaptate sodium (BSH) (chemical formula: $Na_2B_{12}H_{11}SH$) and boron phenylalanine (BPA). The challenge is to achieve a boron concentration sufficient to deliver therapeutic doses of radiation to the tumor cells with minimal toxicity to surrounding normal tissue. This concentration can be estimated to be at least ≈ 20 µg boron per gram tumor mass. Another issue is the purity of borocaptate sodium. Impurities lead to additional and uncontrolled toxicity. In any case it must be guaranteed that BSH readily enters most malignant tumor cells, and that some retention of the drug occurs in the tumor cells over a reasonable amount of time for radiation treatment, while BSH remains virtually excluded from normal tissue. Given these conditions, BNCT is the only nuclear radiation therapy that has the source of radiation directly in the cell after neutron capture. It therefore has some resemblance to brachytherapy, which is discussed in Chapter 12 but more so with therapeutic nanoparticles presented in Chapter 14.

BNCT appears to be a very promising radiotherapy for some specific tumors and melanomas. It is highly effective if tumors are local and close to the surface. However, in the medical community the acceptance level is presently not very high. After encouraging experiments in Japan, the USA and Europe, BNCT treatments have essentially ceased. The decline of interest is mostly due to the still insufficient concentration of ^{10}B at the tumor site, and only at the tumor. Here progress is awaited from better suited boron carriers. Furthermore, in the past BNCT could only be performed at nuclear research reactors and the number of operational reactors has decreased during the last couple of years. Due to the off-clinical locations of treatment centers, the number of clinical studies has been rather limited, and therefore the advantages of BNCT irradiation modality compared to other radiotherapies are not well documented. Because of this comparably low clinical activity, the development of boron-containing pharmaceuticals is also at a low level. However, with the advent of accelerator-based facilities that can be integrated into a clinical infrastructure there is hope that the BNCT irradiation modality becomes more widely accepted. At CERN (Geneva, Switzerland) a new development has taken place for constructing a tabletop RFQ proton accelerator [7], which might revolutionize the application of proton therapy as well as proton/neutron converters, to make epithermal neutrons more viable in radiotherapy clinics. A review of the present status of BNCT treatment and future perspectives is published in [8].

11.5 Summary

1. Neutrons for radiation therapy are produced in nuclear reactors by fission of ^{235}U or by proton-neutron converters via proton accelerators.
2. For accelerator-based fast neutron production three different nuclear reactions can be used: ^{9}Be(p,n)^{9}B, ^{9}Be(d,n)^{10}B, and ^{3}H(d,n)^{4}He.
3. Tumors with suitable (not too deep) location can be irradiated directly with fast neutrons, combining high LET, high RBE, and low OER.
4. Fast neutrons have no range. LET is highest for 2 MeV neutrons.
5. The number of radiation fractions required for a survival chance of 1 % is less than for low LET radiation by a factor of 3.
6. Some tumors can be treated by boron neutron capture therapy (BNCT) which requires thermal neutrons that are absorbed by ^{10}B isotopes with subsequent emission of α-particles and ^{7}Li particles.
7. The fragments of ^{10}B, α-, and ^{7}Li-particles have a very short range in the order of the size of a cell.
8. BNCT is the only nuclear RT which has the source of radiation directly in the tumor cell.
9. The standard boron delivery agents are borocaptate sodium (BSH) and BPA, which are preferentially taken up by tumor cells.
10. Neutron radiation therapy is a promising modality not practiced much at present.

References

[1] Genreith C. MEDAPP: Fission neutron beam for science, medicine, and industry. Journal of Large-scale Research Facilities. 2015; 1: A18.

[2] Lee CL, Zhou XL. Thick target neutron yields for the 7Li(p,n)7Be reaction near threshold. Nucl Instr and Meth in Phys Res B. 1999; 152: 1–11.

[3] Karki A. A study of ^{9}B spectroscopy via the ^{9}Be(p,n)^{9}B reaction using the neutron time-of-flight technique. PhD Theses. Ohio University; 2013.

[4] Chichester DL. Production and applications of neutrons using particle accelerators. Idaho National Laboratory Idaho Falls, Idaho 83415; 2009.

[5] Shin JW, Pat TS. New charge exchange model of GEANT4 for ^{9}Be(p,n)^{9}B reaction. Nuclear Instr and Meth B. 2015; 342: 194–199.

[6] Yamamoto T, Tsuboi K, Nakai K, Kumada H, Sakurai H, Matsu A. Boron neutron capture therapy for brain tumors. Transl Cancer Res. 2013; 2: 80–86.

[7] Miniature accelerator at: http://home.cern/about/updates/2015/07/miniature-accelerator-treat-cancer

[8] Moss RL. Critical review, with an optimistic outlook on Boron Neutron Capture Therapy (BNCT). Applied Radiation and Isotopes. 2014; 88: 2–11.

Further reading

Barth RF, Vicente MG, Harling OK, Kiger WS, Riley KJ, Binns PJ, Wagner FM, Suzuki M, Aihara T, Kato I, Kawabata S. Current status of boron neutron capture therapy of high grade gliomas and recurrent head and neck cancer. Radiation Oncology. 2012;7:146.

Sauerwein WAG, Wittig A, Moss R, Nakagawa Y, editors. Neutron capture therapy – principles and applications. Springer Verlag; 2012.

Useful website

https://www-ad.fnal.gov/ntf/what_is/index.html

12 Brachytherapy

12.1 Introduction

Brachytherapy is the application of radioisotopes through direct contact with tumors. The Greek name “brachys” implies “close”, in contrast to “tele” which stands for “distant”. Consequently all *external beam radiotherapies* (EBRT) discussed so far in Chapters 9–11 can be characterized as *teletherapies*. For brachytherapy, appropriate radioisotopes are required that are encapsulated in small tubes and either incorporated by a physician directly into the tumorous tissue, or applied on the outside. Typical capsule size is demonstrated in Fig. 12.1. Such a therapy promises success if the tumor is well defined, localized and if the tumor can be reached either by penetration through the skin (interstitial), through open cavities (intracavitary) like the rectum for prostate cancer or the vagina for cervical cancer, or in direct contact on the skin. Those listed here are the main applications of brachytherapy, but other carcinogenic tissues can also be treated by brachytherapy such as breast cancer, cancer in the gall bladder, intraocular, skin, and thyroid cancer.

Fig. 12.1: Stainless steel capsules containing radioactive isotopes for brachytherapeutic treatment of tumors (reproduced from https://en.wikipedia.org/wiki/Brachytherapy, © Creative Commons).

Brachytherapy is an important alternative to surgery, chemotherapy, and teletherapy via γ-, p-, α-, or neutron irradiation and it has many advantages compared to teletherapies. It guarantees:

- an overall comparatively short treatment time;
- local application of radiation;
- conserves healthy tissue;
- lower probability for incontinence and impotence in case of prostate cancer;
- conserves skin and tissue in case of breast cancer;
- patient’s presence is required only once and time spent in hospital is rather short.

DOI 10.1515/9783110553116-014

However, there are also disadvantages:

- extended or metastasized tumors cannot be treated by brachytherapy;
- implantation of radioactive capsules can be a radiation hazard to the clinic staff;
- only *Low Dose Rate* (LDR) capsules can be implanted by hand;
- *High Dose Rate* (HDR) capsules are difficult to handle and require special procedures to be described in Section 12.3.

The main applications of brachytherapy are for the treatment of the following cancers: uterus, cervix, prostate, intraocular, skin, thyroid, and bone.

12.2 Radioisotopes for brachytherapy

12.2.1 General considerations

Some important issues for selecting radioisotopes suitable for brachytherapy need to be considered first. These are summarized as follows:

- Half-life should neither be too short nor too long. Half-lifes between a few days and a few months are most suitable.
- The emitted radiation should have energies sufficient to treat the tumor but not too high to make radiation protection a severe problem for patients, clinic staff, and environment.
- Radioactive material should be in a solid form that can be encapsulated without danger of spilling. This applies to mother and daughter radioisotopes and excludes any gaseous or liquid form of isotopes. But pellets, wires, or plaques are appropriate.
- The specific activity, i.e., the activity per mass unit, should be high, so that small amounts of radioactive material already provide sufficient radiation in a short exposure time.

Considering these requirements, Tab. 12.1 lists the most common radioisotopes for brachytherapy. All of them are β- and γ-emitters, usually from the same mother isotope. Other types of emitters are not used. Because of their short range (see Fig. 3.12), β-radiation is already stopped within thin stainless steel container walls, while even “soft” γ-radiation of only a few tens of keV photon energy can penetrate through the walls. In practice, most (β,γ)-sources are effectively only γ-sources, unless the β-energy is very high as is the case for ^{106}Ru, and ^{90}Sr/^{90}Y. Since β-decay always comes with a broad energy distribution (see discussion in Section 2.4), only the maximum energy $E_{\max}$ is listed in Tab. 12.1. The broad β-energy distribution entails a broad γ-emission spectrum, in which case only the mean γ-energy is quoted in Tab. 12.1.

All isotopes listed in Tab. 12.1 are either fission products of ^{235}U in a nuclear reactor (^{137}Cs, ^{106}Ru, ^{90}Sr), or they are generated by thermal neutron capture in (n, β(γ))-

Tab. 12.1: Most frequent radioisotopes for brachytherapy, including form of capsules, type of radiation, half-life, and energy of emitted particles.

Radionuclide	Form	Type	$T_{1/2}$	Energy [MeV]
^{137}Cs	Tubes, needles, afterloading	γ	30.2 a	0.662
^{60}Co	Tubes, afterloading	γ	5.26 a	1.17, 1.33
^{192}Ir	Wires, afterloading	γ	73.8 d	0.38 (mean)
^{125}I	Seeds	γ	59.6 d	0.0274, 0.0314, 0.0355
^{103}Pd	Seeds	γ	17.0 d	0.021 (mean)
^{198}Au	Seeds, grains	γ	2.7 d	0.412
^{106}Ru ^{106}Rh	Plaques	β^-	1.02 a 30 s	0.039 3.54
^{90}Sr ^{90}Y		β^-	28.7 a 64 h	0.546 2.27

reactions (^{60}Co, ^{192}Ir, ^{125}I). Only in a few cases charged particle activation (CPA) is used via proton or deuterium accelerators (^{103}Pd). The standard application of CPA is for production of light isotopes that are inaccessible to neutron activation. However, these light isotopes have too short half-life times to be useful for brachytherapy.

Fission products of ^{235}U must be radiochemically separated from all other radioisotopes that are present in irradiated uranium pellets so that the emission spectrum is “clean” when clinically applied and free of spectral lines from contaminated radioisotopes.

Extensive and risky radiochemistry can be avoided by thermal neutron activation in nuclear reactors (see Section 2.6.4 for the general background). But the use of isotopically clean targets to avoid any contaminations by other isotopes is a stringent prerequisite. The activity of the final product can then be controlled by extrinsic parameters: neutron flux, exposure time, density; and by intrinsic parameters: thermal neutron capture cross section and lifetime.

In the following, specific nuclear reactions are discussed for the isotopes listed in Tab. 12.1.

12.2.2 Radioisotope ^{137}Cs

The radioisotope ^{137}Cs is a fission product of ^{235}U with a half-life of 30.17 years. The decay branches into a pure β^--emission (5.4 %) and for the majority of isotopes (94.6 %) to β^--emission followed by γ-radiation. The branching scheme is shown in Fig. 3.5.

Formally:

$$
\begin{aligned}
{}^{137}_{55}\mathrm{Cs} &\Rightarrow {}^{137}_{54}\mathrm{Ba} + \beta^- \ (1.174\,\mathrm{MeV}),\ 5.4\,\% \\
{}^{137}_{55}\mathrm{Cs} &\Rightarrow {}^{137}_{54}\mathrm{Ba} + \beta^- \ (0.512\,\mathrm{MeV}) \\
&\qquad + \gamma\,(0.6617\,\mathrm{MeV}),\ 94.6\,\%
\end{aligned}
$$

For brachytherapy only the γ-emission is of interest, as the β^--radiation is stopped in the container walls. The isotope ^{137}Cs is prepared in the form of pellets and encapsulated in stainless steel beads. The beads are then inserted in a stainless steel rod through an inlet, the cross section is shown in Fig. 12.2. The total activity is controlled by the number of beads filled in. The specific activity remains low, which is ideal for low dose rate (LDR) applications. Furthermore, because of the long half-life of 30 years the source can be used over a long period of time and is well suited for long term LDR implants.

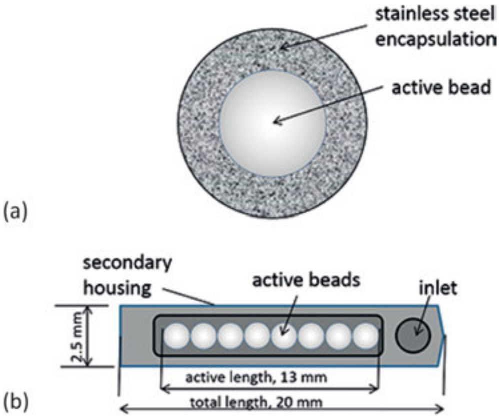


Fig. 12.2: (a) Encapsulation of ^{137}Cs isotopes in stainless steel beads; (b) cross section of a stainless steel rod accepting beads through an inlet.

12.2.3 Radioisotope ^{60}Co

^{60}Co is a radioactive isotope produced by thermal neutron activation. First the stable ^{59}Co isotope is irradiated by neutrons and forms ^{60}Co which is unstable with $T_{1/2} = 5.26$ years. The decay via β- and γ-radiation features two main γ-lines at 1.173 MeV and 1.33 MeV. The ground state isotope ^{60}Ni is stable. The reaction path is as follows:

$$
\begin{aligned}
{}^{59}_{27}\mathrm{Co} + \mathrm{n} &\Rightarrow {}^{60}_{27}\mathrm{Co}\ (\text{radioactive}) \\
{}^{60}_{27}\mathrm{Co} &\Rightarrow {}^{60}_{28}\mathrm{Ni} + \beta^- + \gamma\,(1.173 + 1.33)
\end{aligned}
$$

The decay scheme is shown in Fig. 12.3. ^{60}Co is a pure γ-source as the β-radiation is too soft and completely absorbed in the container walls. The main use of ^{60}Co in brachytherapy is in HDR afterloading applications. But in order to get a high specific activity, long neutron irradiation times are required as the thermal neutron absorption cross section is with 37 barn rather low.

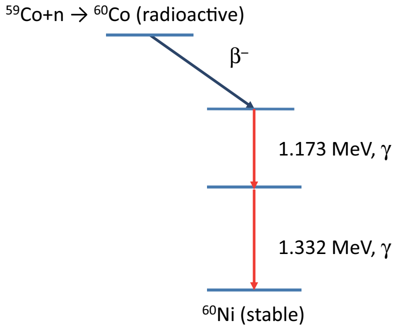


Fig. 12.3: Decay scheme of ^{60}Co after neutron activation.

12.2.4 Radioisotope ^{192}Ir

^{192}Ir is also produced by neutron activation. The reaction path for the isotope production is as follows: ^{191}Ir captures a thermal neutron and is converted to radioactive ^{192}Ir, which decays with a half-life of 74 days by β^-- (1.459 MeV) and γ-emission to the ground state of ^{192}Pt, formally:

$$\begin{aligned} {}^{191}_{77}\mathrm{Ir} + \mathrm{n} \Rightarrow {}^{192}_{77}\mathrm{Ir}\ (\text{radioactive}) \\ {}^{192}_{77}\mathrm{Ir} \Rightarrow {}^{192}_{78}\mathrm{Pt} + \beta^- + \gamma\,(0.38\,\mathrm{MeV}, 74\,\mathrm{d}) \end{aligned}$$

The γ-emission lines are a complex mixture of transitions, whose average energy is 0.38 MeV. ^{192}Ir radioisotopes are usually fabricated in the form of wires for HDR applications and afterloading insertion. The wire is surrounded by a Pt coating and can be cut into smaller pieces as required for the specific application. The very high specific activity up to TBq/g and the half-life of about 74 days makes this isotope one of the most favorable for HDR applications.

12.2.5 Radioisotope ^{125}I

^{125}I is produced by neutron activation of ^{124}Xe converting it by EC to radioactive ^{125m}Xe. The excited intermediate state decays by various γ-emissions to radioactive ^{125}Xe, which, in turn, decays by another EC process with a half-life of 16.9 h to ^{125}In. The isotope ^{125}I is not stable, but decays again by EC with a half-life of 59.4 days to ^{125}Te. Formally:

$$\begin{aligned} {}^{124}\mathrm{Xe} + \mathrm{n} \Rightarrow {}^{125\mathrm{m}}\mathrm{Xe}\ (\text{radioactive}) \\ {}^{125\mathrm{m}}\mathrm{Xe} \Rightarrow {}^{125}\mathrm{Xe} + \gamma\,(56.9\,\mathrm{s}) \\ {}^{125}\mathrm{Xe} + \mathrm{e} \Rightarrow {}^{125}\mathrm{I} + \gamma\,(16.9\,\mathrm{h}) \\ {}^{125}\mathrm{I} + \mathrm{e} \Rightarrow {}^{125}\mathrm{Te} + \gamma\,(35.5\,\mathrm{keV}, 59.4\,\mathrm{d}) \end{aligned}$$

The ^{125}I transition to ^{125}Te features a soft γ-line at 35.5 keV, and two characteristic x-ray lines at 27 keV and 31 keV. These soft γ- and x-ray lines can easily be shielded by thin lead sheets, making radiation protection effective. The radioisotope ^{125}I is mainly used for prostate implants. For this purpose ^{125}I is incorporated into implantable seeds such as silver rods. An example is shown in Fig. 12.4.

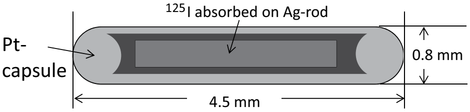


Fig. 12.4: Capsule for encasing ^{125}I radioisotopes. Several other capsule designs are in use.

12.2.6 Radioisotope ^{103}Pd

Alternatively, for prostate implants the radioisotope ^{103}Pd is used with the decay scheme:

$$^{103}\mathrm{Pd} + e \Rightarrow {}^{103}\mathrm{Rh} + \gamma\,(21\,\mathrm{keV},\,17\,\mathrm{d})$$

The shorter half-life and lower γ-emission energy is often considered an advantage over ^{125}I treatment. For the preparation of ^{103}Pd there exist several routes, using either thermal neutron capture (NC) or charged particle activation (CPA). Some are briefly described:

1. Proton irradiation of rhodium foils with a cyclotron causes the reaction: ^{103}Rh (p,n)^{103}Pd. The maximum cross section for this reaction is found at a protron energy of about 10 MeV. Activation is followed by a difficult chemical separation process of the radionuclides from the expensive target material [1].
2. Alternatively, activation can also be achieved by deuterium bombardment of an Rh target via the reaction: ^{103}Rh(d, 2n)^{103}Pd.
3. Neutron activation via the capture reaction: ^{102}Pd(n,γ)^{103}Pd.
4. Photon activation of a Pd foil via γ, n reaction: ^{104}Pd(γ,n)^{103}Pd.
5. Proton irradiation of natural isotopic mixture of Ag foils with highly energetic protons of about 66 MeV following the reaction: $^{\mathrm{nat}}$Ag(p, 2p + xn)^{103}Pd [2].
6. Neutron activation via ^{103}Ag(n,γ)^{103}Pd.

Many more reactions are known and are still being developed. An overview of different CPAs for ^{103}Pd production can be found in [3].

12.2.7 Radioisotopes ^{198}Au and ^{106}Ru

^{198}Au produced by neutron activation of ^{197}Au and decaying with a half-life of 2.7 days via emission of β^-- and γ-radiation of 0.4 MeV photons was used in the past in platinum encapsulated grains for the treatment of tumors in the head and neck region.

However, this isotope is only of historical interest, but is no longer in use, mainly because of the short half-life requiring frequent replacements and because of a low specific activity. It has been replaced by ^{103}Pd and ^{125}I isotopes.

^{106}Ru is a fission product and decays by β^-- and γ-emission to ^{106}Rh with a half-life of 1.02 years. The β^--emission energy of this transition has rather low energy of 39.4 keV. However, the daughter nucleus ^{106}Rh is also a β^--emitter with a much higher energy of 3.54 MeV and a half-life of 30 s. The active material is mounted into a silver sheet forming the surface of a plaque for exposure on one side and extra radiation shielding on the opposite side. The shape, form, and size can be adapted to specific applications on the skin. ^{106}Ru/^{106}Rh sources are no longer in use and have been replaced by ^{90}Sr/^{90}Y-β^--emitter.

12.2.8 Radioisotope ^{90}Sr

^{90}Sr is also a fission product that decays by β^-- and γ-radiation to ^{90}Y with a half-life of 28.8 years. Similar to ^{106}Ru, the β^--energy (0.546 MeV) is too low to be useful. However, the daughter nucleus ^{90}Y also decays by β^--, γ-emission into stable ^{90}Zr, where the β^--radiation has a maximum energy of 2.28 MeV. Formally:

$$\begin{aligned} {}^{90}\mathrm{Sr} &\Rightarrow {}^{90}\mathrm{Y} + \beta^- \ (E_{\max} = 0.546\,\mathrm{MeV}, 28.8\,\mathrm{a}) \\ {}^{90}\mathrm{Y} &\Rightarrow {}^{90}\mathrm{Zr} + \beta^- \ (E_{\max} = 2.28\,\mathrm{MeV}, 2.66\,\mathrm{d}) \end{aligned}$$

As for ^{106}Ru, the radioactive ^{90}Sr/^{90}Y material is embedded in a silver sheet and formed into a plaque that can be used for cancer treatment via β^--irradiation close to the body surface.

Some more isotopes have been used in the past for brachytherapy but are no longer in use. Others are still being developed (^{169}Yb, ^{170}Tm). There is ongoing research to find new pathways for convenient radioisotope production and isotope separation from targets, which remains an active field in radiochemistry.

12.3 Procedures

12.3.1 Interstitial and contact brachytherapy

There are two main procedures for brachytherapy treatment in terms of placement of the radioactive source: *interstitial* and *contact*.

For interstitial brachytherapy the radioactive source is implanted directly into the carcinoma, either temporarily for the time of irradiation or permanently. Typical organs for interstitial placement are breast (temporarily) and prostate (permanently).

For contact brachytherapy the radioactive source is either placed close to the target tissue through a body cavity such as the vagina, uterus, cervix or the oral cavity. Or the sources may be applied externally on the surface of the skin in case of skin cancer.

Brachytherapy is also distinguished according to the total dose (Gray, Gy) and dose rate (Gray/hour, Gy/h) administered to the patient:

- low dose rate (LDR) of up to 2 Gy/h
- medium dose rate (MDR) between 2–12 Gy/h
- high dose rate (HDR) above 12 Gy/h
- pulsed dose rate (PDR), where each pulse can range from LDR to HDR

For placement of radioisotopes interstitially or in contact an applicator is often used. This is a hollow tube that is inserted into the body and filled with capsules, seeds or wires. Hot loading implies that the applicator is preloaded and contains radioactive sources at the time of placement into the patient. Afterloading, in contrast, is a procedure where the applicator is placed first into the target position and the radioactive sources are loaded later, either by hand (manual afterloading) or by a machine (automatic remote afterloading).

For LDR and PDR applications the radioactive source typically remains in its implanted location for 24 h before removal. In the case of HDR the treatment time is a few minutes only, but may be repeated several times during a period of 1–2 weeks. PDR is applied for a minute or so and repeated several times within 24 h. In addition, LDR seeds may be implanted into the carcinoma, remaining permanently in the body.

LDR sources can be administered by the clinical personnel using their hands or simple tools. However, HDR requires the use of an afterloading machine. For both types of applications, contact or interstitial, first a hollow thin walled metal or plastic tube (applicator) is inserted into the cancerous region. After surgical placement of the tubes or catheters, the hollow tubes are connected to an afterloading machine. The afterloading machine contains radioactive wires, mainly ^{192}Ir, which can be slipped into the hollow tubes and quickly removed again. The insertion of wires is predefined and computer controlled using laser scanners. A typical afterloading machine is shown in Fig. 12.5.

Some selected examples (cervical, prostate, and breast cancer) may help to illustrate brachytherapeutical procedures.

12.3.2 Cervical cancer

Cervical cancer is a typical example where brachytherapy has been successfully applied. Either LDR or HDR brachytherapy may be performed although HDR brachytherapy is preferred since it does not require hospitalization and limits radiation exposure to hospital staff. For cervical cancer, HDR brachytherapy is delivered using intracavitary applicators in direct contact with the cancerous tissue. The applicator consists

Fig. 12.5: Remote afterloading platform for placement of high dose rate wires into predefined tubes (courtesy Elekta AB, www.elekta.com).

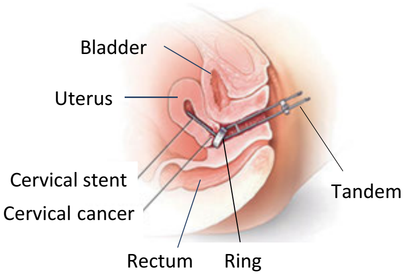


Fig. 12.6: Cervical cancer treatment via brachytherapy (reproduced from www.aboutcancer.com/intracavitary_radiation_treatments.htm).

of a tandem, ring, and a stent (see Fig. 12.6). First a cervical stent is placed into the cervical portion of the uterus for fixation. Following cervical stent placement, the ring and tandem are inserted. The tandem is placed into the uterus where the cervical stent is located. Then, the ring is placed next to the cervix where the tumor is located. Following placement, the applicator is connected to the afterloading machine where a radioactive source (usually ^{192}Ir) is located. When the treatment is initiated, the radioactive wire from an afterloading machine will quickly travel into the tandem and the ring where it will deliver the prescribed dose of radiation to the tumor. Once the treatment is completed, the applicators are removed quickly. This treatment can be combined with other oncological therapies, such as chemotherapy and EBRT.

12.3.3 Prostate cancer

Brachytherapy is widely accepted as a therapy for early stage *prostate cancer*, as long as the cancerous tissue is confined to the prostate gland. The permanent placement of short-lived radionuclide sources, emitting low energy photons is often used as the primary treatment. However, attempts are also being made to use fractionated or single session HDR brachytherapy in combination with EBRT.

Whenever *prostate cancer* is treated via brachytherapy, then usually radioactive seeds are implanted permanently into the prostate to provide a local and homogeneous distribution of radiation. The LDR seeds decay over weeks to months and remain in place even after complete decay. Seeds are implanted according to a well-defined treatment plan and placement is continuously monitored by ultrasound imaging through the rectum or by CT scans. A matrix helps find the proper location for the seeds to be inserted interstitially in the area between the testicles and the anus, i.e., intrasperineal implantation, as shown in Fig. 12.7. The method is known as transperineal interstitial permanent prostate brachytherapy (TIPPB).

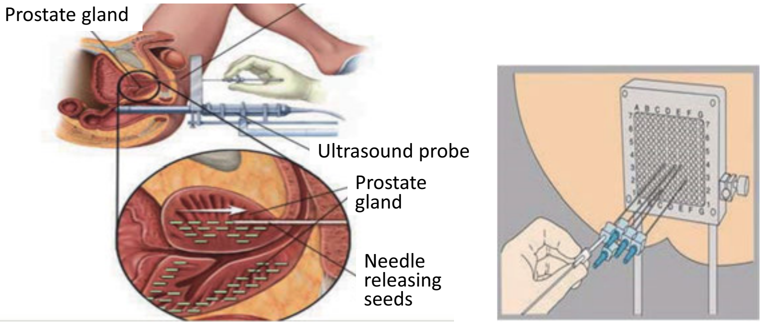


Fig. 12.7: Left: placement of seeds in prostate tissue controlled by ultrasound imaging. Right: matrix for locating seed positions (from https://en.wikipedia.org/wiki/Prostate_brachytherapy, © Creative Commons).

12.3.4 Breast cancer

The last example shows *breast cancer* treatment using interstitial flexible plastic tubes across the extension of the breast, Fig. 12.8 [4]. After placement of the tubes, ^{192}Ir wires are inserted and halted at predefined points for a specified time before moving to the next point. Repeating this procedure several times guarantees a homogenous distribution of radiation across the breast according to a predetermined radiation treatment plan. The main benefit of breast brachytherapy compared to EBRT is that a high dose of radiation can be precisely applied to the tumor while sparing radiation to healthy breast tissues and underlying structures such as the ribs and lungs.

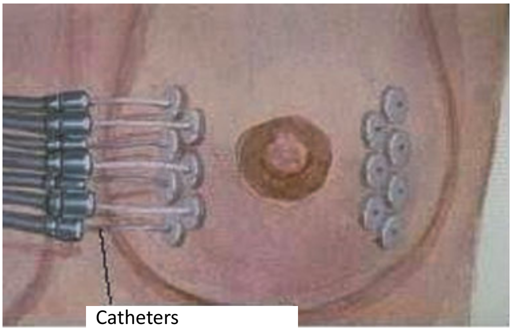


Fig. 12.8: Treatment of breast cancer with brachytherapy by first inserting flexible plastic tubes guiding radioactive wires from an afterloading machine (reproduced from [4], © Creative Commons).

Other less common applications of brachytherapy are the irradiation of intraocular melanoma on the rear side of the eyeball using an eye plaque loaded with ^{125}I seeds as an alternative to PBT (see Section 10.4).

12.4 Dosimetry

Application of LDR or HDR requires a detailed treatment plan determining the total dose to be administered and the dose rate for each fraction. For this purpose treatment planning systems (TPS) are commercially available and can be refined for special conditions by the radiation therapist.

We have already discussed dosimetry aspects in Chapters 4 and 8. However, brachytherapy requires special considerations because of the direct contact of radioactive sources with the cancerous tissue. This is a dosimetry topic with internationally agreed recommendations by the American Association of Physicists in Medicine (AAPM) and the European Society for Radiotherapy and Oncology (ESTRO). The last updated report was published in [5] and earlier reports can be found in this publication. The purpose is to provide dose estimation methods and dosimetric parameters for specific photon-emitting sources used for brachytherapy with photon energies exceeding 50 keV (mainly ^{137}Cs, ^{60}Co, ^{192}Ir). While the dosimetry of LDR sources is well established, new developments of HDR sources, pulsed dose rate (PDR) applications, afterloading techniques, and newly developed isotopes, require frequent updates of the recommended procedures.

In the simplest case of dose simulation, the tissue is replaced by an infinitely extended spherical phantom with the density of water. Radioactive sources are homogeneously distributed inside the phantom and treated as point sources (Fig. 12.9 (a)). Absorption of radiation is only considered by the phantom, not by the other sources. However, scattering of radiation by the other sources is taken into account. The dose calculation is performed according to the superposition principle. Each source radiates with a dose that drops off according to the $1/r^2$ law. The dose rate $\dot{D}(r, \theta)$ at the point $P(r, \theta)$ at distance r and angle θ of a single elongated source of length L in comparison to the dose at the standard position $P_0(r_0, \theta_0)$ and at right angle ($\theta_0 = 90°$) to the elongated source is given by the expression (Fig. 12.9 (b)):

$$\dot{D}(r,\theta) = S_K \Lambda \frac{G(r,\theta)}{G(r_0,\theta_0)} g(r) F(r,\theta).$$

Here S_K is the air kerma strength of the source, Λ is the dose rate constant in water, $G(r,\theta)$ is a geometry function considering the actual shape of the source as compared to a point source, $g(r)$ is the radial dose function, and $F(r,\theta)$ is the anisotropy function. All these factors are explained below in more detail.

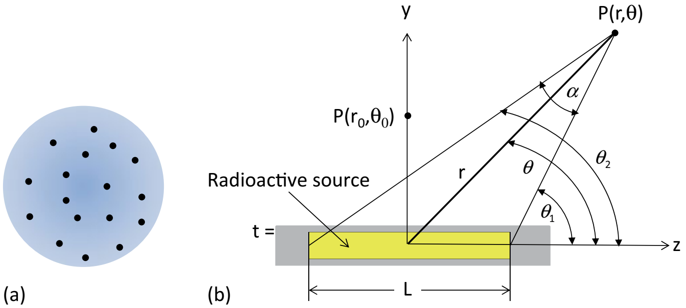


Fig. 12.9: (a) Phantom for calculation of dose. (b) Geometry for calculating dose and dose rate at particular points in the phantom.

The air kerma strength S_K is derived from the reference kerma rate in vacuum $\dot{K}_{\mathrm{vac}}$. As explained in Section 4.3, kerma stands for *kinetic energy release in matter* and is defined as the energy fluence of a photon beam (x-ray or γ-radiation) converted into kinetic energy of charged particles via the photo effect and/or Compton effect. Accordingly, $\dot{K}_{\mathrm{vac}}$ is defined as the dose rate in vacuum attributed to all photons of energy larger than a cut-off energy E_{min} at a distance d from the source center. The distance d can be any distance that is sufficiently large in comparison to the source size, such that the source can be treated as a point source. The term “in vacuum” refers to the fact that an actual measurement in air should be corrected for photon attenuation and scattering in air as well as from encapsulation material and for photon scattering from nearby objects like walls, furniture, ceiling. The cut-off energy E_{min} (typically 5 keV) excludes all low energy photons that increase the overall dose without contributing to the dose in tissue at distances exceeding 1 mm in tissue. The kerma strength S_K is defined as the kerma rate in vacuum at a reference distance d_{ref} times the references distance squared:

$$S_K = \dot{K}_{\mathrm{vac}}(d_{\mathrm{ref}}) \times d_{\mathrm{ref}}^2.$$

The SI unit of S_K is $[S_K] = \mathrm{Gy\,h^{-1}\,m^2}$.

The dose rate constant Λ is defined as the dose rate in water at a distance of 1 cm from the source on the transverse plane ($\theta_0 = 90°$) per unit kerma strength:

$$\Lambda = \frac{\dot{D}(r_0, \theta_0)}{S_K}$$

Thus, the dose rate constant Λ is a conversion factor that converts dose rates in vacuum (air) to dose rates in water. The SI unit of the dose rate constant Λ is $[\Lambda] = \text{Gy h}^{-1}/(\text{Gy h}^{-1}\,\text{m}^2)$.

The geometry function $G(r, \theta)$ takes into account the variation of the relative dose from a spatially extended source as compared to a point source. $G(r, \theta)$ reduces to $1/r^2$ for a point source and approximates to $\alpha/(Lr \sin\theta)$ for a cylindrically shaped line source, L and α are defined in Fig. 12.9.

The radial dose function $g(r)$ accounts for effects of attenuation and scattering in water on the transverse plane of the source ($\theta_0 = 90°$), excluding the normal $1/r^2$ fall-off, which is already included in the geometry function $G(r, \theta)$.

The anisotropy function $F(r, \theta)$ collects all dose anisotropy distributions outside of the source, including absorption in the source encapsulation of thickness t. It is defined as unity on the transverse plane, but decreases outside the transverse plane.

The model calculation outlined above has a number of shortcomings. Variations of composition, finiteness of tissues and organs, neglect of radiation shielding usually applied in clinical practice, etc. are not considered. Nevertheless it is a good starting point that can be refined for more realistic circumstances in advanced treatment planning systems (TPS). A number of specific TPS computer programs for brachytherapy applications are commercially available that help predicting desired dose rates and total prescribed dose with the intention to avoid common treatment mistakes.

12.5 Summary

1. Brachytherapy is an irradiation modality of cancerous tissue and is administered by direct contact of radioisotopes to the tumor.
2. The advantage of brachytherapy over external beam radiation therapy is the comparatively short treatment time and a local application that conserves healthy tissue.
3. In brachytherapy usually only the γ-radiation from radioisotopes is effective, the concomitant β-radiation is absorbed in the surrounding container walls.
4. Brachytherapy can be interstitial into the tumor or in outside contact with the tumor.
5. Most common radioisotopes for brachytherapy are ^{137}Cs, ^{192}Ir, and ^{103}Pd.
6. Most radioisotopes for brachytherapy are generated by neutron capture, but ^{103}Pd is produced by a number of other procedures including proton and γ-irradiation.
7. Most common applications are in the treatment of cervical cancer, prostate cancer, and breast cancer.
8. High dose rates require the use of applicators in combination with afterloading machines.

References

[1] Sonck M, Fenyvesi A, Daraban L. Study on production of ^{103}Pd and characterisation of possible contaminants in the proton irradiation of ^{103}Rh up to 28 Mev. Nuclear Instruments and Methods B. 2000; 170: 281–292.
[2] Mausner LF, Kolsky KL, Awasthi V, Srivastava SC. Production of ^{103}Pd by proton irradiation of silver. Journal of Labelled Compounds and Radiopharmaceuticals. 2001; 44: S775–S777.
[3] Tarkanyi F, Hermanne A, Kiraly B, Takacs S, Ditroi F, Csikai J, Fenyvesi A, Uddin MS, Hagiwara M, Baba M, Ido T, Shubin YN, Ignatyuk AV. New cross-sections for production of ^{103}Pd; review of charged particle production routes. Applied Radiation and Isotopes. 2009; 67: 1574–1581.
[4] Njeh CF, Saunders MW, Langton CM. Accelerated Partial Breast Irradiation (APBI): A review of available techniques. Radiation Oncology. 2015; 5: 90.
[5] Calatayud JP, Ballester F, Das RK, DeWerd LA, Ibbott GS, Meigooni AS, Ouhib Z, Rivard MJ, Sloboda RS, Williamson JF. Dose calculation for photon-emitting brachytherapy sources with average energy higher than 50 keV: Report of the AAPM and ESTRO. Med Phys. 2012; 39: 2904–2929.

Further reading

Hoskin P, Coyle C, editors. Radiotherapy in practice – brachytherapy, Oxford University Press; 2011.
Podgorsak EB, Radiation physics for medical physicists. 2nd edition. Springer Verlag; 2014.
Chao AW, Chou W, editors. Reviews of accelerator science and technology. Vol. 2. Medical applications of accelerators. World Scientific Publishing; 2009.
Baltas D, Sakelliou L, Zamboglou N. The physics of modern brachytherapy for oncology. Taylor & Francis; 2007.

Useful websites

Neutron capture activation calculator: https://www.ncnr.nist.gov/resources/activation/
Overview over all isotopes: www.periodictable.com/Isotopes/039.90/index.p.full.dm.html
Podgorsak EB, technical editor. Radiation Oncology Physics: A Handbook for Teachers and Students. Available from: http://www-pub.iaea.org/mtcd/publications/pdf/pub1196_web.pdf

Part C: Diagnostics and therapeutics beyond radiology

13 Laser applications in medicine

13.1 Introduction

Lasers are nowadays so ubiquitous in our daily life that one might forget how much sophisticated technology is implemented in these devices, which can be as tiny as the tip of a pin. The name *LASER* is an acronym for *Light Amplification by Stimulated Emission of Radiation*. Groundbreaking work which led to the invention of lasers was done in the 1950s using microwaves, called accordingly MASER. Masers as well as lasers are based on the principle of population inversion, an idea that Albert Einstein put forward in one of his publications [1]. In 1964 C. H. Townes, N. Basov, and A. Prokhorov shared the Nobel Prize in Physics for their fundamental work, which led to the technical realization of lasers.

When comparing light from a light bulb, which is a black body radiator, and light from a laser pointer, we notice three fundamental differences, schematically illustrated in Fig. 13.1. The light bulb radiates polychromatic, incoherent, and nondirectional electromagnetic radiation, some of this light is in the visible range of our photoreceptors in the retina. Laser light, in contrast, has a high monochromaticity and the light propagation is unidirectional. Furthermore, laser light features phase coherence. Light emitted from black body radiators has no coherence at all. Light bulbs have been basically abolished in recent years, but as a demonstrator for black body radiation they still serve a good purpose.

All lasers consist of three main components: (1) an active medium; (2) an optical pump; (3) a resonator.

Using these three components, the optical pump is used to populate an unoccupied upper atomic energy level in an active medium with electrons. The resonator warrants that stimulated transitions of electrons from the excited state to the ground

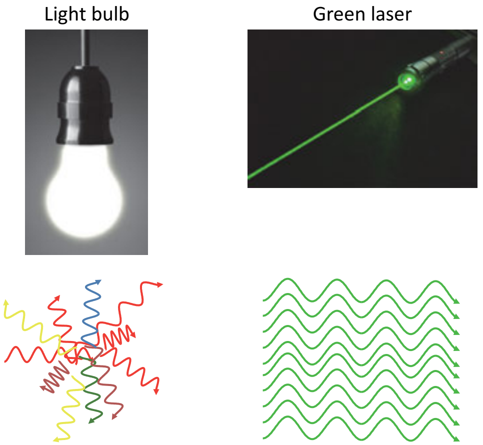


Fig. 13.1: Light from a light bulb and from a laser pointer. Light from the light bulb is polychromatic, propagates in all directions, and lacks phase conherence. In contrast, light from a laser is monochromatic, unidirectional, and features phase coherence.

DOI 10.1515/9783110553116-015

state, in phase with the wave field in the resonator, dominate over spontaneous transitions. The stimulated emission of radiation amplifies the intensity in the resonator, which is then released from the resonator cavity. Lasers come in a huge variety of specifications concerning wavelength, power, tunability, and pulsation. Because of this vast variety we have to confine ourselves to some basic features and main uses in medicine. In medicine we can identify five main areas of laser application:

1. dermatology
2. plastic surgery
3. oncology
4. ophthalmology
5. dentistry

There is a certain amount of overlap between some of these areas. For instance dermatology overlaps with plastic surgery and oncology.

Laser light operating from the far infrared to the ultraviolet spectral range is characterized as nonionizing radiation. Nevertheless potential health hazards are considerable. This is due to the fact that very high intensity is focused on a very tiny spot. An example may illustrate the potential danger. A 100 Watt light bulb may have a quantum efficiency of at most 10 %, i.e., only 10 % of energy is converted into light intensity. The emitted light is distributed over a solid angle of 4π. At 1 m distance from the light bulb the intensity is: $I = P/A = 10\,\text{Watt}/4\pi r^2 = 10\,\text{Watt}/4\pi \times 1\,\text{m}^2 = 0.8\ \text{Watt/m}^2$. In the case of a laser with the same power but focused on an area of 1 mm radius, the intensity is $10\,\text{Watt}/\pi \times 10^{-6}\,\text{m}^2 = 3 \times 10^6\,\text{Watt/m}^2$, which is factor $\approx 10^6$ higher than for the ordinary light bulb. At a 1 m distance the intensity is still the same since laser light is highly parallel. Even for a 1 mW laser (pointer) the intensity is still 380 W/m^2, which is 400 times higher than for the quoted light bulb. Lasers in the visible spectral range are categorized in classes according to their power output and potential danger:

Class 1 is inherently safe when enclosed in light-tight equipment, like CD players or spectrometers;
Class 1M is safe except when passed through magnifying optics;
Class 2 covers lasers with power up to 1 mW. If the eye is hit, the natural blink reflex (closing of eye lids within 0.25 s) prevents damage.
Class 2M is also safe if not viewed through optical instruments;
Class 3R (formerly 3a) includes lasers with power up to 5 mW, carrying a small risk of eye damage within blink time reflex. Staring into such a beam can likely damage a spot on the retina because of the focusing properties of cornea and lens;
Class 3B is likely to cause immediate damage to the retina upon exposure;
Class 4 lasers have the power to burn skin, and even scattered light may cause skin and eye damage. Medical, industrial, and scientific lasers are in this class.

Lasers in the nonvisible infrared and ultraviolet spectral region are particularly dangerous as the eye lid reflex is not operable. On the other hand, laser light in these spectral ranges is highly absorbed and does not reach the retina, but can nevertheless cause severe damage to cornea and skin.

The remainder of this chapter is organized as follows. First we introduce the basic physical concept of a laser and give an overview on their specifications. Then we discuss the interaction of laser light with tissue, which results in different effects depending on intensity, wavelength, and pulse duration. These different effects lead directly to various areas of medical application discussed in the last part. Understanding the laser-tissue interaction is hence particularly important.

13.2 Laser basics

13.2.1 Two level system

We start with a two energy level system as depicted in Fig. 13.2 (a) characterized by an energy difference $\Delta E = E_2 - E_1 = hf$. Furthermore, we assume that the lower energy level E_1 is occupied by one electron, whereas the upper level E_2 is unoccupied. If a photon with the energy $E_{\text{photon}} = hf$ hits this two level system, there is a certain probability that the photon will transfer all its energy to the electron in the ground level and lift it up to the excited upper state. This transition may be called resonant absorption or internal photoelectric effect. It occurs with a certain quantum efficiency, which depends on the atomic or molecular environment the two level system is embedded in. We have already seen such a case in Chapter 11/Vol. 1 for the cis-trans isomerization of retinal in the retina of our eyes, a transition with a quantum efficiency of about 20 %, i.e., on the average any 5th photon generates such a transition. If dipole allowed, the lifetime of the electron in the upper state is likely to be very short, in the order of a few femtoseconds (fs) to picoseconds (ps). Then the electron recombines with the hole in the ground state by emitting a photon of the frequency corresponding to the energy difference and in any arbitrary direction within a solid angle of 4π, as indicated in Fig. 13.2 (b). This transition is not what we aim for. We aim for a stimulated emission, which amplifies the photon field in the direction of the incident photon to yield an amplified and directional photon output, as sketched in panel (c). How can this be

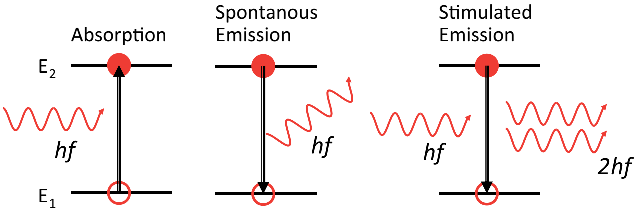


Fig. 13.2: Photon absorption (a) and emission (b) in a two level systems. Stimulated emission requires that the electron is in the excited state before the photon arrives, as indicated in panel (c).

achieved? As this simple illustration suggests, it requires a condition where the electron is already in the excited state before the incident photon arrives. This condition is referred to as *population inversion*. Then the incident photon and the photon from stimulated emission can overlap coherently such that the photon field is intensified.

Now we take a statistical approach, considering a total number of N identical atoms, each one featuring a two level energy system in a photon field of density ρ. There may be N_1 atoms in the ground state and N_2 atoms in the excited state. To determine the population of the levels and in particular the population difference for any given temperature we need to determine the rate of transition from 1 to 2 and vice versa. Level 1 is populated by spontaneous and stimulated emission from 2 to 1 minus stimulated absorption from 1 to 2. Vice versa, level 2 is populated by stimulated absorption from 1 to 2, minus spontaneous and stimulated emission from 2 to 1. The respective rate equations are therefore:

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = (B_{21}N_2\rho + AN_2) - B_{12}N_1\rho,$$
$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = B_{12}N_1\rho - (B_{21}N_2\rho + AN_2).$$

Here B_{21}, B_{12}, and A are the Einstein coefficients for stimulated emission, stimulated absorption, and spontaneous emission, respectively. Assuming that $B_{21} = B_{12} = B$, we notice that in equilibrium:

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = -\frac{\mathrm{d}N_2}{\mathrm{d}t}.$$

In fact we are interested in the population difference $\Delta N = N_1 - N_2$, since population inversion requires $\Delta N < 0$. ΔN can be evaluated from the rate equations:

$$\frac{\mathrm{d}\Delta N}{\mathrm{d}t} = 0 = AN - A\Delta N - 2B\rho\Delta N,$$

which yields for the population difference:

$$\Delta N = \frac{N}{1 + 2\frac{B}{A}\rho} = \frac{N}{1 + 2\frac{\rho}{C}},$$

where $C = A/B$. Even if spontaneous emission could be completely suppressed, the difference ΔN remains positive; meaning that population inversion in a two level system is intrinsically not possible. In fact the limit $C \to 0$ implies $\Delta N = 0$, i.e., equal population of the upper and lower level.

13.2.2 Three level system

In a three level system, shown in Fig. 13.3, we assume that electrons in the ground energy level E_1 can be excited to an upper level E_3 by electron collisions in a plasma (gas laser) or by optical pumping (solid-state laser). The excited state has a short lifetime and decays fast, but not to the ground state. Instead the transition takes place to an intermediate energy level E_2. The lifetime of E_2 is assumed to be much longer than the lifetime of E_3 since the transition $E_2 \to E_1$ is often dipole forbidden. The rate equations for the three level system can then be simplified, because stimulated absorption to level 2 and emission from level 2 via the photon field of the pump is missing. Optical pumping goes immediately and resonantly into level 3. Thus we have:

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = AN_2 - BN_1\rho,$$
$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = BN_1\rho - AN_2.$$

Again $\mathrm{d}N_1/\mathrm{d}t = -\mathrm{d}N_2/\mathrm{d}t$ and $\mathrm{d}\Delta N/\mathrm{d}t = 0$ in equilibrium. For the occupational difference we then obtain:

$$\Delta N = N\frac{1-\rho/C}{1+\rho/C}.$$

This equation can become negative for $\rho > C$, which is the case for a very high photon density in the resonator space. Hence population inversion becomes indeed possible. It should be pointed out that these rate equations are a simplified version of real laser activity since coupling to phonons and cooperative effects are neglected. Nevertheless, the population inversion is properly reproduced.

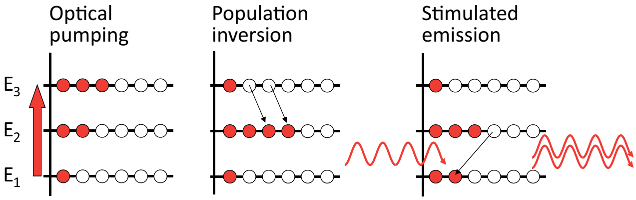


Fig. 13.3: Energy level E_2 is populated by optical pumping to energy level E_3, which is depleted quickly compared to the lifetime of energy level E_2. With this population inversion, stimulated emission from energy level 2 to 1 with corresponding amplification of the photon field becomes possible.

13.2.3 Basic components of lasers

When the transition $E_2 \to E_1$ takes place, the emitted photon with energy $hf_{21} = E_2 - E_1$ is trapped in an optical cavity with mirrors on both ends such that the photon will bounce back and forth forming a standing wave with the boundary condition $L = n\lambda/2$, where L is the cavity length and λ is the wavelength of the laser light. This condition holds for flat parallel mirrors. The emitted photon will then stimulate further $E_2 \to E_1$ transitions and thus amplify the photon density in the cavity. As the photons in the cavity form standing waves, the stimulated emission is in phase with the already existing photons, guaranteeing phase coherence among trapped photons. Furthermore, the optical cavity has an optical axis. Photons that travel parallel to the optical axis are in phase with the standing wave. Photons that have an angle against the optical axis either escape the cavity or are eliminated by destructive interference.

Figure 13.4 depicts a basic outline of a laser consisting of an optically active medium, flash lamps for optical pumping, mirrors on both ends of an optical cavity, one of them semitransparent to let the laser beam out of the cavity. Parallel to the optical axis the photon field is amplified, off-axis photons are lost and do not contribute to the amplification process. The optic medium can be a gas, a crystalline solid, or an organic material.

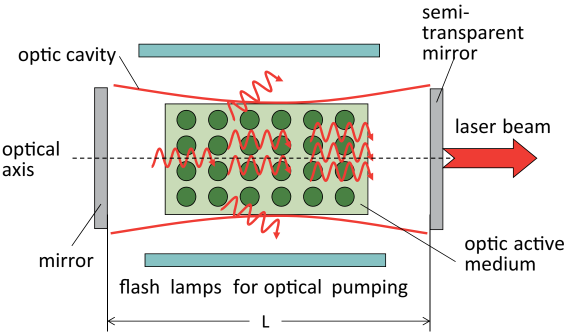


Fig. 13.4: Basic components of a laser, consisting of an optic medium, flash lamps for optical excitation, and mirrors for trapping the photons that form standing waves.

13.2.4 YAG laser

Figure 13.5 shows the energy scheme of a widely distributed Nd-YAG laser. The optic active medium is an yttrium aluminum garnet (YAG) crystal with the chemical composition $Y_3Al_5O_{12}$ doped with neodymium ions. YAG is a transparent colorless crystal. It can be doped with many different substitutional ions on the yttrium and aluminum crystal lattice sites, which then becomes colored depending on the absorption bands in the visible regime. The most prominent example is doping with neodymium Nd^{3+} ions substituting for yttrium. The relevant electronic transitions for the laser process are depicted in Fig. 13.5. The Nd^{3+} ions are optically pumped from the Nd^{3+} ground state $^4I_{9/2}$ to a higher lying pump band, mainly $^4F_{5/2}$. In the past, pumping was achieved with a broadband discharge lamp. Discharge lamps have now been replaced by more efficient narrow band laser diodes. The pump band becomes quickly depopulated by transitions to the $^4F_{3/2}$ atomic state. The lasing transition takes places from the $^4F_{3/2}$ excited state to $^4I_{11/2}$. The final transition from $^4I_{11/2}$ to the ground state $^4I_{9/2}$ is radiation-less, meaning that the energy difference is carried to the lattice in a

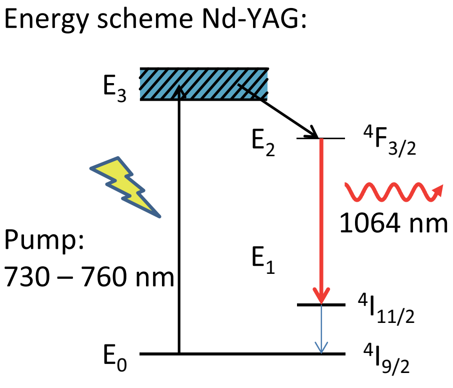


Fig. 13.5: Energy scheme and optical transitions of Nd in YAG.

nonradiative process such as lattice vibrations. So in fact the Nd-YAG laser is characterized by a four level system.

Now we take a closer look at the electronic states and transitions of Nd^{3+} as a representative example for other laser transitions. Nd is one of the lanthanide atoms, which are characterized by sequentially filling of the inner 4f electronic shell from zero (La, [Xe] $4f^0\,5d^1\,6s^2$) to a maximum of 14 electrons (Lu, [Xe] $4f^{14}\,5d^1\,6s^2$). The outer three electrons $5d^1\,6s^2$ are responsible for chemical bonding, while the inner 4f shell remains chemically protected and exhibits almost atomic-like spectroscopic levels. The electronic configuration of lanthanide atoms in the 3+ ionic state is then denoted as: [Xe] $4f^n$. Consequently the electronic configuration of Nd^{3+} is [Xe] $4f^3$. These three electrons in the 4f shell occupy the atomic levels according to Hund's rule (Fig. 13.6 (b)): all spins are parallel for less than half-filled shell and the lowest three m_L levels 3, 2, and 1 are occupied. This yields a total angular momentum of $L = 6$ and a total spin of $S = 3/2$. Combining L and S again according to Hund's rule for half-filled shells the lowest level is $J = L - S = 6 - 3/2 = 9/2$, the next one is $J = L - S + 1 = 11/2$ up to $J = L + S = 15/2$. These four ^{4}I levels are energetically split due to spin-orbit coupling, as shown in Fig. 13.6 (a). They are further energetically fanned out into sublevels due to crystal field effects, which are not shown.

The next higher term in the level scheme is the ^{4}F state, which is also split by spin-orbit coupling. The lowest of the sublevels is the $^4F_{3/2}$ state, which is triply degenerate with the electron configurations 4f((+3, +2, −2), (+3, +1, −1), (0, +2, +1)). Only one of the configurations (+3, +2, −2) is shown in Fig. 13.6 (b), and only "spin-up" configurations are considered.

The lasing transition from $^4F_{3/2}$ to $^4I_{11/2}$, marked in dark red in Fig. 13.6 (a), requires a change of the angular momentum by $\Delta L = 3$. This corresponds to an octopolar transition with a long 1/e lifetime of about 250 µs. This has two advantages. Because of the long lifetime, stimulated emission becomes possible and the transition with a wavelength of $\lambda = 1064$ nm is very sharp according to the Heisenberg uncertainty principle $\Delta E\,\Delta t \leq h$.

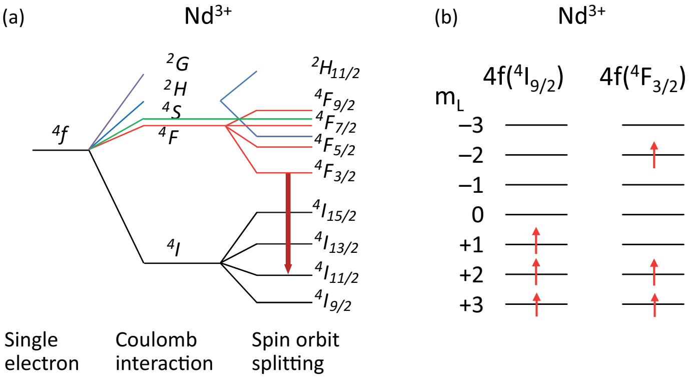


Fig. 13.6: (a) Energy levels of 4f Nd^{3+}; (b) electronic configuration of the $^4I_{9/2}$ and $^4F_{3/2}$ levels. The dark red arrow corresponds to the lasing transition.

There are several other transitions that can be used for laser activity such as the transition from $^4F_{3/2}$ to $^4I_{15/2}$. Furthermore, Nd ions can be replaced by Yb, Er, Ce, or Cr for a selection of additional lasing transitions with different wavelengths. We will refer to these transitions later on for specific applications.

13.2.5 Laser types, wavelengths, and units

As previously stated, there is a large variety of lasers to choose from, a short list is given here as an overview:

1. gas lasers: HeNe, Ar, CO_2
2. dye laser: dyes dissolved in water or in methanol, ethylene, glycol, etc.
3. excimer lasers: noble gas halogenates
4. solid-state lasers: Nd-YAG laser, ruby laser, Ti: sapphire
5. heterostructure semiconductor lasers: GaAs, GaAlAs, GaN
6. free electron laser

The wavelength regions where the various lasers have their fundamental emission lines are indicated in Fig. 13.7. Lasers can also be excited to higher harmonics, filling in the wavelength gaps seen in this chart.

Free electron lasers with extremely high brightness emitting light from the extreme ultraviolet to the hard x-ray regime are presently only of interest for science and technology, and therefore we will not discuss them any further. The other lasers listed may be used in a continuous mode (continuous wave laser or cw laser) or in a pulsed mode with pulse width ranging from milliseconds (ms) to femtoseconds (fs) and variable pulse repeat frequency. Pulsing is achieved by three different methods. For low frequencies it is sufficient to pulse the pumping flash lamp. For intermedi-

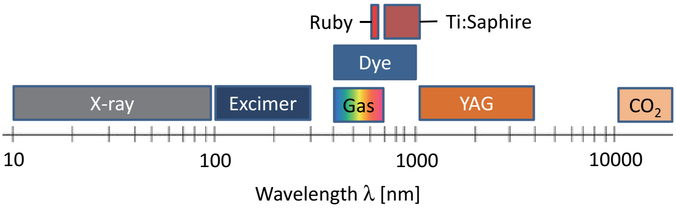


Fig. 13.7: Laser types and their wavelength regions.

ate frequencies the quality factor of the resonator is switched, which is referred to as *Q switching*. For very high frequencies mode coupling is used. These three methods will be explained further below.

In medical practice three types of lasers are most frequently used:

1. Nd-YAG solid-state lasers (Nd doped Yttrium Aluminum Garnet crystal), emitting at 1064 nm. This laser is used for treating hemangioma, glaucoma, removing tattoos, caries, and for fragmentation of stones in the urinary bladder.
2. CO_2 gas laser emit in the far infrared at 10 µm. Application areas are in neurosurgery, dermatology, plastic surgery, ophthalmology, and in general for diagnostics and therapy.
3. Excimer lasers emitting in the UV range are used for photorefractive keratectomy (Section 13.5.1).

For laser applications the following terms and units are used. Laser fluence is the time average energy delivered per unit area:

$$\text{fluence} = \frac{\text{energy}}{\text{area}}; \quad \text{units} = \left[\frac{\text{Joule}}{\text{m}^2}\right].$$

Note that the laser fluence defined here corresponds to the energy fluence defined in Section 4.3 for x-rays. In the case of a Gaussian beam profile, the spot size of the laser beam is taken at the I_0/e intensity level, where I_0 is the intensity at the center.

Intensity is defined as fluence per time, which yields power per area:

$$\text{intensity} = \frac{\text{energy}}{\text{area} \times \text{time}} = \frac{\text{power}}{\text{area}}; \quad \text{units} = \left[\frac{\text{Watts}}{\text{m}^2}\right].$$

The peak power of a pulsed laser is defined:

$$\text{peak power} = \frac{\text{laser pulse energy}}{\text{pulse duration}}; \quad \text{units} = \left[\frac{\text{Joule}}{\text{s}}\right].$$

Often fluence and intensity are quoted in units of J/cm^2 and Watts/cm^2, respectively. It is important to differentiate between time average versus single pulse specification.

13.3 Laser pulsation

The time average interaction of lasers with tissue is roughly speaking a local increase of temperature. However, in a short pulse much higher power can be concentrated leading to very different reactions of the tissue, such as photocoagulation, photoablation, and evaporation. Therefore, the ability to pulse a laser beam opens many more treatment possibilities compared to a cw operation. As mentioned before, there are three methods for pulsing a laser: mechanical, Q switching, and mode coupling.

13.3.1 Mechanical switching

At low frequencies pulsing and pulse shaping of a laser beam can be achieved by mechanically switching the pumping flash lamp on and off. Alternatively, the exit laser beam can be chopped with a rotating shutter or mirror outside of the resonator cavity. In this case each pulse has the same intensity as if the laser were operated in cw mode.

13.3.2 Q switch

Q switching is an alternative switching scheme for higher frequencies up to a few kHz and in particular for gaining much higher pulse intensities in each sharp pulse of a few nanoseconds length. Q switches are fast shutters placed inside of a laser resonator. The name “Q switch” refers to a change of the quality factor of the laser resonator. In the low Q state, the Q switch is closed and the cavity loss is high, preventing lasing. Losses occur at the mirrors and are due to absorption and scattering in the cavity. In contrast, in the high Q state the Q switch is open, losses are negligible, and lasing takes place.

A laser resonator containing a Q switch is schematically shown in the upper panel of Fig. 13.8 and the time sequence of the cavity properties during one operational cycle is plotted in the lower panels. The resonator consists of a pair of mirrors, which guide the continuously pumping light source through the laser crystal. Reflector and semitransparent output coupler complete the basic design. The Q switch is positioned inside the cavity in front of the output coupler.

At the start of a cycle the Q switch is closed, preventing lasing. While the Q switch is closed the pumping light source is continuously on and increases the stored energy (gain) in the laser crystal, i.e., electrons are being pumped into the upper state, but stimulated emission is hindered by interrupting the cavity resonance. When the Q switch opens by an external stimulus, the cavity loss quickly drops to zero. All the stored energy (gain) is now emitted in a short light pulse. The pulse duration is determined by the time it takes to empty the inverted electron state. The buildup of stored energy during the low Q state and short rise and fall times of the Q switch govern high

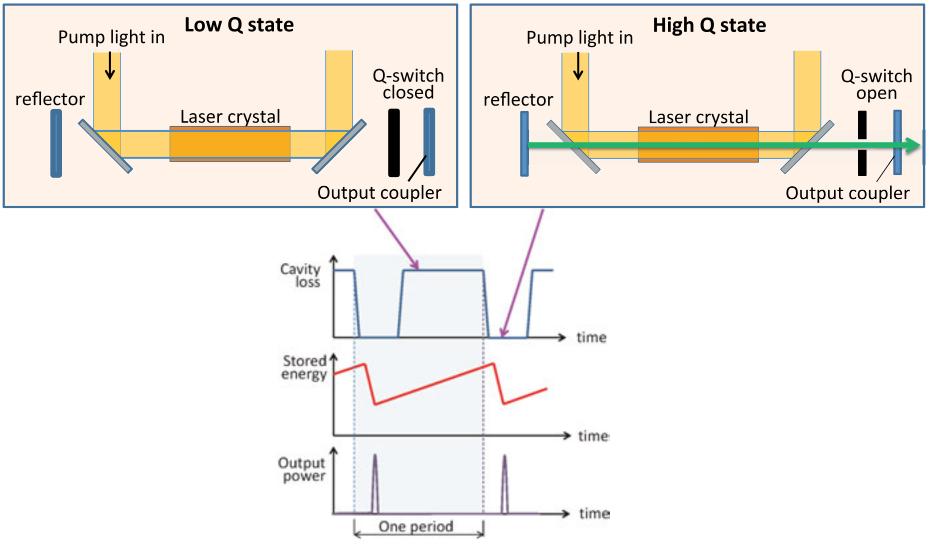


Fig. 13.8: Upper panels: laser cavity in the low Q state with closed Q switch (left) and high Q state with open Q switch (right). Bottom panels: the cavity loss, stored energy, and output power as a function of time in a Q switched laser cavity are plotted; the dashed vertical lines denote one cycle of the operation, which results in a single laser pulse. One cycle typically lasts anywhere from about 10 μs to 1 ms and is controlled by the Q switch (adapted from [2], © 2017 by PennWell).

pulse peak power and short pulse lengths. In general, the lower the repetition rate, the higher is the output power. High power YAG lasers typically have PRF of only 10 Hz and therefore a PRT of 100 ms.

Q switches are distinguished according to being actively or passively switched. The most common Q switches are active devices powered externally: electro-optic (EO) and acousto-optic (AO) modulators. The Q switch in EO modulators is a Pockels cell. The Pockels cell rotates the polarization of light passing through when a voltage is applied. By inserting this crystal between two crossed polarizers P and P′ as shown in Fig. 13.9 (a), light can only pass through this assembly when the applied voltage activates the Pockels cell and rotates the polarization of the light. EO modulators are typically used at low pulse repetition rates (up to a few kilohertz) and high pulse energies, commonly in the range of multiple mJ per pulse.

For higher pulse repetition rates (tens of kilohertz or more) and lower pulse energies, AO modulators are usually preferred. According to Fig. 13.9 (b), in an AO modulator a piezocrystal is excited by radiofrequency to set up a strain wave modulation that acts as a grating. Light passing through the modulator is diffracted at a fixed angle when the modulator is activated. If the light beam is deflected, the Q switch is in the off-state preventing lasing; otherwise in the on-state the beam is transmitted, allowing lasing.

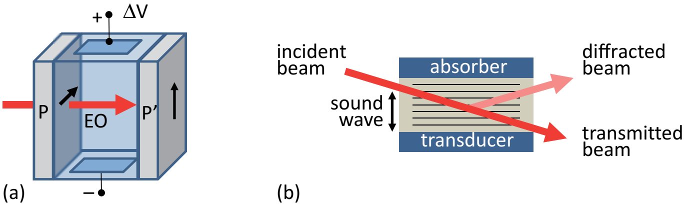


Fig. 13.9: (a) Electro-optic (EO) modulator rotates the polarization of the beam when activated. (b) Acousto-optic (AO) modulator deflects the light when the transducer is activated.

13.3.3 Mode locking

Mode locking is an advanced operational mode of lasers that leads to femtosecond pulsing and is finding more and more applications in technology and in medicine [3, 4]. In simple terms the operational principle can be described as follows. In lasers the standing waves fulfill the boundary condition $L = n\lambda/2$, where L is the cavity length and λ is the wavelength of the laser light (see Fig. 13.4). If the optical transition has a narrow bandwidth, this condition can be fulfilled for only a few discrete wavelengths. Conversely, if the optical transition has a broad band, the boundary condition can be fulfilled by thousands of different wavelength as $L \gg \lambda$. All allowed wavelengths, within a Gaussian width $\delta\lambda$, will overlap and interfere, where the interference of waves with different wavelength and random phase will lead to a beating effect (Fig. 13.10 (a)). The beating effect results in a chaotic sequence of pulses with different intensities and pulse lengths. If it can be managed that these different wavelengths are coupled in a fixed phase relationship, the interference of these coupled modes yields a periodically reoccurring wave packet with a period of $T = 2L/c$ and a pulse length of:

$$t_\mathrm{P} = \frac{2L}{cN} = \frac{T}{N},$$

where N is the number of modes in the cavity fulfilling the boundary condition (Fig. 13.10 (b)) and c is the speed of light. The number of modes that can participate in phase-locked interference is limited by the band width $\delta\lambda$ that still allows stimulated emission:

$$N = \frac{4L(\delta\lambda)}{\lambda_0^2},$$

where $\lambda_0 \gg \delta\lambda$ is the center wavelength. Now we can rephrase the pulse length as follows:

$$t_\mathrm{P} = \frac{2L}{cN} = \frac{\lambda_0^2}{2c(\delta\lambda)}.$$

Thus the pulse becomes shorter the broader the bandwidth $\delta\lambda$ is. The argument is similar to the one for diffraction gratings in optics. The higher the number of slits in a grating, the sharper the central diffraction maximum is and the lower the intensity

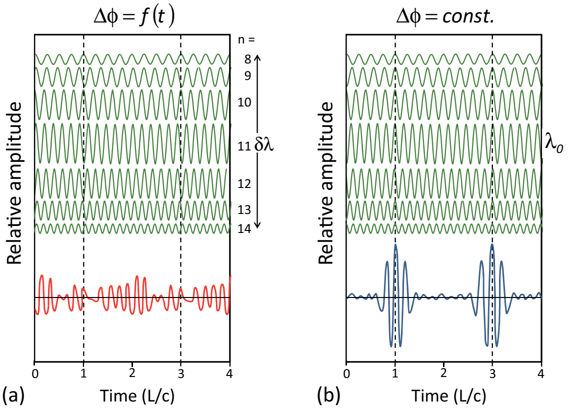


Fig. 13.10: Interference of waves with random and time-dependent phase (a) and with fixed phase relationship (b) are compared. The dashed lines are guides for identifying the phase relation of waves with different wavelengths. The time scale is in units of L/c. n is the number of wavelengths fitting between the vertical dashed lines, $\delta\lambda$ is the wavelength spread of participating modes with center wavelength λ_0 (adapted from [3, 4]).

of side maxima is. This very short pulse travels around the cavity in a time $T = 2L/c$ and every time it reaches the output mirror the laser emits part of the pulse intensity. Therefore the pulse repeat frequency PRF is given by the reciprocal roundtrip time of the laser cavity: RPF $= 1/T = c/2L$. The PRF corresponds to the frequency difference of two neighboring modes that fulfill the boundary condition: $n\lambda_n = (n+1)\lambda_{n+1} = 2L$. Rephrasing in terms of frequency differences, we find: $\Delta f = f_{n+1} - f_n = c/2L$. As an example we assume a central wavelength of 500 nm and a bandwidth of 5 nm. Then the pulse width t_p becomes 83 fs, and for a cavity length of 2 m the PRF is 75 MHz. For a bandwidth of 100 nm the pulse width becomes reduced to 4 fs while the PRF remains the same. A sequence of pulses with intrinsic pulse width t_p of a mode coupled laser is schematically shown in Fig. 13.11.

A continuous laser can be turned into mode locking pulsed operation if the gain is higher in mode locking than in cw mode. Mode locking can be achieved by using active or passive switches, similar to the Q switched lasers. Active mode locking requires gating of the cavity at the intrinsic PRF, determined by the roundtrip time of the pulses

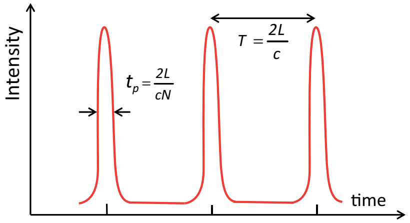


Fig. 13.11: Pulse sequence of a mode coupled laser.

using for instance an AO modulator. Passive mode locking, on the other hand, can be performed by using the so called Kerr lens effect in the gain material, Fig. 13.12. The Kerr lens effect occurs through nonlinear optical properties of the laser crystal at very high light intensities. The high peak intensity of the mode-locked pulse changes the refractive index of the laser crystal locally. The refractive index is higher at the center of the crystal than at the edges according to the beam profile. This refractive index gradient acts as a lens and leads to self-focusing of the beam, enabling it to pass through an internal aperture. In contrast, in cw operation the beam hits the aperture and the beam intensity suffers from clipping.

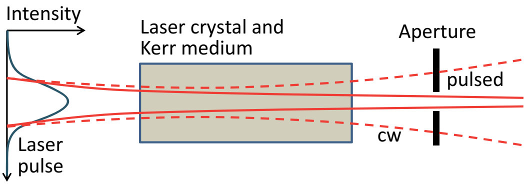


Fig. 13.12: A mode-locked laser with beam focusing due to nonlinear optical properties of the laser crystal at very high intensities.

13.4 Laser interaction with tissue

13.4.1 Overview

When a laser beam hits the surface of the skin, three main effects occur: the laser beam is partially transmitted, partially reflected, and partially scattered. As the surface of the skin is rough on the length scale of the wavelength used, (specular) reflection can be neglected. The diffuse scattering of photons at rough surfaces is the reason why we are able to recognize opaque objects (Fig. 13.13). With specular reflection our visual perception could easily be deceived. Scattering also occurs below the surface and this is the main reason for the finite penetration depth of laser light in the skin or tissue. However, the details depend on the wavelength of the light, the intensity, the area of exposure, and the time of exposure. For instance, if the laser beam causes photocoagulation, the penetration depth changes over time from deep to shallow. When UV light strikes the skin, it produces melanin known as tanning of the skin to protect it from genetic damage. Tanning reduces the penetration depth of light.

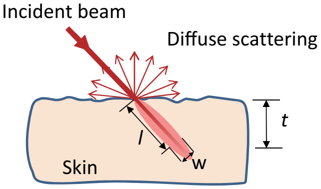


Fig. 13.13: Incident laser light causes diffuse scattering outside and penetration inside. t is the penetration depth and l is the penetration length. The width w is due to scattering of light in the tissue.

In spite of the huge parameter space for light interaction with skin/tissue, the penetration depth of laser light is often described with the usual Lambert–Beer law, according to which the transmitted intensity at depth l is expressed as:

$$I(\lambda, l) = I_0(\lambda) \exp(-\alpha(\lambda) c l) = I_0(\lambda) \exp(-l/L).$$

Here L is the penetration length, α is the attenuation coefficient, and c is the molar concentration of a solute in a solvent. Applied to tissue, c is an ensemble average over different molecules (proteins, membranes, cartilage, etc.). As the penetration depth and the attenuation coefficient cannot be determined in vivo, the best alternative is a water solution containing the main chromophores of the tissue: melanin and hemoglobin. Melanin is responsible for the color of the skin and the tan developing under sunshine. The attenuation coefficient and the penetration depth of water, HbO_2 and melanin are shown as a function of wavelength in Fig. 13.14. Water has a transparency window in the visible range of the optical spectrum. In this window the attenuation coefficient of HbO_2 and of melanin is particularly high. Oxygenated HbO_2 and deoxygenated Hb (not shown) can clearly be distinguished at the wavelength 560 nm and between 600 nm and 700 nm. The absorbance properties of Hb and HbO_2 are discussed in Section 8.6.5/Vol. 1. Also indicated in Fig. 13.14 are the wavelengths of the main lasers used for various applications discussed further below. According to Fig. 13.14 the main absorption of laser light in the visible regime is due to blood and melanin. Obviously the blood vessels can be better seen under white skin than dark skin. Typical penetration depths for different lasers in tissue are listed in Tab. 13.1.

Tab. 13.1: Emitted wavelength and penetration depth in tissue of typical lasers used in medicine.

Laser type	Emitted wavelength [nm]	Penetration depth [mm]
CO_2	10.60	0.10
Nd:YAG	1.064	6.00
Ar+	0.4880, 0.5145	2.00
Excimer	0.126–0.351	0.01

Considering laser-tissue interactions, usually five types are distinguished and listed as follows [4, 5]:

1. photomechanical interaction
2. plasma-induced ablation
3. photoablation
4. photothermal interaction
5. photochemical interaction

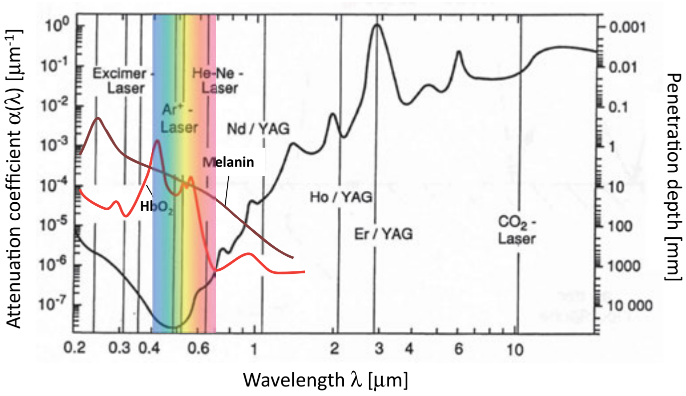


Fig. 13.14: Attenuation coefficient of water, melanin, and oxygenated hemoglobin as a function of wavelength. Note that all values plotted are on a logarithmic scale.

These main interaction types are related to the characteristic properties of lasers with respect to average power density and pulse length according to the graph shown in Fig. 13.15 [5]. Note the double logarithmic scales. The dashed blue lines indicate the time average laser power within a corridor of 1 J/cm^2 (left dashed line) and 10^3 J/cm^2 (right dashed line). In the top left area the laser flux of single pulses is very high due to very short pulses. This is the area of fs lasers. As we continue along the diagonal

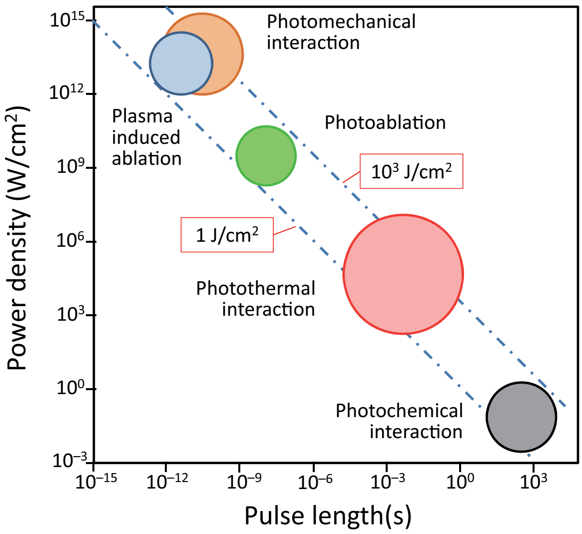


Fig. 13.15: Main types of laser-tissue interaction in relation to laser characteristics. Note the double logarithmic plot of laser power density of a single pulse versus pulse length (adapted from [5] by permission of Springer-Verlag, 2017).

dashed lines, the power density decreases and the pulse length increases. In the bottom right corner we find lower power cw lasers. The circles indicate areas where one of the above listed processes predominantly takes place and these are described in more detail in the following subsections. We start at the lower end and successively increase the laser flux up to the top left corner. Photochemical interaction is a special interaction mediated by uptake of drugs and will be presented at the end of this section.

13.4.2 Photothermal interaction

If the intensity is not too high, photons of the laser couple to the vibrational modes of molecules and increase the temperature locally. From sunbathing we know that photons are converted to heat in the skin. Photons are scattered in the tissue over some diffusion length and finally become absorbed by molecular vibrations, which increases their kinetic energy recognized as heat. The same processes also occur with a laser beam, but in a much smaller area with a much higher intensity. The result can be a denaturation or destruction of the tissue. Figure 13.16 shows an example of a laser beam hitting the surface of the skin at normal incidence. While the laser beam intensity is attenuated, it spreads out to the sides by scattering and heat conduction, yielding the indicated isointensity contour lines.

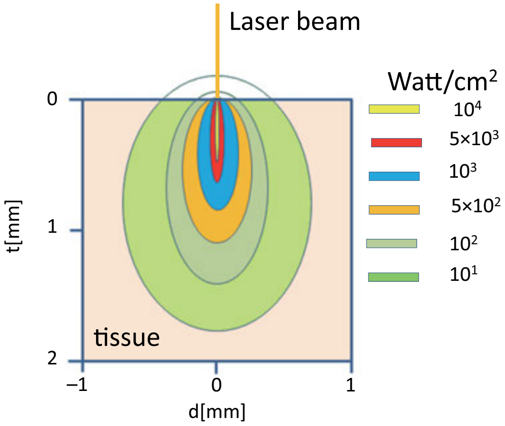


Fig. 13.16: Light scattering and heat conduction enlarges the volume of the laser interaction.

Heating by the laser beam due to absorption has three main effects, depending on the intensity and duration of laser beam exposure:

1. Hyperthermia implies a local and moderate tissue heating to several degrees above body temperature for some tens of minutes. This local hyperthermia causes cell death due to changes in enzymatic processes. Hyperthermia by bare laser beam exposure is difficult to control, but has effectively been applied medi-

ated by plasmonic [6] and magnetic nanoparticles [7], which will be discussed in Chapter 14.

2. Coagulation occurs when local temperatures reach 50–100 °C within a matter of seconds, resulting in a complete and irreversible destruction of tissue via denaturation of proteins and fibers. Boiling an egg is a vivid example of the denaturation process that occurs during coagulation.
3. Volatilization implies loss of tissue material by evaporation. If this loss is local, there is a temperature gradient between the hot spot and neighboring tissue. Within the edges of the hot spot a zone of coagulation necrosis, i.e., cell death by coagulation, can be recognized. With volatilization as well as with coagulation tumors can be destroyed preferentially in the skin due to the limited penetration of the laser light.

Two examples may illustrate how coagulation and volatilization can be used for specific treatments. Cutaneous angiomas, also known as port-wine stain, are a reddish skin color due to abnormal blood vessels in the upper dermis, sometimes occurring upon birth. These stains can be removed by a laser beam which is absorbed more by hemoglobin causing coagulation than by the dermis. According to the plot in Fig. 13.14 an Ar laser would be most suitable for such a treatment. The second example concerns the treatment of virus infected warts, which may appear anywhere on the skin, from hand, feet, to genitals. They can be treated by volatilization with a CO_2 laser. Upon volatilization the lesion disappears into smoke, and the coagulation necrosis seals the blood vessels in the skin without bleeding.

13.4.3 Photoablation

Laser-induced processes which cause vaporization of tissue material without thermal lesions at the margins are referred to as photoablation. This effect requires very energetic lasers emitting at short wavelengths in the UV. Two properties are combined to make photoablation effective: the high electric field at short wavelengths and high flux ($\approx 10^9$ Watts/cm^2) of the laser field breaks chemical bonds without generating heat at the edges. The resultant molecular fragments then expand in a plasma plume carrying the thermal energy away without destroying the surrounding tissue. Secondly, the interaction of the short wavelength laser light is very superficial as the absorption of UV radiation in the range of 200–300 nm is very strong. For this reason excimer lasers are used for photoablation which emit at 193 nm (ArF), 248 nm (KrF), or 308 nm (XeCl). Photoablation is mainly used in ophthalmology and in particular for photorefractive keratectomy (LASEK and LASIK), which is discussed further below.

13.4.4 Plasma-induced ablation

Increasing further the laser flux, plasma-induced ablation occurs for pulses shorter than nanoseconds. Plasma-induced ablation is solely based on tissue ionization in contrast to plasma creation followed by shockwave propagation, discussed in the next section. Nd: YAG and Ti: sapphire lasers are used for plasma-induced ablation with ps to fs pulse length and flux on the order of 10^{12} W/cm^2. Such a high power density leads to electrical fields of 10^7 V/cm comparable with electric fields that orbiting electrons experience in atoms. Therefore, the Q switched or mode-locked lasers can ionize molecules in the biological tissue. This creates a very high density of free electrons during a very short time with values reaching 10^{18}/cm^3 due to avalanche amplification. Hot free electrons from ionization are accelerated to high energies and collide with other molecules leading to further ionization. Light electrons and heavy ions move with different velocities typical for highly ionized plasma. The plasma allows well-defined removal of tissue by “optical breakdown” (plasma explosion) without evidence of thermal or mechanical damage to surrounding tissue. The applications are similar to those of photoablation.

13.4.5 Photomechanical interaction

Photomechanical effects may be caused either by the creation of a plasma, by an explosive vaporization, or by cavitation. In any of these cases the photomechanical effect is associated with the generation of high pressure shock waves.

Plasma creation: using nano- or picosecond pulsed Nd: YAG lasers, very high flux (same units as intensity: J cm^{-2} s^{-1}) can be concentrated over a very small area reaching flux values between 10^{11} to 10^{14} W/cm^2. This high flux creates a plasma by ionizing atoms. At the border of the ionized area a steep pressure gradient occurs that causes the shock wave to propagate outwards. The expansion of the shock wave causes mechanically destructive effects. This effect is, for instance, used for removing a thin gray coating on an artificial lens that may develop after cataract treatment. By carefully focusing the laser, the tissue membrane on the back side of the lens can be vaporized without harming the retina.

Explosive evaporation: if the exposure time of the laser is shorter than the characteristic time for heat diffusion in the tissue, the laser accumulates heat in a local spot that finally causes an explosive vaporization of the target tissue. As thermal diffusion is rather slow, the laser pulse width to be chosen can be comparatively long for this effect to occur. A Q switched Nd: YAG laser with a pulse length of 100 µs is typically employed for removal of tattoos via explosive vaporization. The large ink particles in the skin explode and are redistributed into smaller particles which are reabsorbed by the skin and eventually removed. It is important that the laser beam is absorbed by the ink pigments. In the case of colored tattoos, the application of several laser lines may be required.

Cavitation: if a laser pulse creates a hot spot which is not only spatially confined but also mechanically hindered from exploding, then the high pressure gas bubble will implode as soon as the laser pulse is turned off. The subsequent implosion is known as cavitation. Using this phenomenon, urinary stones can be fragmented by a pulsed laser beam with µs pulse length. For this purpose an optical fiber is placed under endoscopic control through the urethra into the bladder, and the stone is broken into pieces small enough for the fragments to be cleared out naturally through the urethra.

13.4.6 Photochemical interaction

Photochemical interaction of lasers with tissue involves the uptake of photosensitive drugs [8, 9]. The treatment steps are explained in Fig. 13.17. First, the tumor area must be identified. Then a photosensitizer is injected in the area of the tumor. The drug should remain and eventually accumulate in the tumor cells while clearing from normal tissue. Clearance may take a few hours up to a few days. Then the area is irradiated by a low power laser beam, lifting the photosensitizer from the ground state to an excited level $S \rightarrow S^*$. The laser light has to be chosen so as to cover the absorption maximum of the photosensitizer S. The excitation energy is then transferred to free and ambient oxygen molecules in the tissue. Free oxygen in the ground state is in a spin triplet state, i.e., the highest occupied molecular level is occupied by two unpaired electrons, one from each O atom, see Fig. 13.18. The spin triplet state is reponsible for the paramagnetic properties of oxygen gas as well as oxygen liquid. Upon energy transfer from the photosensitizer, O_2 molecules are switched from the triplet to the excited singlet state. In the singlet state, $O_2{}^*$ is extremely reactive, as we have learned already in Chapter 8 on cancer treatment. The procedure, termed *photodynamic treatment* (PDT), is confined to the area of the administered drug, contrasting treatments that involve dissipation of heat. Unlike cancer treatment by ionizing radiation aiming at the destruction of DNA in the tumor cell, PDT relies on the attack of cellular components by phototoxic singlet oxygen. If sufficient oxidative damage has been done, the target cells will react by necrosis within the illuminated area and eventually total tumor recession. Necrosis sets in within a few days after laser exposure. PDT is a noninvasive cancer treatment method for tumors just under the skin or at the surface of inner organs that can be reached by a light source, for instance with the help of an endoscope. This excludes cancers that have grown deep into an organ or which have already spread. A number of photosensitizers exist for PDT, which can be divided into porphyrins, chlorophylls, and dyes.

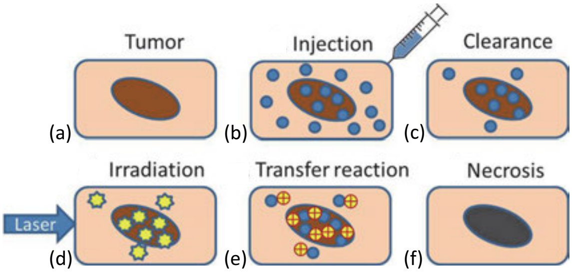


Fig. 13.17: Photodynamic cancer treatment with a photosensitizer that transfers the excitation energy from a laser to ambient oxygen, which transforms to phototoxic singlet oxygen, killing cancer cells. (a–f) are the different steps in the process.

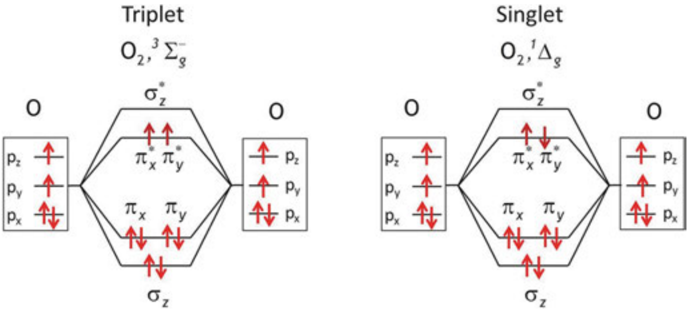


Fig. 13.18: Atomic and molecular orbitals of oxygen in the triplet and singlet state.

Tab. 13.2: Laser-tissue interaction for different lasers and laser flux.

Interaction	Laser type	Flux [Watts/cm^2]	Pulse length [s]	Tissue reaction	Treatment
Photothermal	CO_2	10^6	10^0–10^{-3}	Coagulation	Hyperthermia, stains, warts
Photoablation	Excimer	10^7–10^8	10^{-7}–10^{-8}	Material removal	LASIK, LASEK
Plasma induced ablation	Nd: YAG Ti: sapphire	10^{12}	10^{-12}–10^{-15}	Removal of tissue by optical breakdown	Refractive corneal surgery
Photo-mechanical	Nd: YAG Ti: sapphire	10^{11}–10^{14}	10^{-7}–10^{-13}	Shock wave, Plasma cavitation	Lens and stone fragmentation, tattoo removal
Photochemical		1	cw	Photodynamic therapy	Skin tumors

Necrosis, in contrast to apoptosis, is a sudden and unprogrammed cell death due to trauma, photocoagulation, or in the present case due to phototoxic singlet oxygen molecules. Necrosis does not follow an apoptotic signal transduction pathway. Instead, various receptors are activated, resulting in the loss of cell membrane integrity and an uncontrolled release of cell content into the extracellular space. Hence, the cell fragments are not properly disposed of by phagocytosis. Therefore it often appears necessary to remove necrotic tissue by surgical means, a procedure known as debridement. However, recent studies indicate that the mechanisms of apoptotic and necrotic cell removal are not as different as assumed in the past [10], which will likely change debridement procedures in the future.

Table 13.2 gives an overview on the different laser tissue interactions discussed in this section, including their main applications in medicine.

13.5 Laser applications in ophthalmology

A number of laser applications in medicine have already been mentioned. Apart from tumor treatment the main applications are in ophthalmology and these methods are discussed here to some more extent.

13.5.1 Photorefractive keratectomy (PRK)

Myopia and hyperopia (see Chapter 11/Vol. 1) can be corrected by laser application. In both cases the curvature of the cornea is changed in order to bring the focus on the retina: for myopia the curvature is decreased, for hyperopia the curvature is increased. Before discussing the different corrective laser methods that are applied for these corrections, we first have a closer look at the structure of the cornea shown as a cross section in Fig. 13.19. The cornea has an average thickness of 555 μm and is thinner at the center than at the perimeter. The central and most important layer is the stroma

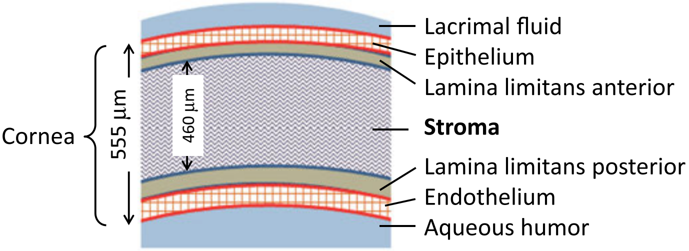


Fig. 13.19: Structure of the cornea, composed of several layers that protect the central stroma. The schematic outline is not to scale, but the respective thicknesses are indicated.

with an average thickness of 460 µm. The stroma is protected on the outside and inside by lamina layers (lamina limitans anterior and posterior) and by epithelium as well as endothelium layers. Photorefractive keratectomy implies a change of the stroma. In order to get to the stroma, a cut through the top layers is required. Two distinct methods are usually applied: *Laser-Assisted Interstitial Keratomileusis* (LASIK) and *Laser Epithelial Keratomileusis* (LASEK). Both methods use ArF excimer lasers with a wavelength of 193 nm. The merits of this laser type are a high absorption coefficient in the UV, the possibility to sharply focus the laser beam for precise ablation of tissue, no development of heat in the treated tissue, and no transparency of the laser light in the different layers and components of the eye, i.e., the laser light cannot reach or harm the retina. Both methods, LASIK and LASEK, and their differences will be explained in the following.

LASIK is a method for correcting the refractive properties of the cornea, which contributes to 75 % of the total refractive power of the eye. In the first step a thin and sharp 100 to 140 µm deep cut is performed by an oscillating metal blade, called microkeratome, to open the cornea cover and to get access to the stroma. This is shown schematically in panels 2 and 3 of Fig. 13.20. The cut contains components of the epithelium and the stroma. The layer cut, called flap, is not lifted off. Instead a hinge is left on the side (panel 4), so that the 'flap' can be closed again after surgery. Then the stroma is ablated with an excimer laser according to the predefined refractive properties aimed for. Plano-LASIK implies a flattening of the stroma within the area of the laser beam using one broad beam shown in panel 5 of Fig. 13.20. In contrast, with a focused excimer laser beam the exact required and predefined curvature can be ablated, as indicated in frame 6 of the same figure. After finishing ablation, the flap is closed again and sealed (panels 7 and 8). The opening of the cover can also be done with a femtosecond laser, which provides a sharper cut and avoids complications that may occur with the use of metal blades. Furthermore upon closing the flap, better adhesion between flap and stroma is achieved.

LASEK: In contrast to LASIK, in LASEK the 50 µm epithelium is dissolved in alcohol, moved to the side, and disposed of. The disposed of part does not contain stromal components. Then the treatment of the stroma follows in the same manner as for LASIK. Since there is no flap for closing, the eye needs to be protected for a few days by a contact lens until the epithelium has regenerated. LASIK and LASEK

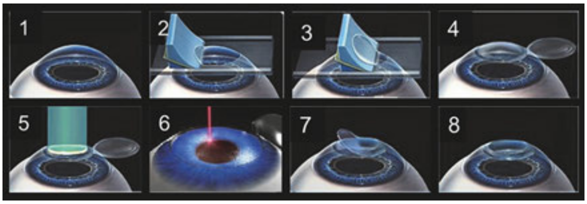


Fig. 13.20: Surgical procedure in LASIK (adapted from http://eyeclinicpc.com/lasik.html).

are only distinct in the last step of the treatment: LASIK recycles the epithelium layer whereas LASEK disposes of the epithelium and protects the eye by a contact lens. The epithelium has the capability to regenerate, whereas the stroma does not. Changes to the stroma are permanent.

Correcting myopia requires a flattening of the cornea achieved by material removal. Hyperopia is corrected for by an increase of the refraction power of the eye and entails material addition. However, this cannot be realized by ablation methods. Nevertheless, hyperopia can be corrected with LASIK methods using a “trick”. Instead of ablating the center of the stroma like for correcting myopia, a ring around the center is ablated which results in an increased curvature at the center by elastic relaxation. The depth of the ring determines the curvature of the center and thereby the refraction power that can be corrected for. Both correction methods for myopia and hyperopia are compared in Fig. 13.21.

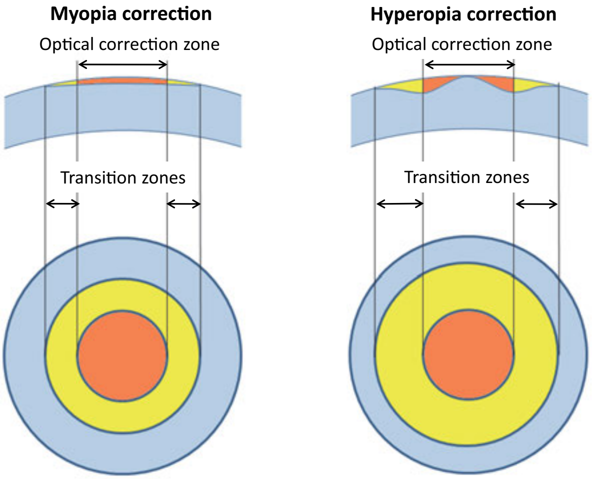


Fig. 13.21: Myopia and hyperopia correction with LASIK are compared. Myopia requires a flattening of the cornea curvature at the center within the optical zone, hyperopia requires a steepening of the curvature. Myopia correction is achieved by removal of material in the center; for hyperopia correction a ring around the center is removed.

In all cases the success of the treatment depends on a careful and thorough characterization of all optical properties of the eye, which includes the refraction power, astigmatism, thickness and thickness gradients of the cornea (pachymetry performed by ultrasound), and the entire topography of the eye surface. During surgery markers are placed on the eye that allow tracking of any eye movement and corresponding correction of the laser delivery.

Although LASIK/LASEK surgery has a high success rate, one problem that may occur after surgery should be mentioned. The epithelium layer consists of basal cells, which have the capability to divide and therefore to regenerate. In contrast, the stroma consists of fibrous collagen material that holds a large amount of water controlling the transparency of the cornea. Stroma and the bordering membranes (lamina limitans) do not contain regenerative cells. This is an important property for a lasting LASIK/LASEK refractive correction. However, both epithelium and stroma are infiltrated by nerve fibers (innervation). In fact, the cornea holds the densest sensory innervation of the body, more than the skin. These nerve fibers have a multitude of tasks, such as the control of lacrimal fluid and the cornea reflex that closes the eye lids in the case of lacrimal fluid surplus or any other irritation of the epithelium. Upon LASIK or LASEK surgery, a portion of nerve fibers are damaged and patients may experience persistent eye pain. Some of the nerve fibers regenerate only within a year after surgery. Further information on LASIK and LASEK can be found in the review of [11].

13.5.2 Diabetic retinopathy

As a result of diabetes two conditions may develop on the retina summarized as diabetic retinopathy: diabetic macular edema and proliferative retinopathy. Diabetic retinopathy is the most frequent cause of blindness for ages from 25–65 years in Europe and North America. In both cases the application of lasers promises curative treatment.

Diabetic macular edema (DME): macula edema refers to some leakage in the macula that potentially causes decreased vision. The normal retinal blood vessels develop tiny little blebs along the blood vessel walls, see Fig. 13.22. These blebs leak blood that can cause fluid to accumulate in the macula degrading vision and color sensitivity. The leakage points can be sealed off with a focal laser. A small lens is placed on the cornea and an argon laser is focused on the blebs near the macula but not within the macula, which would cause blindness. Argon lasers are used for this treatment because of the high absorption of the emission lines at about 500 nm by hemoglobin. During

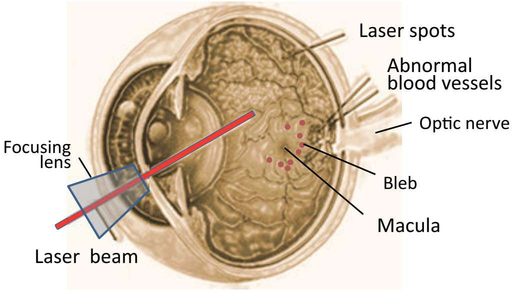


Fig. 13.22: Laser treatment of diabetic retinopathy removing blebs close to the macular and creating laser spot patterns in peripheral part of the retina for reducing oxygen demand (adapted from www.ferchoeyeclinic.com/).

treatment, the eye must be kept as still as possible for a few minutes to ensure accurate application of the laser beam and to avoid damage to the macula. Focal laser treatment has temporary success but may be repeated as new blebs show up. When treatment is started, the patient experiences bright flashes of light but no pain.

Proliferative diabetic retinopathy (PDR): the appearance of diabetic retinopathy is associated with the proliferation of a protein called *vascular endothelial growth factor* (VEGF) in the retina. VEGF stimulates the production of new blood vessels in the retina to bring more oxygen to the tissue to compensate for insufficient retinal blood circulation because of diabetes. Blindness may develop from massive proliferation of these abnormal and leaking blood vessels causing either retinal detachment or a rare type of painful glaucoma that is not to be mistaken for the regular type of glaucoma. Proliferation can be stopped by suppressing production of VEGF. Production of VEGF is suppressed indirectly by decreasing the oxygen demand. The oxygen demand, in turn, is reduced by destroying small portions of the peripheral retina, which are less important for vision. With a focused argon laser a pattern of laser spots is imprinted, causing local photocoagulation. The procedure is known as *Panretinal Photocoagulation* (PRP). The treatment consists of a greater number of larger, more intense burns, more rapidly applied than in macular laser therapy, which may cause pain. PRP is not only used to stop proliferation of abnormal blood vessels but also for welding back together retina and choroidal layer in case of detachment. New developments aim at minimizing retinal damage by delivering laser micropulses (micropulse laser). These micropulses appear to cause fewer retinal injuries. It has been speculated that the success of PRP may likely be due to a significant reduction of the choroidal thickness under the fovea, reducing blood flow [12].

For both laser methods, focal and PRP, to be successful in the long run, a general treatment of the diabetes that is responsible for the development of diabetic retinopathy in the first place is necessary.

13.5.3 Cataract and glaucoma

Cataract: nonlaser procedures for cataract and glaucoma treatment are discussed in Chapter 11/Vol. 1. There are a few additional treatments that make use of lasers. As already mentioned in Section 11.2.6/Vol. 1, nowadays for cataract lens replacement a femtosecond laser may be used to photomechanically destroy the lens and the fragments are then sucked out through a tiny hole in the cornea. Several years after cataract surgery a membrane can develop on the back side of the new lens that impairs vision in the long run. In this case an Nd:YAG laser is used to explosively vaporize and remove the membrane.

Glaucoma: This is an eye condition expressed by an abnormal intraocular pressure (IOP). If not treated in time, it will lead to blindness. The IOP is controlled by a balance between production rate and outflow of aqueous humor. If the production rate is higher than the drainage, IOP will rise. To cure this problem either the production rate has to be reduced or the depletion needs to be increased. The production rate can be controlled to some extent by pharmaceuticals. If this is not sufficient, there are three laser-assisted methods to support glaucoma treatment:

1. Laser trabeculoplasty (LTP) (Fig. 13.23): the laser beam is pointed onto the trabecular network to increase drainage through the canal of Schlemm. For this treatment a double frequency Nd: YAG laser is used, which emits at a wavelength of 532 nm in the green visible region.
2. Controlled cyclophotocoagulation (Coco) (Fig. 13.24): instead of increasing drainage, IOP can alternatively be lowered by suppressing the production rate of aqueous humor. To this end a laser beam is directed to the ciliary body, where aqueous humor is produced. Two methods are applied: either the laser beam is directly focused on the ciliary body causing photocoagulation, or a narrow endoscope is used to deliver the laser beam through the cornea. The endoscope containing a diode laser is coupled to a video imaging system for direct visualization and control of the ciliary processes caused by the laser [13, 14]. This results in an efficient and secure procedure for reducing IOP.

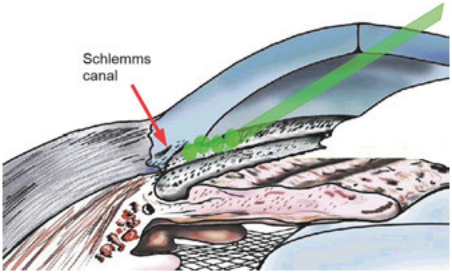


Fig. 13.23: Trabeculoplasty by laser surgery for increasing the drainage of aqueous humor (courtesy A.R.C. Laser Technology).

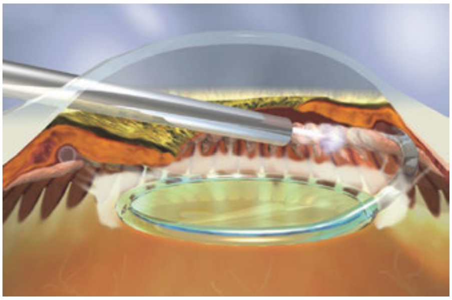


Fig. 13.24: Endoscope-controlled photocoagulation of the ciliary body for suppressing the production rate of aqueous humor (reproduced from [13] by permission of John Wiley & Sons, 2017).

13.6 Summary

1. Laser light is monochromatic, coherent, and directional.
2. All lasers consist of three main components: (1) active medium; (2) optic pump; and (3) resonator.
3. Laser safety is an important issue, in particular for laser light that is invisible.
4. Lasers can be described by a three or four level system.
5. Lasers are characterized according to their active medium as gas lasers, dye lasers, excimer lasers, solid-state lasers, and semiconductor lasers.
6. Lasers can be operated in a continuous mode or in pulsed mode.
7. Pulsing of lasers is achieved by mechanical switching, by Q switching, or by mode locking.
8. Q switched lasers require a fast switch in the resonator, which is either an electro-optic or an acousto-optic modulator.
9. Laser penetration into the skin and eye depends on the wavelength.
10. In the visible wavelength regimes hemoglobin and melanin mainly contribute to the absorption of laser light.
11. Laser-tissue interaction can be categorized according to average power density, peak pulse power density, and pulse length.
12. The main types of interaction are photomechanical, photochemical, photothermal, and photoablative.
13. Photochemical interaction requires administering a photosensitizer.
14. Photochemical reactions produce highly oxidizing singlet oxygen molecules.
15. Photorefractive keratectomy can be used for correcting myopia and hyperopia.
16. Two methods are applied: LASIK and LASEK. They are distinguished by the cut through the epithelium and stroma and by the recycling of the epithelium.
17. Diabetic retinopathy either removes blebs on the retina or reduces the oxygen supply in the retina.
18. Enhanced intraocular pressures can be reduced by either increasing the drainage of aqueous humor through the canal of Schlemm, or by reducing the production rate in the ciliary body.

References

[1] Einstein A. Zur Quantentheorie der Strahlung. Physikalische Zeitschrift. 1917; 18: 121–128.
[2] Arrigoni M, Bengtsson MS, Schulze M. Solid state lasers: laser pulsing: the nuts and bolts of Q-switching and modelocking. Laser Focus World. Issue 06/01/2012.
[3] Abramczyk H. Introduction to laser spectroscopy. Elsevier; 2005.
[4] Lakowicz JR, editor. Topics in fluorescence spectroscopy. Vol. 1. Plenum Press; 1991.
[5] Niemz MH, Laser-tissue interactions: fundamentals and applications. 2nd edition. Springer-Verlag; 2007.
[6] Carroll L, Humphreys TR. Laser-tissue interactions. Clinics in Dermatology. 2006; 24: 2–7.
[7] Hirsch LR, Stafford RJ, Bankson JA, Sershen SR, Rivera B, Price RE, Hazle JD, Halas NJ, West JL. Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. PNAS. 2003; 100: 13549–13554.
[8] Dolmans DE, Fukumura D, Jain RK. Photodynamic therapy for cancer. Nature Reviews Cancer. 2003; 3(May): 380–387.
[9] Huang Z. A review of progress in clinical photodynamic therapy. Technol Cancer Res Treat. 2005; 4: 283–293.

[10] IKH Poon, Hulett MD, Parish CR. Molecular mechanisms of late apoptotic/necrotic cell clearance. Cell Death and Differentiation. 2010; 17: 381–397.

[11] Ambrosio R, Wilson S. LASIK vs LASEK vs PRK: advantages and indications. Semin Ophthalmol. 2003; 18: 2–10.

[12] Okamoto M, Matsuura T, Ogata N. Effects of panretinal photocoagulation on choroidal thickness and choroidal blood flow in patients with severe nonproliferative diabetic retinopathy. Retina. 2016; 36: 805–811.

[13] Siegel MJ, Boling WS, Faridi OS, Gupta CK, Kim C, Boling RC, Citron ME, Siegel MJ, Siegel LI. Combined endoscopic cyclophotocoagulation and phacoemulsification versus phacoemulsification alone in the treatment of mild to moderate glaucoma. Clin Experiment Ophthalmol. 2015; 43: 531–539.

[14] Richter GM, Coleman AL. Minimally invasive glaucoma surgery: current status and future prospects. Clinical Ophthalmology. 2016; 10: 189–206.

Further reading

Silfvast WT. Laser fundamentals. 2nd edition. Cambridge: Cambridge University Press; 2004.

Thyagarajan K, Ghatak A. Lasers: fundamentals and applications. Springer-Verlag; 2011.

Welch AJ, van Gemert MJC, editors. Optical-thermal response of laser-irradiated tissue. Springer-Verlag; 2011.

Bhattacharya B. Laser in ophthalmology. Step by step. JAYPEE; 2009.

Jelinkova H, editor. Lasers for medical applications, diagnostics, therapy and surgery. Woodhead Publishing; 2013.

Bäuerle DW. Laser processing and chemistry. 4th edition. Springer-Verlag; 2011.

14 Nanoparticles for nanomedical applications

14.1 Introduction

Nanomedicine is a multidisciplinary science and technology field that has emerged over the past 10–20 years. It involves medical physics, materials sciences, biochemistry, biomedicine, pharmaceutics, polymer sciences, clinical sciences, and possibly further fields. Nanomedicine is commonly defined as “*The medical application of nanotechnology for diagnosis, treatment and the general management of human health*” [1]. As such, nanomedicine promises more sensitive diagnostics and more precise treatment of certain diseases, in particular of cancer.

In this short chapter we will mainly focus on physical aspects and methods in nanomedicine. References at the end of this chapter guide to review articles and books that cover further aspects of this rapidly developing field.

From a physics and materials science point of view nanomedicine uses *nanoparticles* (NP) for contrast enhancement of various imaging modalities, for early diagnostics of cancer cells other than imaging, for hyperthermia of tumor cells, and for delivery of targeted pharmaceuticals. These four main applications of nanoparticles can be divided into two categories: diagnostics and therapy.

What is implied with the term ‘*nano*’? The prefix *nano* refers to the length scale from (sub)-nanometer (10^{-9} m) up to about 200 nm. Atoms, molecules, and proteins fall into this range. But the science of atoms, molecules and proteins is not per se nanoscience. In addition to their extension, nanomaterials are usually produced artificially. As materials are scaled down from macrosize to nanosize they feature new physical properties, including mechanical, electrical, optical, and magnetic properties. Often these properties can be fine-tuned by the size and shape of NPs. Furthermore, as the surface/volume ratio increases, surfaces and interfaces start to play a prominent role in terms of surface reactivity and catalytic activity. This distinguishes nanomaterials from atoms, molecules, and proteins which feature their intrinsic size and their properties cannot be further scaled.

Nanoparticles used for nanomedical applications range in diameter from about 5 nm to 200 nm; their surface is protected by a polymer or silica coating and a functionalized shell serves for targeting specific receptors, a basic design is sketched in Fig. 14.1. There are large variations of this basic design, depending on the specific task to be fulfilled.

The expectations that nanomedicine has raised are huge, especially when nanoparticles feature multifunctional properties for diagnostics and therapy. The “dream” nanoparticle finds its own way to the target tissue, its location is monitored by multimodality imaging methods, and after accumulating in the target area it either delivers personalized drugs or it can be stimulated to hyperthermal or chemical activity. In short, the “dream” nanoparticle is expected to “find, tell, and cure”. For nanopar-

DOI 10.1515/9783110553116-016

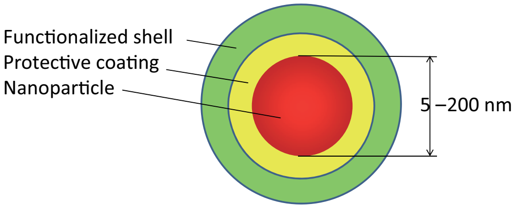


Fig. 14.1: Schematic of a nanoparticle used for nanomedical applications composed of an inorganic core, a polymer coating, and functionalized shell containing affinity ligands for specific targets.

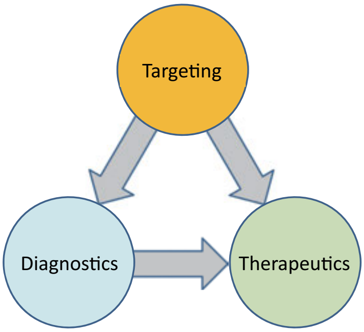


Fig. 14.2: Nanomedical triangle consisting of the main goals: diagnostics of diseases, targeting diseased tissue, and providing a therapy. Ideally, multifunctional nanoparticles can address all three modalities.

ticles that feature this multimodality, a new term has been coined: *theranostic nanoparticles* [2]. The interconnectivity of medical nanoparticles is illustrated in Fig. 14.2. Precise targeting is the prerequisite for all other modalities. In most studies, at least two modalities, such as imaging and therapy, are combined.

Before theranostic nanoparticles succeed in nanomedical treatment protocols, a number of issues first need to be considered and resolved. Precise knowledge of their pathway through the body determines their final efficiency [3]. The pathway includes potential biodegradation, bio-assimilation, elimination through liver, spleen and kidneys, or permanent deposition in the tissue. Toxicity to healthy tissue, reaction of the immune system, and long-term health effects call for attention. The parameter space is huge and encompasses particle size, shape, core-shell structure, materials choice, charge, and functional coatings. Once the main issues are resolved, the benefits of nanoparticles for diagnostics and therapeutics have to be verified in vivo by clinical studies. Obviously there are many tasks for many scientists and practitioners in this new field to make nanomedicine a success story. In this chapter we do not emphasize the synthesis aspects of nanoparticles but rather focus on their physical properties.

In Section 14.2 we will present some general biological aspects that apply to all kinds of NPs, in Section 14.3 we discuss magnetic properties of NPs and their applications in nanomedicine, followed by a treatment of nonmagnetic metallic NPs in Section 14.4.

14.2 Pathway of nanoparticles through the body

14.2.1 Reticuloendothelial system (RES)

After intravenous injection, NPs inevitably encounter the *reticuloendothelial system* (RES) of the body [4]. RES is part of the immune system and consists – among others – of phagocytic or scavenger cells, more specifically monocytes in the blood stream and macrophages in the tissue. Macrophages in the liver are known as Kupffer cells. Once NPs are administered, they are swiftly recognized as aliens by these surveillance phagocytes. Upon recognition, the NPs are coated by plasma proteins (opsonization), which increases the appetite of monocytes to swallow them up, like packmen. After incorporation, the NPs are transported to and disposed of in the liver or spleen, where macrophages and Kupffer cells decompose the coating and excrete the rest through the fecal route (see Fig. 14.3). The opsonization and elimination process can be rapid and varies between 1 to 30 hours. Hydrophobic particles, negatively charged particles, and particles with size bigger than about 50 nm are eagerly scavenged by monocytes. However, long circulation times and low elimination rate is desirable to keep the NPs in the blood stream to reach organs and targeted tissues for diagnostics or therapeutics. Hence, it is necessary to prolong the circulation time, for instance by evading the RES system. One important strategy of NP design is therefore to make them sufficiently small, hydrophilic, and charge neutral. Alternatively, the NPs can be camouflaged (stealth NP) to avoid recognition as aliens by the RES. Still another strategy is to design NPs for being scavenged but as Trojan horses to deliver their drug load to targeted sites.

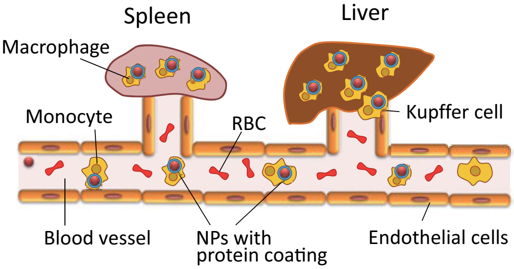


Fig. 14.3: Nanoparticles are rapidly recognized by the reticuloendothelial immune system (RES) that first coats them with proteins before monocytes carry them off to spleen and liver where they are decomposed and finally excreted. RBC = red blood cell.

14.2.2 Clearance

Contrast agents, such as iodinated iopromide (Fig. 5.17) for x-ray imaging or Gd chelates (Fig. 15.37/Vol. 1) for MRI have a short circulation half-life in the body because of their rapid renal clearance [5]. They are too small to be recognized by the RES and not large enough to be kept in the blood stream. All molecules or particles smaller than 10 nm are cleared by the glomerular filter in the kidneys (see Section 10.6/Vol. 1). To keep nanoparticles in the blood circulation, they should exhibit a total hydrodynamic diameter (HD) between 10–50 nm. On the other hand, for perfusion studies of liquids through the kidneys using radioisotopes, as for the MAG_3 scan (see Section 6.4), rapid clearance is advantageous.

14.2.3 Enhanced permeation and retention effect (EPR)

After being intravenously administered, NPs circulate throughout the body and accumulate passively in tumor tissue through the *enhanced permeation and retention* (EPR) effect [6]. The EPR effect is the result of a leaky and ruptured blood vessel system in the tumor volume combined with porous tissue, schematically illustrated in Fig. 14.4. The leaky vasculature lets the NPs penetrate the endothelial wall from where they percolate through the porous tumor tissue. Moreover, the lymphatic drainage system in cancerous tissues is underdeveloped, enhancing the retention time of NPs in those areas. Hence, NPs have sufficient time for either discharging their anticancer drug load or for hyperthermal treatment. In any case, the passive accumulation of NPs already helps identifying location and extension of carcinoma.

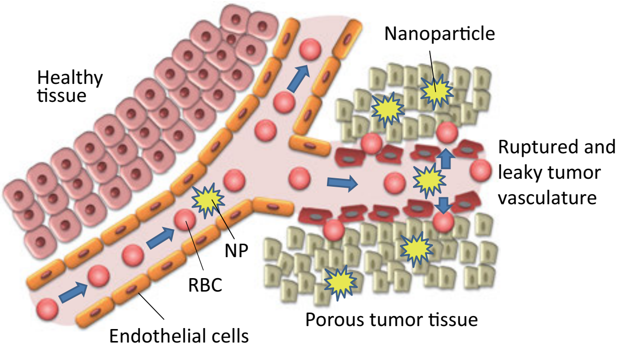


Fig. 14.4: Vasculature in healthy and tumorous tissue. In tumorous tissue the vasculature is ruptured and leaky, allowing nanoparticles to penetrate through and percolate into the tumorous tissue even without specific targeting. Hence tumorous tissue provides enhanced permeation and retention of nanoparticles (adapted with permission from [6]).

14.2.4 Coatings

Whatever the core of a NP contains – metals, oxides, or semiconductors – the NP should be water solvable and biocompatible. These conditions can be well met by silica coatings. Silica consists of an amorphous or nanocrystalline network of silicondioxide (SiO_2) with dangling bonds on the surface saturated by hydroxyl groups. Silica coated NPs have excellent water solubility properties, the shell thickness can be readily adjusted, and biocompatibility is very good. However, the dielectric silica shell collects charges and is sensitive to pH values of the embedding medium, which may cause precipitation and gel formation. Therefore silica coatings must be protected with another organic shell, usually polyethylene glycol (PEG) for enhancing the hydrophilicity and colloidal stability. PEG with the general formula $C_{2n}H_{4n+2}O_{n+1}$ is a biocompatible and water soluble polymer of varying chain length, which is also used in pharmacy for drug delivery and cosmetics. Furthermore, PEG shells hinder protein adsorption, which suppresses opsonization and therefore bypasses RES. PEG coatings can further be modified for targeting specific organs or tumors, as presented in the next section. A generic nanoparticle following the design principles outlined so far is sketched in Fig. 14.5. Many more design features are discussed in the review article of [7].

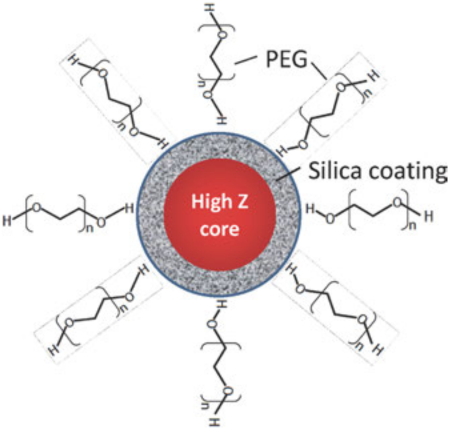


Fig. 14.5: Generic nanoparticle for x-ray contrast enhancement, containing a high *Z* material in the core, coated with silica and covered with a polyethylene glycol (PEG) shell, consisting of a chain with varying length that allows for adapting the hydrodynamic diameter of NPs. PEG also hinders protein adsorption and opsonization.

14.2.5 Antigen-antibody

Viruses and bacteria have antigens on their surface which are different from those antigens that the body produces. Antigens consist of proteins or carbohydrate chains. Once these particles invade the body, they are recognized by antibodies as being foreign. Antibodies, in turn, are proteins that are capable of recognizing specific foreign particles by the shape of their antigens. Antibodies fit together with antigens like lock and key. Antibodies are Y-shaped molecules. The basic unit of an antibody is a tetramer of four polypeptides: two identical light chains and two identical heavy chains, each consisting of a constant and a variable region, see Fig. 14.6 (a). The chains are held together by disulfide bonds.

The body does not produce antibodies that bind to its own antigens. Therefore all particles that are bound to antibodies are foreign. The concept of antigen-antibody has been used to fight cancer cells. Extensive research in recent years has established that antigen structures on tumor cells are different from antigens of normal cells [8]. The next task is then to find antibodies that uniquely target the tumor specific antigens. However, antibodies that predominantly bind antigens in cancer cells instead of normal tissues could not yet be identified. Despite this lack of absolute specificity for cancer cells, antibodies with preferential cancer cell reactivity have very high tumor recognition. Hence, the concept of targeted NPs conjugated with antibody that bind to tumor specific antigens is promising and frequently used. Figure 14.6 (b) schematically shows an NP with a silica coating and PEG shell that also contains antibodies conjugated to the silica coating. NPs can be loaded with additional drugs, contrast agents, radioisotopes, fluorescence tags, enzymes etc. for multifunctional use.

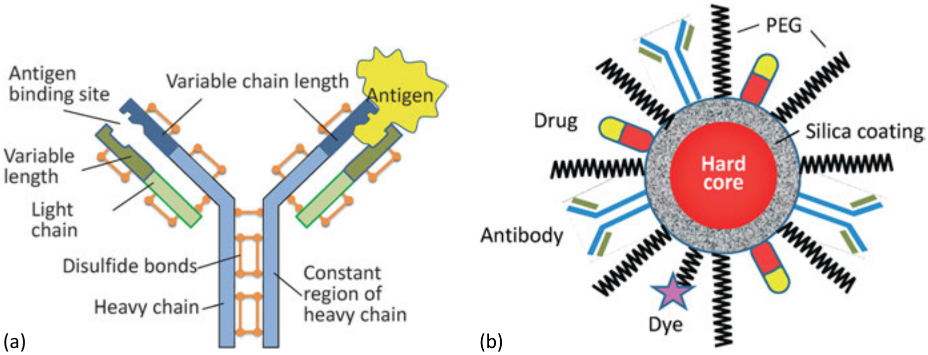


Fig. 14.6: (a) Schematic structure of an antibody. The Y-shaped molecule consists of two heavy chains and two light chains, both have a fixed part and a variable length part. The chains are held together by disulfide bonds. Each arm of the antibody contains one antibody binding site, which acts according to the lock and key principle. (b) Nanoparticle coated with silica and conjugated with PEG, antibodies, drugs, and possibly further contrast agents or targeting ligands.

14.2.6 Targeting

We distinguish between physical targeting, passive targeting, and active targeting of NPs for diagnostics and therapeutics. Physical targeting refers to the fact that after administering to the vascular system NPs are distributed across the body by the blood flow. There is no specific binding site, but depending on their size particles may penetrate through leaky vessels or may get trapped in regions of a stenosis. Active targeting is the opposite of physical targeting. In the latter case, NPs are covered with targeting ligands to dock to specific sites in tissues or organs. For active targeting either antigen-antibody interaction or ligand-receptor interaction is exploited. Passive targeting utilizes differences in vasculature and physiological properties of tumors versus healthy tissue that affects uptake behavior and retention times. This effect is known as the EPR effect and passive targeting is based on this effect.

14.2.7 Size

In nanomedicine size matters. We differentiate between core size, size including coating and shell, and the *hydrodynamic diameter* (HD). The size of the core determines the physical properties of NPs, in particular their magnetic and optic properties. Coating and shell are responsible for biocompatibility, colloidal stability, and biofunctionality. The HD is the size of NPs in watery solution where the shell is covered by attached water molecules and possibly by plasma proteins. The HD size determines biodistribution and circulation half-life in the body. Longer circulation half-lives make NPs more interesting for blood pool investigations. The synthetic capabilities presently available are very advanced, providing complete control over all sizes, shapes, and shells. NPs can be synthesized with spherical or rod-like shape; cores can be solid, porous or cage like; shells fulfill a number of tasks already mentioned before, and more will be highlighted below.

14.2.8 Biocompatibility

Biocompatibility is the opposite of toxicity. Toxicity is easier to define than biocompatibility. Toxic chemicals or materials are capable of causing injury, cell death, or impairing functional body parts. Toxicological hazard is the potential for a compound or a material to cause an adverse biological reaction, taking into account the nature of the reaction and the dose required to provoke it [9]. In this context, biocompatibility can be defined as a “biological performance that in a specific application is judged as suitable for that particular situation” [10]. This implies that materials or chemicals that come into contact with biological tissue do not cause adverse effects in the tissue, such as inflammation or lesions. Vice versa, materials reacting by biodegradation while in contact with tissue should not release or leak out toxic or radioactive substances, unless intended to do so. Biocompatibility has also been defined in more general terms as the “ability of a material to perform with an appropriate host response in a specific situation” [11].

14.3 Magnetic nanoparticles for diagnostics and therapeutics

Gadolinium (Gd) and magnetite (Fe_3O_4) are the most widely used contrast agents for magnetic resonance imaging (MRI) and therefore we will discuss the magnetic properties of these two materials in more detail. Magnetite is also used for hyperthermia applications. We first want to clarify the atomic origin of Fe and Gd magnetism and cooperative effects that lead to ferromagnetism on the macroscale and superparamagnetism on the nanoscale.

14.3.1 Ferromagnetism

(a) Magnetic properties of gadolinium

Gd has a half-filled 4f shell and three electrons in the $5d^1 6s^2$ orbitals. The outer electrons determine the ionicity and contribute to the bonding of Gd in oxides (Gd_2O_3) or in chelate molecules. The inner 4f electrons are essentially unaffected by the chemical bond and retain their atomic levels. In the ground state the spins of the seven electrons in the 4f shell line up to a total spin $S = 7/2$ and an orbital moment $L = 0$ (Fig. 14.7 (a)), resulting in a Gd magnetic moment of $8\mu_{\text{Bohr}}$. This moment is stable, temperature independent, and independent of whether Gd is in a gaseous, molecular, or solid state. In metal form, Gd is a ferromagnet below the Curie temperature of 292 K, i.e., all magnetic moments of different Gd atoms within the hexagonal crystal structure line up in parallel and form an ordered magnetic state (Fig. 14.7 (c)). This ferromagnetic order is mediated by the outer $5d^1 6s^2$ conduction electrons that are magnetically polarized by the inner 4f electrons (Fig. 14.7 (b)). Above the Curie temperature Gd is paramagnetic, meaning that the parallel alignment of the magnetic moments is lost, their orientation is increasingly random as the temperature goes up. However, when a magnetic field H is applied, the magnetic moments easily turn into the magnetic field direction. This phenomenon is referred to as magnetic susceptibility. The magnetic susceptibility χ_{m} is defined by:

$$\chi_{\text{m}} = \frac{\mathrm{d}M(T)}{\mathrm{d}H}, \quad \text{with } M(T) = \frac{1}{V}\sum_i m_i,$$

where $M(T)$ is the magnetization, i.e., the volume average of local magnetic moments m_i. The magnetic susceptibility χ_{m} has a maximum at the Curie temperature, but drops off at higher temperatures T. In Gd^{3+} chelates the Gd^{3+} ions are in an isolated atomic state. Any magnetic interaction is suppressed by the large separation between them. Hence, Gd^{3+} chelate molecules are paramagnetic at all temperatures. In a magnetic field the 4f magnetic moments are partially aligned. As soon as the magnetic field is turned off, all magnetic moments relax back into a disordered state.

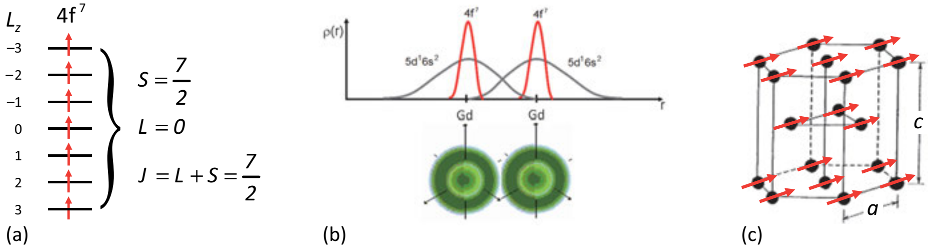


Fig. 14.7: Magnetism of gadolinium. (a) Level scheme of the 4f shell occupied by seven electrons with spin 1/2; (b) 4f wave functions from neighboring atoms do not overlap, in contrast to 5d6s wave functions. Magnetic interaction between 4f electrons is mediated by the polarization of 5d6s electrons; (c) ferromagnetic structure of hexagonal Gd metal.

(b) Magnetic properties of magnetite

In nanomedicine magnetite is used for MRI contrast enhancement and for hyperthermia. Magnetite (Fe_3O_4) is one of three known iron oxides. The other two are maghemite (γ-Fe_2O_3) and hematite (α-Fe_2O_3) [12]. Magnetite, the only oxide we want to discuss here in some detail, has an inverse spinel crystal structure with 24 Fe ions and 32 oxygen ions per unit cell. The 24 Fe ions are grouped in 8 Fe^{3+} ions on lattice sites with tetrahedral local symmetry (T) and another 2×8 Fe ions are on sites with octahedral local symmetry (O). Fe ions on the T sites have 3+ valence, on the O sites the valence fluctuates: 8 Fe ions have valence 3+, the remaining 8 Fe ions have valence 2+. With a half-filled 3d shell, Fe^{3+} has a total spin $S = 5/2$ and orbital moment $L = 0$. The resulting magnetic moment is $5.9\mu_{\mathrm{Bohr}}$. In contrast, Fe^{2+} has a sum spin $S = 4/2$ and a crystal field quenched orbital moment, yielding a magnetic moment of $4.9\mu_{\mathrm{Bohr}}$/atom. The eight Fe^{3+} ions on T sites and the equivalent ones on O sites have antiparallel spins and their magnetic moments are compensated by antiparallel or antiferromagnetic ordering. Only the remaining 8 octahedral sites occupied by Fe^{2+} contribute to the magnetism of magnetite. In total, magnetite is referred to as a *ferrimagnet* with a Curie temperature of 850 K. The spin structure is shown in Fig. 14.8. Maghemite (γ-Fe_2O_3) is also a ferrimagnet. In contrast to Fe_3O_4 there are fewer Fe ions on O sites ($8 + 5$) and they all have 3+ valence: 8 for compensating the magnetic moments of Fe^{3+} on T sites, and 5 more with spin 5/2 and ferromagnetic order. Often γ-Fe_2O_3 and Fe_3O_4 coexist. A distinction is possible, but difficult and elaborate for tiny nanoparticles. Hematite, in contrast, is an antiferromagnet and not useful as CA in MRI. Magnetite has a magnetic moment per formula unit of $4.9\mu_{\mathrm{Bohr}}$, maghemite theoretically should have a

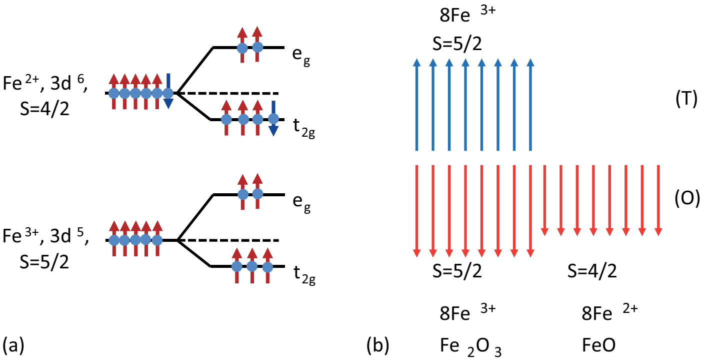


Fig. 14.8: Spin structure of ferrimagnetic magnetite Fe_3O_4. (a) Electronic shells of Fe^{2+} and Fe^{3+}. The degenerate 3d shell is split by local crystal electrical fields into t_{2g} and e_g levels. The resulting total spin for Fe^{2+} is $S = 4/2$ and $S = 5/2$ for Fe^{3+}. (b) Schematic occupation of Fe^{2+} and Fe^{3+} with their respective total spin angular momentum on crystal lattice sites with local tetrahedral (T) and octahedral (O) site symmetry. Note that Fe ions on octahedral sites have fluctuating valence. Fe^{3+} magnetic moments on O and T sites compensate each other by antiparallel orientation. Only Fe^{2+} ions on T sites contribute to the macroscopic magnetism of magnetite.

magnetic moment of $2.4\mu_{\mathrm{Bohr}}$ per formula unit. In practice these values are considerably lower for nanoparticles than for bulk materials due to defects and finite size effects. The magnetic moment of magnetite NPs is found to be in the order of $2\mu_{\mathrm{Bohr}}$. Often manganese ferrite ($MnFe_2O_4$) NPs are used in nanomedicine for their confirmed higher magnetic moment per formula unit as compared to magnetite.

14.3.2 Superparamagnetism

In spite of the highly complex magnetic spin structure, magnetite acts as a normal rather soft ferromagnetic material. In fact it is the oldest one ever known and – according to the saga –discovered by shepherds from the Greek town Magnesia, which gave magnetism its name. One universal characteristic of ferromagnets is their magnetic hysteresis when reversing the magnetization in an external magnetic field. Magnetic hysteresis signifies the presence of a magnetic domain state during reversal, which is the hallmark of ferromagnets. Magnetization causes stray fields. In order to reduce stray fields and to keep the magnetic flux inside of ferromagnets, domains are generated, separated by domain walls, as shown schematically in the left panel of Fig. 14.9. When such a ferromagnet is exposed to an external magnetic field, magnetic hysteresis can be measured as a function of field strength. Magnetic hysteresis is characterized by a saturation magnetization M_{s} at high fields, remnant magnetization M_{r} at zero field, and a coercive field H_{c} when the magnetization switches sign. However, when ferromagnets are reduced in extension to the size of nanoparticles, domains can no longer form, since domain formation would cost too much (exchange) energy. Then each nanoparticle exhibits a single ferromagnetic domain, schematically indicated in the right panel of Fig. 14.9 by an array of arrows within a nanoparticle. All magnetic moments in a nanoparticle act together as a gigantic magnetic moment called the magnetic moment of a *macrospin*. This macrospin is surrounded by a magnetic dipole field. Macrospins can only form in the ferromagnetic state of NPs, i.e., for temperatures below the Curie temperature. Magnetite NPs have macrospins at the body temperature of 310 K, whereas Gd-NPs do not, since the Curie temperature for Gd is 19 K below body temperature.

Now we consider an ensemble of MNPs, each macrospin having a total magnetic moment m_{macro}. In zero external magnetic field the macrospins are randomly oriented resulting in zero magnetization. Turning on a magnetic field, the macrospins will respond to the field and partially orient in the field direction, yielding a finite magnetization ($M = \chi_{\mathrm{m}} H > 0$). The temperature and field dependence of the magnetization of an ensemble of MNPs can be described by the classical Langevin function $L(x)$:

$$M(T, H) = \frac{N}{V} m_{\mathrm{macro}} L(x),$$

$$L(x) = \coth(x) - \frac{1}{x},$$

where $x = \mu_0 m_{\mathrm{macro}} H / k_B T$. x is the ratio of the Zeeman energy $\mu_0 m_{\mathrm{macro}} H$ to the thermal energy $k_B T$, and μ_0 is the magnetic permeability of the vacuum. Because of its resemblance to a classical paramagnetic system, the magnetization of nanoparticles is called superparamagnetism. However, it is the magnetism of tiny ferromagnets in the core of nanoparticles, randomly distributed in space, that yield a paramagnetic response when exposed to a magnetic field. Further details on the superparamagnetism of MNPs can be found in the review of [13]. If superparamagnetism is based on iron-oxide magnetite, this system is referred to as *superparamagnetic iron oxide* (SPIO) nanoparticles.

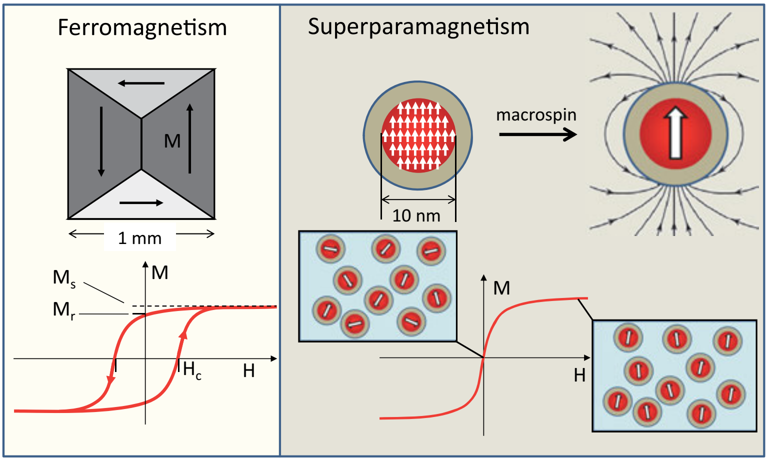


Fig. 14.9: Left panel: ferromagnet in a domain state. The magnetization of a ferromagnet is characterized by a saturation magnetization Ms in high external magnetic fields, a remnant magnetization at zero field, and coercive fields when the magnetization changes sign on the ascending or descending branch of the hysteresis. Right panel: in sufficiently small nanoparticles there are no magnetic domains. The tiny grains are in a single domain ferromagnetic state represented by a macrospin. Without an external field turned on, an ensemble of magnetic nanoparticles (MNPs) will be randomly oriented yielding zero magnetization. Turning on a magnetic field, the MNPs will partially align in the direction of the field, resulting in a finite magnetization. The magnetization versus field can be described by a Langevin function, resembling the magnetization curve of isolated paramagnetic ions, therefore the term *superparamagnetism*.

If nanoparticles get close to each other, they couple via magnetic dipole-dipole interaction and try to align each other. Suspended in a fluid (ferrofluid), the MNPs have a tendency to agglomerate. Agglomeration can be suppressed by sufficiently thick shells surrounding NPs, keeping them at a distance where magnetic dipole-dipole interaction becomes negligible.

14.3.3 Blocking temperature

As indicated in Fig. 14.9, the magnetic moments within MNPs are aligned parallel. The direction in which they line up is called the *easy axis*. This axis is determined by the magnetocrystalline anisotropy and is an intrinsic material property. If all moments in an MNP would decide to orient themselves in another direction, for instance in the direction opposite to the original direction, they would have to overcome a potential barrier, indicated in Fig. 14.10 (a). The potential barrier is given by the magnetic anisotropy energy density K times the magnetic volume V_m of the particle: $E_{ani} = KV_m$. In thermal equilibrium it is a question of temperature and time whether the magnetization within a particle will indeed succeed and overcome the barrier. From a statistical mechanics point of view this will take place on average after the relaxation time τ:

$$\tau = \tau_0 \exp\left(\frac{KV_m}{k_B T}\right),$$

where τ_0 (10^{-9}–10^{-10} s^{-1}) is the attempt frequency. From this equation we notice that the relaxation time, i.e., the average time between two flips, increases exponentially with the ratio of anisotropy energy versus thermal energy. However, for small particles and high temperatures, the flipping time may be quite short.

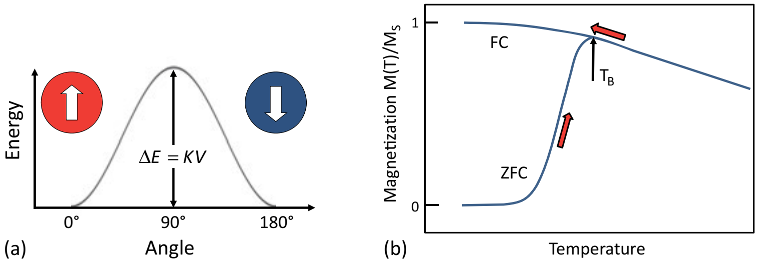


Fig. 14.10: (a) Potential barrier for flipping of magnetic NPs. (b) Magnetization measurements during cooling in an applied field (FC) and during warming up after the system has been cooled in zero field (ZFC). Both curves meet at the blocking temperature T_B.

MNPs appear to be static when during the time of observation τ_{ob} no flip occurs. This may be the case below a certain temperature T_B, called *blocking temperature*. For temperatures $T > T_B$ the ensemble of particles flips frequently, such that the time average magnetization becomes zero. Hence the blocking temperature T_B divides regions of static versus dynamic magnetization. Rephrasing the equation for the relaxation time we find for the blocking temperature:

$$T_B = \frac{KV_m}{k_B \ln(\tau_{ob}/\tau_0)}.$$

Here we notice that the blocking temperature scales linearly with the volume of MNPs. Figure 14.10 (b) shows schematically the result of a magnetization measurement protocol for determining T_B. First the MNPs are exposed to a small magnetic field and cooled from high to low temperature, called field cooling measurement (FC). In this case the average magnetization of all NPs steadily increases with decreasing temperature and saturates at the lowest temperature. In a second measurement the same MNPs are first cooled to the same low temperature but without field. Without field the magnetization is zero for all temperatures. This is referred to as zero field cooling (ZFC) condition. Now starting at the lowest temperature the same field is turned on that was used for the FC experiment. Upon warming, the magnetization will steadily increase and meet the magnetization curve of the FC curve at a particular temperature, which is the blocking temperature T_B. Again, below T_B magnetic NPs are blocked with respect to their magnetization orientation, above T_B the magnetization is free to rotate. For $T < T_B$ the magnetization reversal shows a hysteresis with a finite coercive field H_c.

The blocking temperature increases with increasing NP size. According to [14], the blocking temperatures for Fe_3O_4 NPs with diameter 3.8 nm, 4.5 nm, and 5.5 nm are 100 K, 150 K, and 250 K, respectively. From this trend we conclude that SPIO NPs with a diameter of 10 nm have a blocking temperature above body temperature, i.e., above 310 K. For contrast enhancement of MRI, MNPs should be designed such that T_B is above body temperature. In contrast, for hyperthermia, the MNPs should have a T_B at or below body temperature.

14.3.4 MRI imaging contrast

(a) Basics of MRI contrast

MRI contrast relies on differences in spin-tissue relaxation time T1 and spin-spin relaxation time T2. T1 is also referred to as the longitudinal relaxation time or the time for M_z magnetization recovery; T2 is the transverse relaxation time or M_{xy} magnetization dephasing time (Fig. 15.4/Vol. 1). T1 depends on how fast the spin excitation energy can be transported off into the surrounding molecules, whereas T2 is affected by magnetic field inhomogeneities generated by neighboring spins, field gradients in magnets, or magnetic dipole fields emanating from magnetic CAs.

In MRI the contrast is not as rigid as in XRT: MRI contrast can be fine-tuned with the help of the parameters TR (time to repeat) and TE (time to echo) to generate T1- or T2-weighted images, or for emphasizing differences in the proton density. MRI “intensity” is not comparable with the intensity of a photon or particle beam. MRI intensity is an inductive voltage generated by a precessing in-plane magnetization M_{xy}, referred to as free induction decay (FID). The larger the z component of the magnetization M_z after relaxation time T1, the larger is the signal from FID after time TE. Larger signals produce brighter pixels on MRI images after Fourier transformation (see Chapter 15/

Vol. 1 for further details). T1-weighted images are taken with short TR and short TE, whereas T2-weighted images require long TR and long TE (see Tab. 15.3/Vol. 1).

Shortening T1 by using a CA leads to faster M_z recovery and a stronger FID signal, producing brighter pixels. CAs that decrease T1 are called *positive contrast agents*. Shortening T2 has the opposite effect. A shorter T2 relaxation time yields a weaker FID signal which translates into a darker pixel on the image. CAs tending to shorten T2 are called *negative contrast agents*. Gd^{3+} chelates (Section 15.6.2/Vol. 1 and Fig. 15.36/Vol. 1) belong to the class of positive contrast agents, whereas magnetite (Fe_3O_4) NPs are negative contrast agents. We will return to the discussion of these differences later. CAs usually have an effect on both, T1 and T2. For judging their main characteristics, the reciprocal relaxation times $R_i = 1/T_i$, called relaxivities, are considered:

$$R_i = r_i x_{\mathrm{CA}} + R_{i0}; \quad i = 1, 2.$$

Here r_i is the specific relaxivity (units: $\mathrm{mmol}^{-1}\,\mathrm{s}^{-1}$) and R_{i0} is the baseline relaxivity without CA. For a positive CA, the effect on T1 should be larger than on T2. This implies that the ratio r_2/r_1 is small. For negative CAs the ratio r_2/r_1 becomes large. Koenig and Kellar have calculated the relaxivities of protons in watery environment affected by the presence of MNPs in high fields typical for MRI conditions [15]. The relaxivities are proportional to:

$$r_{1,2} \approx \frac{1}{d_{\mathrm{NP}} D} \gamma^2 m_{\mathrm{macro}}^2 x J(\omega, \tau_{\mathrm{D}}).$$

Here d_{NP} is the NP diameter, D is the diffusion constant of solvent and solute, γ is the gyromagnetic ratio of protons, m_{macro} is the magnetization of a macrospin in the core of an MNP, x is the concentration of NPs in solution, and J is the spectral density function discussed in Section 15.2.5. This relation shows that aside from the obvious dependence on the magnetization of the nanoparticle, relaxivity is affected by NP size and diffusion constant. Here ω refers to the Larmor frequency of protons in the external field, and the relaxation time $\tau_{\mathrm{D}} = (d_{\mathrm{NP}})^2/4D$ is the interaction time of protons with NPs. It may be that several encounters of the protons with MNPs are required for protons to relax.

(b) Gd^{3+} chelates

As discussed in Section 14.3.1 (a), Gd^{3+} ions feature a stable paramagnetic moment of $8\mu_{\mathrm{B}}$ in the 4f shell. Exposed to a magnetic field, the Gd moments line up partially in the field direction and against the randomizing thermal energy. In zero field, the Gd moments relax to $M = 0$. This spin-lattice relaxation is the same process as that for protons in watery medium exposed to a high field within an MRI scanner. However, in the case of Gd ions the magnetic moment is 5300 times bigger than the magnetic moment of one proton. When Gd^{3+} chelates are administered, then these large Gd moments act as an additional local magnetic field, assisting the proton spins to align again after TR and TE pulses, which effectively shortens the T1 relaxation time. For this purpose it is advantageous that the Gd^{3+} ions in the chelate molecules are in close contact with surrounding protons. In contrast, in the core of magnetic NPs the magnetic moments are at some distance from protons in watery environment and therefore the direct magnetic contact is weaker.

Gd^{3+} chelates are presently the only T1 contrast agents that have been clinically approved. Nevertheless, Gd-based chelates have a number of shortcomings that limit their usefulness [16]. After intravenous injection and because of their low molecular weight, the molecules rapidly leak out from the blood pool to the extracellular space with a circulation half-life of only 5 min. From there the molecules are cleared by the kidneys with a clearance half-life of 80 min. Therefore chelate-based CAs do not offer a full volume distribution, they are not useful as blood pool CAs, and owing to the partial volume distribution the targeted contrast may be missed and/or the contrast may be misinterpreted. Compounding the problems is the fact that the T1 relaxivity r_1 drops in magnetic fields between 1 T to 3 T by almost a factor of 3. Due to the short circulation lifetime, poor r_1 in high fields, and toxicity concerns, research focus has shifted to T2 contrast agents and specifically to superparamagnetic iron oxide Fe_3O_4 NPs to be discussed next.

(c) Fe_3O_4 as contrast agent for MRI

What is the difference between Gd^{3+} chelates and superparamagnetic Fe_3O_4 (SPIO) nanoparticles? Both are characterized by a paramagnetic response (susceptibility) in an external field. The main difference is the total moment per particle. In Gd^{3+} chelates we have $8\mu_B$ per molecule. The sum of all magnetic moments in a 20 nm diameter magnetite nanoparticle results in a macrospin of $2.3 \times 10^5 \mu_B$. Even in a 10 nm particle the Fe moments sum to $2.8 \times 10^4 \mu_B$. The magnetic dipole fields of these nanoparticles are strong and random and they severely disturb the local proton spins. Furthermore, contact with surrounding protons in water is indirect via dipolar magnetic fields. Therefore, SPIOs yield negative contrast in distinction to Gd^{3+} chelates. Negative contrast is sometimes not as favorable for imaging as positive contrast. Negative contrast darkens the pixel and makes it less distinguishable from the anatomic background. But it is beneficial for enhancing contrast in T2-weighted images. Moreover, SPIOs are preferred CAs because of their superior biocompatibility. In fact, they are the only clinically approved nanoparticle CA at present.

For comparing the magnetization of different materials, the magnetization measured in units of A/m ($1\,\text{A/m} = 10^{-3}\,\text{emu/cm}^3$) is usually converted to electromagnetic units per gram, also called mass magnetization: emu/g. This is easily done by normalizing the volume magnetization in units of emu/cm^3 with the mass density ρ: emu/g = emu/ρ cm^3. For bulk Fe with a magnetization of 1.7×10^3 emu/cm^3 and an ideal density of 7.93 g/cm^3 we find 217 emu/g independent of the size. Bulk magnetite has a mass magnetization of 118 emu/g, which is reduced by defects to about 90–92 emu/g. For magnetite NPs with a diameter of 5 nm the mass magnetization is further reduced to about 50 emu/g. With increasing diameter up to 10 nm and beyond, the bulk value is again recovered. This tells us that in the case of very small nanoparticles the ferro- or ferrimagnetic state is not preserved. Indeed, in magnetic nanoparticles it can be observed that at the surface spin canting and spin disorder take place that reduce the overall mass magnetization of the NPs. The effect can be used for fine-tuning the r_2 relaxivity and consequentially the effect on T2: smaller magnetite NPs have lower r_2 than larger ones. r_2 ranges from 100 $\mathrm{mmol}^{-1}\,\mathrm{s}^{-1}$ for 5 nm NP up to about 200 $\mathrm{mmol}^{-1}\,\mathrm{s}^{-1}$ for 10 nm and larger NPs [15, 16].

(d) Size of magnetic nanoparticles

Size of NPs matters in nanomedicine as already discussed in Section 14.2.7. This is especially true for magnetic NPs: size determines the biodistribution of NPs in the body, the circulation half-life, and the relaxivity r_2. MNPs have been classified according to their size [17]: (1) ultrasmall superparamagnetic iron oxide NPs (USPIO; < 50 nm); (2) superparamagnetic iron oxide NPs (SPIO; ≈ 100 nm); and (3) micrometer-sized paramagnetic iron oxide particles (MPIO; several µm). Like NPs for x-ray contrast, MNPs are coated by a protective organic or inorganic shell and are colloidally stabilized by PEG. None of the NPs would fit through the glomerular filter, which requires a total HD size smaller than 10 nm. SPIOs are recognized by macrophages in the blood stream and selectively deposited in liver, spleen, and bone marrow. This has a subtractive effect on MRI diagnostics. If the normal liver structure is destroyed by hepatic disease, such as primary liver tumor or liver metastasis, the region is depleted of Kupffer cells. Therefore normal SPIO uptake is hindered and a strong contrast can be observed in T2-weighted images between normal and diseased parts of the tissue. USPIOs have longer plasma circulation times as they are not caught by macrophages in contrast to larger SPIOs. Because of their rapid distribution in the intravascular space and long circulation lifetime, USPIOs are considered ideal for *blood pool* CAs. Hence, USPIOs can be used for angiography, passively targeting tumors, and possibly for functional MRI (fMRI) if they are prepared to pass the blood brain barrier (BBB).

(e) Coating of magnetic nanoparticles

In general, the core provides magnetic sensitivity and contrast enhancement. The coating has a number of additional tasks to fulfill: (1) biocompatibility, (2) colloidal stabilization, (3) preventing agglomeration due to magnetic dipole interaction by steric hindrance, (4) providing reversible binding sites for drug delivering molecules, (5) attaching targeting ligand.

Fe-oxide NPs are inherently biocompatible and would not need a coating since the body stores Fe in the form of iron ferrihydrite crystals encapsulated into a protein hollow sphere called ferritin. Ferritin is stored in liver and spleen to be supplied to hemoglobin on demand. Nevertheless, when magnetite is used as a CA, coating is preferred due to all the merits that coating offers. The most frequently used coating is PEG and dextran. Both improve colloidal stability and blood circulation time.

(f) Alternative magnetic nanoparticles

Apart from magnetite a number of alternative magnetic NPs have been fabricated and tested in animal studies, such as Fe, FeCo and FePt alloys, mixed oxides $CoFe_2O_4$, $MnFe_2O_4$, and $NiFe_2O_4$, hydroxyapatite (HAP, $Ca_5(PO_4)_3(OH)$) doped with Co^{2+}, Fe^{2+}, Fe^{3+}, and Gd^{3+} ions [18]. HAP is a bioconformal mineral that is also present in bones and teeth (see Fig. 3.8/Vol. 1). It could be shown that Gd^{3+} doped HAP features a high magnetic moment together with hemocompatible properties, which might be a good alternative to magnetite as MRI CA. Overviews on the use of MNPs as CAs for MRI are presented in [17, 19].

14.3.5 Magnetic hyperthermia

(a) Relaxation mechanisms

Magnetic hyperthermia refers to a moderate and local rise in temperature generated by the relaxation of MNPs excited in an oscillating (AC) magnetic field at frequencies up to MHz. Hyperthermia is used for cancer treatment since moderate heating to about 44 °C leads to cell death (apoptosis). For understanding the physical basis, we examine once more the relaxation mechanism of MNPs in solution and without interparticle interaction. We have seen that MNPs can overcome the orientational potential barriers at temperatures $T > T_B$ after a relaxation time τ. When turning off an external field, the magnetization of the MNPs relaxes to zero within the same relaxation time τ. The question is now how the relaxation takes place. There are three possible mechanisms that lead to an orientational relaxation, sketched in Fig. 14.11: Néel, Brown, and hysteretic. Néel relaxation assumes a coherent rotation of the magnetization within the particle without physical rotation of the particle. Brown relaxation assumes a physical rotation of the particle without rotation of the magnetization internally. Hysteretic relaxation assumes the formation of domains that propagate through the NP for complete re-

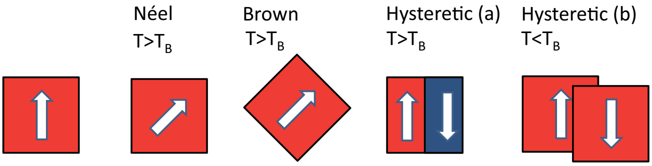


Fig. 14.11: Different relaxation processes of magnetic nanoparticles depending on size and temperature: Néel, Brown, hysteretic domain formation above T_B, and hysteretic magnetization reversal below T_B.

versal. These are three extreme cases. Hysteretic relaxation can be excluded for very small single domain MNPs, but for particles larger than 50 nm domain formation and domain propagation may occur. Néel relaxation requires MNPs with negligible magnetocrystalline anisotropy. Brownian relaxation is the opposite case, requiring high magnetocrystalline anisotropy. For temperatures below the blocking temperature T_B, magnetization reversal is hysteretic, independent of how the reversal actually proceeds, because there is little assistance from thermal energy to overcome the potential barrier KV_m.

If we now expose MNPs to an AC magnetic field, the magnetization in MNPs is rotated back and forth. In the case of Brownian motion, not only the magnetization but also the whole particle is sterically rocked. Since the particle is embedded in a watery fluid, for this type of motion we expect the most effective dissipation of energy and hence heating of the local environment. The Brownian relaxation time τ_B depends on the viscosity η of the fluid, the effective hydrodynamic volume of the particle $V_\mathrm{HD} = \pi d_\mathrm{HD}^3/6$, which is bigger than the magnetic volume V_m, and the thermal energy $k_\mathrm{B}T$ [17]:

$$\tau_\mathrm{B} = \frac{\pi \eta d_\mathrm{HD}^3}{2k_\mathrm{B}T}.$$

In the case of Néel relaxation the damping is first due to spin-lattice relaxation within the MNPs, and in a second step the lattice heat is transferred to the surrounding fluid. Therefore the Néel relaxation time depends on the magnetic volume and is given by [20]:

$$\tau_\mathrm{N} = \tau_0 \sqrt{\frac{\pi k_\mathrm{B}T}{4KV_\mathrm{m}}} \exp\left(\frac{KV_\mathrm{m}}{k_\mathrm{B}T}\right).$$

In general, both relaxation processes will contribute but with different weight depending on the particle size. In general, the effective relaxation time is composed of Néel and Brownian relaxation:

$$\frac{1}{\tau_\mathrm{eff}} = \frac{1}{\tau_\mathrm{B}} + \frac{1}{\tau_\mathrm{N}}.$$

The power dissipation to the surrounding fluid can be determined from a Debye relaxation model according to [20]:

$$P = \frac{1}{2}\omega \mu_0 \chi_0 H_0^2 \frac{\omega\tau}{1 + (\omega\tau)^2}.$$

Here H_0 is the AC magnetic field amplitude and ω is its angular frequency of the AC field, χ_0 is the equilibrium susceptibility determined from the initial slope of the Langevin equation for the magnetization given above and is a materials related property, expressed as:

$$\chi_0 = \frac{\mu_0 m_\mathrm{macro}^2 V_\mathrm{m}}{k_\mathrm{B}T},$$

where V_m is the volume and m_macro is the magnetization of the MNP, respectively.

The expected temperature raise ΔT after time Δt before reaching equilibrium can be expressed as:

$$\Delta T = \frac{P\,\Delta t}{C},$$

where C is the average heat capacity of MNPs and fluid. Numerous experiments have tested the basic dependencies of the power dissipation P as a function of frequency, susceptibility, magnetocrystalline anisotropy, magnetic and hydrodynamic particle volume and shape, particle size distribution, and as a function of fluid viscosity [21, 22]. In all these studies it is silently assumed that the blocking temperature is below the probing temperature. However, for particles with diameter of 10 nm and more, this may not be the case. If the blocking temperature is above the measurement temperature, the particles are in a blocked state. In this case the magnetization curve opens up, showing a remnant magnetization and finite coercive fields. An open magnetic hysteresis entails hysteretic energy loss that corresponds to the enclosed area by the hysteresis. This hysteretic energy loss adds to Néel and Brown energy losses already accounted for.

(b) Specific loss power

Magnetic hyperthermia requires magnetic nanoparticles that preferentially accumulate in cancerous tissue. This can be accomplished either by physical or by active targeting. In the latter case the MNPs need to be conjugated with antibodies that recognize the antigens of cancer cells. The MNPs may be injected intravenously or directly into the tumor tissue. In either case it is important to reach a homogeneous distribution of MNPs across the tumor volume for a uniform temperature topography avoiding hot or cold spots. There are many technical challenges that need attention for a successful hyperthermia treatment. First, the concentration of MNPs to be administered should be kept as low as possible to avoid any adverse biomedical effects. Next, optimized parameters for the power applied are to be identified in a huge parameter space: choice of magnetic material, particle size, shape, and distribution, field amplitude, and AC frequency. Not all parameters are independent. Particle size relates to the relaxation mechanism and the relaxation mechanism determines the efficiency of heat transfer. Particle size also determines whether the relaxation process takes place above or below the blocking temperature. Taking all these considerations into account, the main aim is to maximize the specific heating power of MNPs in a magnetic AC field to reach the targeted temperature with a minimal concentration of MNPs. The critical physical parameter is the loss power density. Rephrasing the expression given above, we find for the loss power density in units of Watts/m^3:

$$P = (\mu_0 M_s H_0)^2 \left(\frac{\omega V}{2k_B T}\right)\left(\frac{\omega\tau}{1+(\omega\tau)^2}\right).$$

According to this equation, for low frequencies $P \approx \omega^2$, while for $\omega\tau \gg 1$, P becomes frequency independent with the saturation value:

$$P = (\mu_0 M_\mathrm{s} H_0)^2 \left(\frac{V}{2k_\mathrm{B} T\tau} \right).$$

Keeping field amplitude H_0 and frequency ω fixed, the strong size dependence of the relaxation time leads to a sharp maximum in the power loss density within a narrow size range of MNPs [20, 22]. Therefore one may select the particle size or the frequency and then adjust all other parameters accordingly. For conversion to *specific loss power* (SLP), also called *specific absorption rate* (SAR), in units of Watts/g, the power density has to be divided by the density of the nanoparticles. Values for SLP vary widely in the literature from 10 to 100 W/g for field amplitudes of 10 kA/m and AC frequencies of about 400 kHz. A record high SLP has recently been reported for 18 nm $MnFe_2O_4$ NPs embedded within an amphiphilic copolymer shell. The magnetic nanoparticles loaded in these polymer nanospheres had an SLP of 580 W/g at a frequency of 435 kHz and field amplitude of 4 kA/m [23]. With such a high SLP the concentration of MNPs to be administered in vivo can be drastically reduced.

Interest in MHT has taken off after a report by A. Jordan and coworkers in 2001 on clinical studies of very promising prostate cancer treatment results [24]. In these studies MNPs were dispersed in ferrofluids and injected into the tumor volume. Since then numerous in vitro and animal studies have been performed, but approval for clinical applications has so far only been granted in Germany. Remaining problems of magnetic hyperthermia are targeting hyperthermia to specific tumor areas, and achieving within those areas a homogeneous heat distribution without affecting adjacent cells. Neighboring tissue can be affected by hyperthermia if MNPs are not completely cleared out from healthy tissue, by diffusion of MNPs from tumor areas to surrounding tissue, or by heat conduction. If these issues are resolved, MNPs have the potential of multifunctional use in diagnostics and therapy of various diseases.

14.3.6 Magnetomechanical cell destruction

Again using magnetic nanoparticles, a completely different approach for cancer cell destruction has recently been taken, based on mechanical force exerted by magnetic particles on cell membranes in low magnetic fields and at low frequencies [25]. This is in contrast to magnetic hyperthermia, which uses high fields and high frequencies at several hundred kHz [22]. We have already seen that a ferromagnetic multidomain state collapses into a single domain state when reducing the lateral extension down to the nanometer length scale. In circular disks of specific size and aspect ratio a spin vortex state develops. The critical diameter and aspect ratio of the vortex state is governed by the exchange interaction, the magnetocrystalline anisotropy, and the shape anisotropy. For an $Ni_{0.8}Fe_{0.2}$ alloy, known as soft magnetic permalloy, the critical diameter is about 1 µm. The magnetic vortex state in zero external magnetic field

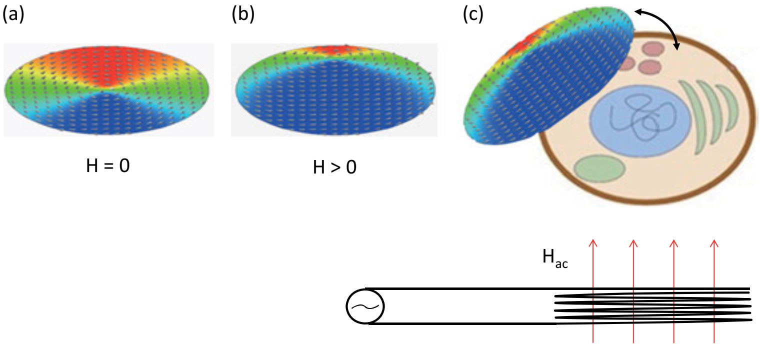


Fig. 14.12: Magnetic microdisks used for magneto-mechanical destruction of cancer cells. (a) In zero magnetic field the magnetization of the disk is characterized by a magnetic vortex state with vanishing magnetization and a vortex core at the center; (b) by applying a magnetic field, the vortex core moves out of the center and the magnetization increases proportional to the external field; (c) magnetic disks immersed in the extracellular space of tumor tissue can be rocked in an ac magnetic field. The low frequency oscillation exerts mechanical force onto cellular membranes that initiates apoptosis.

is illustrated in Fig. 14.12 (a). In the vortex state the magnetic flux is completely closed within the disk, no stray fields emanate from the disks, and the magnetization M integrated over the disk volume is zero. This has decisive advantages as these vortex disks in the remnant state do not tend to agglomerate in watery solution due to the lack of dipole-dipole interaction, unlike single domain magnetic nanoparticles.

When a magnetic field H is applied to the magnetic disk, then the vortex core moves out of the center in the direction normal to the applied field direction, as shown in Fig. 14.12 (b). This causes an asymmetry in the spin distribution and a finite magnetization. For small fields the magnetization is linearly proportional to the external field: $M = \chi_\mu H$, where χ_μ is the magnetic susceptibility of the permalloy microdisk. At the same time, the magnetic field exerts a torque T on the disk $\vec{T} = \vec{M} \times \vec{H}$, which lets the disk turn into the field direction. This effect can be used for rocking the magnetic disks back and forth in the extracellular space of tumor volumes by an external AC magnetic field H_{AC}. It turns out that small field amplitudes of less than 100 Oe and small frequencies of about 10–20 Hz already have a dramatic effect on the integrity of cell membranes, initiating apoptosis within a short exposure time of only 10 minutes [25]. The mechanism of induced cell death is not entirely clear. The damage may cause a disruption of the membrane or of the ion channels. In any case, it appears that the mechanical oscillation of the magnetic disks attached to the cancer cells is transduced to a transmembrane ion current that triggers apoptosis. For targeting cancer cells the microdisks are first coated with a gold layer to enhance biocompatibility and then conjugated to antibodies that bind to specific antigens overexpressed on the surface of targeted cancer cells.

14.4 Metal nanoparticles for diagnostics and therapy

14.4.1 X-ray imaging contrast

In x-ray radiography and CT imaging a contrast problem is often encountered. X-ray contrast solely depends on lateral differences in electron density. The biomolecular structure from one part of tissue to another may be very different. However, if there is no difference in electron density, x-ray radiography based on x-ray attenuation will not detect a difference. Bones can clearly be distinguished because of their high Ca content with an above average electron density. But gaining contrast of soft tissues and organs is more challenging. For imaging the gastrointestinal tract it is common practice to first swallow a higher contrast Ba_2SO_4 slurry before taking an x-ray radiograph. Similarly, problems occur when recording radiographs of the lung, which is additionally hampered by the thoracic skeleton in the foreground (see Section 5.3.3 for a discussion), of inner organs (kidneys, liver, spleen, etc.), or when imaging blood vessels (angiography). In all these cases contrast enhancement with the application of high atomic number nanoparticles may provide a solution.

At present the most frequently used contrast enhancing agents are small nonionic iodine ($Z = 53$) containing molecules, which come in a variety of configurations differing in viscosity, osmolality, and number of iodine ions per formula unit. The molecular structure of one of the iodine contrast agents (CAs) is shown in Fig. 5.17. There are a number of problems associated with these CAs. Large quantities have to be injected to compensate for the low iodine concentration. These small molecules have a short circulation half-life in the body because of rapid renal clearance, which – in turn – can cause renal problems. To overcome the adverse effects of iodine-containing CAs, a number of alternative NPs have been engineered in recent years, the main idea is briefly presented here. More extensive reviews can be found in [26, 27].

In general, when engineering NPs for x-ray contrast enhancement they should fulfill at least part of the following conditions: (1) high contrast efficacy; (2) excellent biocompatibility; (3) proper size; (4) long in vivo circulation time; (5) long colloidal stability in the body; (6) reasonable cost effectiveness.

High contrast efficacy implies the use of high Z materials at a high density to lower the total amount of NPs needed for the required contrast. Size of NPs and long in vivo circulation are interconnected requirements, as already discussed in Section 14.2. If the nanoparticles are smaller than 10 nm, they will be cleared out through the kidneys and the circulation time is shortened. Simultaneously, such small NPs may constitute a potential hazard to the glomerular filtration system. If they are of the size of 100 nm and above, they may be eliminated by monocytes and deposited in the liver and spleen. So, proper size matters and is somewhere between 20–80 nm. But the size effect can also be used for preferentially imaging liver and spleen using particles that are bigger than 100 nm and recognized by RES. Vice versa, if those NPs do not yield a high contrast of liver and spleen, this would indicate damaged macrophages and Kupffer cells as a result of diseased respective organs.

High contrast efficacy can be achieved by using NPs containing a core of high Z metal ions, such as Au ($Z = 79$), Bi (83), or lanthanide ions from Gd (64) to Yb (70). The high Z atoms should be nonoxidizing and nonsolvable. Since some of the metals with the exception of Au are highly oxidizing, the metal core has to be protected by a tight and impermeable shell surrounding the heavy metal core. This is usually achieved by a silica coating and PEG shell that feature additional biocompatible properties, as indicated before and schematically illustrated in Fig. 14.5.

In standard x-ray radiography resonant K-edge absorption of x-ray photons does not play any role since the K-edges of soft materials have too low energies with respect to the mean photon energy of the bremsstrahlung spectrum when employing excitation energies between 80–140 kVp. Even for Ca (4 keV) and I (33.2 keV) the absorption edges are too low to be of any relevance. However, K-edge absorption becomes more significant for nanoparticles containing heavy metal cores, such as Au (80.7 keV) or Bi (90.5 keV) and even for the lanthanide ions K-edge absorption is significant (61.4 keV for Yb). For these CAs K-edge absorption contributes markedly to the total attenuation of the x-ray beam [28]. Using this fact, NPs can be designed to contain heavy metal alloys that cover an energy range of highly absorbing K-edges. Furthermore, K-edge contrast imaging can be applied by taking two sequential pictures, one with weighted average x-ray photon energy above an absorption edge and another one below the same absorption edge. How the weighted average photon energy can be adjusted is explained in Section 5.3.3. At present, Au NPs are favored for K-edge contrast imaging because of their superior biocompatibility, their nontoxic and nonreactive chemical properties, and because of their comparatively easy synthesis.

NPs containing magnetic materials, such as Gd, can also be used for MRI contrast enhancement. NPs containing Au may also be used for plasmonic resonance in the near-infrared regime for photothermal therapy. These are but two examples for NPs with bimodal properties: x-ray and MRI diagnostics for Gd-containing NP; x-ray diagnostics together with photothermal therapy using Au NPs. Along the same lines, multimodal NPs have been designed with sensitivity to several imaging modalities and functionalities, such as Au NPs coated with Gd chelates or Fe_3O_4 for x-ray contrast, MRI contrast, and photothermal therapy. Multimodal NPs are discussed in more detail below.

There is a large variety of NPs that can be used to replace iodine-containing CAs. Many have been tested in vivo but only few have been approved for clinical use. The main remaining problem is the cost effectiveness. Nanoparticles containing Au or lanthanide ions tend to be much more expensive than iodine-based CAs that are presently mass produced at a lower cost. For a normal CT the cost can be 10 to 20 times higher when using high Z metal NP CAs instead of iodine-containing molecules, which prohibits their routine use. But for research and for patients with iodine intolerance the NP route is a viable alternative.

14.4.2 Plasmon resonance

The optical properties of metal NPs are of considerable interest in various fields of science because of their tunable absorbance of electromagnetic waves by size and aspect ratio of the NPs and hence their color variation. It turns out that the optical properties of metal NPs are also of interest for nanomedical applications in a three-fold way: they enhance the fluorescence of fluorophores; they can be used for hyperthermia, and they are used for photochemical therapy to be discussed below.

When light (electromagnetic waves, EM waves) strikes molecules or particles, the electrons in the particles are excited to dipole radiation. If the particle diameter d is much smaller than the wavelength of light ($d \ll \lambda$), the emitted EM wave radiates homogeneously in all directions (4π). This type of radiation is called Rayleigh scattering and is responsible for the blue sky during daytime and red sky in the morning and evening. With increasing particle size, light scattering becomes preferentially directed in the forward direction. This scattering is known as Mie scattering. Both Rayleigh and Mie scattering are illustrated in Fig. 14.13.

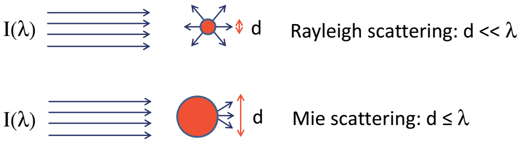


Fig. 14.13: Rayleigh and Mie scattering of light at particles of various size.

If the particles consist of a metal containing a “free” electron gas, the incident EM wave can set these electrons into “plasma” oscillations: the free negative electrons oscillate against the rigid positive ionic background like in a tiny antenna. This has a decisive consequence for the interaction of EM radiation with metal particles: metal particles strongly scatter *and* absorb EM waves. Hence the refractive index has to be treated as a complex and wavelength dependent number. Gustav Mie worked out in a seminal paper the absorption and scattering cross sections of EM waves at spherical metal nanoparticles [29]. To his honor the scattering at metal particles is referred to as *Mie scattering*. Here we outline a simplified version of the main results.

In a static case the dipolar polarization $\vec{P}$ of matter in an electric field $\vec{E}$ is expressed by: $\vec{P} = \alpha\vec{E}$, where α is the polarizability characterizing the material, here the metal NP. For a metal sphere of radius $r = d/2$ in an oscillating EM field the polarizability can be expressed as [30]:

$$\alpha = \frac{\varepsilon_\mathrm{p} - \varepsilon_\mathrm{m}}{\varepsilon_\mathrm{p} + 2\varepsilon_\mathrm{m}} r^3 = Qr^3 .$$

Here ε_p and ε_m are the dielectric permittivities of metal particle (p) and surrounding medium (m), respectively. This leads to the scattering and absorption cross sections [26]:

$$\sigma_{\text{scat}} = \frac{8\pi}{3}k^4 r^6 |Q|^2; \quad \sigma_{\text{abs}} = 4\pi k r^3 \,\text{Im}(Q),$$

where $k = 2\pi/\lambda$ is the wavenumber of the EM radiation. Note that the absorption cross section scales with r^3, while the scattering cross section scales with $(r^3)^2$. Thus size determines which cross section dominates. The scattering cross section contains the typical k^4 dependence that is characteristic for Rayleigh scattering. Absorption and scattering together contribute to the total extinction cross section in the forward direction that determines the color appearance of metal NPs in transmission.

The Q factor in the equation above is strongly wavelength dependent and determines the frequency ω of the absorption maximum. According to the Drude model the dielectric response of a free electron gas is [31],

$$\varepsilon_{\text{p}} = 1 - \left(\frac{\omega_{\text{p}}}{\omega}\right)^2 \quad \text{with } \omega_{\text{p}} = \sqrt{\frac{n_{\text{e}} e^2}{\varepsilon_0 m^*}}$$

where ω_{p} is the plasma frequency of the free electron gas, specific for each metal, n_{e} is the free electron density, e is the electrical charge, ε_0 is the vacuum permittivity, and m^* is the effective electron mass in the specific metal. In a quantum picture the collective plasma oscillation of electrons in metals is referred to as plasmon in analogy to phonons representing lattice vibrations. In first approximation an absorption maximum is expected for the condition $\varepsilon_{\text{p}} + 2\varepsilon_{\text{m}} = 0$ in the denominator of the polarizability. From this we can estimate the frequency for the *surface plasmon resonance* (SPR) peak: $\omega_{\max} = \omega_{\text{p}}/\sqrt{2\varepsilon_{\text{m}} + 1} \approx \omega_{\text{p}}/2$ or $\lambda_{\max} \approx 2\lambda_{\text{p}}$ for a watery environment of the NPs. Figure 14.14 shows the absorption maxima for spherical Au NPs of different sizes, reproduced from [32]. Au particles have an absorption maximum at 525 nm but low absorbance in the red regime from 600 to 700 nm. Therefore Au particles appear red for particle sizes below 100 nm. At particle sizes of 100 nm the absorbance peak is shifted to 600 nm, while low absorbance occurs in the 400–500 nm regime. These Au particles therefore appear bluish. Ag particles show an absorption maximum in the UV range at 375 nm, shifting to longer wavelengths with increasing particle size. The shift of the absorbance peak with increasing size is the generally observed trend for the SPR peak position. Although the size dependence of the SPR frequency shift is poorly understood [33], it is an important property for nanomedical treatments. Most likely the shift can be understood in terms of effective electron density variations in the surface layer of NPs.

While there is still some uncertainty with respect to the SPR frequency shift as a function of spherical NP size, there is a clear dependence of SPR frequency with respect to their aspect ratio. In addition to dipolar extinction, spherical NPs exhibit strong quadrupole contributions [30]. By changing the shape from spherical to prolate, the quadrupole contribution vanishes. In Au nanorods, the absorption maximum becomes a clear function of the aspect ratio (AR) as seen in Fig. 14.15 [34]. In fact, there are two resonances: one fixed and one changing with the AR. The changing one is due to the *longitudinal* SPR in the direction parallel to the rod, while the fixed one refers

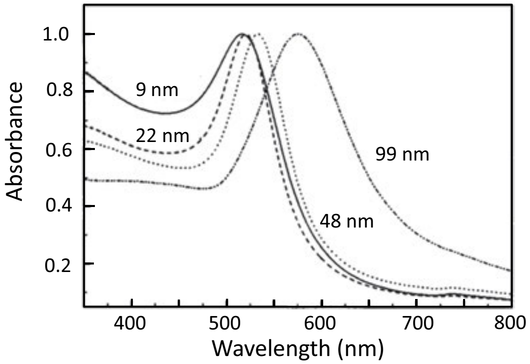


Fig. 14.14: Absorbance of Au NPs with different size as a function of incident wavelength. All curves are normalized to the maximum absorbance (reproduced from [32] by permission of ACS).

to the *transverse* SPR in the direction normal to the rod axis. The tunability of the SPR by the AR is a very interesting property to be exploited in photothermal therapy. It is also important that these SPR peaks coincide with the *water window* for tissues, which stretches from 650 nm to about 1300 nm. In this region light has the highest penetration depth in tissue, compare Fig. 13.14.

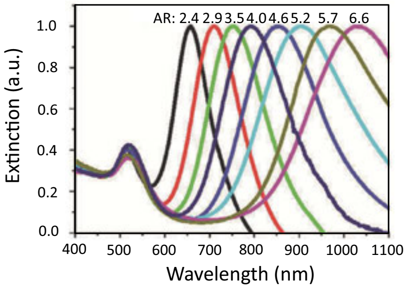


Fig. 14.15: Longitudinal and transverse surface plasmon resonance peaks in Au nanorods as a function of aspect ratio (AR). All spectra are normalized to maximum extinction. The colors of the curves have no relation to the actual color appearance of the nanoparticles in watery solution. The latter ones vary from bluish for AR = 2.4 to reddish for AR = 6.6 (reproduced from [34] by permission of Elsevier Publishing Inc.).

Au particles with optical tunability in the infrared regime have also been fabricated in the form of shells coating silica cores. It turns out that here the SPR is solely determined by the ratio shell to core thickness, independent of the particle size [35]. This has great advantages for biomedical applications: first, the particle size can be adapted to the required biodistribution in the body; and second, reshaping of particles under heat load is not relevant.

The cross sections quoted above indicate that the scattering cross section increases more rapidly with increasing radius than the absorption cross section. This is indeed the case as can be seen in Fig. 14.16 for Au nanospheres [36]. At a diameter of 80 nm the scattering cross section approaches the one for absorption; and for NP size larger than 100 nm scattering completely dominates over absorption. This dependence guides nanomedical applications: smaller sized NPs are preferentially used for absorptive applications where EM energy can be converted into heat of NPs that is transferred to the embedding tissue; larger sized NPs are more beneficial for imaging because of their strong scattering.

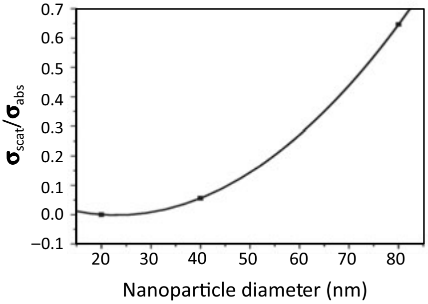


Fig. 14.16: Ratio of scattering cross section to absorption cross section. For NPs of about 20–30 nm in diameter, absorption completely dominates over scattering. With increasing diameter, the scattering cross section becomes comparable to the absorption cross section (adapted from [36] by permission of ACS).

14.4.3 Photothermal therapy

Some basics of photothermal therapy (PTT) by laser exposure are discussed in Section 13.4.2. PTT with the assistance of metal NPs is much more effective than without. The most favored NPs for PTT are Au and Ag nanorods due to their relative inertness and absorption bands in the *near-infrared* (NIR) region. Figure 14.14 shows the extinction peak of Au NPs which can be tuned by their size and aspect ratio. The laser wavelength, on the other hand, is selected by the required penetration depth. According to Fig. 13.14 skin treatment is best performed with a YAG or CO_2 laser, whereas deeper tissues can be reached by an Ar laser. Once the laser wavelength is selected, the size and shape of Au or Ag NPs can be adjusted for maximum surface plasmon extinction. However, there is an intrinsic stability problem with nanorod particles due to their shape. Nanorods are thermodynamically unstable because of their high surface energy. To lower the surface energy, NPs with high aspect ratio tend to reshape themselves towards sphericity. Surface diffusion is enhanced and reshaping accelerated upon high heat load while being in plasmonic resonance. Reshaping, in turn, shifts the plasmon resonance energy and also the photothermal conversion efficiency.

Au nanoparticles can be fabricated with a solid Au core (Au NP) or by coating a spherical core of dielectric material forming a nanoshell (Au NS). In the latter case the SPR is determined by the outer diameter and the shell thickness. The tunability of the SPR appears to be the result of hybridization between inner cavity and outer surface plasmon resonance. A red shift of 300 nm has been achieved for variation of the shell thickness between 20 nm and 5 nm [37].

The therapeutic benefit of Au NPs or NSs results from resonant light absorption at the respective plasmon energy. Photon excitation of the electron gas in the metal NPs is followed by a relaxation on the femtosecond time scale, mainly due to electron-electron scattering, leading to a rapid increase of the surface temperature. Electron-phonon coupling then causes heating of the lattice on the picosecond time scale [38]. The sequence of events is illustrated in Fig. 14.17. NPs in contact with tissue dissipate their thermal energy to the surrounding medium conductively and by infrared radiation. When Au NPs in a cellular medium are illuminated by laser light, the large temperature difference between hot Au NP surface and cooler surrounding tissue leads to cell death. Heating rate and cooling rate must be in equilibrium, otherwise melting of the NPs may occur. Thermal equilibrium can be better controlled with a cw laser than with a pulsed laser. As already mentioned, intense photothermal heating may result in melting or reshaping of the nanostructure, changing its optical properties irreversibly. In conclusion, Au NPs have excellent properties for hyperthermia, including nontoxicity and tunable plasmon resonance for optimal absorption. The only limitation is the penetration depth of laser light, which is in the order of 10 mm in the infrared regime. For most applications this is sufficient. Alternatively, hyperthermia with MNPs should be used which is not depth limited.

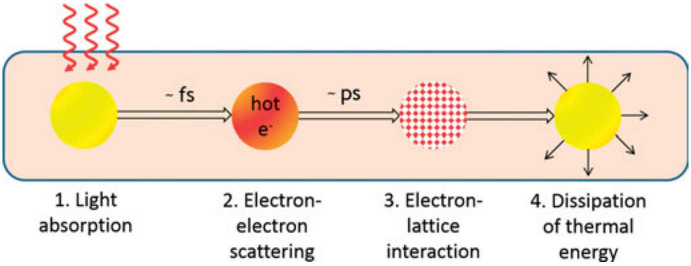


Fig. 14.17: Sequence of events after resonant light absorption by Au nanoparticles in tissue: 1. light absorption at plasmonic resonance; 2. electron-electron scattering heats up the electron gas in the metal; 3. electron-lattice interaction heats up the nanoparticle; 4. dissipation of thermal energy by radiation and heat conduction to the surrounding tissue.

14.4.4 Imaging

As stated in Section 14.4.2, Au NPs scatter light strongly, the details depend on size, shape and geometry. The large scattering cross section for particles larger than 50 nm makes them very promising for imaging carcinoma, provided that they are conjugated to antibody receptors and targeted to cancer cells. Upon attachment to antigens, the NPs reveal the location of cancerous cells by their strong and colorful scattering properties. The SPR scattering from Au NPs can be imaged particularly well by a dark field microscope, which collects only scattered light against a dark background [39]. The principle of a dark field microscope is shown in Fig. 14.18 in comparison to a normal bright field microscope. The main difference is the aperture, which blocks direct light from the white lamp. So the objective lens images only those rays that have been scattered at the sample. Figure 14.19 shows an image taken with a dark field microscope from Au nanorods accumulated in cancer tissue. In comparison, Au nanorods in heathy tissue are dispersed at low density but are not agglomerated. The gold nanoparticles show a red color due to SPR in the NIR region. Cancer imaging is made possible by using Au NPs conjugated to antibodies which specifically and homogeneously bind to the surface of the cancer cells with greater affinity than to the noncancerous cells. The intensity difference can be as high as a factor of 6 [39].

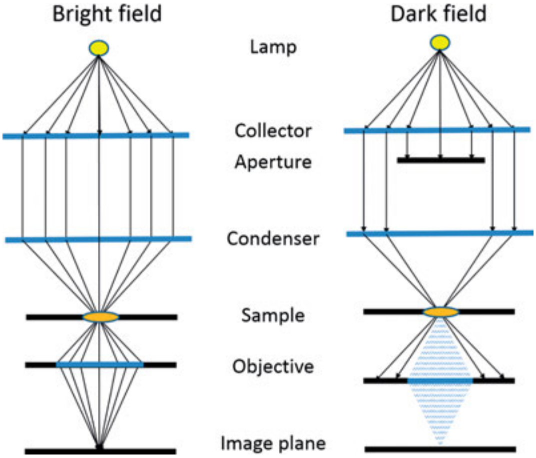


Fig. 14.18: Comparison of a bright field microscope and a dark field microscope. In a dark field microscope particles in the sample that scatter light appear bright against a dark background.

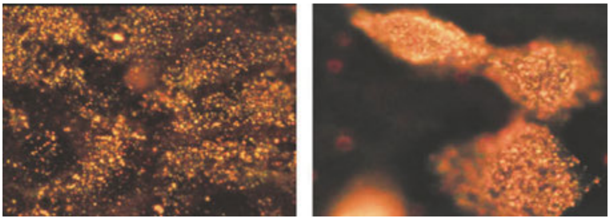


Fig. 14.19: Dark field light scattering images of Au nanorods conjugated with antibodies in healthy tissue (left) and cancerous tissue (right). In healthy tissue the Au NPs are randomly distributed, whereas in cancer cells the NPs are agglomerated. The Au nanorods show a red color due to SPR in NIR (reproduced from [39] by permission of ACS).

14.4.5 Surface enhanced Raman scattering

Metal NPs exhibit further remarkable optical properties upon excitation of their surface plasmon resonance. At SPR conditions a considerable electromagnetic field enhancement occurs close to the NP surface. It has been shown that due to this overlap the fluorescence emission of fluorophores can be enhanced by a factor of 50 [40]. It is, though, important to keep a proper distance of a few nanometer between fluorophore and NP surface. The enhancement effect can be utilized for imaging chemical sensors and/or biosensors with specific resonances overlapping with the SPR when conjugated to metal NPs. Similarly, Raman scattering of biomolecules experiences a dramatic enhancement under SPR conditions and can also be used for imaging [41, 42].

Raman scattering is a spectroscopic method for studying vibrational excitations of molecules. When incident light is scattered at atoms or molecules, the outgoing wave may have the same frequency or may be blue or red shifted. In the first case we talk about Rayleigh scattering. Rayleigh scattering is an elastic scattering process without any energy transfer to the specimen. In the graph of Fig. 14.20 Rayleigh scattering is symbolized by an upscattering of the samples' electronic state to a virtual and nonresonant energy level E_i indicated by a dashed line, and subsequently by a downscattering to the initial state. The energy of the incident and scattered waves are identical: $E_\mathrm{i} = E_\mathrm{f}$. Elastic or Rayleigh scattering is the dominating process when scattering light at molecules or nanoparticles.

Raman scattering, in contrast, is an inelastic scattering process and under normal circumstances comparatively rare. On average only about 1 ppm scattering events are of Raman type. Some energy $\Delta E = \pm\hbar\Omega$ is transferred between light and internal excitations of the sample. In Raman spectroscopy these small energy shifts are analyzed. Raman spectroscopy has to be distinguished from *photoelectron emission spectroscopy* (PES). In PES the entire photon energy is transferred to potential and kinetic energy of electrons and the remaining kinetic energy of the emitted electron is analyzed. In the case of Raman scattering, energy transfer is only partial. Raman scattering can best be described in a quantum picture. Let us assume a monochromatic incident photon with energy $E_\mathrm{i} = \hbar\omega$. This photon modulates the molecule such that it becomes excited to a vibrational or rotational energy level of energy $E_\mathrm{vib} = \hbar\Omega$. Because of energy conservation the final photon energy is $E_\mathrm{f} = \hbar\omega - \hbar\Omega$. The excitation of a molecule leads to a red shifted *Stokes* line. Vice versa, the photon may also take up the vibrational energy of a molecule during the scattering process, in which case the final photon energy is blue shifted (anti-Stokes line): $E_\mathrm{f} = \hbar\omega + \hbar\Omega$. Alternatively, these internal energy levels can be excited directly and resonantly by electromagnetic waves in the terahertz frequency regime (THz). In principle, momentum conservation also has to be taken into account in these scattering processes. However, since the vibrational excitations are smaller than the incident light energy by a factor of 10^3, momentum conservation can be neglected.

Rayleigh scattering and Raman scattering are fast processes on the time scale of femto- to picoseconds. In contrast, fluorescence is a much slower process on the time scale of nanoseconds, and luminescence is even slower. First, molecules are excited to a higher energy level. Often a nonradiative transition takes place to an intermediated vibrational level, followed by a transition to the ground state. Fluorescence emission is recognized by colored light after exposing the specimen to white light or UV radiation, sometimes even long after exposure. The fluorescence spectrum is characteristic for the (bio-)molecule exposed to light but independent of the wavelength of the exciting radiation as long as $\lambda_{\text{excitation}} < \lambda_{\text{fluorescence}}$. Fluorescence spectroscopy is similar to Raman spectroscopy and can be used for discriminating malignant skin tumors from benign ones. The schematics of these different spectroscopic transitions are illustrated in Fig. 14.20.

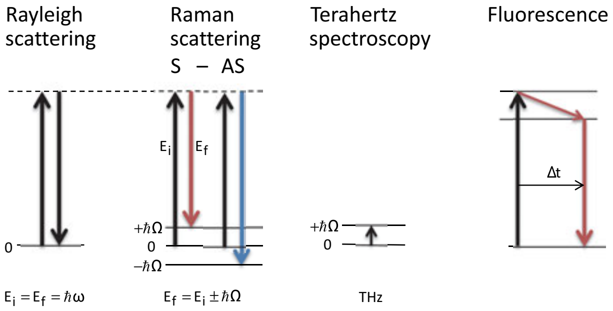


Fig. 14.20: Comparison of different interactions of light with matter: elastic Rayleigh scattering, inelastic Raman scattering, direct level excitation with terahertz radiation, and fluorescence transition. Fluorescence light is not immediate but occurs with some time delay Δt. S = Stokes, AS = anti-Stokes.

Raman spectroscopy is a widely used method in condensed matter physics and physical chemistry for studying vibrational excitations of solids and molecules. It is also used in medicine, and in particular for histological examinations of healthy versus pathological cells. The Raman spectrum of molecules is a fingerprint of those molecules and can be used for their identification and characterization. When scanning a monochromatic beam across the surface of a specimen with the Raman analyzer tuned to a specific and characteristic transition, the distribution of this molecule in the specimen can be imaged [41].

As already stated, under normal circumstances Raman scattering is a rare process, which occurs in the order of 1 ppm with respect of ordinary Rayleigh scattering. However, the Raman scattering cross section becomes drastically enhanced in the presence of Au or Ag NPs. When these NPs are excited to SPR, the strong electrical field at their surfaces enhances the Raman cross section, which scales with the square of the electrical field amplitude. This enhancement process is termed *surface enhanced Raman* *scattering* (SERS). It is more effective for nanorods than for nanospheres, as nanorods act like local antenna. The enhancement factor can be in the order of 10^4–10^5 for adsorbed molecules on nanorods [40]. The current status and future perspectives of Raman imaging in medicine are discussed in [41, 42].

14.5 Multimodality of theranostic nanoparticles

One and the same nanoparticle may feature several beneficial properties for diagnostics and therapeutics. Best examples are Au NPs and magnetite nanoparticles. Au NPs can be used for hyperthermia, as CT contrast agent, and for dark field imaging because of their strong scattering properties of light. Magnetite NPs are used for hyperthermia and for MRI contrast enhancement. To further increase the multimodality of nanoparticles, different materials can be combined. An example is shown in Fig. 14.21 (a) from [2]. Here the nanoparticle has a gold shell deposited on a dielectric core like silicon oxide. The shell diameter and thickness is selected on the one hand for surface plasmon resonance in the near-infrared, and on the other hand for good light scattering in dark field imaging. Magnetite nanoparticles for MRI contrast enhancement are embedded in the Au nanoshell. At a certain distance to the Au shell surface, fluorophores emitting in the green spectral range are deposited. The nanoshell is then functionalized with proteins to bind antibodies for targeting cancer cells. Finally the nanoparticle is coated with PEG to support biodistribution in the body and to hinder protein adsorption and opsonization and antibodies are attached to the surface for targeting cancer cells. This nanoparticle has, apart from protective coatings, seven functionalities: targeting with antibodies; therapy with PTT (and possibly magnetic hyperthermia); diagnostic with CT and MRI contrast enhancement, SERS, fluorescence, and dark field imaging. In panel (b) and (c) of Fig. 14.21 the specific relaxivity provided by magnetite NPs embedded in the Au nanoshell and the enhanced fluorescence intensity in the infrared regime are shown.

While the development of NPs for cancer diagnostics and therapy has progressed tremendously over recent years, clinical studies are mostly lacking. In animal studies accumulation of targeted NPs in cancer tissue has been confirmed and regression of tumor volume upon application of hyperthermia has been observed. However, long-term survival tests are still largely lacking and therefore it is presently difficult to judge the success rate of a cancer treatment based on nanomedical methods. On the other hand, nanoparticles can also be used in other areas of medicine but oncology. For instance, gold coated nanoshells activated by SPR can be used for fusing together arteries, or NPs can be functionalized for directly binding pathogens and toxins in the blood stream. This emerging field holds much promise.

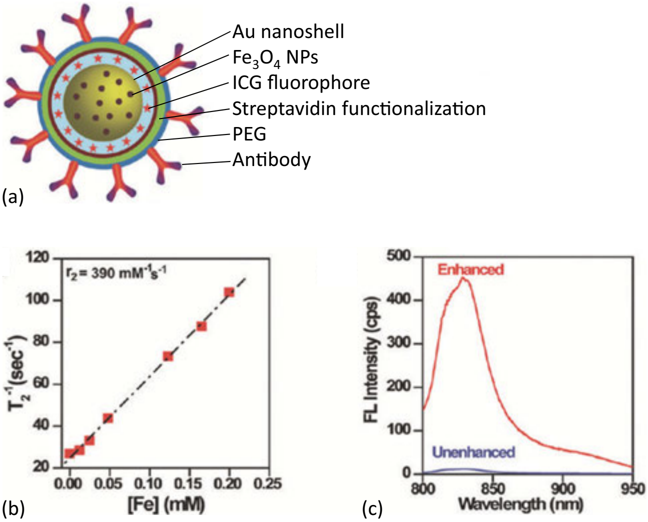


Fig. 14.21: Multimodal nanoparticle for targeting, imaging, and photothermal therapy. (a) Au nanoshell contains magnetite nanoparticles with adjustable density. The nanoshell is coated with an indocyanine green (ICG) fluorophore; the nanocomplex is then functionalized with the protein streptavidin for binding antibodies. The nanocomplex is finally passivated with PEG to reduce nonspecific binding. (b) Transverse relaxation rate (T_2^{-1}) as a function of Fe concentration in the magnetic NPs. The specific relaxivity r_2 obtained from the slope. (c) Fluorescence (FL) spectra of enhanced ICG from nanoparticles and unenhanced control (adapted from [2] by permission of ACS).

14.6 Summary

1. Nanomedicine is a multidisciplinary field with contributions from physics, materials science, biochemistry, polymer sciences, and many others.
2. Nanomedicine uses nanoparticles for diagnostics and therapeutics.
3. Nanoparticles that have both capabilities and purposes are called theranostic nanoparticles.
4. When using nanoparticles the pathway through the body has to be considered with respect to glomerular filtering, opsonization, and deposition in the liver.
5. The reticuloendothelial system (RES) is part of the immune system and consists of scavenger cells that help filter out dead and toxic cells.
6. Enhanced permeation and retention (EPR) effect of nanoparticles is the result of a leaky and ruptured blood vessel system in the tumor volume combined with a porous tissue.
7. For keeping NPs in the bloodstream their hydrodynamic diameter (HD) should be between 10–50 nm.
8. PEG coating consists of a biocompatible and water soluble polymer of varying chain length. PEG shells hinder protein adsorption, which suppresses opsonization and therefore bypasses RES.
9. We distinguish between physical targeting, passive targeting, and active targeting of NPs for diagnostics and therapeutics.

10. We differentiate between core size, size including coating and shell, and the hydrodynamic diameter (HD).
11. Superparamagnets are nanoparticles containing single domain ferromagnets.
12. The magnetization of superparamagnets fluctuates above the blocking temperature, but is stable below the blocking temperature.
13. The blocking temperature depends on the crystal anisotropy and particle size.
14. Magnetic contrast agents can be positive or negative. Positive contrast agents shorten T1, negative contrast agents shorten T2.
15. Magnetic nanoparticles can be used as contrast agent in magnetic resonance imaging and for hyperthermia.
16. High x-ray contrast efficacy requires the use of high *Z* elements.
17. The size and shape of gold nanoparticles can be tuned for maximum light absorption in the near infrared water window.
18. The therapeutic benefit of Au nanoparticles or nanoshells results from resonant light absorption at the respective plasmon energy.
19. Larger Au nanoparticles of about 100 nm diameter scatter light strongly, which can be used for dark field imaging.

References

[1] Nanomedicine, An ESF–European Medical Research Councils (EMRC) Forward Look Report; 2005.

[2] Bardhan R, Lal S, Joshi A, Halas NJ. Theranostic nanoshells: from probe design to imaging and treatment of cancer. Acc Chem Res. 2011 Oct 18; 44(10): 936–946.

[3] Kolosnjaj-Tabia J, Lartiguea L, Javed Y, Luciania N, Pellegrino T, Wilhelma C, Alloyeaub D, Gazeaua F. Biotransformations of magnetic nanoparticles in the body. Nano Today. 2016; 11: 280–284.

[4] Rao L, Xu JH, Cai B, Liu H, Li M, Jia Y, Xiao L, Guo SS, Liu W, Zhao WZ. Synthetic nanoparticles camouflaged with biomimetic erythrocyte membranes for reduced reticuloendothelial system uptake. Nanotechnology. 2016; 27: 085106.

[5] Longmire M, Choyke PL, Kobayashi H. Clearance properties of nano-sized particles and molecules as imaging agents: considerations and caveats. Nanomedicine (Lond). 2008 Oct; 3(5): 703–717.

[6] Ranganathan R, Madanmohan S, Kesavan A, Baskar G, Krishnamoorthy YR, Santosham R, Ponraju D, Rayala SK, Venkatraman G. Nanomedicine: towards development of patient-friendly drug-delivery systems for oncological applications. Int J Nanomedicine. 2012; 7: 1043–1060.

[7] De Crozals G, Bonnet R, Farre C, Chaix C. Nanoparticles with multiple properties for biomedical applications: A strategic guide. Nano Today. 2016; 7: 435–463.

[8] Scott AM, Wolchok JD, Old LJ. Antibody therapy of cancer. Nat Rev Cancer. 2012 Mar 22; 12(4): 278–287.

[9] Biological evaluation of medical devices – Guidance on the conduct of biological evaluation within a risk management process, ISO TR 15499, 2012.

[10] Black J. Biological performance of materials: fundamentals of biocompatibility. 4th edition. CRC Press, Taylor & Francis; 2006.

[11] Williams DF, editor. Definitions in biomaterials: Proceedings of a Consensus Conference of the European Society for Biomaterials, Chester, England. Elsevier; 1987.

[12] Benitez MJ, Mishra D, Szary P, Badini Confalonieri GA, Feyen M, Lu AH, Agudo L, Eggeler G, Petracic O, Zabel H. Structural and magnetic characterization of self-assembled iron oxide nanoparticle arrays. J Phys Condens Matter. 2011 Mar 30; 23(12): 126003.

[13] Bedanta S, Barman A, Kleemann W, Petracic O, Seki T. Magnetic nanoparticles: a subject for both fundamental research and applications. Journal of Nanomaterials. 2013: 952540.

[14] Fang M, Ström V, Olsson RT, Belova L, Rao KV. Particle size and magnetic properties dependence on growth temperature for rapid mixed co-precipitated magnetite nanoparticles. Nanotechnology. 2012 Apr 13; 23(14): 145601.

[15] Koenig SH, Kellar KE. Theory of $1/T_1$ and $1/T_2$ NMRD profiles of solutions of magnetic nanoparticles. Magn Reson Med. 1995 Aug; 34(2): 227–233.

[16] Estelrich J, Sánchez-Martín MJ, Busquets MA. Nanoparticles in magnetic resonance imaging: from simple to dual contrast agents. International Journal of Nanomedicine. 2015; 10: 1727–1741.

[17] Bin Na H, Song IC, Hyeon T. Inorganic nanoparticles for MRI contrast agents. Adv Mater. 2009; 21: 2133–2148.

[18] Laranjeira MS, Moc A¸ Ferreirad J, Coimbrae S, Costag E, Silvag AS, Ferreirah PJ, Monteiroa FJ. Different hydroxyapatite magnetic nanoparticles for medical imaging: Its effects on hemostatic, hemolytic activity and cellular cytotoxicity. Colloids and Surfaces B: Biointerfaces. 2016; 146: 363–374.

[19] Chaughule RS, Purushotham S, Ramanujan RV. Magnetic nanoparticles as contrast agents for magnetic resonance imaging. Proc Natl Acad Sci India Sect A Phys Sci. 2012; 82: 257–268.

[20] Rosensweig RE. Heating magnetic fluid with alternating magnetic field. J Magn Magn Mater. 2002; 252: 370.

[21] Wang X, Gub H, Yang Z. The heating effect of magnetic fluids in an alternating magnetic field. Journal of Magnetism and Magnetic Materials. 2005; 293: 334–340.

[22] Fortin JP, Wilhelm C, Servais J, Ménager C, Bacri JC, Gazeau F. Size-sorted anionic iron oxide nanomagnets as colloidal mediators for magnetic hyperthermia. J Am Chem Soc. 2007 Mar 7; 129(9): 2628–2635.

[23] Liu XL, Guang Choo ES, Ahmed AS, Zhao LY, Yang Y, Ramanujan RV, Xue JM, Fan DD, Fan HM, Ding JD. Magnetic nanoparticle-loaded polymer nanospheres as magnetic hyperthermia agents. J Mater Chem B. 2014; 2: 120–128.

[24] Jordan A, Scholz R, Maier-Hau K, Johannsen M, Wust P, Nadobny J, Schirra H, Schmidt H, Deger S, Loening S, Lanksch W, Felix T. Presentation of a new magnetic field therapy system for the treatment of human solid tumors with magnetic fluid hyperthermia. Journal of Magnetism and Magnetic Materials. 2001; 225: 118–126.

[25] Kim DH, Rozhkova EA, Ulasov IV, Bader SD, Rajh T, Lesniak MS, Novosad V. Biofunctionalized magnetic-vortex microdiscs for targeted cancer-cell destruction. Nat Mater. 2010 Feb; 9(2): 165–171

[26] Liu Y, Ai K, Lu L. Nanoparticulate x-ray computed tomography contrast agents: from design validation to in vivo applications. Acc Chem Res. 2012 Oct 16; 45(10): 1817–1827.

[27] Lusic H, Grinstaff MW. X-ray-computed tomography contrast agents. Chem Rev. 2013 Mar 13; 113(3): 1641–1666.

[28] Lee N, Choi SH, Hyeon T. Nano-sized CT contrast agent. Adv Mater. 2013 May 21; 25(19): 2641–2660.

[29] Mie G. Contributions to the optics of diffuse media, especially colloid metal solutions. Ann Phys. 1908; 25: 377–445.

[30] Lance Kelly K, Coronado E, Zhao LL, Schatz GC. The optical properties of metal nanoparticles: the influence of size, shape, and dielectric environment. J Phys Chem B. 2003; 107: 668–677.

[31] Kittel C. Introduction to solid state physics. 7th edition. John Wiley & Sons; 1995.

[32] Link S, El-Sayed MA. Size and temperature dependence of the plasmon absorption of colloidal gold nanoparticles. J Phys Chem B. 1999; 103: 4212–4217.

[33] Peng S, McMahon JM, Schatz GC, Gray SK, Sun Y. Reversing the size-dependence of surface plasmon resonances. Proc Natl Acad Sci USA. 2010 Aug 17; 107(33): 14530–14534.

[34] Huang X, El-Sayed MA. Gold nanoparticles: Optical properties and implementations in cancer diagnosis and photothermal therapy. Journal of Advanced Research. 2010; 1: 13–28.

[35] Jain PK, El-Sayed MA. Universal scaling of plasmon coupling in metal nanostructures: extension from particle pairs to nanoshells. Nano Lett. 2007; 7: 2854–2858.

[36] Jain PK, Lee KS, El-Sayed IH, El-Sayed MA. Calculated absorption and scattering properties of gold nanoparticles of different size, shape, and composition: applications in biological imaging and biomedicine. J Phys Chem B. 2006 Apr 13; 110(14): 7238–7248.

[37] Prodan E, Radloff C, Halas NJ, Nordlander P. A hybridization model for the plasmon response of complex nanostructures. Science 2003 Oct 17; 302(5644): 419–422.

[38] Link S, Hathcock DJ, Nikoobakht B, El-Sayed MA. Medium effect on the electron cooling dynamics in gold nanorods and truncated tetrahedral. Adv Mater. 2003; 15: 393–396.

[39] Huang X, El-Sayed IH, Qian W, El-Sayed MA. Cancer cell imaging and photothermal therapy in the near-infrared region by using gold nanorods. J Am Chem Soc. 2006 Feb 15; 128(6): 2115–2120.

[40] Bardhan R, Grady NK, Cole JR, Joshi A, Halas NJ. Fluorescence enhancement by Au nanostructures: nanoshells and nanorods. ACS Nano. 2009 Mar 24; 3(3): 744–752.

[41] Mikla VI, Mikla VV. Raman spectroscopy in medicine. Medical Imaging Technology. 2014; 8: 129–141.

[42] Keating ME, Byrne HJ. Raman Spectroscopy in Nanomedicine: Current Status and Future Perspective, Mark E Keating and Hugh J Byrne Nanomedicine. Nanomedicine (Lond). 2013 Aug; 8(8): 1335–1351.

Further reading

Howard AK, Vorup-Jensen T, Peer D, editors. Nanomedicine. Springer Verlag; 2016.
Jain KK, editor. Handbook of nanomedicine. Springer Verlag; 2012.
Hehenberger M. Nanomedicine. CRC Press, Taylor and Francis; 2016.
Nanomedicine, Wikipedia.

15 Prosthetics

15.1 Introduction

When body parts and functions fail there are two possibilities for remedy: regeneration or replacement. Regeneration is always the better option. However, if by any reason this option is not available, replacement by a donor or by artificial parts is the second best choice. In this chapter we consider artificial replacements, called prostheses. Prostheses may replace limbs or organs and are supposed to take over as well as possible the function of the healthy organ. We distinguish between two types of prosthesis:

- Exoprosthesis, prosthesis external to the body;
- Endoprosthesis, prosthesis within the body, also called implants.

In general, all prostheses have an internal and external connection to the body. In contrast, wheel chairs, walkers or crutches are not considered prosthesis as they lack body connection.

Exoprostheses demand a high degree of biomechanics, materials science, bionics, electronics, computer technology, and biomedicine.

Endoprostheses cover a wide range of specifications. Any implant has its specific requirements and therefore a more general characterization would be inappropriate at this point. However, the overarching prerequisite for all implants is *biocompatibility* (see Section 14.2.8). Replacement of entire organs by artificial parts is presently not possible and may never be. But temporary replacements, for instance by ventricular assistant devices (VAD) such as the *Berlin heart* can be very useful and successful [1].

Orthopedics is a discipline that treats musculoskeletal disorders. In contrast to prosthetics, in orthopedics lost organs are not replaced by artificial parts, but implants may be used to support existing organs.

Prosthetics and orthopedics are a rapidly developing multidisciplinary science and technology field, but it is not an invention of modern times. The daughter of a priest in ancient Egypt, Tabaketenmut, who lived between 950 und 710 BC, had an artificial toe. The prosthesis was found as a burial object that showed signs of wear. Advanced materials science, microelectronics and bioengineering offer new solutions to prosthetic challenges. In this short overview only a glimpse on this field can be provided, illustrated with a few examples. For more information literature is provided at the end of this chapter.

DOI 10.1515/9783110553116-017

15.2 Exoprostheses

In most cases exoprostheses replace parts of the upper or lower limbs and are closely connected with the kinematics and elastomechanics of the body as presented in Chapters 2 and 3/Vol. 1. Paralympics are an excellent showplace for the development of new exoprostheses.

15.2.1 Lower limb prostheses

Lower limb prostheses mimic as well as possible the functions of legs and feet: standing, walking, running, turning, climbing, hunkering down, jumping, etc. The purpose of lower limb prosthetics is to enable amputees to resume or improve their mobility level to some fraction of the original. Amputations are distinguished according to their location, as illustrated in Fig. 15.1. Walking, as discussed in more detail in Section 2.3.4/Vol. 1, is the most basic mobility. At the same time it is a highly complex motion, involving the activation and coordination of many muscles, bones, and joints: rolling the toes, swinging the femur by lifting the leg, bending the knee, rolling the heels, stance of the foot, and repeating again with the other leg in well-defined rhythmic sequence. During this movement there are many other features that are being taken care off while walking, such as shock absorption when hitting the ground with one leg, supporting the weight of the entire body when lifting the other leg, and balancing the upright posture. For a lower limb prosthesis it is important for obvious reasons whether the prosthesis starts above (transfemoral) or below the knee (transtibial). We first discuss prostheses that start below the knee.

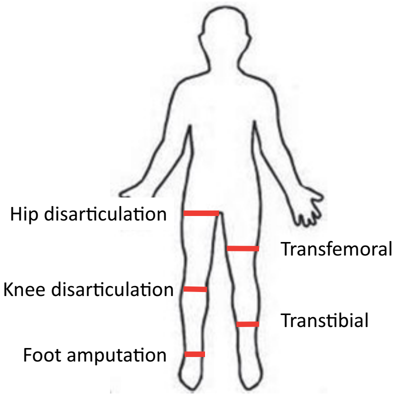


Fig. 15.1: Amputations of the lower limb.

(a) Transtibial prostheses

The basic prosthesis design for transtibial amputations consists of three parts: a socket at the top fits to the stump; a height adjustable shin connects the socket at the top and the foot together with the ankle at the bottom (see Fig. 15.2 (a)). Prosthetic foot and ankle serve a number of important functions during walking such as shock absorption during loading, stability during weight bearing, and smooth progression of the limb during walking. Furthermore, durability and structural stability are basic requirements for lower limb prosthesis. Some versions are shown in Fig. 15.2. The most basic design is the so called *solid ankle cushioned heel* (SACH) design, shown in Fig. 15.2 (a). The SACH foot includes a solid ankle and a rigid keel that runs along the length of the prosthetic foot sole. The heel is composed of a foam wedge that provides cushioning in the heel section during heel strike. This foot is commonly used on pediatric or geriatric amputees as an early version after amputation and before fitting more advanced prostheses. An advanced design is shown in Fig. 15.2 (b), providing good connection between stump and prosthesis. An inside liner at position (1) distributes pressures to create a precise fit and to increase connection between stump and prosthesis. Furthermore, a vacuum system in position (2) draws the stump into the socket for a tight fit. Further down, a torsion and shock absorber at position (3) increases the walking comfort. Carbon fiber springs at position (4) absorb energy by landing on the heels and release that energy to give a little push when leaving the ground. And finally a split toe design at position (5) adapts to uneven terrain, adding safety and stability when walking or standing. The toe design is shown in the inset without foot cover. Most of this prosthesis is mechanical using advanced materials. Only the vacuum pump requires power during fitting.

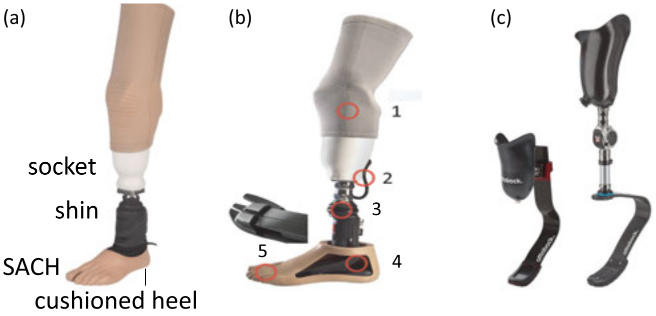


Fig. 15.2: Different prosthesis designs for lower limb transtibial amputation. (a) Basic design consisting of socket, shin, and a solid ankle cushioned heel (SACH) foot. (b) Advanced design featuring inside liner at position (1), vacuum pump at position (2), torsion and shock absorber at position (3), carbon fiber springs at position (4), and split toe at position (5). (c) Carbon reinforced blades for sport activities in two versions for transtibial and transfemoral amputations (reproduced from www.ottobockus.com/ by permission of Otto Bock).

The designs shown in Fig. 15.2 (a) and (b) are cosmetically adapted to a real foot. For sports activities, however, functionality is more important than cosmetics. In this case flexible carbon reinforced blades are preferred by athletes (Fig. 15.2 (c)). These blades are fabricated from carbon fiber reinforced polymers (CFRP), and unlike all previously discussed foot prostheses, they transform kinetic energy from walking or jumping into potential energy, like a spring, allowing the wearer to run and jump. CFRP is a strong and lightweight composite material, usually made from epoxy mixed with carbon fibers, which is also used for molding car bodies and components of aircrafts. In the CFRP blade manufacturing process, 30 to 90 sheets of impregnated materials are pressed onto a mold to produce the final shape. The mold is then autoclaved to fuse the sheets into a solid plate. The carbon reinforced blades have no motors, sensors or microprocessors. The blades solely use elastic response returning 80 % of the energy stored during compression.

(b) Transfemoral prostheses

Amputations above the knee are called transfemoral amputations. They require prostheses which are considerably more complex because of the additional degrees of freedom that the knee provides. Present day high prosthetics technology includes computer control of the entire course of motion during gait. Sensors measure the knee angle, the angular velocity of the knee joint, and load of knee and foot. The acquired data are fed into a microprocessor that controls hydrologic valves for steering the knee joint and for determining the angular velocity of the lower leg, aiming at a normal gait pattern. The inbuilt microprocessor allows for mimicking the actions of the knee and adapts to all walking speeds in real time.

Microprocessor controlled prostheses are called *C-legs*, an example is shown in Fig. 15.3. They were introduced in the late 1990s and are now in the fourth generation development stage. People with a C-leg perform significantly better on staircases compared with mechanical knees during stair descent, but there is not much of a difference in stair ascent [2].

Although C-leg prostheses perform better than mechanical prostheses, C-legs come with certain disadvantages. Most microprocessor controlled knee joints are powered by a battery housed inside the prosthesis, which adds to the weight and needs to be recharged every one or two days depending on the use. Nevertheless, in a comparative study of people wearing C-legs and various nonmicroprocessor control (NMC) prosthetic knees it could be shown that C-leg users have increased functional mobility and ease of performance improving their quality of life and lessening the impairment [3].

Fig. 15.3: Microprocessor controlled prosthesis (C-leg) for transfemoral amputations. The enlargement shows the valve system inside the lower C-leg (reproduced from www.ottobockus.com/ by permission of Otto Bock).

15.2.2 Upper limb prostheses

In contrast to the lower limb system, the upper limbs have increased flexibility and functionality but no standard motion pattern. This makes the construction of prostheses very complex. We distinguish three types of upper limb prostheses:

1. cosmetic prosthesis should look as natural as possible but have no function aside from holding or counteracting on an object (Fig. 15.4 (a));
2. mechanical or kinematic prosthesis that translates remaining muscle contraction in the remaining arm stump into movement of artificial arm and fingers;
3. myoelectric controlled prostheses that offer the ultimate combination of function and natural appearance.

The first functional arm prosthesis was designed by Ferdinand Sauerbruch (1875–1951), a German anatomist and surgeon. His design enabled simple movements to be executed with the remaining muscles (Fig. 15.4 (b)). Kinematic prostheses are contin-

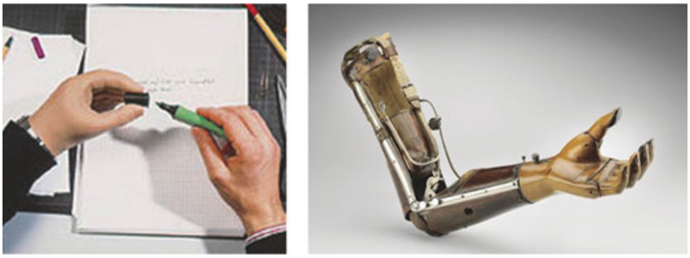


Fig. 15.4: Cosmetic hand (left) and original kinematic prosthesis designed by Sauerbruch (right).

uations of Sauerbruch's prosthesis allowing the amputee to open and close a gripper hand through motion of the shoulder. Thanks to continuous technological developments, the *myoelectric controlled prostheses* are the most advanced ones for the upper limbs that feature astounding capabilities to control translational and rotational motion with the help of several tiny motors.

Myoelectric prostheses use remaining nerve fibers, which run from the brain to the arm. The surgical procedure aimed at improving the control of upper limb prostheses is referred to as *targeted muscle reinnervation* (TMR). The residual nerve fibers after amputation of the upper arm are attached to muscles fibers on the chest, shown schematically in Fig. 15.5. Those nerve endings act as a phantom of the real arm. When deciding to lift the nonexisting arm, a corresponding muscle on the chest will contract. With the help of electromyography (EMG), explained in Section 6.9/Vol. 1, the electrical potential generated by muscle cells can be detected with electrodes on the skin. This EMG signal is amplified and processed to drive tiny motors that activate arm or finger motion. But first the EMG signals on the chest must be mapped out and uniquely correlated with brain activities for specific motions: lifting arm, turning hand, grabbing an object, etc. This requires intensive training. Then the EMG signal can be used for myoelectrical control of motors that actually execute the intended motion. Myoelectric hand prostheses with many degrees of freedom and very good mechanical capabilities are now commercially available: elbows that flex and extend with muscle signals so the arm can be extended to reach for a beverage and bring it up to the lips;

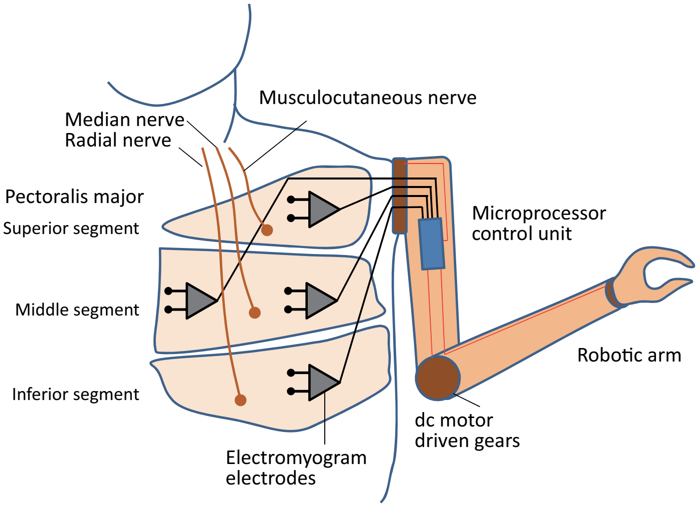


Fig. 15.5: Myoelectric control of upper limb prosthesis also referred to as targeted muscle reinnervation (TMR). Nerve fibers from the brain formerly ending in the arm are connected to different locations of muscles on the chest. When a muscle is activated by nerve fibers from the brain, the corresponding EMG signal is detected with sensors on the skin and translates this signal with help of a microprocessor into specific movements of the arm via DC motor control.

wrists that bend and rotate, allowing to position objects for convenient viewing and handling; hands that can carry a suitcase or hold an egg without cracking it; thumbs that can change orientation to multiple hand positions. If the upper arm still exists, the EMG signal can be taken directly from nerve endings in the upper arm. This facilitates the training process for the prosthesis dramatically.

EMG signals are weak, noisy and the mapping is not very precise. Alternatively, one may record the transmural potential from nerve fibers directly. However, this would require an invasive attachment of electrodes under the skin with all adverse consequences that this procedure bears, such as inflammation and electrode corrosion. For these reasons, recording EMG signals on the skin remains the preferred procedure.

In spite of impressive progress that myoelectric prostheses have seen in recent years, proportional and simultaneous control of a large number of degrees of freedom is presently not available, resulting in swift and precise but unnatural arm movements. Although there is no standard pattern for hand and arm movement unlike the normal gait, attempts are presently being undertaken to recognize patterns and to predict movements of the upper limb that would eventually mimic a more natural course of motion. Commercially available prostheses and scientific/technological developments are changing quickly. The current state is reviewed in [4] and in [5] the performance of different commercially available systems are compared and rated.

The opposite neural direction, i.e., transferring signals from the prosthetic hand to the brain has also been implemented in some cases. Sensors for temperature, pressure, surface topography, vibrations, etc. have been tested and their electrical signal output was transferred to nerve fibers that are ultimately processed and interpreted by the brain. This reverse signal transfer of sensory information provides a limited feedback, which is important for improving the dexterity of the hand and finger movement. There is still a long way to go. The main developments that can be foreseen for the future control of prosthetic arms and hands are the incorporation of pattern recognition and feedback systems.

Research and development of myoelectric control of upper limb prosthesis has also been beneficial to paraplegic patients with *spinal cord injuries* (SCI). In the case of SCI the limbs are still present but the voluntary motor control below the level of the spinal cord lesion is cut off. In one recent study it was shown that intracortically recorded signals from the motor cortex via a microelectrode array could be linked in real time to muscle activation to restore partial movement of a paralyzed patient [6]. In another recent study it was demonstrated that after long-term training with a body-machine interface, spinal cord injured paraplegics could improve somatic sensation and regain voluntary motor control in muscles below the lesion [7]. In this unique study partial motor control could be reactivated by gait training with a brain-controlled and EEG monitored robotic treadmill. This neurological recovery could hint to a hitherto unknown cortical and spinal cord plasticity triggered by long-term training.

15.3 Endoprostheses of the joints

The most frequently applied endoprostheses are those of the musculoskeletal system, i.e., hip, knee, and spinal discs. Other prostheses support the cardiovascular system such as heart valves, pacemaker and stents. A third type of endoprostheses support the sensory system, such as artificial lenses, middle ear implants, and cochlea implants. In this section we discuss endoprostheses of the musculoskeletal system, in particular hip and knee.

15.3.1 Hip replacement

The joint between femur and pelvis is a ball and socket joint (see Section 2.3.4/Vol. 1), providing three rotational degrees of freedom. Figure 15.6 (a) shows a cross section of femur and hip. The hemispherical head (ball) of the femur fits into the socket in the pelvic bone, called acetabulum. The bone surfaces of ball and socket are covered with articular cartilage, a highly specialized connective tissue that protects and cushions the bones and enables them to move easily. The cartilage surface is lubricated by a synovial fluid produced by a thin lining located close to the joint. The cartilage and lubrication is not unique to the hip joint, but is typical for all *synovial joints* in the body such as the knee. The ball-and-socket joint of the hip is tightly connected by a number of ligaments. Not shown in the cross section of Fig. 15.6 (a) is a small centered depression (pit) in the head, called the fovea capitis. The fovea capitis contains a central ligament connecting the head of the femur inside to the acetabulum of the hip bone. Three more ligaments (iliofemoral, ischiofemoral, and pubofermoral) augment the strong linkage of femur and pelvic bone from the outside. The iliofemoral ligament is the strongest ligament of the body with a tensile strength of about 3500 N [9]. These ligaments have to be cut open when replaced by prosthesis. Most “hip fractures” are actually fractures of the femur at the neck position, indicated by a red line in panel (a) of Fig. 15.6. This is also the limiting line for shortening the femur before inserting a metal shaft during hip surgery, as seen in panel (b).

We distinguish between prosthesis of the pivot head of the femur (thigh bone) and prosthesis of the counterpiece in the hip bone of the pelvis, the acetabulum. The replacement of the pivot head only is called *hemi-endoprosthesis* (HEP), the replacement of both head and acetabulum is called *total endoprosthesis* (TEP). Other frequently used terms are *total hip replacement* (THR) or *total hip arthroplasty* (THA).

There are three main reasons that necessitate HEP or TEP:

1. osteoarthritis leading to a breakdown of the cartilage that covers both sides of the hip joint (60 %);
2. fractures of the femoral neck due to osteoporosis (30 %);
3. fractures due to various accidents (8 %).

Other reasons are bone tumors, rheumatoid arthritis, and osteonecrosis (2 %). The percentages given are rough estimates. Osteoarthritis is by far the most prevailing reason for hip replacement.

First hip replacements started in the early 1960s and have steadily increased in number since then. According to a recent study in the US, the number of TEPs more than doubled over a 10 year span from 2000 to 2010 [8]. It has presently reached a record high of 0.3 % of the population that undergo HEP or TEP procedures annually. Prevalence of TEP increases with age and reaches about 10 % among people 80 years and older. More surprising is the fact that TEP has also risen by over 100 % in the age bracket 45–64 years during the same time span. Presently more than 4.5 million people in the USA live with an artificial hip. This large number highlights that a substantial fraction of the population can be kept mobile with a hip replacement despite arthritis thanks to a multidisciplinary effort of physicists, materials scientists, engineers, and clinicians to develop and improve suitable prosthesis.

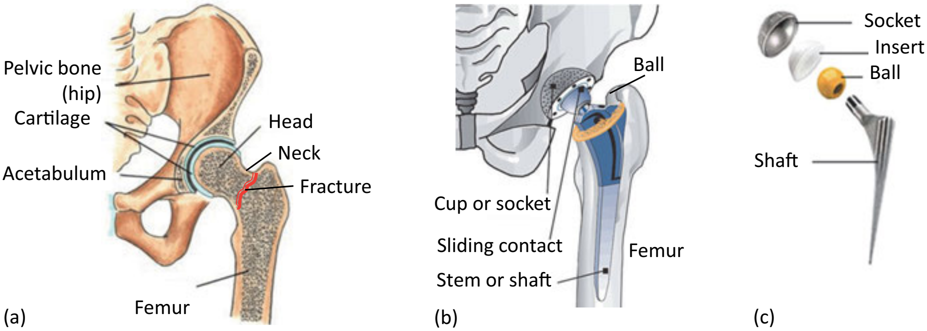


Fig. 15.6: (a) Cross section of femur and pelvic bone. The head of the femur is anchored in the acetabulum, both together forming a ball and socket joint for maximum mobility. The joint is kept together with a number of strong ligaments, not shown in the cross section. The red line indicates the position where hip fractures usually occur; (b) total endoprosthesis involves the replacement of the socket in the acetabulum and the head of the femur, now sitting in the form of a ball on a shaft that is inserted into the femur; (c) parts of a full prosthesis are shown, consisting of socket, ball and shaft. The details depend on the supplier. The head and the socket may be fabricated of metal, ceramics, plastics, or material combinations. In some designs an insert between ball and socket is used. The shaft is usually made of a sintered titanium alloy.

According to Fig. 15.6 (c) the implant for a TEP consists of a shaft, a ball and a socket, and sometimes also an extra insert between ball and socket is introduced. Hip replacement requires careful surgery as many muscles, vessels and nerve fibers intercept the working space. The shaft is a long and bulky mechanical piece that requires careful insertion into the femur. Implants are not custom made, but from a selection of sizes the best fitting one is chosen. With x-ray projection radiography the size of the required implant is roughly estimated and computer simulation aids the procedure indicating, in particular, the position for the incision to access the hip joint. After opening, the existing hip joint is completely removed. The natural socket is hollowed out and cleaned from remaining cartilage. Then a socket is inserted into this hollow space, eventually cemented into the pelvis. The upper part of the femur including the head is removed with a swing saw and a hole is drilled and rasped into the femur for inserting the shaft of the artificial hip stem. One end of the stem is inserted into the trabecular bone of the femur; the other end is topped with a smooth ball to fit into the socket. Ball and shaft allow for a rough length adjustment via a conical fit inside the ball head. In a first test the required length is estimated and later the ball is press fitted onto the shaft without further adjustment or possibilities for correction. Once ball and shaft are tightly attached and inserted into the femur, ball and socket are pressed into place and the incision is closed. There are standard and minimal invasive procedures offered for hip replacements. Minimal invasive procedures conserve all muscles and tendons, such that rehabilitation time of the patient is much shortened compared to the standard procedure.

Although hip replacement has become a standard and in most cases a very successful procedure, there are quite a number of issues that require attention. The success of the implant depends on the biocompatibility of all parts used, a low wear rate between socket and ball, and a biomechanical force distribution from the implant to the surrounding bone under dynamic load (walking, running, etc.), without causing permanent plastic deformation or strain on the femur. Some of the issues are discussed in more detail as follows.

(a) Cemented versus uncemented parts (Fig. 15.7)

In the case of cementation, parts are secured to healthy bone using acrylic cement. Uncemented parts are made from material that features a rough surface allowing bone to grow on to it and holding it in place. Often it is per se not obvious which procedure is the better option, and the decision may change during surgical procedures. In the case of impaired bone quality, cementation is preferred. Some suppliers offer prostheses

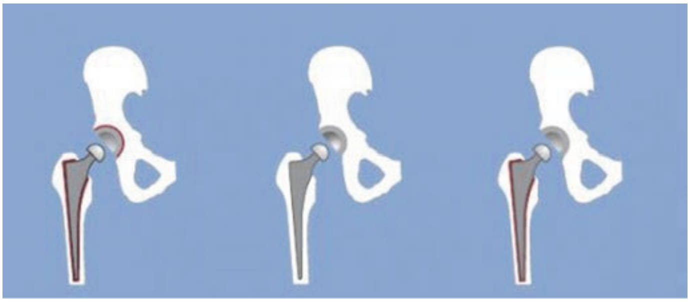


Fig. 15.7: Cemented, uncemented, and partial (hybrid) cemented implants.

with intraoperative flexibility, changing from cementless to cemented implantation without having to change instrumentation. Cementless bone-implant interfaces have the chance to naturally grow a cement line and a bone precursor substance, ensuring bone-like mineralization at the interface [10].

(b) Hip stem

The hip stem has the task to take up the load from the socket via the ball and to distribute the stress homogeneously over the femur. Computer simulations using finite element analysis help to optimize the shape and to predict the maximum stress in the shaft [11, 12, 13]. The implant can settle further after implantation and after carrying static and dynamic load, which may change the stress distribution as well as the final length of the leg. The goal is to achieve a stress distribution in the femur which is as natural and as close as possible to the one before implantation. Therefore, an optimal implant should provide a bone stress pattern that closely replicates the preoperative physiological stress state.

(c) Shafts

Shafts are distinguished by different neck inclination angles (CCD angle, usually chosen between 125° and 135°), by different lengths (short or long shafts), by different shapes and surface coatings, and by different materials (Fig. 15.8). In the past, long shafts were used. Presently there is a tendency to insert shorter shafts. Shorter shafts allow replacement when the original one has worn out after some 15–20 years. Short shafts are therefore preferred for younger patients. An x-ray radiograph comparing short and long shaft implants in the same patient is shown in Fig. 15.9, reproduced from [14]. Shafts are usually made of Co-Cr-Mo alloys, pure Ti, or Ti6Al4V alloys. All these metals and alloys fulfill requirements concerning corrosion resistance, durability, and biocompatibility. The tensile strength exceeds 1000 MPa, which is twice the maximum load ever expected in the femur. The alloy Ti6Al7Nb is mainly used for forging shafts that are used without cementation, as the surface is amenable for regrowth of bone tissue directly onto it, enhancing long-term stability and durability. To further promote bone ingrowth, the blades are often covered by a plasma-spray coating of hydroxyapatite, featuring a unique microporosity. This bioactive surfacing agent will ultimately provide strong bonding as the bone grows into it.

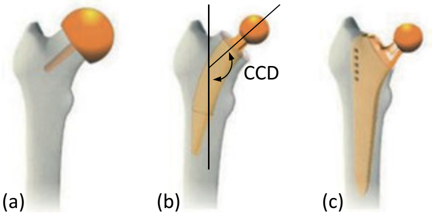


Fig. 15.8: Different shapes of hip stems (a) ball replacement only, (b) short shaft, (c) long shaft.

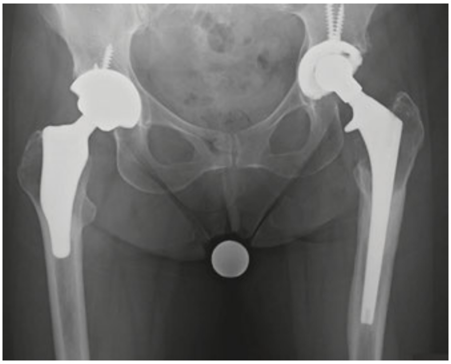


Fig. 15.9: X-ray radiograph of a short and long stem hip implant (reproduced from [14] by permission of SLACK, Inc.).

(d) Material combinations

For the combination of ball and socket (cup) five alternatives are available:

(a) metal on plastic (polyethylene or UHMWPE)
(b) metal on ceramic
(c) metal on metal (MoM)
(d) ceramic on plastic (UHMWPE)
(e) ceramic on ceramic (CoC)

Here UHMWPE stands for ultrahigh molecular weight polyethylene. Metal-on-plastic has the longest tradition and is the most frequently used and least expensive bearing. The plastic consists of highly stable and reliable polymer material with greatly reduced signs of wear. Nevertheless, debris from wear eventually diffuses into the body, and as the body attacks it, osteolysis may develop, which necessitates a replacement of the implant. Therefore, lower wear rate is essential and requires smooth surfaces on either side of the bearing. The metal ball is usually Co-Cr-Mo alloy with a highly polished mirror finish that is traditionally used for articulating bearing surfaces in total hip joint replacements. However, a highly polished mirror finish with a roughness of less than 20 nm does not necessarily guarantee a low wear rate. Surface finish with a roughness of more than 100 nm shows better wettability of lubricants resulting in lower wear rate.

Metal-on-metal bearings were developed for lower wear rate and for increased mobility because of larger size cups that can be fabricated from metals. MoM bearings were therefore targeted to younger and active patients. However, the promise did not hold and metal particles from wear distributed throughout the body causing severe problems. Patients with MoM implants were called back to clinics for survey and ultimately replacement. By now MoM bearings have been taken off the market.

Ceramic on plastic is a good combination of two very reliable materials. Ceramic heads are harder than metal heads and are the most scratch-resistant implant material. The hard, ultra-smooth surface can greatly reduce the wear rate on the polyethylene bearing. The potential wear rate for this type of implant is less than for metal on plastic.

Ceramic on ceramic is a combination of two very hard materials. In the past there have been problems with catastrophic shattering and squeaking noises. Both problems have been overcome. CoC is the hardest implant material that can be used in the body, and has the lowest wear rate of all, to almost immeasurable amounts (1000 times less than metal-on-polyethylene or about 0.0001 millimeters per year). Consequently, inflammation, bone loss, or systemic distribution of wear debris in the body is not an issue. New ceramic materials offer improved strength and more versatile sizing options. But it is also the most expensive option.

(e) Length adjustment

The standard hip prosthesis has only a rough length adjustment capability, since ball and shaft are joined by a conical press fit. Any residual length difference between left and right foot is compensated by orthopedic shoe inserts. Length adjustments can be improved by measuring the length of the leg before and after TEP. For this purpose two LEDs are attached to the leg and by lifting the leg, the center of rotation can be determined, as shown in Fig. 15.10. This procedure is repeated after TEP. If the circles indicate a shifted center of rotation due to a length change, the distance between ball and shaft can be individually adjusted by a screw connection replacing the traditional press fit [15].

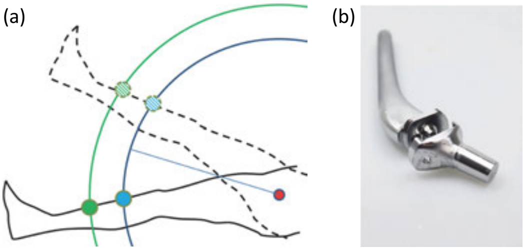


Fig. 15.10: (a) Length measurement of the leg with two LEDs attached to the leg describing a circle by lifting the leg. Measurements are taken before and after hip replacement. If there is a noticeable difference, the length can be changed with an adjustable shaft shown in (b) before closing the incision in the leg (adapted from https://www.fraunhofer.de/en.html).

(f) Load sensor

On a trial basis, endoprostheses were furnished with load sensors in the neck of the implant (Fig. 15.11) for measuring the load during different activities, walking, running, jumping, etc. [16]. The sensors measure 3 force components and 3 moment components using 6 semiconductor strain gauges. The recorded data are sent by high frequency transmission to a computer. The analyzed data may help improve the design of implants. Based on these in vivo sensors it could be confirmed that during walking

Fig. 15.11: Hip shaft with integrated strain sensors in the neck of the implant for measuring forces and torques during various activities (reproduced from [16] by permission of Elsevier Publishing Inc.).

the combined forces acting on the hip joint are about 2.5 × BW, during slow jogging the forces increase to about 5 × BW, standing on one leg causes a load nearly the same as for slow walking, and when climbing a staircase the forces in the joint are similar to those during walking, but the torque acting on the implants increases.

(g) Alternative approaches

Over the years there have been continuous improvements to the materials used for artificial hip replacement, to the instruments used for the surgical procedures, and the surgical procedures themselves have been optimized by computer simulations and monitoring. All these improvements follow the basic idea that the hip joint is to be replaced by hard, solid, durable and biocompatible parts. This is indeed the only solution for femoral fractures. However, if the bone is still in good condition but the cartilage is worn down, then alternative approaches should be considered. Researchers have now reported first success in regrowing cartilage from the patient's own stem cells. Once grown, the new cartilage has to be inserted into the joint. In the long run this bioengineered cartilage may make most solid hip replacements for curing osteoarthritis obsolete [17, 18].

15.3.2 Knee replacement

Knee replacement is a procedure similar to hip replacement. The reasons for knee replacement are the same, i.e., predominantly osteoarthritis, but the replacement rate is twice as high. Also for knee replacements we distinguish between partial and total knee prosthesis. In spite of all the similarity there is one important difference. Knee joints are individual parts of the body, slightly different for each person. Standardized parts such as in hip replacement do not work. Therefore, knee replacements require careful planning with the support of MRI and CT imaging, which is fed into a simulation program to find the optimal shape for a knee joint that is then individually fabricated.

The knee is the largest joint in the body (Fig. 15.12). It is made up of the lower end of the thighbone (femur), the upper end of the shinbone (tibia), and the kneecap (patella). The points of contact of these three bones where they touch are covered with articular cartilage.

The knee joint is a so called modified hinge joint (see Section 2.3.4/Vol. 1) that allows a monoaxial flexion-extension and a limited rotation in the flexed position. Flexion occurs when the tibia swings into a bent position and moves toward the back of the thigh. Normal knee flexion is about 135° and is only limited physically by touching of the calf and leg muscles in the back. Transfemoral prostheses (see Section 15.2.1 (b)) are strictly monoaxial without much loss of mobility.

The menisci of the knee are located between the femur and tibia. These C-shaped menisci wedges act as 'shock absorbers' that cushion the joint. Large ligaments hold the femur and tibia together and provide stability (Fig. 15.12). In particular these are the lateral collateral ligaments (LCL) and the anterior and posterior cruciate ligaments (ACL and PCL). ACL and PCL prevent the tibia and femur from sliding forward, LCLs control sideways movement. All remaining surfaces of the knee are covered by a thin lining, the synovial membrane. Similar to the hip joint, this membrane releases a fluid that lubricates the cartilage reducing friction to nearly zero in a normal knee.

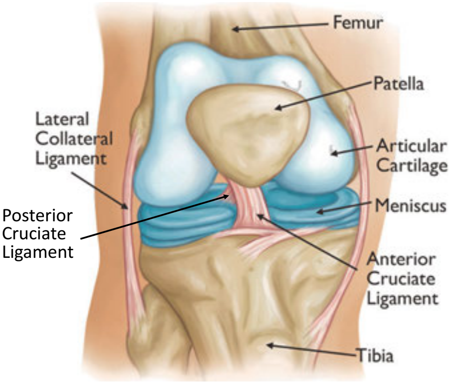


Fig. 15.12: Anatomy of the normal knee joining the lower part of the femur with the upper end of the tibia and the patella in front. The meniscus serves as shock absorber and the ends of the three bones, where they touch each other, are covered with articular cartilage that allows their relative movement. Strong ligaments, in particular the cruciate ligament, keep the knee joint together (reproduced by permission of OrthoInfo. © American Academy of Orthopaedic Surgeons, http://orthoinfo.aaos.org).

Knee replacement – also called knee arthroplasty – is more accurately described as knee “resurfacing” as only the surface of the bones are actually refinished rather than replaced. The different steps of knee replacement surgery can be seen in a video, the link is provided under useful webpages.

During surgery the ACL and the PCL are removed. They are no longer needed because the design of the artificial implant provides the stability of the knee that these ligaments guaranteed before. The LCL ligaments remain in the knee, providing further stability to the artificial knee replacement. Further details depend on whether a full or partial replacement is required; both versions are shown schematically in Fig. 15.13. In a partial replacement only one side (one compartment) of the knee is replaced, preserving all ligaments. The material combination for knee implants is always metal on plastic. For the metal implant either Co-Cr-Mo or Ti6Al4V alloys are chosen, the plastic spacer is made of UHMWPE.

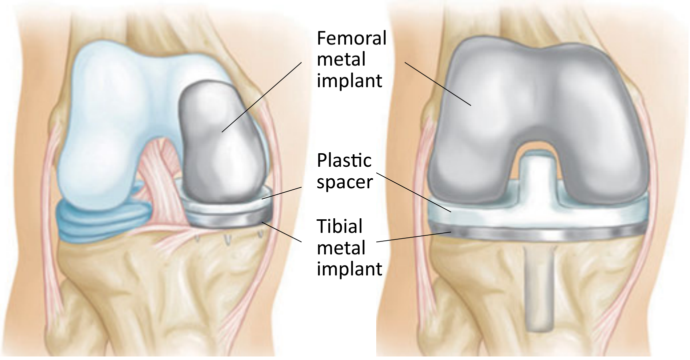


Fig. 15.13: Partial (left) and total (right) replacement of the knee. Between femoral and tibial components is a plastic spacer. The plastic spacer takes over the function of the meniscus (reproduced by permission of OrthoInfo, © American Academy of Orthopaedic Surgeons, http://orthoinfo.aaos.org).

15.4 Endoprostheses of the cardiovascular system

Endoprostheses of the cardiovascular system comprise artificial heart valves, pacemakers and stents. The so called *Berlin heart* is a prosthesis partly inside and partly outside of the body. It is a temporary prosthesis mainly for children with severe heart failure, characterized as a *ventricular assist device* for a few weeks to months until a donor heart for heart transplantation becomes available. The Berlin heart has similar functionalities to a heart lung machine (see Section 9.9/Vol. 1), but it is mobile and transportable by the patient. For more information we refer to [19]. In the following we confine the discussion to stents; pacemakers are discussed in Section 7.6.3/ Vol. 1.

Cardiovascular stents are used as vessel scaffolding to alleviate an arterial narrowing, called stenosis, and to restore blood flow to distal vasculatures. The procedure of placing a stent in an artery is termed *percutaneous coronary intervention* (PCI), also known as *coronary angioplasty*. PCI is a minimal invasive alternative to open heart surgery for setting a bypass. The origin of the name "stent" is not clear. On the one hand, stent is an old English word for pillar or brace; on the other hand, the name may originate from a British dentist Charles Stent (1807–1885), who tested materials for dental casts. Stents are small hollow and meshed tubes made from metals or synthetic fibers that can be slid into constricted arteries for restoring the original blood flow. Stents were primarily used for treating stenosis within the coronary arteries to prevent myocardial infarction. Nowadays stents are also being used for treating peripheral arteries, such as the carotid, the renal and the femoral arteries. Furthermore, stents are inserted to widen the urinary tract, the biliary tract, lacrimal duct, and for increasing the drainage of aqueous humor to lower the intraocular pressure (see Fig. 11.18/Vol. 1).

Figure 15.14 (a–h) shows the main steps of a PCI procedure. After a heart examination that indicates reduced blood flow and insufficient oxygenation of the myocardia,

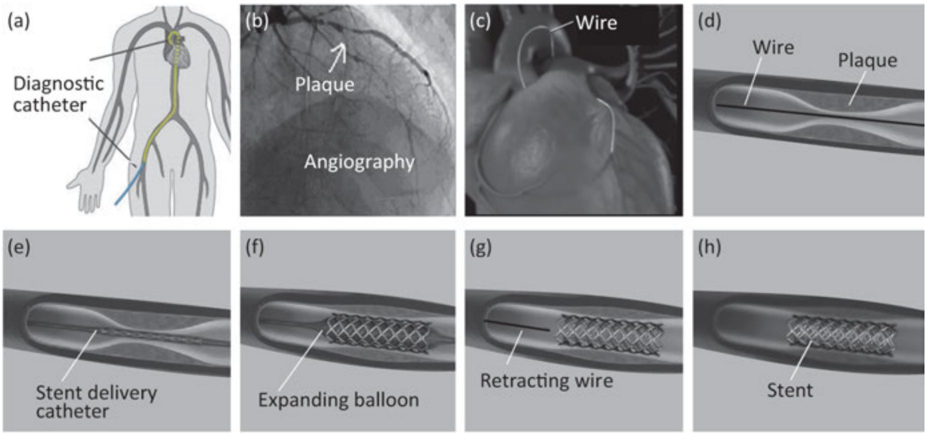


Fig. 15.14: Sequence of procedures for implanting a stent into a cardiovascular artery. (a) A catheter is inserted through the groin and pushed up to the heart. (b) After the catheter has reached the heart, an iodine-containing dye is spread through the catheter into the vasculature to identify obstructions by angiography; (c) once a plaque has been identified, a thin guiding wire is fed through the catheter; (d) the guiding wire is fed onwards through the plaque. (e) While keeping the guiding wire in place, the original catheter is pulled back out and replaced by another catheter that is equipped with a balloon and stent on its proximal end; (f) once the stent is positioned at the location of the plaque, the balloon is inflated through the sheathed catheter. The stent expands, presses against the vessel wall and opens the plaque; (g) the stent is now expanded and plastically stabilized. The balloon is deflated again; the catheter and guiding wire are retracted. (h) The expanded stent remains at the deployed position and keeps the plaque open.

the area of obstruction can be localized by catherization. For this procedure, a catheter (plastic tube) is inserted into the groin or wrist of the patient and threaded through to the heart (a). Once the catheter has reached the heart, an iodine-containing dye is injected into the coronary arteries and with the help of radiographs the obstruction can be visualized (b). Only obstructions of 70 % and more are considered for stenting. If such a plaque has been identified, a metal guiding wire is inserted into the catheter (c) and threaded through the plaque (d). Then the plastic catheter is pulled out again and replaced by another stent delivery catheter that is wrapped at the proximal end with an inflatable balloon embraced by a stent (e). Once the stent is positioned at the plaque site with the help of the guiding wire, the balloon is inflated (f). The stent will expand plastically and keep the obstructed artery open. Then the balloon is deflated again and the catheter together with the guiding wire is pulled back out (g), while the stent remains expanded in place (h). The entire process is monitored by x-ray radiography. The stent delivery catheter has a radio-opaque tip for locating its position, while the guiding metal wire, although very thin, still has an intrinsic x-ray opacity. The stent delivery and especially the balloon expansion phase should be kept as short as possible, as this phase completely blocks blood circulation. On the other hand, insufficient pressure to the balloon leads to premature stent collapse. Therefore a compromise has to be found.

While PCI appears straightforward, there are several issues that warrant closer inspection. These concern the material choice, mesh design of the scaffold, drug eluting stents, biocompatibility, plastic expansion behavior of the stent including elastic recoil, and change of local hemodynamics. There are many more physiological aspects, such as re-stenosis, stent thrombosis, etc. which will not be considered here.

Four main versions of stents are commonly used for PCI: (1) balloon-expandable stents; (2) self-expanding stents; (3) shape memory alloy stents; (4) biodegradable polymer stents.

Balloon-expandable stents deform plastically beyond the elastic limit of the material and are relatively rigid at their expanded diameter. Elastic recoil occurs infrequently and is hindered by the shape of the mesh and large angle deformation of the struts during expansion. As the stent expands, it also shortens, which has to be taken into account when considering the required length for the plaque in question. Bare metal stents (BMS) are usually made of CoCr or PtCr alloys, which are considered biocompatible. The mesh is fabricated either by cutting a thin hollow tube with a laser beam, or by weaving metal wires. There are many mesh patterns offered by various manufacturers and it is not clear which pattern is the best. One example is shown in Fig. 15.15. The thinner the struts and the fewer there are, the less the hemodynamics of the blood flow is affected. On the other hand, thinning the struts has a limitation in plastic stability of the stents and minimal elastic recoil. Furthermore, x-ray opacity becomes a problem with decreasing thickness. The thinnest struts on the market are 74 µm thick and made of PtCr alloy. The hemodynamics is potentially affected by stents in a twofold way: first, erythrocytes in contact with foreign materials such as

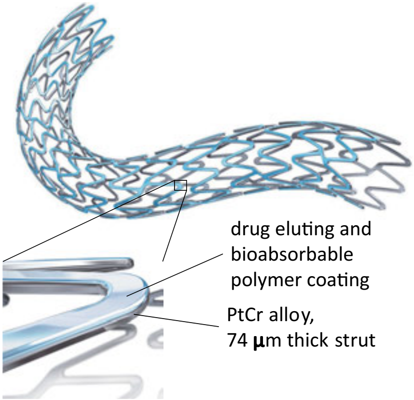


Fig. 15.15: Stent made of a 74 µm thick PtCr alloy strut providing high flexibility and coated with a 4 µm thick bio-absorbable polymer for drug delivery on the luminal side. This "high-tech" stent has recently been released for clinical use (image provided courtesy of Boston Scientific, © 2017 Boston Scientific Corporation or its affiliates, all rights reserved).

metals tend to cause clotting. And second, in the area of the stent the blood vessel is not compliant to pulsatile pressure changes, which may cause turbulent flow. An up-to-date review on BMS stents can be found in [20].

Self-expanding stents store potential energy in a reduced spring-like fashion that opens up when released. Self-expanding stents tend to be more compliant to varying blood flow and blood pressure than balloon-expandable stents. Moreover, they are retrievable and repositionable, in contrast to balloon-expandable stents, but otherwise they are similar in material choice and mesh design.

Shape memory stents are made from NiTi alloys. Their physical principle is based on a martensitic or structural phase transformation that changes the macroscopic shape of the metal without any internal atomic diffusion (Fig. 15.16). The alloy can be deformed to arbitrary shape at low temperatures. However, when heated above the martensitic phase transition temperature, it always assumes the original shape. This property is used in nitinol stents. The action is similar to elastically preloaded and self-expanding stents. But the release mechanism is different. The transformation temperature is set to about 30 °C. When delivered to the targeted area, premature expansion during delivery is constrained by a retractable sheath. Once released, the nitinol stent then conforms to the vessel wall at body temperature.

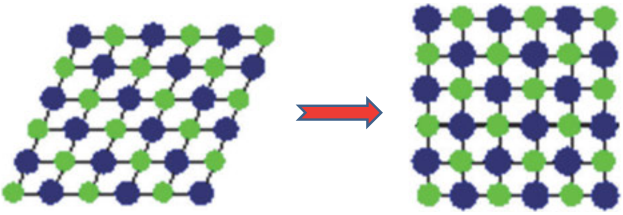


Fig. 15.16: Shape memory NiTi alloy. Left panel: deformed low temperature crystal structure. Right panel: structural transformation to the intrinsic crystal structure after heating the alloy above the martensitic phase transition temperature.

Often BMS are covered with a biodegradable polymer coating containing drugs that prevent cell proliferation, called drug-eluting stents (DES). The antiproliferative drugs are released slowly over time to help prevent tissue regrowth that can reclog the artery.

In addition to the aforementioned three types of metal stents, occasionally biodegradable or bioresorbable polymeric stents (BPS) are used. Bioresorbable polymeric stents are being developed to improve the biocompatibility and the drug reservoir capacity of metal stents. Moreover they offer a transient alternative to the permanent metallic stent implants. In most cases polyhydroxybutyrate (PHB) is used for this purpose. After about three months the obstructed artery has stabilized itself such that the supporting scaffold of the stent is no longer needed, but BMS remain permanently in the body, whereas BPS dissolve themselves. In this respect BPS has an edge over BMS [21].

15.5 Implants of the sensory system

Prostheses for the sensory systems are limited to visual and auditory perception. For the other senses (taste, touch, smell, pressure, etc.) we do not have artificial surrogates. Some peripheral sensors for the prosthetic hand were already mentioned in Section 15.2.2. Artificial lenses and light sensitive devices on the retina are the main implants for visual perception. Hearing aids in the outer ear, middle ear, or in the cochlea aim at restoring auditory sensitivity.

15.5.1 Middle ear implants

We distinguish between implants in the outer ear, middle ear, and inner ear. Concerning implantable hearing aids in the middle ear we further distinguish between bone conduction hearing aids and active middle ear implants. An overview is presented in Fig. 15.17.

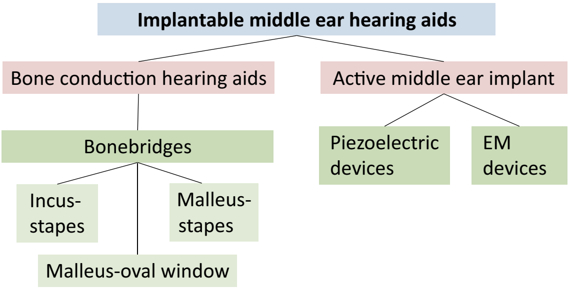


Fig. 15.17: Flow chart for implantable middle ear hearing aids.

The eardrum is the interface between the outer ear and the middle ear. Myringoplasty refers to a reconstruction and augmentation of the eardrum via fascia or perichondrium that improves sound transmission to the middle ear, and from the middle ear to the cochlea. The improved eardrum is indicated by brownish color in Fig. 15.18 (a). If the middle ear is impaired, there are several solutions depending on the severity of the hearing loss and the cause of the loss. If the connection between incus and stapes is malfunctioning, special kinked middle ear prostheses as shown in panel (b) may already alleviate the problem. If the incus is no longer usable, a direct connection between malleus and stapes can transmit the sound, as indicated in panel (c). Even in cases that the stapes is inoperable, a stiff connection between malleus and oval window can reestablish sound sensitivity (d). In all cases the implants are metal parts. 3D printing capabilities will likely revolutionize bone bridge implants in the future, as these will replicate the ossicles more naturally than any metallic implant can do. The main problem is the limited access to the tiny and delicate ossicles. Minimal invasive procedures through the Eustachian tube with the support of endoscopic methods may be a solution. But the small image size provided by the endoscope, difficult orientation and poor illumination have so far hindered further progress. This may change with the availability of more powerful endoscopes, as described in Chapter 14/Vol. 1. At present endoscopy and middle ear surgery is performed either by opening the eardrum or by opening a new channel through the temporal bone behind the ear.

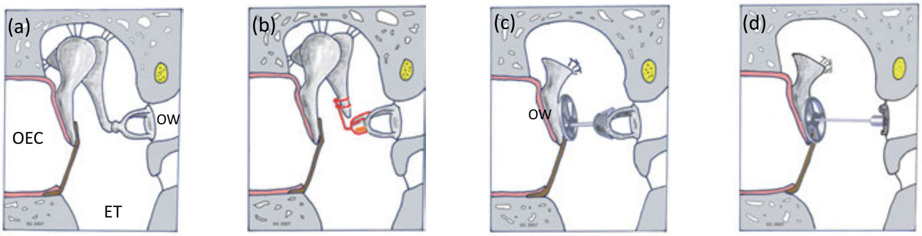
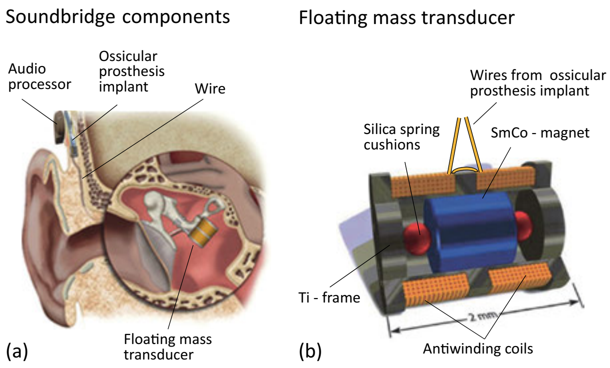


Fig. 15.18: Prostheses of the middle ear. OEC = outer ear canal, ET = Eustachian tube, OW = oval window (adapted from www.hno-arzt-augsburg.de/).

Alternatively, hearing loss in the middle ear may be revitalized via active *middle ear implants* (MEI). There are two main approaches: piezoelectric and electromagnetic sound amplifiers. In contrast to hearing aids which merely amplify the sound pressure in the outer canal, MEIs convert sound from the environment into mechanical vibrations that stimulate bones and structures within the middle ear. One version that is most frequently implanted in case of hearing impairment is the so called vibrating sound-bridge MEI (Fig. 15.19). It consists of outer and inner parts, shown in Fig. 15.19 (a). An audioreceiver and processor are mounted outside on the skin behind the ear. A pick-up coil (ossicular prosthesis implant) is implanted under the skin.


Soundbridge components

Floating mass transducer

Fig. 15.19: Vibrating sound-bridge middle ear implant consists of an outer part and an inner part. Left panel: overview of the system composed of an audio processor and ossicular prosthesis implant behind the pinna. A wire connects the pick-up coil in the implant with the floating mass transducer attached to one of the ossicles in the middle ear. Right panel: cross section of the floating mass transducer. A small SmCo magnet at the center moves back and forth when activated by the coils. Silica balls on either end of the Ti frame cushion the movement (adapted from [22] by permission of Taylor & Francis).

A small magnet keeps outside and inside parts aligned. The signal from the pick-up coil is brought into the middle ear via a wire and activates a transducer attached to one of the ossicles or directly to the oval window. The transducer consists of a floating mass, which – in turn – is a tiny but strong SmCo magnet (b). The magnet sits inside of two antiwinding coils. When activated, the magnet bounces back and forth in the rhythm of the audio signal. The lateral magnet movement is spring-like cushioned on either side by silica balls. The powering battery is mounted outside in the audioreceiver and can easily be replaced without surgical procedures. The vibrating sound-bridge transmits a sound that is perceived as “natural” because of the wide frequency range of up to 8 kHz. Lower frequencies can still be transmitted through the ear canal. Furthermore, the vibrating sound-bridge can be combined with cochlear implants if needed for an increased frequency range.

The successful application of the floating mass transducer depends on an intact and well-ventilated middle ear space and an intact vibrating ossicular chain. If the ossicular chain is interrupted, the floating mass transducer can also be placed directly onto the oval window with good results [22]. The sound perception achieved with the vibrating sound-bridge is more natural than with hearing aids inserted in the outer ear canal. In particular, acoustic feedback noise that is often encountered with in-ear canal hearing aids is not an issue for vibrating sound-bridge MEIs. Because of the magnet in the floating mass transducer, MRI diagnostics might cause adverse effects, such as local heating and demagnetization of the magnet, displacement of the transducer or interruption of the ossicular chain. None of these effects could, however, be detected and therefore the vibrating sound-bridge is permitted for MRI up to 1.5 T.

Piezoelectric MEIs use the piezoelectric effect for converting (or transducing) vibrational energy from soundwaves into an electrical signal. The piezoelectric effect is explained in Section 13.4/Vol. 1. The electrical signal is then used for activating a second transducer in the middle ear that converts the electrical signal back into mechanical vibrations in a fashion similar to the floating mass transducer.

All implants discussed so far aim at restoring a partial loss of hearing ability, also known as conductive deafness. It requires that the cochlea is still operational, at least partially, and that auditory signals can be send via the cochlear nerve bundle to the auditory cortex in the brain for deciphering sound signals. Complete deafness or nerve deafness resulting from an infection, trauma, etc. is presently not curable. With the help of cochlear implants only a very rudimentary sound sensation can be gained. Research groups are presently trying to restore lost cochlear hair cells by genetic means.

15.5.2 Retina implants

In contrast to hearing aids and implants, which improve hearing loss rather than refurbish hearing ability to deaf individuals, retinal implants are aimed at restoring some eyesight that enhances blind people's visual recognition and unguided mobility, which significantly increases their quality of life. Retinal implants require the full or partial existence of the retina, and in particular of the ganglion cells and nerve fibers. Other visual prosthesis that target the optic nerve system directly, the dorsal lateral geniculate nucleus (LGN), or the visual cortex, are not considered here. These implants are in a developmental stage; their performance is presently rather poor and may never succeed because of intrinsic problems. Retinal implants are orders of magnitude more complex than hearing aids and implants, and consequentially their present status is promising but not yet a breakthrough. Recent overviews on the state-of-the-art can be found in [23, 24, 25].

Retinal implants can be classified according to their location as epiretinal or subretinal implants (Fig. 15.20). Epiretinal implants are located on top of the retina and interface directly with the ganglion cell layer beneath. Subretinal implants are positioned below the ganglion and bipolar cell layers in place of the photoreceptors (see Fig. 11.19/Vol. 1). In both cases the implants are pixelated solid-state device arrays that stimulate either ganglion cells or bipolar cells by changing electrical potentials in response to a light stimulus. Any single diode or electrode within an array is bigger than a typical cell size and therefore stimulation proceeds via extracellular potential changes.

The two distinct types of implants are briefly described as follows.

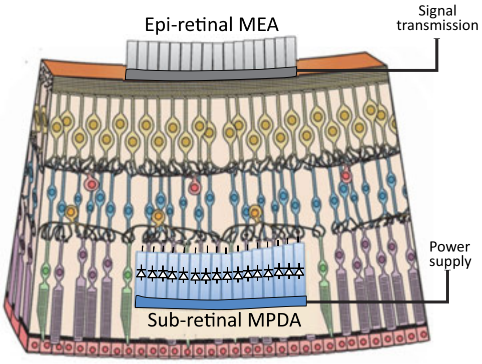


Fig. 15.20: Cross section through the retina showing the positions of retinal implants. Epiretinal implants are placed on top of the retina. Subretinal implants are at the location of the receptors. MEA = microelectrode array, MPDA = Micro-photodiode array.

(a) Epiretinal microelectrode array (MEA)

In case of epiretinal implants the scattered light from the outside world that carries the information on surrounding objects is detected by a digital camera mounted into a pair of glasses. Hence the natural lens, which may still function, is bypassed by the camera lens. The optical pathway of the image and the parts required are shown in Fig. 15.21. The camera video images are processed and pixel-wise transferred wirelessly to a receiver, feeding an electronic case implanted on the outside of the eyeball (bulbus). An epiretinal MEA is tacked onto the retina. The electronic case on the bulbus and the epi-MEA are connected by wires. The epi-MEA stimulates ganglion cells according to the processed images from the video processor. If the ganglion cells respond to electrode potential changes, they will fire an action potential that can be received in the visual cortex. With this method some eyesight can be restored, albeit over a rather small visual field of about 20° combined with low spatial resolution and low dynamic range. Power supply for camera and wireless transmission is carried by the patient in a bag. To focus on an object, the person carrying the visual system has to turn the camera by moving the head instead of turning the eyes. Although the optic pathway

Fig. 15.21: Epiretinal implant of the Argus II type uses a camera incorporated in a pair of glasses. After processing the video images (VPU), they are sent wirelessly to a receiver implanted on the bulbus. From there the signal is transferred by wire to a micro-electrode array, tacked to the retina (reproduced from [27], © Creative Commons).

appears cumbersome, it is presently the only successful implant on the market with about 150 implants worldwide. This general concept is followed and developed by three research-industrial consortia; the version Argus II of Second Sight Medical Products Inc. is shown in Fig. 15.21. It contains an MEA of 10 × 6 dots. Patients were screened with an assessment instrument that requires the fulfillment of a number of tasks. The conclusion from 26 participating patients was a “significantly improved completion of vision-related tasks with the device ON versus OFF” [26], i.e., with a binary on-off dynamic range. Similarly the Argus II retinal prosthesis system was used for studying whether patients can read a dot array normally used for tactile braille reading [27]. Instead of reading a 3 × 2 dot array through the camera, the dots directly stimulate a 3 × 2 subset of the 10 × 6 MEA for visual perception of individual braille letters. The high percentage of correctly read letters (more than 80 %) suggests that text can be stimulated and read as visual braille in retinal prosthesis by blind patients.

(b) Subretinal micro-photodiode array (MPDA)

In contrast to epiretinal MEA, the subretinal implant uses the light that enters through the natural lens. The light is detected by a photodiode array on a chip that is implanted in the subretinal space. Each photodiode has attached to it a tiny electrode for stimulating ganglion cells proportional to the received signal amplitude. An intraocular wire connects power supply with the MPDA. The advantage of this approach is a natural use of eye movement eliminating an elaborate image processor. The signals from the photodiodes are directly transferred to the inner nuclear layer presuming that the signal processing capabilities of the retinal network can at least partially be used for creating a signal output resembling the normal physiological one. The images observed by patients are still too crude to judge whether signal processing takes place in the retina. The disadvantage of the subretinal integrated circuitry is seen in the need of a power supply for the photodiodes from the outside that generates heat. In one design, however, the power is solely delivered by light entering through the lens. Another drawback of MPDA is the sensitivity of the optic chip, which even in the best version is much reduced in comparison to the sensitivity of a normal, healthy eye. Hence a powerful amplifier is mandatory for each photodiode and higher light intensity is needed to achieve limited eyesight. To overcome these problems, it was suggested to develop a hybrid system: use of a camera for primary image recording; then conversion of the visible image to the infrared regime and emission at much higher intensity. The infrared radiation goes through the lens of the eye and is then detected by a photodiode array in the subretinal space. Care must be taken that the high infrared intensity does not heat up the retina causing damage. Like in the previous case, the subretinal MPDA implant only works as long as ganglion cells and the optic nerve fibers are still intact. This may not be the case in the late stage of glaucoma.

The realization of various design concepts may differ in pixel size, pixel density, and total number of pixels. Furthermore there are differences in materials used. Chips have been fabricated using Si technology, conjugated and light sensitive polymers, or carbon nanotubes [28]. Also, microarrays may be patterned on flexible polymer sheets that conform better to the concave inner shape of the bulbus. Conformity is an issue when increasing the lateral size of arrays for larger visual fields. In all cases, biocompatibility of all components in contact with tissue is mandatory, which may necessitate encapsulation of the chip into biocompatible envelops.

Whatever the design and choice of materials, retinal implants face a tremendous challenge with respect to resolution, contrast, sensitivity, dynamic range, visual field and visual acuity. Progress is slow and the performance of retinal implants may not improve dramatically in the near future. Therefore, different strategies have been suggested, such as confocal image recording for better foreground-background recognition or enhanced image processing to simplify complex and cluttered images by emphasizing essential objects in the foreground [29].

15.6 Summary

1. Exoprostheses are prostheses external to the body; endoprostheses are prostheses within the body, also called implants.
2. We distinguish between transtibial prostheses, i.e., prostheses below the knee, and transfemoral prostheses, those above the knee.
3. Microprocessor controlled transfemoral prosthesis are called C-legs.
4. Myoelectric prosthesis use remaining nerve fibers, which run from the brain to the arm. Myoelectrical control of upper limb prostheses is referred to as targeted muscle reinnervation (TMR).
5. There are three main reasons that necessitate hip replacement: (1) osteoarthritis leading to a breakdown of the cartilage that covers both sides of the hip joint; (2) fractures of the femoral neck due to osteoporosis ; (3) fractures due to various accidents.
6. Hip replacements require consideration of shaft, stem, ball, and socket. The material combination is important since it determines wear and lifetime.
7. Knee replacement is better described as knee “resurfacing” as only the surface of the bones is actually refinished rather than replaced.
8. Cardiovascular stents are used as vessel scaffolding to alleviate an arterial narrowing, called stenosis, and to restore blood flow to distal vasculatures.
9. The procedure of placing a stent in an artery is termed *percutaneous coronary intervention* (PCI).
10. Four main versions of stents are commonly used for PCI: (1) balloon-expandable stents; (2) self-expanding stents; (3) shape memory alloy stents; (4) biodegradable polymer stents.
11. Implantable middle ear hearing aids can either passively restore bone conduction, or actively bring the sound from the outside to the middle ear, for instance by a vibrating sound-bridge.
12. The essential part of the vibrating sound-bridge is a floating mass transducer that is activated by a receiver behind the pinna.
13. Retinal implants can be classified according to their location as epiretinal or subretinal implants.
14. Retinal implants provide limited visual perception to blind patients with functional retina.

References

[1] https://en.wikipedia.org/wiki/Berlin_Heart

[2] Wolf EJ, Everding VQ, Linberg AL, Schnall BL, Czerniecki JM, Gambel JM. Assessment of transfemoral amputees using C-Leg and Power Knee for ascending and descending inclines and steps. J Rehabil Res Dev. 2012; 49(6): 831–842.

[3] Seymour R1, Engbretson B, Kott K, Ordway N, Brooks G, Crannell J, Hickernell E, Wheeler K. Comparison between the C-leg microprocessor-controlled prosthetic knee and non-microprocessor control prosthetic knees: a preliminary study of energy expenditure, obstacle course performance, and quality of life survey. Prosthet Orthot Int. 2007 Mar; 31(1): 51–61.

[4] Atzori M, Müller H. Control capabilities of myoelectric robotic prostheses by hand amputees: a scientific research and market overview. Front Syst Neurosci. 2015 Nov 30; 9: 162.

[5] Belter JT, Segil JL, Dollar AM, Weir RF. Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review. J Rehabil Res Dev. 2013; 50(5): 599–618.

[6] Bouton CE, Shaikhouni A, Annetta NV, Bockbrader MA, Friedenberg DA, Nielson DM, Sharma G, Sederberg PB, Glenn BC, Mysiw WJ, Morgan AG, Deogaonkar M, Rezai AR. Restoring cortical control of functional movement in a human with quadriplegia. Nature. 2016 May 12; 533(7602): 247–250.

[7] Donati AR, Shokur S, Morya E, Campos DS, Moioli RC, Gitti CM, Augusto PB, Tripodi S, Pires CG, Pereira GA, Brasil FL, Gallo S, Lin AA, Takigami AK, Aratanha MA, Joshi S, Bleuler H, Cheng G, Rudolph A, Nicolelis MA. Long-term training with a brain-machine interface-based gait protocol induces partial neurological recovery in paraplegic patients. Sci Rep. 2016 Aug 11; 6: 30383.

[8] Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, Jiranek WA, Berry DJ. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg Am. 2015 Sep 2; 97(17): 1386–1397.

[9] Hewitt JD, Glisson RR, Guilak F, Vail TP. The mechanical properties of the human hip capsule ligaments. J Arthroplasty. 2002 Jan; 17(1): 82–89.

[10] Grandfield K. Bone, implants, and their interfaces, Physics Today. 2015; 68: 40–45.

[11] Srimongkol S, Rattanamongkonkul S, Pakapongpun A, Poltem D. Mathematical modeling for stress distribution in total hip arthroplasty. Int J of Math Models and Methods in Appl Sciences. 2012; 6: 885–892.

[12] Jasik A, Mroczek M. The influence of mechanical and material factors on the biological adaptation processes of the femoral bone implants. Arch Metall Mater. 2016; 61: 189–194.

[13] de Freitas Spinelli L, de Souza Macedo CA, Galia CR, Rosito R, Schnaid F, Corso LL, Iturrioz I. Femoral stem-bone interface analysis of logical uncemented stem. Brasilian J of Biomedical Engineering. 2012; 28: 238–247.

[14] Patel RM, Lo WM, Cayo MA, Dolan MM, Stulberg SD. Stable, dependable fixation of short-stem femoral implants at 5 years. Orthopedics. 2013 Mar; 36(3): e301–307.

[15] BMWi – ZIM Kooperationsnetzwerk 027, www.zim-bmwi.de/.

[16] Damm P, Graichen F, Rohlmann A, Bender A, Bergmann G. Total hip joint prosthesis for in vivo measurement of forces and moments. Med Eng Phys. 2010 Jan; 32(1): 95–100.

[17] Pak J, Lee JH, Kartolo WA, Lee SH. Cartilage regeneration in human with adipose tissue-derived stem cells: current status in clinical implications. Biomed Res Int. 2016; 2016: 4702674.

[18] Moutos FT, Glass KA, Compton SA, Ross AK, Gersbach CA, Guilak F, Estes BT. Anatomically shaped tissue-engineered cartilage with tunable and inducible anticytokine delivery for biological joint resurfacing. Proc Natl Acad Sci USA. 2016 Aug 2; 113(31): E4513–E4522.

[19] Stiller B, Lemmer J, Schubert S, Ewert P, Schulze-Neick I, Hübler M, Redlin M, Berger F, Hetzer R. Management of pediatric patients after implantation of the Berlin Heart EXCOR ventricular assist device. ASAIO J. 2006 Sep/Oct; 52(5): 497–500.

[20] Jorge C, Dubois C. Clinical utility of platinum chromium bare-metal stents in coronary heart disease. Med Devices (Auckl). 2015 Aug 27; 8: 359–367.
[21] Eberhart RC, Su SH, Nguyen KT, Zilberman M, Tang L, Nelson KD, Frenkel P. Bioresorbable polymeric stents: current status and future promise. Review. J Biomater Sci Polym Ed. 2003; 14(4): 299–312.
[22] Marino R, Linton N, Eikelboom RH, Statham E, Rajan GP. A comparative study of hearing aids and round window application of the vibrant sound bridge (VSB) for patients with mixed or conductive hearing loss. Int J Audiol. 2013 Apr; 52(4): 209–218.
[23] Ghezzi D. Retinal prostheses: progress toward the next generation implants. Front Neurosci. 2015 Aug 20; 9: 290.
[24] Goetz GA, Palanker DV. Electronic approaches to restoration of sight. Rep Prog Phys. 2016 Sep; 79(9): 096701.
[25] Stingl K, Zrenner E. Electronic approaches to restitute vision in patients with neurodegenerative diseases of the retina. Ophthalmic Res. 2013; 50(4): 215–220.
[26] Geruschat DR, Richards TP, Arditi A, da Cruz L, Dagnelie G, Dorn JD, Duncan JL, Ho AC, Olmos de Koo LC, Sahel JA, Stanga PE, Thumann G, Wang V, Greenberg RJ. An analysis of observer-rated functional vision in patients implanted with the Argus II retinal prosthesis system at three years. Clin Exp Optom. 2016 May; 99(3): 227–232.
[27] Lauritzen TZ, Harris J, Mohand-Said S, Sahel JA, Dorn JD, McClure K, Greenberg RJ. Reading visual braille with a retinal prosthesis. Front Neurosci. 2012 Nov 22; 6: 168.
[28] Bareket L, Waiskopf N, Rand D, Lubin G, David-Pur M, Ben-Dov J, Roy S, Eleftheriou C, Sernagor E, Cheshnovsky O, Banin U, Hanein Y. Semiconductor nanorod–carbon nanotube biomimetic films for wire-free photostimulation of blind retinas. Nano Lett. 2014 Nov 12; 14(11): 6685–6692.
[29] Jung JH, Aloni D, Yitzhaky Y, Peli E. Active confocal imaging for visual prostheses. Vision Res. 2015 Jun; 111(Pt B): 182–196.

Further reading

Edelstein J, Moroz A, editors. Lower-limb prosthetics and orthotics: clinical concepts. SLACK Inc.; 2011.
Santambrogiio L, editor. Biomaterials in the regenerative medicine and the immune system. Springer Verlag; 2015.

Useful website

Video showing schematically a full knee replacement: https://www.youtube.com/watch?v=kAMuP2MNqUo

16 Questions & answers

Questions

Chapter 1

1. What is the relation between x-ray energies and x-ray wavelengths?
2. What types of x-ray radiation are usually distinguished?
3. What instrument can produce x-ray radiation?
4. How is bremsstrahlung generated?
5. How is characteristic x-ray radiation generated?
6. How can x-ray radiation be distinguished from γ-radiation?
7. What are typical x-ray energies for radiography and radiotherapy?
8. What are the essential parts of a medical x-ray tube for radiography?
9. The intensity of x-ray bremsstrahlung is a linear function of which energy difference?
10. What determines the cutoff energy $E_{\max}$?
11. For radiography an Al filter is used. What is the purpose of the Al-filter?
12. What is characteristic x-ray radiation useful for?
13. In what medical discipline is characteristic radiation used as an analytic tool?
14. Which transition corresponds to the $K_{\alpha1}$-x-ray radiation?
15. The accelerating voltage of an x-ray tube may be 50 kV. With this voltage is it possible to excite tungsten K_α radiation?
16. Which energy is higher, the excitation energy or the corresponding energy of the x-ray characteristic radiation? As an example, compare W-K_α excitation (absorption edge) with characteristic W-K_α radiation?
17. What are the differences between the Mosely law and Bohr's model for the hydrogen atom?
18. What are the characteristic design features of an x-ray tube for radiography?
19. Are x-ray tubes for radiography actively cooled?
20. What are the advantages of the Straton rotating x-ray tube compared to tradional ones?
21. What are the principle design features of linear accelerators?
22. Synchrotron radiation is radiation that is emitted from orbiting electrons in a storage ring. When these electrons radiate x-ray photons at high intensity, do they stop orbiting?
23. What is synchrotron radiation mainly used for?

DOI 10.1515/9783110553116-018

Chapter 2

1. Nuclei are composed of which particles?
2. How many quarks are in a nucleon?
3. How is the atomic number A defined?
4. What are isotopes?
5. When are isotopes stable, and when are they unstable?
6. What is the proper ratio of protons to neutrons for stable isotopes?
7. Unstable isotopes are radioactive isotopes. Name the two types of radioactive isotopes.
8. What is the definition of atomic mass unit?
9. Why is the atomic weight of chemical elements not an integer number of the atomic mass unit?
10. What types of nuclear decays occur during radioactive decay of unstable isotopes?
11. Which of the decay products do not change the atomic number of the radioisotope?
12. What is the difference between lifetime of a radioisotope and half-life?
13. How is the activity defined?
14. What is the unit of the activity?
15. In case of decay chains how are the activities of mother and daughter radioisotopes related?
16. How can radioisotopes be produced artificially? Name at least three methods.
17. Which machines are required for the various activation processes?
18. How can the radioisotope ^{99m}Tc be produced, which is heavily used for scintigraphy?

Chapter 3

1. What are the two physical mechanisms that attenuate a particle beam interacting with a target?
2. What is the definition of fluence?
3. How is intensity related to fluence?
4. When photons interact with matter, what are the four different cross sections contributing to the attenuation of the intensity?
5. Coherent scattering has different names according to the size of the objects scattered from. Name a few of them.
6. What are the two steps that take place during photoelectric absorption?
7. What is the difference between generating characteristic x-ray radiation by electrons and generating them by photons?

8. What is the essential difference between the photoelectric effect and the Compton effect?
9. What is the difference between coherent x-ray scattering and Compton scattering?
10. Is pair production an important process?
11. Why do alpha particles suddenly stop?
12. What is the range of 5 MeV alpha particles in the normal atmosphere?
13. What is the Bragg peak for alpha particles?
14. Do photons have a range similar to alpha particles?
15. What is the range of beta-particles?
16. What are the main interactions of fast and slow neutrons with matter?
17. What is the best moderator for fast neutrons?
18. In which energy range do neutrons experience resonance absorption?

Chapter 4

1. What is dosimetry good for?
2. How is dose defined?
3. Quality factor Q_R and relative biological effectiveness RBE are identical factors characterizing what?
4. What is the difference between dose and equivalent dose, and what are their units?
5. What is the difference between the quality factor Q_R and weighting factor w?
6. What is energy fluence in comparison to fluence?
7. What does the acronym Kerma stand for, and how is Kerma defined?
8. Does the definition of Kerma hold for all types of radiation?
9. How can the mass energy transfer coefficient be distinguished from the mass attenuation coefficient?
10. What dosimeters are available for measuring the dose?
11. What is the difference between proportional counters and Geiger–Mueller detectors.
12. What is meant by “dead time”, and how does it affect the count rate?
13. What are the main sources of radiation to which people are exposed?
14. What is the tolerable dose without further control?
15. What is the ALARA principle?
16. What is the best protection measure against radiation?

Chapter 5

1. How is x-ray attenuation contrast defined?
2. What is an attenuation profile?
3. What are the two reasons for a blurry x-ray radiograph?
4. How can the contrast in a blurry radiograph be sharpened?
5. How can the x-ray spectrum be hardened, meaning that the average x-ray energy is shifted to higher energies?
6. For recording x-ray radiographs, what has completely replaced the exposure of x-ray films?
7. Does the digital recording of x-ray intensity have the same spatial resolution as x-ray films?
8. What is the main advantage of digital recording x-ray radiographs?
9. Why does mammography use different x-ray sources than in standard x-ray radiography?
10. Contrast enhancement modalities are frequently used in x-ray radiography. Which ones are available?
11. Does phase contrast imaging utilize the real or the imaginary part of the x-ray refractive index?
12. What is the main advantage of phase contrast imaging?
13. Give a brief motivation for CT imaging.
14. How is the Hounsfield scale defined?
15. In Hounsfield weighted images, what appears black and what appears white?
16. In today's CT-scanners a number of detectors work in parallel defining a slice in the z direction. How many detectors are mounted in parallel for scanning which slice thickness at once?
17. What is the voxel size of a standard CT scan?
18. What is the dose delivered during a standard CT scan?
19. When is x-ray CT imaging preferred over MRI imaging?

Chapter 6

1. In scintigraphy the source of the electromagnetic radiation is in the body. How do they get there?
2. What type of radioactive emitter is used for scintigraphy: alpha, beta, or gamma emitters?
3. Why are heavy radioisotopes required?
4. How are the radioisotopes produced for use in scintigraphy?
5. Why is the gamma emission of ^{137}Cs not used for scintigraphy?
6. What are the most frequent radioisotopes used for scintigraphy?
7. What is photon energy of the gamma emission of ^{99m}Tc?

8. Are the radiopharmaceuticals used for scintigraphy authentic or analog?
9. What is the basic working principle of a ^{99}Mo/^{99}Tc generator?
10. After what amount of time can the Tc–elute be extracted from the generator?
11. What detection system is used for recording scintigrams?
12. What does the acronym SPECT stand for?
13. How many detector arrays are used for SPECT scanning?
14. How does SPECT compare with x-ray CT?
15. What is SPECT used for?

Chapter 7

1. PET requires specific radioisotopes. What are the requirements?
2. How are PET radioisotopes produced?
3. Positron annihilation is the reverse process of which process?
4. Why do the emitted gamma photons fly apart at 180°?
5. What is the typical straggling distance from positron decay to positron annihilation?
6. How is the gamma radiation from positron annihilation detected?
7. How does coincidence detection work?
8. What are possible artefacts of coincidence counting?
9. What is PET mainly used for?
10. What is the difference between perfusion and diffusion?
11. How does the F-18 metabolism work?
12. Which method has higher spatial resolution, SPECT or PET?
13. What is the main disadvantage of PET?

Chapter 8

1. What are the two main phases of the cell cycle?
2. When does cell replication take place?
3. Which protein is known to be a tumor suppressor?
4. Tumors develop in six stages. Name those six stages.
5. At what stage does tumor development become noticeable by imaging modalities?
6. When does a tumor turn from a benign to a malign tumor?
7. How is the survival curve S defined as a function of the dose D?
8. When are cells most sensitive to radiation, and when are they least sensitive to radiation?
9. Why is radiation treatment fractionated?
10. What are the main effects of radiation to a cell?

11. How is the relative biological effectiveness (RBE) defined?
12. When is the relative biological effectiveness (RBE) highest?
13. Which types of radiation are high LET radiation?
14. Which radiation is low LET radiation?
15. What is OER?
16. For which type of radiation is OER important or not important?
17. What is the effect of oxygen for tumor control?
18. What is the effect of low-dose versus high-dose application?
19. What is the effect of chemotherapy?

Chapter 9

1. What are typical x-ray energies for XRT?
2. How are high energy x-ray photons generated?
3. A 10 MeV incident photon energy hits a water target. What is the percentage of photons undergoing Compton scattering as compared to pair production?
4. The dose-depth profiles of electrons and photons in soft matter are very different. Why?
5. Why is the Kerma-depth profile different from the dose-depth profile?
6. How far below the surface of the skin does the peak of the depth dose profile lie?
7. What is the physical reason for the dose build up below the surface?
8. What is the difference between SSD and SAD?
9. What are the advantages and disadvantages of SSD versus SAD?
10. What is the difference between GTV, CTV, and PTV?
11. What is image guided radiation therapy good for?
12. XRT does not have range; the depth dose profile is continuously dropping off. Why is it nevertheless possible to deposit a high dose at one specific target volume inside the body?
13. What is IMRT-MLC?
14. What are the essential ingredients of cyberknife technology?
15. What is gamma-knife used for?

Chapter 10

1. What is the main interaction of protons with matter?
2. When is the LET highest for protons, at the beginning or at the end of its track?
3. Do protons have a range?
4. Why does the range of protons increase almost by the power of four with energy?
5. What is the RBE for protons?
6. Is proton beam therapy considered to be low or high RBE treatment?

7. Is PBT high or low LET radiation?
8. How can a spread out Bragg be formed?
9. What is the main advantage of PBT?
10. The dose-depth profile for carbon ions exhibits a tail beyond the Bragg peak. What is the reason for the tail?
11. What are the 4 essential parts of a conventional proton radiation facility?

Chapter 11

1. What are the three main methods of neutron radiation therapy?
2. How are neutrons distinguished according to their energies?
3. Why are neutrons used for cancer treatment?
4. Which types of carcinoma are mainly treated by fast neutrons?
5. Which are the main accelerator based nuclear reactions for the production of neutrons?
6. What is the shape of the depth-dose profile for high energy neutrons?
7. What is the reason for dose build-up below the surface in case of high energy neutrons?
8. What is the effect of high LET radiation on the fractionation of dose delivered to the patient?
9. In BNCT the source of radiation is in the tumor cells. What is the role of neutrons and what type of radiation occurs?
10. What are the dose relevant particles in BNCT treatment?

Chapter 12

1. What kind of radiation therapy is brachytherapy?
2. What is the antonym of brachytherapy?
3. What are the main applications of brachytherapy?
4. How is brachytherapy applied?
5. What are the main advantages of brachytherapy compared to external beam therapies?
6. Which type of radiation is mainly active in brachytherapy?
7. Which are the main isotopes used for brachytherapy?
8. How are the isotopes produced?
9. How are the isotopes applied?
10. What are the two main procedures by brachytherapy treatment?

11. What is typical low dose rate (LDR) application, what is a high dose rate (HDR) application in terms of Gray per hour?
12. The dosimetry for Brachytherapy requires special consideration. What is different compared to standard dosimetry.

Chapter 13

1. What does the acronym LASER stand for?
2. What are the basic characteristics of laser light?
3. What are lasers mainly used for in medicine?
4. Lasers can be operated in continuous or in pulsed mode. What is the reason for laser pulsation?
5. What are the three methods for pulsation of a laser?
6. Which types of Q-switches do you know?
7. How is mode locking achieved?
8. What are the main lasers used in medicine?
9. What laser flux is required to cause plasma induced ablation and what is it used for?
10. How does photodynamic treatment proceed?
11. What is the difference between LASEK and LASIK?
12. Which types of laser treatments are common in ophthalmology?

Chapter 14

1. What are the three main tasks of nanoparticles in medicine?
2. When nanoparticles enter the body they are recognized by the immune system. What is it called, and how does it act on the nanoparticles?
3. The size of nanoparticles is an important parameter from a physiological and a physical point of view. Please explain both.
4. What types of targeting possibilities do nanoparticles have?
5. What is the Curie temperature?
6. What is a macrospin?
7. Why is the magnetic moment of Gd $8\mu_B$ and not $7\mu_B$, as one would expect from $m = gS\mu_B$ with $S = 7/2$ and a Landé-factor of 2 for spin only magnetic moment?
8. What is the blocking temperature of nanoparticles?
9. For contrast enhancement of MRI, MNPs should be designed such that T_B is above body temperature. In contrast, for hyperthermia the MNPs should have a T_B at or below body temperature. Why?
10. What are the three relaxation mechanisms for magnetic nanoparticles used for hyperthermia?

11. What is meant by positive contrast in MRI as compared to negative contrast?
12. Why do magnetite nanoparticles provide negative contrast?
13. What is the reason that Gd^{3+} chelates act as positive contrast agent?
14. What is K-edge contrast imaging?
15. Compare Rayleigh scattering and Mie scattering of electromagnetic waves. Which type of scattering prevails as function of wavelength?
16. What is the preferential direction of Mie scattering?
17. How does the optical absorbance of metal nanoparticles vary with size and aspect ratio?
18. Describe briefly the essential points of photothermal therapy (PTT).
19. How can the accumulation of Au nanoparticles in tumorous tissue be made visible?
20. What are theranostic nanoparticles?

Chapter 15

1. What is the discipline called which provides support of the musculoskeletal system but not a replacement?
2. What are the main considerations for the construction of a computerized leg (C-leg)?
3. What is the challenge of the myoelectric control of the upper limbs?
4. What is the best material combination in the case of total hip replacement?
5. Stents are used for the biliary tract, lacrimal duct, and for increasing the drainage of aqueous humor to lower the intraocular pressure.
   Where are these tracts and ducts located in the body?
6. A stent design should fulfill which physiological and mechanical requirements?
7. A sound bridge is an implant serving as a hearing aid. Where is it implanted, and where does it act on?
8. Retinal implants are distinguished according to their location.
   Which ones are offered?
9. Which group of patients can be helped with retinal implants?

Answers

Chapter 1

1. Energies are inversely related to wavelength: $E \sim 1/\lambda$.
2. Bremsstrahlung, characteristic radiation, and synchrotron radiation.
3. X-rays can either be produced by x-ray tubes, by linear accelerators, or by electron synchrotrons.
4. Bremsstrahlung occurs upon deacceleration of fast electrons in the Coulomb field of nuclei.
5. Characteristic x-ray radiation is generated by creating an electron hole in a core shell, which is filled by an electron from shells with less binding energy. The energy difference is emitted as x-ray radiation.
6. Both are highly energetic electromagnetic waves. Distinction is only possible by the generation: γ-radiation is a byproduct of radioactive decay stemming from nuclei. X-rays are artificially generated by accelerators and electronic excitations.
7. For radiography acceleration voltages up to 150 pKV are used, for radiotherapy the accelerating voltages are in the range of 2–20 MV.
8. A cathode for the production of electrons, a high voltage supply for accelerating electrons in a potential gradient, and a target (anode) for stopping the electrons and production of bremsstrahlung
9. The x-ray intensity is proportional to the difference $E_{\max}$ of the high voltage applied and the energy of the photons: $I \sim E_{\max} - E_{\text{photon}}$.
10. The cut-off energy is determined by the accelerating voltage V_{acc} : $E_{\max} = eV_{\text{acc}}$.
11. The purpose of the Al-filter is to recuce soft x-ray radiation by absorption that would otherwise increase the dose but would not contribute to the image quality of radiographs.
12. Characteristic radiation has sharp spectral lines, which are characteristic for the element that emits this radiation. Therefore by analyzing the energy of the characteristic radiation the element can be identified which emitted the radiation.
13. In forensic medicine.
14. The transition from the $2p_{1/2}$ (L_2) to 1s (K_1) energy level.
15. No. The W-K_α radiation requires an excitation energy of 69.5 keV.
16. The excitation energy is always higher than the corresponding x-ray characteristic radiation. For tungsten the excitation energy of the W-K_α radiation is 69.5 keV, the energy of the characteristic W-K_α radiation has an energy of 59.3 keV.
17. The Mosely law is similar to Bohr's model for the energy levels in hydrogen atoms modified by the charge of the nucleus and the hole in the K-shell after excitation.

18. The cathode is offset, and the electrons hit the rim of a rotating anode. Cathode and anode sit in an evacuated glass tube. The anode is driven by an outside stator that couples to a rotor inside the glass tube.
19. No, they are only cooled by radiation.
20. The Straton tube is continuously cooled, the focus can be controlled by a magnetic field bending the electron beam, and the x-ray radiation is automatically filtered by the Al housing.
21. Linear accelerators require an electron source (klystron), a pulse generator and modulator, and microwave cavities for accelerating the electrons.
22. No. The electrons loose only part of their kinetic energy by radiation. For instance an electron with 1 GeV kinetic energy may emit a 1 keV photon, which is only 1 ppm of the kinetic energy. Nevertheless, the lost energy is boosted up in the ring, so that the electrons stay in the orbit.
23. Synchrotron radiation is mainly used for the analysis of materials, including biomaterials, via scattering and spectroscopy.

Chapter 2

1. Nuclei are composed of the nucleons: protons and neutrons.
2. Three. Two up and one down in protons, two down and one up in neutrons.
3. A is the sum of protons and neutrons.
4. Isotopes are nuclei with the same number of protons but different number of neutrons.
5. Isotopes with the proper number of neutrons are stable. To many neutrons or to less neutrons make isotopes unstable.
6. For light atoms stable isotopes have about the same number of protons and neutrons. In heavy atoms there are more neutrons than protons.
7. Natural isotopes and artificial isotopes.
8. The atomic mass unit is defined as 1/12 of the mass of the isotope ^{12}C.
9. There are two reasons. One is the binding energy of nucleons in nuclei, which causes a mass deficiency. The other reason is the isotope mixture that is usually present in chemical elements. The atomic weight is then an average weighted by the natural abundance.
10. β^-, β^+-decay, EC, and γ-emission.
11. Lifetime is the time T after which $1/e$ or 63 % of radioisotopes have decayed. Half-life $T_{1/2}$ is the time after which half of the radioactive isotopes or 50 % have decayed.
12. Activity is the number of decays (events) per second.
13. Becquerel; 1 Bq is one event per second.

14. The activity of the daughter radioisotope depends on the production rate of the daughter nucleus via decay of the mother nuclei minus the decay rate of the daughter nuclei.
15. Proton bombardment, nuclear fission with slow neutrons, activation with slow neutrons.
16. Cyclotron for proton activation. Nuclear reactor for fission production and for thermal neutron activation.
17. The most efficient way is by fission of 235-uranium.

Chapter 3

1. Absorption and scattering.
2. Number of particles per area.
3. Intensity is fluence per time.
4. Photoelectric effect, Compton effect, Pair production, Coherent scattering.
5. Thomson scattering for elastic scattering at bound electrons; Rayleigh scattering from electron clouds of atoms; x-ray scattering from crystal structures; Mie scattering from nano- to microsize objects. Small angle x-ray scattering is the same as Mie scattering.
6. First step: x-ray photon is absorbed and all photon energy is transferred to a core electron. Second step: the core hole is filled and another fluorescence photon is emitted.
7. Generating characteristic x-ray radiation by electrons: there is one electron going in, two electrons and one photon coming out. Generating characteristic x-ray radiation via the PE effect: one photon goes in, one electron and one photon come out.
8. In photoelectric effect the photon emitted is characteristic for the atom from which it is emitted. The photoelectron has no directional preference. In Compton scattering the scattered photon has no relation to the atom it is scattered from. The Compton electron has a directional preference in the forward direction.
9. Coherent scattering is elastic, i.e. the photon energy does not change during the scattering process. Compton scattering is an inelastic scattering process. The photon transfers energy to the Compton electron.
10. Pair production does not occur below 1 MeV. However, for photon energies above 1 MeV, pair production becomes increasingly important.
11. High energy alpha particles loose little energy per collision. But at the end of their straight path they loose their remaining kinetic energy over a very short distance.
12. 5 cm.
13. The Bragg peak is a peak in linear energy transfer just before the end of the range of alpha particles.

14. No. Photons have a half thickness where half of the incident intensity has diminished.
15. The range of beta particles depends on the energy and is typically meters in air.
16. Fast neutrons interact via collisions, slow neutrons may be absorbed by target nuclei.
17. Light and heavy water.
18. In the energy region of a few eV.

Chapter 4

1. Dosimetry provides precise radiation measurements, defines radiation limits, and provides protection against unsafe radiation levels.
2. Dose is defined as deposited radiation energy per mass unit.
3. Q_{R} and RBE characterize different types of radiation with respect to their effectiveness of damaging biological tissue.
4. Dose is the deposited radiation energy per mass unit (unit Gray), equivalent dose is radiation energy weighted by the quality factor (unit Sievert).
5. Both are weighting factors: Q for the ionization effectiveness of radiation, and w for the sensitivity of body parts to radiation.
6. Fluence is the number of particles per area. Energy fluence considers only fluence from photons multiplied by their energy.
7. Kerma stands for kinetic energy released in matter. It is defined as the product of energy fluence and mass energy transfer coefficient.
8. No. It is a definition that is solely applied for electromagnetic radiation, i.e. x-rays and γ-radiation that transfer kinetic energy to electrons.
9. The mass attenuation coefficient accounts for all kinds of interactions of photons with matter, including those which do not transfer kinetic energy to the matter. The mass energy transfer coefficient takes into account only those attenuation processes, which result in kinetic energy of charged particles.
10. Film dosimeters, finger dosimeters, pocket dosimeters, and radiation monitors.
11. The design of these counters is identical, but the operational voltage is different. Proportional counters operate at lower voltage than Geiger–Mueller counters. The output signal of proportional counters is proportional to the energy of incident particles. Geiger–Mueller counters are event counters without energy resolution.
12. Dead time is the time when a counter is insensitive to radiation. The measured count rate is lower than the true count rate because of detector dead time.
13. Cosmic radiation, terrestrial radiation, and radiation from industry and medical applications.
14. 5 mSv/a.

15. ALARA stands for “as low as reasonably achievable”.
16. Shorten the exposure time and increase the distance to radioactive sources.

Chapter 5

1. Contrast is defined as the difference intensity of spatially separated regions divided by the sum of the intensities.
2. An attenuation profile is a lateral profile of intensity variations due to spatially varying mass absorption coefficients.
3. Reasons are an extended x-ray source and Compton scattering.
4. By the use of an oscillating collimator (grid).
5. Beam hardening can be achieved by two methods: Either by using a filter that absorbs the low energy end of the spectrum, or by increasing the accelerating voltage.
6. X-ray films have been replaced by flat panel thin film transistors.
7. No, the resolution of flat panel displays is slightly lower.
8. Sharing x-ray radiographs instantly worldwide and recording life videos of surgical procedures.
9. Mammography uses softer x-rays than in standard x-ray radiography, because of the low x-ray absorption of the female breast for hard x-rays.
10. Injection of contrast agents for imaging the GI tract or blood vessels; dual energy x-ray imaging; phase contrast imaging.
11. The real part.
12. Phase contrast imaging emphasizes topological differences even though absorption contrast may be lacking. Therefore edges are more pronounced in phase contrast imaging than in absorption imaging.
13. CT imaging can localize organs in a slice that would not be visible in projection radiography.
14. The Hounsfield scale is a relative scale of x-ray absorption with respect to water.
15. Air appears black, and bones appear white.
16. At present 256 detectors are mounted in parallel to cover a slice in the z direction of about 120 mm.
17. $0.5 \times 0.5 \times 0.5\,\mathrm{mm}^3$.
18. About 2–3 mSv.
19. For recognizing bone fractures, lung diseases, and early cancer developments.

Chapter 6

1. By injection.
2. Gamma emitters.

3. Because light radioisotopes do not emit gamma radiation.
4. By nuclear fission and proton irradiation.
5. Because Cs can substitute for K and Na and would disturb the electrolyte household of the body.
6. ^{99m}Tc and ^{123}I.
7. 140 keV.
8. They are analogous pharmaceuticals.
9. $^{99}\mathrm{MoO}_4^{2-}$ is doubly charged and strongly bond to alumina. When decaying to ^{99m}Tc, single charged $^{99}\mathrm{TcO}_4^{-}$ is less bonded to alumina and can therefore be washed out.
10. After 23 hours.
11. In the past photo plates, like for radiography. Presently scintillation counters.
12. Single photon emission computed tomography.
13. Two to three detector banks.
14. X-ray CT has higher spatial resolution and does not require isotopes. SPECT is better than x-ray CT in terms of functional studies, metabolism, and perfusional investigations.
15. For myocardial stress test, bone metabolism, and brain activity.

Chapter 7

1. The radioisotopes must be positron emitters, and they should have short half life.
2. They are produced by proton bombardment of light elements.
3. Pair production.
4. Because of momentum conservation.
5. A few millimeters.
6. By coincidence detection.
7. If one gamma quant is detected in one counter, a second one should be detected in a counter positioned 180° apart and within a narrow time window.
8. Coincidence by scattering and random accidental coincidence.
9. Cancer recognition and perfusion studies.
10. Perfusion is the flow of gases or liquids through the body driven by a concentration gradient, pressure gradient, or by active transport; diffusion is particle transport driven by concentration gradient only.
11. The positron emitter 18F is combined with glucose-like molecules yielding FDG. In the body FDG follows the same metabolic pathway as true glucose but is not metabolized to CO_2 and water, instead remains in tissue with high glucose demand. Those parts of the tissue indicate potential cancerous cells.
12. PET.
13. High cost for isotope production and for radiochemistry.

Chapter 8

1. Interphase and mitosis.
2. During the S phase.
3. P53.
4. Mutation, extension, attack, penetration, angiogenesis, and invasion.
5. Somewhere between penetration and angiogenesis.
6. During angiogenesis with additional oxygen supply.
7. S is an exponential function of the dose D.
8. They are most sensitive during the M phase and least sensitive during the S phase.
9. The probability to irradiate tumor cells during the most sensitive M phase is higher when the total dose is fractioned in smaller amounts over several sessions. In the interval time healthy tissues have a chance for self-repairing.
10. Direct action by single strand break or double strand break of a DNA. Indirect action via the generation of free toxic radicals.
11. RBE is defined as the ratio of x-ray dose to the dose of other types of radiation for the same survival fraction.
12. The RBE is highest when the interaction distance between two hits matches the diameter of the DNA. This is the case for radiation with a linear energy transfer of about 100 keV/μm. At the maximum the RBE reaches values of about 8–10.
13. Fast neutrons, alpha particles, carbon ions.
14. X-rays or gamma rays, protons, slow neutrons
15. OER is the oxygen enhancement ratio, meaning that the radiation effect is enhanced in the presence of oxygen within the cancerous volume.
16. For low LET radiation OER is important; for high LET radiation OER is not important.
17. The presence of oxygen supports the radiotherapy of LET radiation.
18. Low dose causes tumor growth delay but no growth control. High dose provides tumor control.
19. Chemotherapy enhances the radiation sensitivity.

Chapter 9

1. 4–25 MeV.
2. By an electron linac. At the end of the accelerator path the electrons hit a tungsten target for converting electrons to x-ray photons.
3. 80 % of photons undergo Compton scattering, the remaining 20 % cause pair production.

4. Electrons straggle for a certain distance and then stop. Photons are first converted into Compton electrons and electron/positrons and then build up the dose by ionizing soft matter.
5. For Kerma only the transfer of photon energy into kinetic energy of electrons is important, which occurs immediately starting beneath the skin, while the dose develops only after these charged particles straggle until they stop.
6. At roughly 1 to 3 mm below the surface depending on the photon beam energy.
7. The reason is the conversion of photons into Compton electrons and electron/positron pairs, which straggle over a certain distance and deposit kinetic energy for ionization of molecules.
8. SSD refers to constant source-surface distance, the reference point is the skin surface. SAD refers to constant source axis distance, which is in fact the constant source target volume distance.
9. For SSD simulation of the treatment plan is easier, but movement of treatment couch is required to meet the condition. SAD does not require patient movement, but dose-to-patient simulation is more difficult.
10. GTV is the target volume to be irradiated. CTV is an extra clinically defined volume that may be affected by proliferation of cancer cells. PTV allows for any patient or organ movement and encloses GTV and CTV.
11. With IGRT the PTV can be considerably reduced, sparing healthier tissue from high dose radiation.
12. This is only possible by multiple exposures from different angles or orientations. The target volume accumulates a high dose by exposure from all different orientations, while the tissue between surface and target is exposed only once.
13. IMRT-MLC stands for intensity modulated multileaf collimation. It is presently the best way to define a tumor volume with the prescribed dose.
14. Cyberknife technology overcomes most of the shortcomings of traditional XRT with a coplanar gantry. The linac is short, light weight and can be positioned at any angle by a robotic arm. Patient and organ movement is compensated for by an infrared and x-ray tracking system. The linac produces a maximum of 8 MeV photons, staying below the photon-neutron production threshold. MLC is not required as the fine x-ray beam automatically exposes a well-defined target volume.
15. The gamma-knife was used for treating brain tumors. The main justification is a simultaneous irradiation of a target volume with high energy gamma photons that can penetrate the skull. Gamma-knife technology requires Co isotopes. This technology has become obsolete by the introduction of cyberknife technology.

Chapter 10

1. Ionization of core electrons, Coulomb scattering at nuclei, elastic recoil scattering.
2. At the end of the track.
3. Yes, the range is equivalent with the position of the Bragg peak.
4. If the initial LET were independent of the proton energy, range and energy should be linearly related. But LET becomes less effective with increasing energy. Therefore the range increases further with higher proton energy.
5. About 1.1
6. PBT is low RBE treatment.
7. Very low in the beginning, increasing LET in the Bragg region. But still rather low compared to fast neutrons.
8. SPOB can be formed by tuning the incident energy and the proton fluence.
9. The proximal side receives much less radiation than in case of XRT. The distal side is not exposed at all. Therefore healthy tissue is better spared, i.e. the non-target dose is significantly lower.
10. The tail of the dose profile is due to fragmentation of carbon ions.
11. Ion source, linac, storage ring (synchrotron), and beam delivery.

Chapter 11

1. (1) Direct neutron irradiation with fast neutrons as fission product in a reactor; (2) irradiation with neutrons produced by (p,n) conversion via an accelerator; (3) thermal neutron capture by ^{10}B, producing alpha particles in tumor cells referred to as boron-neutron capture therapy.
2. Cold neutrons, thermal neutrons, epithermal neutrons, fast neutrons, and high energy neutrons.
3. Fast neutrons have high LET and high RBE, while their OER is 1. These are favorable properties for cancer treatment.
4. Inoperable tumors, tumors with radiation resistance to XRT and PBT, extended tumors close to the surface.
5. ^{9}Be(p,n)^{9}B, ^{9}Be(d,n)^{10}B, ^{3}H(d,n)^{4}He
6. The shape is similar to the one for high energy x-rays. Dose increases below the surface, goes through a maximum, and drops off linearly with increasing depth.
7. High energy neutrons first need to slow down to about 2 MeV before they develop their maximum LET. This occurs a few millimeters below the surface, depending on the initial neutron energy.
8. The number of fractions can be reduced in case of high LET; the dose per fraction is kept constant.

9. The role of thermal neutrons is to be captured by ^{10}B that causes a nuclear reaction releasing fast alpha particles. The alpha particles with a short range destroy cancer cells.
10. Alpha particles and Li ions.

Chapter 12

1. Brachytherapy is a radiation therapy where radioisotopes are in direct contact with carcinoma.
2. External beam therapy.
3. Prostate, cervical, breast, and skin cancer.
4. Either in form of capsules containing radioisotopes, radioactive wires, or plaques.
5. Very local treatment with high dose over short time, saves healthy tissue.
6. Mainly gamma radiation, beta radiation is stopped in the container wall.
7. 137-Co, 192-Ir, 103-Pd.
8. Mainly by neutron activation.
9. Either by hand or by using an applicator, allowing the insertion of a wire.
10. Interstitial (implantation) or contact on the inner or outer skin.
11. LDR is typically 2 Gy/h, HDR is typically 12 Gy/h.
12. In standard dosimetry there is one source and the target is at a certain distance from the source. In Brachytherapy there are multiple of sources; the radiation from all these sources overlap and they go in 4π.

Chapter 13

1. Light amplication by stimulated emission of radiation.
2. Laser light is monochromatic, unidirectional, and the electromagnetic waves have phase coherence.
3. For removing tumors close to inner or outer surfaces, lens corrections, removal of tattoos, fragmentation of stones, glaucoma surgery.
4. Mainly for increasing the laser flux that enables photoablation processes.
5. Mechanical switching, Q-switch, and mode coupling.
6. Electro-optical and acousto-optical
7. By forming wavepackets with a broad band of wavelengths that constructively interfere at specific nodes traveling in space and time.
8. CO_2, Nd:YAG, Ar^+, and Excimer lasers
9. In the order of 10^{12} W/cm^2, used mainly for LASIK and LASEK.
10. PDT is a procedure for cancer treatment with reactive O_2 molecules in the singlet state. First a photosensitive drug is administered, and after accumulation in

tumor cells a laser is used for excitation the photosensitizer that transfers its energy to oxygen, thereby transforming triplet oxygen to singlet oxygen.

11. LASIK conserves the epithelium layer of the cornea, LASEK disposes the epithelium layer.
12. LASEK, LASIK, diabetic macular edema removal, proliferative diabetic retinopathy, cataract, and glaucoma treatments.

Chapter 14

1. Precise targeting of local areas of tumors, support of imaging diagnostics, providing cures. Short: targeting, diagnostics, and therapeutics.
2. This part of the immune system is called reticuloendothelial system and consists of monocytes and macrophages that deposit the nanoparticles in the liver and spleen for excretion.
3. From a physiological point of view, the size of nanoparticles determines the RES and the EPR and therefore the lifetime of nanoparticles in the body. From a physical point of view the size of nanoparticles determines the blocking temperature for magnetic NPs and plasmon frequency in the case of metal nanoparticles.
4. Active and passive targeting. Physical targeting by the size of particles is passive targeting. Active targeting requires ligands that can dock to specific sites.
5. Curie temperature characterizes ferromagnets. Curie temperature is an ordering temperature. Below the Curie temperature the magnetic moments in a solid are in an ordered state. Above the Curie temperature the magnetic moments are disordered, also called paramagnet.
6. In the ordered state below the Curie temperature the magnetic moments in a nanoparticle are ordered and form a tiny single domain of ferromagnetic state. The collection of these ordered moments are called macrospin.
7. The total magnetic moment of the half-filled Gd-shell does not follow from $gS\mu_B$. The correct equation is $g(S(S+1))^{1/2}$, yielding $8\mu_B$.
8. The blocking temperature separates low temperature behavior of magnetic nanoparticles from a high temperature behavior. At low temperatures nanoparticles do not reorient within a typical observation time; at high temperatures macrospins are not fixed in space but fluctuate strongly.
9. In MRI it is important that the magnetization of the NPs align in the field of Helmholtz coils in the MRI machine and are stable in the field along the z direction. This is only the case for nanoparticles whose blocking temperature is below body temperature. For hyperthermia the magnetization of the nanoparticles should follow the external ac field. This is only possible for NPs in the superparamagnetic state above the blocking temperature.
10. Néel relaxation, Brown relaxation, and hysteretic relaxation.

11. In the case of positive contrast, T1 is shortened. In the case of negative contrast, T2 is shortened.
12. Magnetite nanoparticles with large magnetic dipole fields contribute to the fluctuation of local fields that shortens T2.
13. Gd^{3+} chelates are in direct contact with protons and support the external field with their large magnetic moment. This shortens the T1 relaxation time and therefore acts as a positive contrast agent.
14. K-edge contrast imaging is a radiography modality by taking two sequential pictures, one with a mean x-ray photon energy above the K x-ray absorption edge of a heavy metal particle like Au, and another one taken below the K-edge of the same element.
15. Rayleigh scattering occurs for objects such as molecules which are small compared to the wavelength of electromagnetic waves. Mie scattering is the scattering of EM waves at objects with a diameter smaller or equal the wavelength of EM radiation.
16. Forward direction with respect to the incoming wave.
17. The absorbance maximum moves to longer wavelength with increasing particle size and with increasing aspect ratio.
18. Using plasmonic excitation of metal nanoparticles with a laser, heat is generated in particles that is transferred to the tissue by thermal conductance and by radiation.
19. The illuminated nanoparticles can be best made visible by a dark field microscope.
20. Theranostic nanoparticles are those which serve the purpose of diagnostics, for instance by contrast enhancement, and for therapeutics, such as hyperthermia. In any case, theronostic nanoparticles serve a multimodality purpose.

Chapter 15

1. Orthopedics.
2. The C-legs have to mimic the complete knee, leg, and foot movement. The motor control of all movements aims at reproducing a normal gait pattern.
3. The main challenge is the training of remaining nerve fibers attached to new muscles to generate EMG signals, which can be translated into a smooth and accurate control of upper limb prostheses.
4. The best combination with the lowest wear is a ceramic on ceramic combination. It is also the most expensive one.
5. These are the tracts of the gall, of the tear duct, and the outflow duct of the aqueous fluid.

6. A stent should not disturb the hemodynamics and should not initiate blood clotting. Stents should be plastically deformable and rigid in the expanded shape.
7. Sound bridges are implanted in the middle ear and act on the oval window.
8. Epiretinal microelectrode array and subretinal microphotodiode array.
9. Blind patients whose retina and in particular the ganglion cells together with the optic nerve bundle is still intact.

List of acronyms used in this book

ADC	analog digital converter
ADC	apparent diffusional constant
ADP	adenosine diphosphate
ALARA	as low as reasonably achievable
amu	atomic mass units
ATP	adenosine triphosphate
AV	atrioventricular node
AVV	atrioventricular valve
BBB	blood-brain barrier
BF	breathing frequency
BMD	bone mineral density
BMR	basal metabolic rate
BMS	bare metal stents
BOLD fMRI	blood oxygen level dependent fMRI
BPS	biodegradable polymeric stents
BSA	beam-shaping assembly
BSA	body surface area
BW	body weight
c.m.	center of mass
CA	contrast agent
CAP	cardiac action potential
CBF	cerebral blood flow
CCC	continuous curvilinear capsulorhexis
CCD	caput-collum-diaphyseal angle
CCD	charge coupled device
CIRT	carbon ion radiation therapy
CLE	confocal laser endoscopy
CO	cardiac output
COE	caloric oxygen equivalent
CPA	charged particle activation
CPB	cardiopulmonary bypass
CS	Compton scattering
CSF	cerebrospinal fluid
CT	computed tomography
CTV	clinical Target Volume
cw	continuous wave
dB	decibel
DBT	digital breast tomosynthesis
DCE	dynamic contrast enhancement
DES	drug elusion stent
DNP	dynamic nuclear polarization
DOV	depth of view
DR	digital recording
DSA	digital subtraction angiography
DSB	double strand break
DTI	diffusion tensor imaging

DOI 10.1515/9783110553116-019

DTL	drift tube linac
DWI	diffusion weighted imaging
EBRT	external beam radiotherapy
ECC	extracorporal circulation
ECG	electrocardiography
EDP	end diastolic pressure
EDV	end diastolic volume
EEG	electroencephalography
EF	ejection fraction
EM	electromagnetic
EMG	electromyography
EPI	echo planar imaging
EPP	end plate potential
EPR	enhanced permeation and retention
ERBT	external radiation beam therapy
ERV	expiratory rest volume
ESP	end systolic pressure
ESV	end systolic volume
ETL	echo train length
FCRM	fiber optic confocal reflectance microscope
FE	fractional excretion
FF	flattening filter
FF	filtration fraction
FFDM	full-field digital mammography
FFF	flattening filter free
FFT	fast Fourier transform
FID	free induction decay
fMRI	functional magnetic resonance imaging
FOV	field of view
FRC	fractional rest volume
FSE	fast spin echo
GFR	glomerular filtration rate
GTV	gross tumor volume
Hct	hematocrit value
HD	hydrodynamic diameter
HDR	high dose rate
HEP	hemi-endoprosthesis
hMRI	hyperpolarization magnetic resonance imaging
HSLS	high spin - low spin
IAEA	International Atomic Energy Agency
ICRP	International Commission for Radiological Protection
IGRT	image guided radiotherapy
IHC	inner hair cell
IMRT	intensity modulated radiation therapy
IRV	inspiratory rest volume
ITD	interaural time difference
Kerma	kinetic energy release in matter
kHz	kilohertz
kV	kilovolt

kVp	peak kilovoltage
kW	kilowatt
LASEK	laser epithelial keratomileusis
LASER	light amplification by stimulated emission of radiation
LASIK	laser-assisted interstitial keratomileusis
LCI	low coherence interferometry
LDR	low dose rate
LED	light emitting device
LET	linear energy transfer
Linac	linear accelerator
LSO	lateral superior olive
MEG	magnetoencephalography
MEI	middle ear implants
MET	mechanoelectric transduction
MeV	mega-electronvolt
MHR	metabolic heat production
MHT	magnetic hyperthermia
MHz	megahertz
MI	mechanical index
MLC	multileaf collimator
MNP	magnetic nanoparticle
mpMRI	multiparameter MRI
MRI	magnetic resonance imaging
MRSI	magnetic resonance spectroscopy imaging
MRT	magnetic resonance tomography
MSFP	mean systemic filling pressure
MSO	medial superior olive
MUAP	motor unit action potential
MV	minute ventilation
NBI	narrow band imaging
NIR	near-infrared
NMJ	neuromuscular junction
NP	nanoparticle
NRT	neutron radiation treatment
NTD	nontarget dose
OCT	optical coherence tomography
OER	oxygen enhancement ratio
OHC	outer hair cell
PAH	para-aminohippuric acid
PCI	percutaneous coronary intervention
PCI	phase contrast imaging
PCV	packed cell volume
PD	proton density
PDR	proliferative diabetic retinopathy
PDR	pulse dose rate
PDT	photodynamic therapy
PE	photoelectric effect
PES	photoelectron emission spectroscopy
PET	positron emission tomography

PHIP	parahydrogen induced polarization
PI	pulsatility index
PPD	percentage photon dose
PRF	pulse repeat frequency
PRK	photorefractive keratectomy
PRP	panretinal photocoagulation
PRT	proton radiotherapy
PRT	pulse repeat time
PTV	planning target volume
PWV	pulse wave velocity
PZT	$PbTiO_3$
Q	quality factor
RBC	red blood cell
RBE	relative biological effectiveness
RBF	renal blood flow
RC	respiratory coefficient
Re	Reynolds number
RES	reticuloendothelial system
RF	radio frequency
RF	respiratory fraction
RMR	resting metabolic rate
RPF	renal plasma flow
RT	radiotherapy
RTR	real-time radiography
RV	residual volume
SA	sinoatrial node
SAD	source to axis distance
SATP	standard ambient temperature and pressure
SAXS	small angle x-ray scattering
SCI	spinal cord injury
SE	spin echo
SERS	surface enhanced Raman scattering
SLAC	Stanford linear accelerator
SNR	signal-to-noise ratio
SOBP	spread-out Bragg peak
SPE	single photon emission
SPECT	single photon emission computed tomography
SPIO	superparamagnetic iron oxide
SPR	surface plasmon resonance
SSB	single strand break
SSD	source-to-surface distance
SV	stroke volume
TE	time of (spin, acoustic) echo
TEP	total endoprosthesis
TGC	time gain compensation
THz	terahertz
TIPPB	transperineal interstitial permanent prostate brachytherapy
TLC	total lung capacity
TMR	targeted muscle reinnervation

TPFR	total peripheral flow resistance
TR	time of repetition
TT	transfer time
TV	tidal volume
UHMWPE	ultrahigh molecular weight polyethylene
ULFMRI	ultralow field magnetic resonance imaging
US	ultrasound
VC	vital capacity
VEGF	vascular endothelial growth factor
XRT	x-ray radiotherapy
YAG	yttrium-aluminum garnet
ZFC	zero field cooling

Index