Waldemar Nawrocki

Introduction to Quantum Metrology

Quantum Standards and Instrumentation

Springer

Introduction to Quantum Metrology

Waldemar Nawrocki

Introduction to Quantum Metrology

Quantum Standards and Instrumentation

Springer

Waldemar Nawrocki
Faculty of Electronics
and Telecommunications
Poznan University of Technology
Poznan
Poland

The work was first published in 2007 by Wydawnictwo Politechniki Poznańskiej (Publisher of Poznan University of Technology) with the following title: Wstęp do metrologii kwantowej.

ISBN 978-3-319-15668-2 ISBN 978-3-319-15669-9 (eBook)
DOI 10.1007/978-3-319-15669-9

Library of Congress Control Number: 2015932516

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Quantum metrology is a field of theoretical and experimental study of high-resolution and high-accuracy methods for measurement of physical quantities based on quantum mechanics, particularly using quantum entanglement. Without equivalent in classical mechanics, quantum entanglement of particles or other quantum systems is an unusual phenomenon in which the state of a system can be determined better than the state of its parts. Attempts are made to use new measurement strategies and physical systems in order to attain measurement accuracy never achieved so far.

Quantum metrology sprang into existence at the beginning of the twentieth century, along with quantum mechanics. After all, the Heisenberg uncertainty principle, together with the Schrödinger equation and the Pauli exclusion principle constituting the basis of the formalism of quantum mechanics, is also the fundamental equation of quantum metrology. The uncertainty principle sets limits to measurement accuracy, but has no relation to the technical realization of the measurement.

Quantum metrology only started to develop in the latter half of the twentieth century, following the discovery of phenomena of fundamental importance to this field, such as the Josephson effect, the quantum Hall effect or the tunneling of elementary particles (electrons, Cooper pairs) through a potential barrier. Using new important physical advances, quantum metrology also contributes to progress in physics. In the past 50 years the Nobel Prize was awarded for 16 achievements strongly related to quantum metrology. In 1964 Townes, Basov and Prokhorov received the Nobel Prize *for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser-laser principle*. Presently masers constitute a group of atomic clocks, installed in metrology laboratories as well as on satellites of GPS and GLONASS navigation systems. Gas lasers are basic instruments in interferometers used in metrology of length, and semiconductor lasers are used in industrial measurements. Haroche and Wineland were awarded the Nobel Prize in 2012 *for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems*, that is, for studies in the field of quantum metrology. In 1964–2012 the Nobel Prize was awarded for important discoveries such as *the Josephson effect* (Josephson

1973), *the quantum Hall effect* (von Klitzing 1985), and *the scanning tunneling microscope* (Röhrer and Binnig 1986). Thus, scientific achievements in quantum metrology or relevant to this field are considered very important for science in general.

Currently, the major field of practical application of quantum metrology is the development of standards of measurement units based on quantum effects. Quantum standards are globally available universal primary standards that allow to realize a given unit at any place on the Earth by measurements which, in appropriate conditions, will yield equal results anywhere. The creation of quantum standards of the base units of the International System (SI) is in accordance with the objectives set by the International Committee for Weights and Measures and realized in collaboration with the International Bureau of Weights and Measures (BIPM).

This book provides a description of selected phenomena, standards, and quantum devices most widely used in metrology laboratories, in scientific research, and in practice.

The book opens with a discussion of the theoretical grounds of quantum metrology, including the limitations of the measurement accuracy implied by theoretical physics, namely the Heisenberg uncertainty principle and the existence of energy resolution limits, discussed in Chap. 1. Providing the rudiments of systems of measurements, Chap. 2 discusses the currently adopted standards for the realization of SI units, and the changes in the classical system of units allowed by quantum metrology. Chapter 3 is devoted to the activities and proposals aimed at the development of a new system of units to replace the SI system, with units of measurement defined in relation to fundamental physical and atomic constants. Chapters 4, 6, 9, 10, and 12 present the theory and practical realizations of quantum standards of units of various quantities: the voltage standard using the Josephson effect, the resistance standard based on the quantum Hall effect, the atomic clock-based frequency standard, the length standard using laser interferometry, and the mass standard based on the masses of atoms and particles. Chapter 11 is devoted to scanning probe microscopy. Chapters 5 and 8 discuss sensitive electronic components and sensors based on quantum effects and including, among others, superconducting quantum interference magnetometers (SQUIDs), single electron tunneling transistors (SETT), and advanced quantum voltage-to-frequency converters based on the Josephson junction. Presented in Chap. 5 along with many application systems, SQUIDs are *the most sensitive of all sensors of all physical quantities*.

The description of the discussed devices and the underlying physical effects is complemented by a presentation of standardization methods and principles of comparison between quantum standards (with the time standard used as an example) in accordance with the hierarchy of the system of units.

Intended to serve as a textbook, this book also represents an up-to-date and hopefully inspiring monograph, which contributes to scientific progress. As a scientific survey it puts in order the fundamental problems related to electrical metrology, the universal standards, and the standardization methods recommended by BIPM. As an academic textbook it propagates a new approach to metrology, with more emphasis laid on its connection with physics, which is of much importance for the constantly developing technologies, particularly nanotechnology.

Large parts of this publication represent a translation from my book *Introduction to Quantum Metrology* published in Polish by Publishing House of Poznan University of Technology in 2007 and used here in translation with the publisher's permission.

I thank all those who helped me collect material for this book. Special thanks go to my colleagues at the Polish Central Office of Measures in Warsaw and at Physikalisch-Technische Bundesanschtalt in Braunschweig.

Poznan, Poland Waldemar Nawrocki

Contents

Chapter 1
Theoretical Background of Quantum Metrology

Abstract This chapter outlines the history of quantum mechanics and presents its fundamental formulas: the Schrödinger equation with the interpretation of the wave function, the Pauli exclusion principle, and the Heisenberg uncertainty principle. The latter is illustrated with sample numerical calculations. We briefly discuss the application of quantum effects in metrology, and present and compare the limits of accuracy and resolution of classical and quantum standards. Prospects for a new system of units are discussed as well.

1.1 Introduction

Measurement consists in comparing the measured state A_x of a quantity to its state A_{ref} considered a reference state, as shown schematically in Fig. 1.1. Thus, the accuracy of the measurement cannot be better than the accuracy of the standard.

For many years metrologists have been working on the development of standards that would only depend on fundamental physical constants and atomic constants. With such standards units of measurement can be realized on the basis of quantum phenomena. One of the directions of research in metrology is the creation of a new system of measurement, based on a set of quantum and atomic standards, that would replace the classical SI system.

In the present chapter we shall focus on measures and the development of the system of units of measurement. We shall discuss the history of measurement standards and present in detail the currently used International System of Units (SI). The set of the base units of the SI system of measurement will be discussed as well. The following chapters present the quantum phenomena that are the most commonly used in metrology: the Josephson effect is discussed in Chap. 4, the quantum Hall effect in Chap. 6, and the tunneling of single electrons in Chap. 8. These three phenomena are considered of particular importance not only for electrical metrology,

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_1

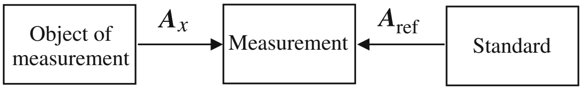


Fig. 1.1 Measurement: a comparison between the object and standard

but also for science as a whole. For his theoretical studies predicting the effect of voltage quantization Brian David Josephson was awarded the Nobel Prize in 1973, and Klaus von Klizing received the Nobel Prize in 1985 for his discovery of the quantum Hall effect.

Standards for a quantum system of measurements based on quantum mechanical phenomena have been implemented in the past 25 years. Quantum phenomena are described in terms of concepts of quantum mechanics. The beginning of quantum mechanics is conventionally set at 1900, when Max Planck proposed a new formula for the intensity of emission of electromagnetic radiation in the thermal and visible wavelength ranges of the spectrum. In his analysis Planck assumed that the energy changed by quanta proportional to a constant denoted as h, which was later named the Planck constant. The dependence proposed by Planck described the measured emission of electromagnetic radiation much better than the models in use at that time, based on the rules of classical physics. As a set of rules and formulas quantum mechanics was developed in the 1920s by Erwin Schrödinger, who formulated the Schrödinger equation, and Werner Heisenberg, the author of the uncertainty principle, with major contributions by Louis de Broglie, Max Born, Niels Bohr, Paul Dirac, and Wolfgang Pauli. Unlike classical physics, quantum mechanics often leads to renounce the common-sense comprehension of physical phenomena and utterly astounds the researcher or reader. One of its latest surprises is the observation of electric current flowing simultaneously (!) in both directions on the same way in the circuit [4]. As Niels Bohr, the author of the model of the hydrogen atom, said: *Anyone who is not shocked by quantum theory probably has not understood it.*

Quantum mechanical phenomena are employed in at least three fields of metrology:

- The construction of quantum standards of units of physical quantities: standards for electrical quantities such as voltage, electrical resistance or electrical current, and non-electrical standards, including the atomic clock and the laser standard for length;
- The determination of the physical limits of measurement precision by the Heisenberg uncertainty principle;
- The construction of extremely sensitive electronic components: the magnetic flux sensor referred to as SQUID (Superconducting Quantum Interference Device) and the SET transistor based on single electron tunneling.

1.2 Schrödinger Equation and Pauli Exclusion Principle

The development of quantum mechanics was preceded by a discovery by *Max Planck* in 1900. On the basis of measurements of the intensity of black-body radiation Planck derived a formula based on the assumption that the energy was exchanged in a noncontinuous manner between particles and radiation, and emitted in quanta proportional to a constant, which is now known as the Planck constant h ($h = 6.626 \times 10^{-34}$ J s), and the radiation frequency f:

$$E = h \times f. \tag{1.1}$$

The results of measurements of the energy density of the radiation as a function of the temperature and frequency in the thermal, visible and ultraviolet ranges of the spectrum (spanning the wavelengths from 200 nm to 10 μm) could not be explained by the rules of classical physics. Established on the basis of classical physics, the Rayleigh-Jeans formula for the energy density, although formally correct, only described accurately the studied phenomenon in the far infrared spectral range, i.e., for wavelengths above 5 μm. For shorter wavelengths the results obtained from the Rayleigh-Jeans formula diverged so much from the measurement data that the theoretical dependence in this range, predicting infinite energy density, was named the ultraviolet catastrophe. Only the formula derived by Planck on the assumption that energy was quantized corresponded well to the measurement data in the whole wavelength range [7], as indicated by the plot in Fig. 1.2.

The Planck law is expressed as a dependence $u(f, T)$ of the spectral radiant emission on the frequency f and temperature T (1.2), or a dependence $u(\lambda, T)$ of the spectral radiant emission on the wavelength λ and temperature T:

$$u(f, T) = \frac{4hf^3}{c^3} \frac{1}{\exp\left(\frac{hf}{k_B T}\right) - 1}, \tag{1.2}$$

where $u(f, T)$ is the spectral radiant emission of a perfect black body, f is the radiation frequency, T—the absolute temperature of the black body, h—the Planck constant, k_B—the Boltzmann constant, and c—the speed of light in vacuum.

The Planck lecture on the subject, presented at a meeting of the Berliner Physikalische Gesellschaft on December 14, 1900, is considered the beginning of quantum mechanics. Five years later, in 1905, *Albert Einstein*, analyzing the photoelectric effect, came to the conclusion that not only the emitted energy, but also the energy E of absorbed light was quantized. Einstein was the first to affirm that the energy of light was carried in portions hf proportional to the frequency f and to the Planck constant [2]. Contrary to the common opinion at that time, Einstein presented light as discrete. This result of theoretical considerations was received with general disbelief. When the quantum nature of the energy of light was confirmed experimentally by the American physicist Robert A. Millikan in 1915, he himself was very surprised at that result. Moreover, to light quanta, later named

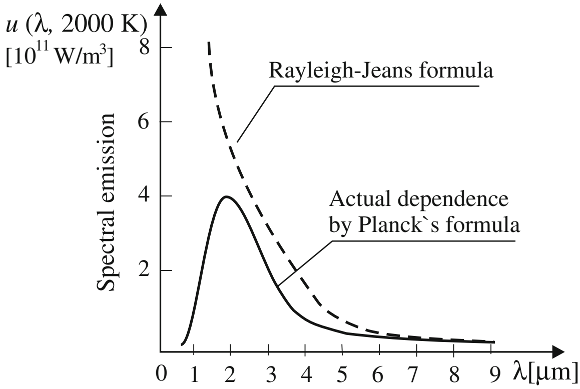


Fig. 1.2 Spectral radiant emission of a black body in the thermal and visible spectral ranges at the temperature of 2,000 K according to the Rayleigh-Jeans formula (*dashed line*) and by the Planck formula (*solid line)* [6]

photons, Einstein attributed properties of material particles with zero rest mass. Evidence for the quantum character of light was provided, among others, by the experiments by Walther Bothe and Arthur H. Compton.

Einstein's studies on the photoelectric effect were so momentous for theoretical physics that they brought him the Nobel Prize in 1921. Albert Einstein is often portrayed as a skeptic of quantum mechanics, in particular its probabilistic description of phenomena. However, by his studies on the photoelectric effect (1905) and the specific heat of solids (1907) he unquestionably contributed to the development of quantum mechanics. In his publication on specific heat Einstein introduced elements of the quantum theory to the classical theory of electrical and thermal conduction in metals, which was proposed by Paul Drude in 1900.

A hypothesis that the whole matter had a particle-wave nature was put forward by *Louis de Broglie* in 1924 in his PhD Thesis, in which he considered the nature of light [5, 8]. At that time it was already known that light exhibited the properties of both particles and waves. The evidence provided by experiments included the observation of the bending of light rays in the gravitational field of stars, an effect predicted by Einstein. De Broglie's hypothesis proposed that, if light had the properties of particles, besides those of waves, it might be that also the elementary particles that constituted matter had characteristics of both particles and waves. According to de Broglie, the movement of a material particle is associated with a wave of length λ and frequency f:

$$\lambda = \frac{h}{p} = \frac{h}{mv}, \quad f = \frac{E}{h}, \tag{1.3}$$

where λ is the wavelength corresponding to the position of the particle, p the momentum of the particle, m its mass, E its energy, and h the Planck constant.

The movement of an electron at a speed of 10^3 m/s is associated with a wave of wavelength $\lambda \approx 7 \times 10^{-7}$ m (ultraviolet radiation), as calculated from (1.3). The movement of a neutron with the same speed (10^3 m/s) is associated with a wave of wavelength $\lambda \approx 4 \times 10^{-13}$ m. In other words, a neutron moving with that speed can be considered a *de Broglie wave* with a wavelength of 4×10^{-13} m. Similar wavelengths are characteristic of cosmic rays. Particles with a mass much larger than that of the neutron, even when moving at a much lower speed (1 m/s), would be associated with waves so short that their measurement would be impossible. Thus, the wave properties of such particles cannot be confirmed. For example, a particle with a hypothetic mass of 10^{-3} g and a velocity of 1 m/s would correspond to a de Broglie wave of wavelength $\lambda \approx 7 \times 10^{-28}$ m. Since neither waves with such a wavelength, nor elementary particles with a mass of the order of 1 mg have ever been observed, we do not know if they exist.

A good example of the particle-wave duality of matter is the electron, discovered by *John J. Thomson* in 1896 as a particle with an electric charge that was later established to be $e = 1.602 \times 10^{-19}$ C, and a mass $m = 9.11 \times 10^{-31}$ kg. The wave properties of the electron were demonstrated experimentally by the diffraction of an electron beam passing through a metal foil, observed independently by George P. Thomson (son of J.J. Thomson, the discoverer of the electron), P.S. Tartakovsky [8] and the Polish physicist Szczepan Szczeniowski.

One of the milestones in the history of physics, the Schrödinger equation was formulated (not derived) by the Austrian physicist *Erwin Schrödinger* in 1926 on the speculative basis by analogy with the then known descriptions of waves and particles. Since the validity of the formula has not been challenged by any experiment so far, it is assumed that the Schrödinger equation is true. If the Heisenberg uncertainty principle sets the limits to the accuracy with which parameters of a particle can be determined, the Schrödinger equation describes the state of an elementary particle. The *Schrödinger equation* features a function referred to as the wave function or state function and denoted as Ψ (*psi*), and expresses its complex dependence on time and the position coordinates of the particle:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + A\Psi = j\hbar\frac{\partial\Psi}{\partial t}, \tag{1.4}$$

where Ψ is the wave function, A a function of time and the coordinates of the particle, m the mass of the particle, t denotes time, ∇ the Laplace operator, $\hbar = h/2\pi$ is reduced the Planck constant, i.e. the Planck constant divided by 2π, and j is the imaginary unit.

When the function A is independent of time, it expresses the potential energy of the particle. In such cases the Schrödinger equation takes the form:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + A\Psi = E\Psi, \tag{1.5}$$

where A is the potential energy of the particle, and E denotes its total energy.

A physical interpretation of the wave function was first provided in 1926 by *Max Born* (born in Wroclaw, then Breslau, in 1882). The wave function describes the probability that the particle is present within a certain region, specifically in a volume dV. The probability is proportional to the square of the module of the wave function:

$$p = k|\Psi|^2 \mathrm{d}V, \tag{1.6}$$

where p denotes the probability, k is the proportionality coefficient, and V the volume of space available to the particle.

Pauli's exclusion principle says that in an atom no two electrons can have the same quantum state i.e. no two electrons can have the same set of four quantum numbers. The Pauli exclusion principle must be taken into account when separate atoms or nanostructures are analyzed (2-dimensional electron gas—see Chap. 6, nanostructures—see Chap. 7 and single electron tunneling—see Chap. 8).

1.3 Heisenberg Uncertainty Principle

A description of elementary particles which was equivalent to the Schrödinger equation had been proposed before, in 1925, by the German physicist *Werner Heisenberg* (who was 24 years old at that time!). Two years later Heisenberg formulated the uncertainty principle [3, 8], which is one of the foundations of quantum mechanics. The now famous relations expressing the uncertainty principle were proposed by Heisenberg in 1927 in an article *Über den Inhalt der anschaulichen quantentheoretischen Kinematik und Mechanik*. The uncertainty principle is closely related to the particle-wave duality of matter. It defines the limits of accuracy with which the state of a particle can be determined. The uncertainty principle has nothing to do with the accuracy of the instruments used for the measurement. When the position x of an electron, regarded as a particle, is established with an uncertainty (or margin of error, in the language of metrology) Δx, the state of this electron can also be represented in the wave image as a wave beam consisting of waves with different wavelengths. The electron is assigned a wavelength λ, the value of which is related to the momentum of the electron. According to the de Broglie formula:

$$\begin{gathered} \lambda = \frac{h}{p} \\ p = mv, \end{gathered} \tag{1.7}$$

where m is the mass of the electron, and v its speed.

Along the segment Δx corresponding to the uncertainty in the position of the particle a wave has n maximums and the same number of minimums:

$$\frac{\Delta x}{\lambda} = n. \tag{1.8}$$

A wave beam with zero amplitude beyond the segment Δx must include waves that have at least $(n + 1)$ minimums and maximums along this segment:

$$\frac{\Delta x}{\lambda - \Delta\lambda} \geq n + 1. \tag{1.9}$$

From (1.8) and (1.9) it follows that:

$$\frac{\Delta x \times \Delta\lambda}{\lambda^2} \geq 1.$$

From the de Broglie formula we get:

$$\frac{\Delta\lambda}{\lambda^2} = \frac{\Delta p}{h}.$$

Finally, we obtain the formula for the *uncertainty principle*:

$$\Delta x \times \Delta p \geq \hbar/2, \tag{1.10}$$

where, $\hbar = h/2\pi$—reduced the Planck constant.

According to the (1.10), the product $\Delta x \times \Delta p$ of the uncertainty Δx in the position of the particle (e.g. electron) in one dimension and the uncertainty Δp in its momentum p in simultaneous determination of x and p (which is very important to note) is not less than the reduced Planck constant divided by 2. This means that even in the most accurate measurements or calculations for simultaneous determination of the position x of the particle and its momentum p, if the uncertainty Δx in the position is reduced, the uncertainty Δp in the momentum must increase, and vice versa. When the position of the particle is defined in three dimensions by coordinates x, y, z, a set of three inequalities applies in place of the single inequality (1.10):

$$\begin{aligned} \Delta x \times \Delta p_x &\geq \hbar/2 \\ \Delta y \times \Delta p_y &\geq \hbar/2 \\ \Delta z \times \Delta p_z &\geq \hbar/2. \end{aligned} \tag{1.11}$$

The uncertainty principle is of much practical importance in nanometer-sized structures. For example, if the uncertainty in the determination of the position of the electron is 2×10^{-10} m (the order of magnitude corresponding to the dimensions of an atom), according to the formula (1.10) the velocity of the electron can be determined with an uncertainty Δv:

$$\Delta x\,(m\;\times\Delta v) \geq \hbar/2$$
$$\Delta v \geq \frac{h}{4\pi\,m\,\Delta x} = \frac{6.626\times 10^{-34}\,\mathrm{Js}}{4\pi\times 9.1\times 10^{-31}\,\mathrm{kg}\times 2\times 10^{-10}\,\mathrm{m}} = 2.9\times 10^{5}\,\mathrm{m/s}.$$

This is a wide range, about three times wider than the velocity of the electron, v_{th}, related to the thermal energy k_BT at room temperature ($v_{\mathrm{th}} \approx 10^5$ m/s).

If we give up the simultaneous determination of the parameters of the motion of the particle and assume that its position is fixed, the formula (1.12) below will provide the limits to the uncertainty ΔE in the energy of the particle and the uncertainty Δt in its lifetime or the observation time:

$$\Delta E\times\Delta t \geq \hbar/2. \tag{1.12}$$

For instance, let us calculate the time Δt necessary for a measurement of energy with the uncertainty $\Delta E = 10^{-3}$ eV = 1.6×10^{-22} J:

$$\Delta t \geq \frac{\hbar}{2\,\Delta E} = \frac{6.63\times 10^{-34}}{4\pi\times 1.6\times 10^{-22}} \approx 3.3\times 10^{-13}\,\mathrm{s} = 0.33\;\mathrm{ps}.$$

1.4 Limits of Measurement Resolution

A question worth considering in measurements of low electrical signals by means of quantum devices is the existence of limits to the energy resolution of the measurement. Can signals be measured in arbitrarily small ranges? By the energy resolution we understand here the amount of energy or the change in energy that is possible to measure with a measuring instrument. Well, the limits of measurement resolution are not specified. Its physical limitations result from:

- The Heisenberg uncertainty principle in the determination of parameters of elementary particles;
- The quantum noise of the measured object, which emits and/or absorbs electromagnetic radiation;
- The thermal noise of the measured object.

The thermal noise power spectral density in an object at an absolute temperature T is described by the Planck equation:

$$\frac{P(T,f)}{\Delta f} = hf + \frac{hf}{\exp(\frac{hf}{k_BT})-1}, \tag{1.13}$$

where k_B is the Boltzmann constant.

The Planck equation (1.13) takes two extreme forms depending on the relationship between the thermal noise energy k_BT and the quantum hf of energy of electromagnetic radiation. For $k_B T \gg hf$ the Planck equation only includes the thermal noise component and takes the form of the Nyquist formula:

$$E(T) = \frac{P(T)}{\Delta f} \cong k_B T. \tag{1.14}$$

For $k_B T \ll hf$ the Planck equation only includes the quantum noise:

$$E(f) = \frac{P(f)}{\Delta f} \cong hf. \tag{1.15}$$

It can be noticed that thermal noise plays a dominant role in the description of the spectral power at low frequencies, whereas quantum noise predominates at high frequencies. The frequency f for which both components of the spectral noise power are equal, $k_B T = hf$, depends on the temperature; for example, $f = 6.2 \times 10^{12}$ Hz at the temperature of 300 K, and $f = 56$ GHz at 2.7 K, which is the temperature of space [3].

The (1.15) also describes energy as a physical quantity quantized with an uncertainty conditioned by the uncertainty in the measured frequency (currently of the order of 10^{-16}), and the accuracy with which the Planck constant is determined (presently of the order of 10^{-9}).

Below we present an attempt to estimate the lower bound of the measurement range in the measurement of electric current. Electric current is a flow of electrons, and its intensity in the conductor is defined as the derivative of the flowing electric charge $Q(t)$ with respect to time t, or as the ratio of the time-constant electric charge Q to the time t over which the charge is transferred:

$$I = \frac{\mathrm{d}Q(t)}{\mathrm{d}t}. \tag{1.16}$$

On the microscopic scale we can consider, and, what is more important, record, the flow of single electrons and calculate the intensity of the related electric current. Thus, for example, the flow of one billion (10^9) of electrons over 1 s is a current of 1.6×10^{-10} A, or 160 pA. It is now possible to measure electric current of such intensity with present-day ammeters, and sometimes even with multimeters, which combine a number of measurement functions for different physical quantities. A flow of electric charge at a much lower rate, such as one electron per second or one electron per hour, will no longer be regarded as electric current. In this case it is preferable to consider the motion of individual electrons and count the flowing charges, since the averaging of individual electrons over time tends to be useless.

Sensors measuring physical quantities respond to the energy or change in energy of the signal. Thus, the sensitivity of the measurement is limited by the Heisenberg uncertainty principle. The best energy resolution to date has been achieved by SQUID sensors; its value, equal to $0.5h$, reaches the physical limit (see Chap. 5). In measurements of linear displacement the best linear resolution, of 10^{-6} Å $= 10^{-16}$ m, has been obtained with the X-ray interferometer [1]. This wonderful achievement in metrology is worth comparing with atomic dimensions: for example, the diameter of the copper atom is 3.61 Å, and that of the gold atom is 4.08 Å. In measurements of displacement and geometrical dimensions the best linear resolution, $\Delta a = 0.01$ Å $= 10^{-12}$ m and $\Delta b = 0.1$ Å in vertical and horizontal measurements, respectively, has been obtained with the scanning tunneling microscope (STM) in the study of the atom arrangement on the surface of a solid (see Chap. 11). The operation of the STM is based on the quantum effect of electron tunneling through a potential barrier.

References

1. R.D. Deslattes, Optical and x-ray interferometry of a silicon lattice spacing. Appl. Phys. Lett. **15**, 386 (1968)
2. A. Einstein, Űber einen die Erzuegung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. Ann. Phys. **17**, 132–148 (1905)
3. R.H. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 1. (Addison-Wesley, Reading, 1964)
4. J.R. Friedman, V. Patel, W. Chen, S.K. Tolpygo, J.E. Lukens, Quantum superposition of distinct macroscopic states. Nature **406**, 43–45 (2000)
5. S. Hawking, *A Brief History of Time* (Bantam Books, New York, 1988)
6. V. Kose, F. Melchert, *Quantenmaβe in der elektrischen Meβtechnik* (VCH Verlag, Weinheim, 1991)
7. W. Nawrocki, *Introduction to Quantum Metrology (in Polish)* (Publishing House of Poznan University of Technology, Poznań, 2007)
8. I.W Savelev, *Lectures on Physics*, vol. 3. (Polish edition, PWN, 2002)

Chapter 2
Measures, Standards and Systems of Units

Abstract In this chapter we discuss the history of measurements and standards across centuries, presenting the development of standards of units of length, time, temperature and electrical quantities from historical to contemporary units. Measurements of temperature and temperature scales, to which this book does not devote a separate chapter, are discussed at some length here. We also present the development of systems of measurement, from an early Chinese system consisting of units of length, volume and mass, through the first international systems to the SI system. The SI system is defined along with its seven base units: meter, kilogram, second, ampere, kelvin, candela and mole. We refer to the Meter Convention as the starting point of a wide international cooperation aimed at the unification of measurement units and the preservation of quality of standards. We discuss the organization of the international and national measurement services, and the role of the International Bureau of Weights and Measures (BIPM) and the national metrological institutes (NMIs).

2.1 History of Systems of Measurement

Measurement is based on a comparison of the actual state of the measured quantity with another state of this quantity, the reference state, which is reproduced by a standard—see Fig. 1.1.

People have been carrying out measurements and using them since the beginning of civilization. Three physical quantities are of most importance for human existence: length, mass and time. The oldest known evidence of measurement dates back to times 8000–10 000 years ago. Also since very long people have tried to combine units of measurement in a system. A complete system of measurement, based on the dimensions of the bamboo rod and comprising units of length, volume and mass, was introduced by the Emperor Huang Ti in China around 2700 BC. In this system of measurement the unit of length was the distance between two neighboring nodes in a mature bamboo rod, the unit of volume was the space inside

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_2

the bamboo rod between two such nodes, and the unit of mass was represented by the mass of 1200 grains of rice (the approximate amount contained in the unit volume).

A fourth physical quantity of importance for knowledge and production, the temperature, has only been measured since less than 400 years.

Note that in the present-day civilization measurements are necessary not only in trade, as they were at the beginning of civilization, but also play an essential role in the control of the majority of technological processes, such as manufacturing, processing or farming, in the experimental sciences, and in many commonly used devices. For example, an average car has about 100 sensors for the measurement of the temperature and pressure, the levels and flow of liquids, and other quantities. The progress in physics led to the creation of new, quantum standards and instruments of research, which in turn contributed to further scientific discoveries.

The first international system of measurement was created by the Romans, which spread it in the countries they conquered. The system included a unit of length, the Roman foot (295 mm), and the Roman mile, equal to 5000 feet. The Roman measures of length reached as far as the Iberian Peninsula, the North Africa, the British Islands and the Balkans, and might have spread also to the territory of the present-day Poland. Note that the contemporary British foot (equal to 304.8 mm) is longer than the Roman foot, which indicates anthropological changes.

Trade in commodities measured by weight and length involved a substantial development of measures, and standards in particular. In the eighteenth and nineteenth centuries a huge increase in the mobility of the population caused an increased need for replacement of local standards of measurement, such as the cubit, foot, pound or the Russian pood, for centuries used independently in different cities and regions, by standards that would be accepted internationally or at least in a larger area.

The systematization of units of different physical quantities in a system of measurement seems obvious today, but it was in fact a very long process. The currently used International System of Units (SI) was initiated by the deposition of standards of the meter and the kilogram in the Archives of the French Republic in Paris on June 22, 1799. Both standards were made of platinum. The adopted standard for the length of one meter represents 1/10 000 000 of the segment of the Earth's meridian passing through Paris from the North Pole to the equator. The prototype of the kilogram is the mass of 1/1000 m^3 of pure water in its state of maximum density.

In 1832 Carl Gauss proposed a coherent system of measurement based on units of length, mass and time, from which units of magnetic and electrical quantities were derived. The unit of time, the second, was defined as a part of the solar day. In the Gaussian system of measurement the unit of length was the millimeter and the unit of mass the gram. In the following years Gauss and Weber incorporated also units of electrical quantities into the proposed system of measurement. The principles of a coherent system of measurement consisting of *base units* and *derived units* were developed by James Clerk Maxwell and William Thomson under the auspices of the British Association for the Advancement of Science (BAAS). Each derived unit is derived from the base units and therefore depends on them. The base units play a crucial role in the system of measures. The accuracy of the base units determines the accuracy of the derived units. In 1874 the BAAS introduced the CGS system, a coherent system of measurement developed in accordance with these principles, with units of three mechanical quantities, length, mass and time, adopted as the base units. The adopted base units were the centimeter, the gram and the second. Earlier attempts of systematization of units of measurement had been carried out independently in France, England and Prussia.

On the occasion of meetings at the world expositions in Paris (1855 and 1867) and London (1862) a number of bodies put forward resolutions to implement an international system of units. The French government proposed to set up an international commission that would work on the development on an international metric system of measurement. The commission was created, and representatives of thirty countries participated in its works. These finally led to the signing of the international Meter Convention by seventeen countries on May 20, 1875. The Meter Convention adopted the *meter* as the unit of length and the *kilogram* as the unit of mass, and set up the International Bureau of Weights and Measures (French: *Bureau international des poids et mesures*, or BIPM), which was given custody of the international prototype of the meter and the international prototype of the kilogram (IPK). The current tasks of the International Bureau of Weights and Measures include:

- The establishment of basic standards and scales for the measurement of the most important physical quantities, and preservation of the international prototypes;
- Comparisons of national and international standards;
- Coordination of the measurement techniques relevant to calibration;
- Performance and coordination of measurements of fundamental physical constants.

On the national scale, state metrological institutions known as National Metrological Institutes (NMIs) have the same function as the BIPM internationally. The best-known and most outstanding NMIs include the National Institute of Standards and Technology (USA), Physikalisch-Technische Bundesanstalt (Germany), and the National Physical Laboratory (UK). As an intergovernmental organization, the Convention of the Meter is composed of fifty-six member states (as for 2014) and forty-one associated states (including Belarus, Lithuania and Ukraine).

Built around three base units, the meter, the kilogram and the second, the first international system of units is known as the MKS (meter-kilogram-second) system. It is a set of units rather than a system, since it does not define the links between the units, as shown schematically in Fig. 2.1.

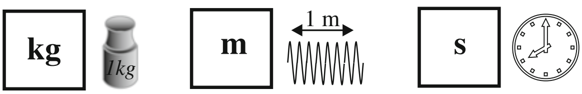


Fig. 2.1 Base units of the MKS system, the first international system of units

An arbitrary unit of length, the meter, was established following the French experience, and defined as the distance between two lines on a bar made of an alloy of 90 % platinum and 10 % iridium (Pt90Ir10), known as the international prototype of the meter. The adopted unit of mass, the kilogram, was defined as the mass of 1/1000 m^3 of water at its point of maximum density. Also the international prototype of the kilogram (IPK) was made of the Pt90Ir10 alloy. The MKS system uses an astronomical definition of the time unit, the second, defined as 1/86 400 of a mean solar day. With the exception of the units of time and angular measure, the decimal system is applied to the units in the MKS system and in all subsequent international systems of measurement.

Thirty prototypes of the meter were commissioned by the BIPM and fabricated in years 1878–1889. Comparable measurements of temperature were necessary for studying and comparing these prototypes, which involved the need for a well-defined temperature scale. The temperature was already known to affect the parameters of the standards, which therefore were kept in controlled temperature conditions. It should be noted that temperature was considered an auxiliary physical quantity at that time. In 1889 the first General Conference on Weights and Measures (CGPM, from the French *Conférence Générale des Poids et Mesures*) approved the meter and kilogram as physical standards. The original international prototype of the meter in the form of a Pt90Ir10 bar, sanctioned by the first CGPM in 1889, is still kept at the International Bureau of Weights and Measures under conditions specified in 1889. Of even greater importance, the prototype of the kilogram adopted in 1889 is still used as the international standard of the unit of mass. The first CGPM adopted also a temperature scale known as the hydrogen scale and realized with a constant-volume hydrogen gas thermometer. The hydrogen gas thermometer was calibrated at two fixed points, at 0° (the temperature of melting ice) and 100° (the temperature of boiling water). In 1946 the International Committee for Weights and Measures (CIPM, from the French *Comité international des poids et mesures*) recommended another system of units, known as the MKSA system, built around four base units, with the ampere added to the meter, kilogram and second, as illustrated in Fig. 2.2.

The MKSA system with four base units has never been formally adopted by the CGPM. In 1954 the 10th CGPM adopted a later version of the MKSA system, with six base units, the number of which had increased by the addition of kelvin and candela.

2.2 The International System of Units (SI)

The system of measurement in force today is the International System of Units, known as the SI system (from the French *Système international d'unités*). The SI system was adopted by the 11th General Conference on Weights and Measures in 1960 as an extension of the MKSA system used at that time. At the beginning the system included six base units (valid since 1954), two supplementary units (radian

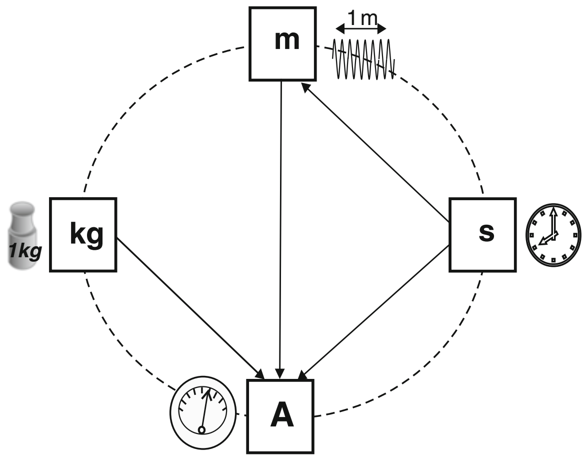


Fig. 2.2 Base units of the MKSA international system of measurement of 1946 and their interrelations

and steradian), and derived units. In 1971 the 14th CGPM increased the number of base units of the SI system to seven by adding the mole as the unit of amount of substance. Since then the base units of the SI system are: the meter, the kilogram, the second, the ampere, the kelvin, the candela, and the mole.

Base units play a crucial role in the system of measurement [1]. The accuracy with which the derived units are represented depends on the accuracy of representation of the base units. The base units provide a basis sufficient for the definition of the derived units of geometrical, mechanical, electrical, thermal, magnetic, optical and acoustic quantities as well as the units of ionizing radiation (the twenty-two named derived units with their names and symbols are listed in [1]. Metrological institutions strive to represent the base units with a minimal uncertainty, which is an understandable, but costly endeavor. Of the seven base units in the present version of the SI system only three are fully independent: the kilogram, the second, and the kelvin. The definitions of the other four, the meter, the ampere, the candela and the mole, involve the kilogram and the second (see Fig. 2.3).

An advantage of a system of measurement is the interrelation of the physical quantities. The unit of a physical quantity can be represented and defined by the units of other quantities. For example, the ampere, the unit of electric current, can be defined by measurement of force and length. The only base unit of the SI system not linked to any other base unit is the kelvin.

As stated in BIPM's SI Brochure: *Derived units are defined as products of powers of the base units. […] The base and derived coherent units of the SI form a coherent set, designated the set of coherent SI units* [1]. The SI system is coherent, which sometimes leads to its misinterpretation as a "comprehensive system". The coherence of the derived units in the SI system means that the derived units are defined by the base units multiplied or divided with no factor other than one. For example, the derived units pascal and joule are defined as follows:

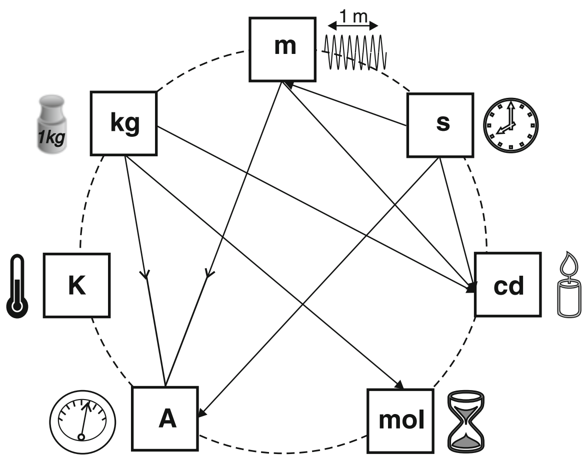


Fig. 2.3 Base units of the SI system and their interrelations

$$1\,\mathrm{J} = (1\,\mathrm{m})^2 \times 1\,\mathrm{kg} \times (1\,\mathrm{s})^{-2}$$
$$1\,\mathrm{Pa} = 1\,\mathrm{kg} \times (1\,\mathrm{m})^{-1} \times (1\,\mathrm{s})^{-2}.$$

Although the SI system of units was adopted in 1960, the definitions of the base units were adopted or revised before or after that date by the General Conferences on Weights and Measures. Below we provide definitions of the base units of the International System of Units.

The metre is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.

This definition was adopted by the 17th CGPM in 1983. It implies that the speed of light in vacuum is exactly 299 792 458 m/s: $c_0 = 299\,792\,458$ m/s. The metre is realized on the international scale with an uncertainty of 10^{-12} [1].

The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram.

Adopted by the 3rd CGPM in 1901, this definition of the kilogram implies that the mass of the international prototype of the kilogram is always one kilogram exactly, $m(K) = 1$ kg. However, due to the inevitable accumulation of contaminants on the surface the international prototype is subject to irreversible surface contamination that is close to 1 μg per year in mass. The uncertainty of realization of the kilogram on the international scale is 2×10^{-9} [1].

The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom. This definition refers to a Cs atom at rest at a temperature of 0 K.

The above definition was adopted by the 13th CGPM in 1967. It implies that the hyperfine splitting in the ground state of the cesium 133 atom is of exactly

9 192 631 770 Hz: ν(hfs Cs) = 9 192 631 770 Hz. The uncertainty of realization of the second on the international scale is better than 10^{-15} [1].

> The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} N/m of length.

Adopted by the 9th CGPM in 1948, this definition implies that the exact value of the magnetic constant μ_0, also known as the vacuum permeability or the permeability of free space, is $4\pi \times 10^{-7}$ H/m. The uncertainty of realization of the ampere on the international scale is 9×10^{-8} [1].

> The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

From this definition, adopted by the 13th CGPM in 1967, it follows that the thermodynamic temperature of the triple point of water is exactly 273.16 K: $T_{tpw} = 273.16\,\text{K}$.

The above definition of the kelvin refers to water with the isotopic composition defined exactly by the following proportions: 0.000 155 76 mol of ^{2}H per mole of ^{1}H, 0.000 379 9 mol of ^{17}O per mole of ^{16}O, and 0.002 005 2 mol of ^{18}O per mole of ^{16}O. The uncertainty of realization of the kelvin on the international scale is 3×10^{-7} [1].

> The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} Hz and that has a radiant intensity in that direction of 1/683 W per steradian.

This definition was adopted by the 15th CGPM in 1979. It implies that the spectral luminous efficacy for monochromatic radiation of a frequency of 540×10^{12} Hz is exactly 683 lm/W, K = 683 lm/W = 683 cd sr/W.

1. The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12; its symbol is "mol".
2. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.

Adopted by the 14th CGPM in 1971, this definition refers to unbound atoms of carbon 12 at rest in their ground state. The definition implies that the molar mass of carbon 12 is exactly 12 g per mole, M(^{12}C) = 12 g/mol. The uncertainty of realization of the mole on the international scale is 2×10^{-9} [1].

The International System of Units has been revised repeatedly over the 50 years following its adoption. The most significant modification was the inclusion of the mole in the set of base units. The latest, eighth edition of the SI Brochure, published in 2006 by the Intergovernmental Organization of Meter Convention, has 180 pages, including three appendices following the main chapters and specifying the parameters of the standards, the conditions for their maintenance and the detailed measurement techniques for comparison of standards. However, even after the revisions that have been made so far the SI system still requires radical changes. Its replacement with an entirely new system of units is not officially considered, though.

In 2005 the CIPM recommended the development of entirely new definitions of as many as four of the seven base units of the SI system, namely the kilogram, the ampere, the kelvin and the mole [12, 13, 14]. The new definitions should relate the base units with fundamental physical constants [19]. In accordance with the recommendation of the CIPM, the kilogram is to be related to Planck's constant h, the ampere to the elementary charge e, the kelvin to the Boltzmann constant k_B, and the mole to the Avogadro constant N_A.

Realizations of the base units of the SI system differ much in quality, determined by the uncertainty with which the units are reproduced. The best uncertainty, of the order of 10^{-16}, is offered by the atomic standard of time/frequency; the unit of luminous intensity, the candela, is realized with the worst uncertainty, of the order of 10^{-3}. In many countries even the national laboratories do not keep standards of all the base units of the SI. For example, the Polish Central Office of Measures (*Główny Urząd Miar*, or GUM) does not have standards of the ampere and the mole. However, the GUM preserves standards of other important units: the volt, the ohm, the metre and the second at a good international level.

It should be stressed that each of the above-discussed definitions of the base units refers to a specific physical constant, the value of which is considered to be both precisely known and constant in time. The conventionality of this assumption is particularly evident in the definition of the kilogram. However, also the values of the other physical constants in these seven definitions are only assumed to be exact. Their stability in time is unknown. The time invariance of the speed of light in vacuum, for instance, was questioned directly by João Magueijo, professor of theoretical physics at the Imperial College in London. Professor Magueijo believes that at the beginning of the universe the speed of light was higher than it is today [9].

We are awaiting a new system of units based on fundamental physical constants (see Chap. 3). The development of such a system, using fundamental physical constants and quantum standards, seems to be a foregone conclusion [8].

2.3 Measurements and Standards of Length

Although in everyday life we have to do with measurements of distance or length ranging from a fraction of a millimeter to hundreds of kilometers, in general the range of measured lengths is much broader. According to Richard Feynman, it spans from 10^{-15} m (the radius of the hydrogen nucleus, the smallest of atomic nuclei) to 10^{27} m (the estimated diameter of the universe, to current knowledge) [4].

First measurements of distance and length were performed with natural standards, such as the foot, which was the length of a human foot, the cubit, which was the length of the arm from the tip of the finger to the elbow, the inch, representing the width of a finger, or the stride. Larger units of measurement included the bow or crossbow shot, the verst, which was the maximal distance from which voice could be heard, and even larger units such as the distance covered in 1 day on foot or horseback. The latter standards were even more subjective than the natural standards based on parts of human body. Longer distances were expressed in multiples of these standards. A unit of length smaller than an inch was the barleycorn, or the length of a grain of barley, and still smaller, the poppy seed, equal to the diameter of a seed of poppy. Sailors, familiar with astronomy and traveling long distances, used the nautical mile, representing 1 min of arc of the Earth's circumference: 40 000 km/360º/60′ = 1852 m.

The obvious advantage of the natural standards of units of length such as the inch or cubit was the permanent availability of the standard, which was in everyone's possession. A major drawback, however, was that the standards varied widely with individual. If the unit of length was represented by the length of the forearm or the foot, the question arose: whose forearm, whose foot were these? In ancient Egypt, the cubit was the length of the pharaoh's forearm, and in the medieval empire of Charlemagne (742–814) the royal foot served as the unit of length. The spirit of democracy in the setting of measurement standards can be traced in the definition, of 1584, of the length of a measuring rod (16 feet long) in the German town of Oppenheim (Fig. 2.4):

> Let sixteen people, tall and short, at random as they come out of the church, put their foot one after another. This length is and should be the length of a reliable measuring rod.

In France dozens of different standards of the foot and cubit were used at the end of the eighteenth century. In Paris the standard of the unit of length was the royal foot, *pied du roi*, which represented 32.484 cm and was subdivided into smaller units: 1 foot = 12 inches = 144 lines = 1728 points. The large number of measurement standards was a major hindrance to the development of trade.

At the end of the seventeenth and in the eighteenth century the adoption of a universal unit of length based on the laws of nature was discussed in France, England and the United States. Two standards of the unit of length were proposed and considered:

- A standard based on the size of the Earth, specifically its circumference or radius. The measurement of the circumference of the Earth was first reported by the Greek Erastosthenes in his work *On the Measurement of the Earth* of 230 BC);
- The length of a pendulum with a 1-s period of free oscillation.

The constancy of the period T of free oscillation of a pendulum of a fixed length l was demonstrated in the studies of Galileo (1583) and the Dutch scientist Christiaan Huygens (1656) [20]. Both have tried to exploit this finding to build an accurate clock, but only Huygens succeeded. However, later studies (Jean Richter 1675) showed that the period of oscillation of a pendulum [20] slightly depended on the acceleration of gravity g in the place of measurement. The value of the acceleration of gravity g on the surface of the Earth is a function of the distance R_c from

Fig. 2.4 Setting of the length of a measuring rod in Oppenheim (Germany, 16th century); 1 rod was equal to 16 feet

the Earth's center of gravity. In practice, the value of g depends on the latitude on which the measurements are performed and the elevation above sea level. The results obtained by Richter significantly reduced the attractiveness of the pendulum as a standard of the unit of length. The length of a pendulum with a 1-s period of oscillation is 248.562 mm, according to the formula:

$$l = \frac{g}{4\pi^2} T^2, \tag{2.1}$$

where l is the length of the pendulum, T its period of free oscillation, and g the acceleration of gravity (e.g., $g = 9.81225$ m/s^2 in Poznan, Poland, on the latitude of 52° and 70 m above sea level).

In May 1790 the National Constituent Assembly of France passed a resolution on the development of units of measurement based on the laws of nature. The units were to be developed by the Academy of Sciences. Less than a year later, in March 1791, a committee of eminent French scholars, including, among others, Lagrange, Lavoisier and Laplace, proposed a standard of the unit of length, the meter, as 1/(10 000 000) of the segment of the Earth's meridian passing through Paris from the North Pole to the equator. The proposition was followed by precise measurements of the meridian. The artifact standard of the meter defined in this way was a platinum bar of rectangular cross section; the unit of length was realized by the length of the bar, i.e. the distance between its end surfaces. This end standard of the meter was made in 1799 and deposited in the Archives of the French Republic.

During its conference in Berlin in 1867 the International Association of Geodesy (French: *Association Géodésique Internationale*) put forward a proposal to introduce a uniform unit of length in Europe. The proposal was supported by the Russian Academy of Sciences in Saint Petersburg, the Academy of Sciences of France and the French Bureau of Longitude (French: *Bureau des Longitudes*). Consequently, the French government proposed to establish an international commission for the development of an international metric system of measurement. This resulted in the international Meter Convention, which was signed by seventeen countries in 1875. For practical reasons, the meridian standard was abandoned in favor of an arbitrary standard of the meter based on the French experience. According to the definition approved by the 1st General Conference on Weights and Measures in 1889:

> One metre is the distance between the middles of the width of the main lines in the zone delimited by the directional lines on the international prototype at the temperature of 0 °C, supported at points at a distance of $0.22L$ from the ends, where L is the length of the prototype.

The physical prototype of the meter was a standard bar made of an alloy of 90 % platinum and 10 % iridium. The above definition of the meter was valid 70 years, until the arbitrary standard was replaced by an atomic standard. By definition, the temperature of the platinum-iridium standard bar was 0 °C. Such temperature can be easily obtained and precisely maintained by using a mixture of ice and water.

It is noteworthy that in 1887 the American physicist Albert Michelson (born in Strzelno, Poland) proposed the use of optical interferometers for precision measurements of length. Michelson used an interferometer for measuring the international standard of the meter in the BIPM. Successive interferometer measurements of this standard confirmed very good measurement resolution and repeatability. For his contribution to the development of metrology Michelson was awarded the Nobel Prize in 1907.

Units of length derived from natural standards, but based on the metric system are still used in the Anglo-Saxon countries and the United States. The latest version of the definition of the yard was adopted by the British Parliament in the Weights and Measures Act of 1963. According to this law the yard is equal to 914.4 mm. The contemporary English system of measurement comprises the following units of length:

1 inch = 25.4 mm;
1 foot = 12 inches = 304.8 mm;
1 yard = 3 feet = 914.4 mm;
1 furlong = 220 yards = 201.168 m;
1 mile = 8 furlongs = 1609.34 m.

These units are still used along with the metric system of units.

In the mid-twentieth century the meter was redefined in terms of a specific wavelength of light. The constancy of optical wavelengths had been already suggested by Newton, and later confirmed by experiments that included Michelson interferometer measurements. The best source of light available at the time of creating the new standard of the meter was a krypton 86 gas-discharge lamp, emitting orange light with a wavelength of 606 nm. The 11th CGPM in 1960, which adopted the SI system, also redefined the unit of length:

> The metre is equal to 1 650763.73 wavelengths in vacuum of light corresponding to the transition between the levels 2p and 5d of the krypton 86 atom.

On the basis of the result of high-accuracy measurements of the speed of light in vacuum, $c = 299\ 792\ 458$ m/s, and the *conventional* assumption that the value of the speed of light is a *precisely known* physical constant, the meter was redefined again in 1983, when the 17th CGPM adopted the following new definition:

> The metre is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.

This is the current definition of the meter as a base unit of the SI system. In contrast to the definitions of the other six base units of the SI system, it seems that the definition of the unit of length will continue to apply for many years to come.

The scanning tunneling microscope (STM) allows precise surface studies, in particular measurements of the distribution of molecules and atoms on the surface of solids (see Chap. 11). The best linear resolution obtained in STM measurements of displacement and geometrical dimensions was $\Delta a = 0.01$ Å $= 10^{-12}$ m and $\Delta b = 0.1$ Å $= 10^{-11}$ m in vertical and horizontal measurements, respectively. The STM is one of the most excellent instruments of research in solid state physics.

2.4 Measurements and Standards of Mass

The concept of mass was introduced to science by Isaac Newton. The measured values of mass range from the mass of elementary particles (such as the electron mass $m_e = 9.11 \times 10^{-31}$ kg) to the mass of celestial bodies (e.g. the Earth, with the mass $m_E = 6 \times 10^{24}$ kg). Measurements of the mass of astronomical objects are based on the laws of gravity. First attempts to determine the mass of the Earth were calculations using the previously determined average density of the Earth.

In the first centuries of commerce bulk commodities, liquid or dry, e.g. grain, were measured by volume in units such as gallons, bushels or barrels. Only in the trade in precious metals and precious stones the weighing scale was used for measurement of mass. Not surprisingly, the first weights had a relatively small mass, and their role was probably played by coins of precious metal. In the Frankish empire of Charlemagne a 1-pound weight, with a mass of 367.2 g, was made of mint alloy, of which 240 denarii were produced. Both 1-pound weights and 1.53 g silver coins were used in trade. The first Polish denarius coin of Mieszko I of

Poland appeared in the latter half of the tenth century. As established by historical studies, Mieszko's denarius had a mass of 1.53 g, which was exactly the mass of the denarius of Charlemagne.

The Meter Convention of 1875 adopted only two units of measurement: the meter as the unit of length, and the *kilogram* as the unit of mass. The mass of one kilogram is the mass of 1/1000 m^3 of water at its point of maximum density (that is, at a temperature of 4 °C). The international prototype of the kilogram (IPK) and its copies (one of them is shown in Fig. 2.5), are made of an alloy of 90 % platinum and 10 % iridium (the same composition as the prototype of the meter), and has the form of a cylinder with the diameter equal to the height, $a = D = 39$ mm. The IPK is the only arbitrary artifact standard of the seven standards of the base units of the SI system. The first General Conference on Weights and Measures in 1889 defined the role of the IPK as an artifact standard:

> This prototype shall henceforth be considered a unit of mass.

In order to resolve the question as to the use of the terms of mass or weight, the 3rd CGPM in 1901 confirmed the validity of the previous definition:

> The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram.

Fig. 2.5 The standard of the unit of mass in the Polish Central Office of Measures (GUM) in Warsaw, the official BIPM copy no. 51. The prototype is covered by two glass bells (courtesy of the GUM)

Despite all care taken in its storage, the international prototype of the kilogram has begun to "put on weight" at a rate of 1 μg per year due to the deposition of impurities on its surface. Therefore, the International Committee of Weights and Measures (CIPM) stated in 1988: *The mass of the international prototype of the kilogram is the reference mass immediately after cleaning and washing the prototype by the special method* [1]. The method includes three stages of preparation of the standard: cleaning, washing and drying. The prototype must be cleaned with a chamois leather cloth, which must be decreased, deacidified and soaked twice for 48 h in a bath of ethanol and ether. The cleaning must be carried out by hand and the chamois leather cloth applied with a pressure of ca. 10 kPa. The water used for washing the IPK should be double-distilled. The washing should be carried out with hot steam directed under pressure to the prototype from a distance of 5 mm from its surface. The drying process includes three phases: collecting water droplets with a filter paper, drying with compressed air, and passive drying. In the latter phase the IPK is left for 2 weeks in a safe with air access. Thus, the mere preparation of the IPK for comparisons takes nearly 3 weeks.

The IPK is kept at the International Bureau of Weights and Measures (BIPM) in Sèvres. Almost 100 platinum-iridium copies of the international prototype of the kilogram have been made to date. Six of these are referred to as official copies of the IPK and kept together with the IPK at the BIPM. Eight other platinum-iridium artifact standards are used by the BIPM for calibrations. The remaining eighty platinum-iridium copies of the international prototype of the kilogram have been distributed to national laboratories all over the world. Copy No. 51 is possessed by the Central Office of Measures (GUM), the national metrological laboratory of Poland.

The definition of the kilogram is important not only for the unit of mass, but also for three other base units, namely the ampere, the mole, and the candela. Thus, any uncertainty in the definition of the kilogram affects these three units as well.

The present standard of the kilogram involves a numerous of difficulties. The international prototype can be used in Sèvres only. It can be damaged or even destroyed. It is subject to surface contamination of nearly 1 μg per year in mass [5]. However, the most important disadvantage of the international prototype is a long-time drift of its mass. The international prototype of the kilogram has been used since 1889. The results of the second (1939–1953) and third (1989–1992) periodic verifications of national prototypes of the kilogram indicate a change in mass of the international prototype by 50 μg in 100 years [5]. This means a relative change of 5×10^{-8} per 100 years. The reason of this drift is unknown. Changes in the mass of platinum-iridium artifacts with respect to fundamental constants became obvious after the third verification period. Thus, the unit of mass must be redefined and a new standard must be adopted. After many proposals and discussions it seems that the prospective new definition of the kilogram will refer to one of two fundamental constants, the Planck constant h or the Avogadro constant N_A [19].

At its 10th meeting in 2007 the Consultative Committee for Mass and Related Quantities (CCM) at the CIPM discussed new definitions of the kilogram and the mole. Seven new definitions of the kilogram have been proposed, each referring to one or more physical constants.

In Poland the national standard of the unit of mass is a BIPM-certified platinum-iridium artifact standard of the kilogram purchased by the Central Office of Measures (GUM) in 1952 as the prototype of the kilogram No. 51 (on the list of BIPM). The national prototype is stored and used by the Laboratory of Mass of the GUM in Warsaw. Since its purchase the prototype No. 51 was used in two latest of three international comparisons (the first comparison was carried out in years 1899–1911). The results of the comparisons for the Polish national standard were: a mass of 1 kg + 185 μg in the second international comparison (from 1939 to 1953), and a mass of 1 kg + 227 μg in the third comparison (from 1988 to 1992); the mass of the prototype was measured with an uncertainty of 2.3 μg [5].

For years efforts have been made to relate the unit of mass, the kilogram, redefined with reference to a fundamental physical constant with other units of measurement, such as those of time, length or voltage. The required standard of the unit of mass, apart from being related to a fundamental physical constant, should have an uncertainty better than the uncertainty of the present international prototype of the kilogram, 5 parts per 10^8 (long-term stability). It seems unlikely that the kilogram will be redefined with reference to the Planck constant h by the 27th GCPM. Haste is inadvisable in the metrology of base units.

2.5 Clocks and Measurements of Time

Periods of astronomical phenomena, such as days, years or monthly cycles of the moon, have always served as a measure of time. Therefore, a common unit of time is easier to accept in an international system of measurement than a common unit of length or mass. Obelisks used as *sundials* in ancient Egypt have remained since 3500 BC. The shadow cast by the sun from the obelisk indicated the time of day, and a division of the arc followed by the shadow allowed more precise indication. The Egyptians, as well as the Maya, used a 12-h division of the day from sunrise to sunset. However, since the length of the day changes over the year (with the exception of places situated on the equator), also the hours represented changeable intervals of time. A summer hour was longer than a winter hour. Also in Europe hours of changeable length were used until the fifteenth century of the common era, approximately. Portable sundials dating back to 1500 BC are known to have been used in ancient Egypt. In Europe there were even pocket sundials in the early Middle Ages. One of the elegant garden sundials purchased for the Imperial Palace in Beijing is shown in Fig. 2.6.

It is also in Egypt that the first *water clocks* were made around 1500 BC. Water clocks were widely known also in China and probably in Babylon as well. A water clock is a vessel with an opening for water outflow in the bottom. The time is indicated by the water level falling steadily as the water flows out of the vessel. Advanced water clocks had a gear train mechanism to drive the hands and

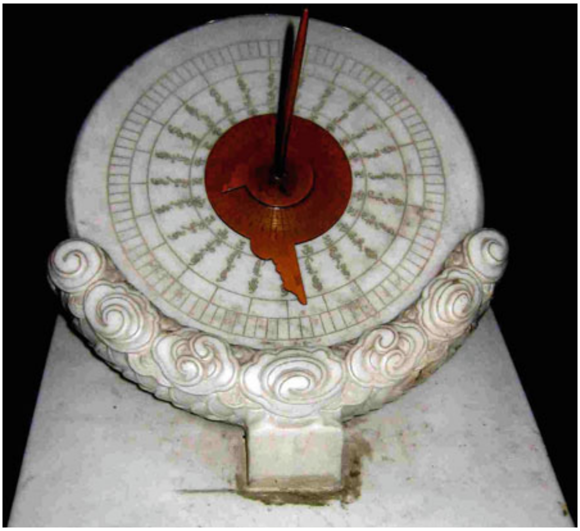


Fig. 2.6 A garden sundial with a 24-h dial-plate (Collection of the Imperial Palace Museum, Beijing)

sometimes also a mechanism that signaled the units of time elapsed. Thus, they were equivalents of the modern timer. Water clocks with metal balls falling on a tin plate were used as well. The number of falling balls indicated the time elapsed. The Greeks used both sundials and water clocks, as well as sand hourglasses (as documented since 300 BC) [7].

The time-measuring instruments used in China included sundials (Fig. 2.6), water clocks and *candle clocks*. A candle clock is a candle with a constant burn-out rate. The introduction of hours to the measurement of time is strongly related to religion. Muslims as well as Christian monks needed clocks to indicate the time of prayer. Five times a day at a fixed hour a muezzin, praising Allah, calls the faithful to prayer from a minaret, a tower on a mosque, and in the morning encourages them with the verse: *Prayer is better than sleep*. The Persian scholar Muhammad ibn Ahmad al-Biruni was the author of the first book on the measurement of time, *On the reckoning of time*, written in the first half of the eleventh century. Since the early Middle Ages bells rang canonical hours, which were seven in a day, in Christian monasteries and churches. The custom of ringing bells at the canonical hours was introduced by Pope Sabinian about 605 CE.

Units of time shorter than the hour, the minutes and seconds, result from the division of the day, but their measurement was not accurate until the invention of mechanical clocks. The accuracy of sundials was improved by the Arabs. Around the year 1000 Ibn Yunus, a court astronomer of the caliph of Egypt, wrote tables of length of the sundial shadow, specified with an accuracy of minutes. Four hundred years later independent calculations for sundial tables were performed by the great Uzbek astronomer Ulugh Beg, the author of an atlas of the sky, the ruler of Samarkand and a grandson of Tamerlane.

The first mechanical clocks appeared in Italian and French cities at the beginning of the fourteenth century. The uniform speed of rotation of the clock wheels, necessary for the measurement of equal time intervals, was ensured by a counterweight providing a constant tension on the clock mechanism. Mechanical clocks were installed on towers of town halls and churches. A breakthrough in the improvement of the accuracy of clocks was the construction of a pendulum clock, with a pendulum stabilizing the rotation of the hands, by Christiaan Huygens in 1656. The period T of oscillation of the mathematical pendulum with a free length l is determined by the relation:

$$T = 2\pi\sqrt{\frac{l}{g}}, \tag{2.2}$$

where T is the period of oscillation, l the length of the pendulum, and g the acceleration of gravity.

Measuring time with an accuracy of 1 min per day, and 10 s per day after some improvements, the pendulum clock invented by Huygens was more accurate than any other clock constructed before in the history of civilization. The pendulum clock allowed the division of the day into equal hours, minutes and seconds. It should be mentioned that before Huygens, also Galileo worked on the idea of pendulum clock and even sketched it in 1582. However, the construction of Galileo's clock was never completed. In his studies of the free fall of bodies in the gravitational field of the Earth Galileo used the pulse for measuring time. The astronomer Johannes Hevelius knew from the works of Galileo that the period of a pendulum is fixed, and used this property for measuring time in 1640, but without a clock.

The first unit of time defined in the international system of units was the second. According to the definition of 1901:

The second is 1/86 400 of the mean solar day.

This definition applied until 1956. In the meantime, fluctuations were observed in the duration of the solar day, for which reason the next definition of the second referred to the astronomical ephemeris year (determined by the position of celestial bodies with respect to the Earth). The definition adopted in 1956 stated:

The second is the fraction 1/31 556 925.9747 of the tropical year.

Substantial progress in the improvement of the accuracy of measurement of time was achieved with the development of electronic quartz clocks. The quartz clock could not have been invented without the discovery of the piezoelectric effect by Pierre Curie in 1880. In 1922 Walter G. Cady observed a high stability of the frequency of electrical oscillations in a quartz crystal, which is now the most widely used piezoelectric material. Cady found also that the selection of the crystal axis along which the quartz crystal was cut was of importance for both the frequency of oscillation and the temperature stability of this frequency. The Canadian Warren

Marrison at Bell Laboratories built a crystal oscillator, which immediately became a frequency standard, and, in 1929, constructed the first quartz clock, which was controlled by a signal from such an oscillator. The accuracy of the quartz clock was much better than that of mechanical clocks. Invented in 1969, the quartz watch became the most accurate of measuring devices of common use. The accuracy of a quartz watch exceeds by many orders of magnitude the accuracy of all other commonly used measuring instruments, such as the weighing scale, the tape measure, the thermometer or the electricity meter.

Improvement in the accuracy of the quartz clock was achieved by the improvement in the stability of the frequency of the quartz oscillator. The frequency of oscillation of quartz was soon found to depend on the temperature. Temperature fluctuations impaired the quality of the quartz oscillator as a source of reference frequency. In a quartz oscillator without temperature stabilization (XO) the relative change $\Delta f/f$ in the frequency as a function of the temperature T is of the order of 10^{-7}/°C. By taking into account and compensating the temperature dependence of $\Delta f/f$ the already good accuracy of quartz oscillators was improved still further. The temperature dependence of $\Delta f/f$ was also employed in the construction of quartz thermometers. Select parameters of quartz oscillators are specified in Table 2.1.

Both the long-term stability (in 1 year or a longer period) and the short-term stability (in 1 s, 100 s or 1 day) are of importance from the users' point of view. The long-term stability, described by the accuracy and the frequency drift per year, will be crucial for the quality of the oscillator in a frequency standard to be used in a metrological laboratory or in a telecommunication network. However, also watches need oscillators that keep right time over a long period (though, of course, the quality requirements in this case are much lower than in the case of frequency standards for laboratory use). The short-term stability is important for oscillators operating in a system or device that requires precise timing for one measurement or a series of measurements. An example of such a system is the Global Positioning System.

Table 2.1 Parameters of quartz oscillators [3, 15]

Type of oscillator	XO	TCXO	Oven OCXO	Double oven OCXO	Double oven BVA OCXO
Stability of frequency in $\tau = 1$ s	–	1×10^{-9}	–	$1–5 \times 10^{-11}$	$1–5 \times 10^{-13}$
Frequency drift per year	10^{-6}	5×10^{-7}	$2–40 \times 10^{-8}$	$1–10 \times 10^{-8}$	$1–4 \times 10^{-9}$
Frequency accuracy per year	$2–10 \times 10^{-6}$	2×10^{-6}	$3–90 \times 10^{-8}$	$1.1–11 \times 10^{-8}$	$1.7–4.8 \times 10^{-9}$
Temperature	1×10^{-7}	$5–50 \times 10^{-8}$	$5–400 \times 10^{-9}$	$2–80 \times 10^{-10}$	$5–20 \times 10^{-11}$
Coefficient (1/°C)		−55 to 85 °C	−30 to 60 °C	−30 to 60 °C	−15 to 60 °C

Quartz oscillators are divided into the following types by the method used for temperature stabilization:

- XO, quartz crystal oscillator without temperature compensation; the temperature coefficient k_t of the temperature dependence of the relative change $\Delta f/f$ in the frequency of these oscillators is $k_t > 10^{-7}$/°C;
- TCXO, temperature-compensated quartz crystal oscillator XO; temperature coefficient $k_t > 5 \times 10^{-8}$/°C;
- Oven OCXO, oven-controlled XO quartz oscillator with heating and temperature stabilization using a thermostat; temperature coefficient $k_t > 5 \times 10^{-9}$/°C;
- Double Oven OCXO, oven-controlled XO quartz crystal oscillator with heating and two-stage temperature stabilization using a thermostat; temperature coefficient $k_t > 2 \times 10^{-10}$/°C;
- Double Oven BVA OCXO, oven-controlled XO quartz crystal oscillator with heating and two-stage temperature stabilization using a thermostat and BVA ("electrodeless") quartz; temperature coefficient $k_t > 5 \times 10^{-11}$/°C.

The best accuracy is achieved by BVA quartz oscillators with two-stage temperature stabilization. In a BVA quartz crystal oscillator excitation involves additional crystal planes. The technology has been developed at the University of Besançon in France. The quality of these oscillators is believed to be comparable with the quality of rubidium atomic frequency standards, discussed in Chap. 9. Quartz oscillators of the highest quality are produced by the BVA OCXO technology by Oscilloquartz in Neuchatel, Switzerland. Resonators of the 86007–B series produced by this company have a temperature coefficient $k_t > 5 \times 10^{-11}$/°C and a typical long-term stability of 2×10^{-11}/°C per day, 5×10^{-10}/°C per month and 4×10^{-9}/°C per year [15]. In 2004 the NIST presented a low-cost miniature cesium atomic clock with an accuracy of 10^{-10}, low energy consumption and a small form factor (a volume of 1 cm^3). At present many companies offer both cesium and rubidium miniature atomic clocks with better performance compared to quartz oscillators. For example, the rubidium atomic clock offered by Quartzlock has an accuracy of 5×10^{-11}, a short-term stability of 8×10^{-12} per 100 s, and a long-term stability of 5×10^{-10} per year.

Since the verification of their high stability atomic clocks have been used as time standards and synchronized with one another. In the SI system:

> The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom.

The International Bureau of Time in Paris set the start of the *Atomic Time* on the 1st of January, 1958, 0 h, 0 min, 0 s [1]. The second from a cesium atomic clock was introduced to the international system of units in 1967.

Cesium atomic generators, providing the best uncertainty of 5×10^{-16}, are currently used as *primary time and frequency standards*. Rubidium atomic standards and optical atomic standards are recommended as secondary time and frequency standards by the CIPM Consultative Committee for Time and Frequency

(Recommendation 2 passed in 2006). The potential uncertainty in the realization of the reference frequency by optical frequency standards is estimated at 10^{-18} (see Chap. 9).

2.6 Temperature Scales

The essence of temperature, the fourth most important physical quantity besides length, mass and time, was understood surprisingly late, only in the seventeenth century. Earlier the thermal states of bodies were compared by touch or assessed on the basis of their color. The late introduction of the concept of temperature is pointed at by Terry Quinn in his extensive monograph *Temperature* [17]. According to James Clark Maxwell, *The temperature of a body is its thermal state considered with reference to its power of communicating heat to other bodies*.

Only a body of higher temperature can transfer heat, which is a form of energy, to a body of lower temperature. It should be stressed that the two bodies (or systems) under consideration must be in thermal contact. This can involve physical contact, in which case the bodies touch, but can as well be touchless, with the heat exchanged by thermal radiation. To meet the conditions of Maxwell's definition of temperature the measured object and the temperature sensor must be in thermodynamic equilibrium, if a contact method of measurement is used. In contactless measurements the temperature is determined by measuring the density of the thermal radiation emitted by the measured object.

The thermometer is believed to have been invented by one of two Italian scientists: Galileo Galilei or Santorio Santorio. In his *Comments to Galen* (*Commentaria in artem medicinalem Galeni*), published in 1612 [17], Santorio, a physician at Padua, included a description of a thermometer and the results of measurements in which it had been used. The described instrument was a gas thermometer, using the temperature dependence of the volume of air, and already furnished with a scale. Commenting upon the invention of Santorio, Galileo remarked that he himself had built a similar thermometer much earlier (around 1592). Most historians of science credit Galileo with the first design of the thermometer. Some others, however, are of the opinion that the priority should be given to Santorio, since it is documented by his treatise.

A major step forward in thermometry was made in 1641, when Ferdinand II, the Grand Duke of Tuscany, used alcohol instead of air as a thermometric liquid and put a 50-degree scale on the thermometer. Points of phase transitions of elements and substances were used as reference points of the temperature scale. In 1693 Carlo Renaldini at Padua proposed a temperature scale from the melting point of ice to the boiling point of water and divided it into twelve intervals [11]. In 1742 Anders Celsius assigned numerical values to the end points of the scale: 0° to the boiling point of water and 100° to the melting point of ice. One degree Celsius represents 1/100 of the Celsius temperature scale. Celsius' thermometer was the first measuring instrument with calibration based on the decimal number system and certainly contributed to the spread of the decimal system. The Celsius scale, currently used in most countries around the world, was created in 1850 by swapping the numerical values at the ends of the scale. Since then 0 °C is the temperature of melting ice and 100 °C the temperature of boiling water.

Another temperature scale had been proposed earlier by Daniel G. Fahrenheit, and was later named in his honor the Fahrenheit scale. After many years of work on his temperature scale Fahrenheit published its description in *Philosophical Transactions of the Royal Society* in 1724. For reasons that are partially obscure today, he chose three fixed points on his temperature scale:

- 0 °F, the temperature of a mixture of ice, water and ammonium chloride (0 °F = −17.8 °C); this was the lowest temperature obtained by both Fahrenheit and the Danish astronomer Ole Rømer, from whom Fahrenheit learned and with whom he later cooperated;
- 32 °F, the freezing point of water (0 °C);
- 96 °F, the temperature of human body (35.6 °C), or, in Fahrenheit's words, *the temperature of a healthy human* (not a precise description).

Fahrenheit investigated whether the boiling point of water could be a fixed point of a temperature scale, but concluded it could not because of the dependence of the temperature of boiling water on the atmospheric pressure. Fahrenheit spread the use of mercury as a thermometric liquid and built a number of mercury thermometers (the first one in 1714). He used these thermometers for measurements of the boiling temperature of many liquids in the range up to 600 °F, the temperature at which mercury itself boils. The Fahrenheit temperature scale is still widely used in several countries, including the United States.

In 1854 William Thomson (Lord Kelvin) proposed an absolute temperature scale that used absolute zero, 0 K = –273.15 °C, as the null point, and the freezing point of water, 273.15 K, as a second fixed point. The unit of this scale is the kelvin, equal to 1 °C.

The successive versions of the temperature scale adopted internationally in the history of thermometry were increasingly sophisticated, covered an increasingly wide range of temperature, and were more practical than the Kelvin scale. A temperature scale always has a number of fixed points defined along with the instruments for realization of the temperature between the fixed points, the methods of measurement, and the formulas describing the temperature dependence that provides the basis for mathematical interpolation between the defined points.

The 1st General Conference on Weights and Measures in 1889, which adopted the physical standards of the meter and kilogram, also approved a temperature scale known as the normal hydrogen scale of temperature. An international scale accepted by many countries was necessary for precise measurements of the standards of the units of length and mass and stabilization of the temperature of these standards. The scale was realized by a constant-volume hydrogen gas thermometer calibrated at two fixed points on the scale:

- 0°, the melting point of water;
- 100°, the boiling point of distilled water under the normal atmospheric pressure, $p_a = 760$ mm Hg (the effect of the pressure on the boiling temperature of liquids had already been known for 200 years).

The nominal temperature range of the normal hydrogen scale was narrow, as it only spanned 100°.

The next scale, adopted by the 7th General Conference on Weights and Measures, was the International Temperature Scale of 1927, known as the ITS-27. The ITS-27 was the fruit of an exchange of experience between the world's major thermometry laboratories: the NPL in UK, the Bureau of Standards in the US, PTR in Germany, and KOL in the Netherlands. The development of the new scale was coordinated by the International Bureau of Weights and Measures (BIPM). The ITS-27 is fundamentally different from the scale adopted previously. It has a much wider range, from −190 to +600 °C, and is realized by a platinum thermometer as a primary standard in the range from −190 to 1063 °C. The platinum thermometer is still used for that purpose. Other thermometers for the ITS-27 were a Pt-PtRh10 thermocouple in the range from 600 to 1063 °C and the optical pyrometer, based on Wien's law of thermal radiation, for temperatures above 1063 °C.

The ITS-27 has six fixed points, which are:

- The boiling point of oxygen at normal atmospheric pressure, −182.97 °C;
- The melting point of ice, 0 °C;
- The boiling point of distilled water at normal atmospheric pressure, 100 °C;
- The boiling point of sulfur, 444.6 °C;
- The freezing point of silver, 960.5 °C;
- The freezing point of gold, 1063 °C.

The International Temperature Scale of 1948 (ITS-48) did not differ much from the ITS-27. For the first time the unit of temperature was referred to as the degree Celsius (°C). The ITS-48 reduced the scale at the lower end by shifting the lower limit from −190 to −182.97 °C, and adjusted the fixed point representing the freezing point of silver to 960.8 °C. Planck's law of radiation replaced Wien's law as the equation for the standard optical pyrometer.

In contrast, the International Practical Temperature Scale of 1968 (ITPS-68) brought a number of changes, including a distinction between the thermodynamic temperature, determined on the basis of fundamental laws of physics, and the temperature measured with reference to a thermometric medium using its empirical parameters. The kelvin was chosen as the unit of thermodynamic temperature.

> The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

The degree Celsius remained an equivalent unit of temperature (1 °C = 1 K). In the IPTS-68 the range of the scale was extended by shifting its lower end down to 13.81 K. A number of new fixed points were added on the IPTS-68 scale, which increased their total number to thirteen. The most important of them are the

Fig. 2.7 The glass cell for the realization of the triple point of water (TPW)

above-mentioned triple point of water, 273.16 K, and the triple point of hydrogen in equilibrium, 13.81 K. A device of key importance for the IPTS-68 scale was a glass cell filled with distilled water (half a liter, approximately), used for the realization of the triple point of water (Fig. 2.7).

The current temperature scale is the International Temperature Scale of 1990 (ITS-90) [1]. It was created with the intention to provide a significantly better approximation to the thermodynamic temperature than that offered by the previous international temperature scales. This objective has been achieved. The ITS-90 covers the range from 0.65 K to the highest temperature that can be measured by a monochromatic pyrometer. The scale has seventeen defining fixed points, specified in Table 2.2.

The fixed points refer to the normal atmospheric pressure, $p_a = 101\ 325$ Pa, unless specified otherwise. In its upper part the ITS-90 has three equivalent defining fixed points: the freezing point of silver (961.78 °C), the freezing point of gold (1064.18 °C) and the freezing point of copper (1084.62 °C). The following standard interpolating thermometers are used for the realization of the ITS-90:

- In the range from 0.65 to 24 K, a helium gas thermometer or a thermometer using the dependence of the boiling temperature of helium on its saturated vapor pressure;
- In the range from 13.8033 to 1234.93 K, a platinum thermometer;
- Above 1234.93 K, an optical pyrometer using monochromatic radiation and based on the Plank law.

A sensation in the ITS-90 was the removal of the already classic boiling point of water from the list of defining fixed points. Its elimination was motivated by the finding that the thermodynamic temperature of the boiling point of water under normal atmospheric pressure was not 100 °C, but 99.975 °C.

The current definition of the kelvin is unsatisfactory. After many years of using glass cells for the realization of the triple point of water (TPW) the thermodynamic temperature of the TPW was found to change. Variations in the thermodynamic temperature of the TPW affect directly to the unit of temperature. In the group of national standards these are random fluctuations by ±50 μK, which corresponds to an uncertainty of $\pm 1.8 \times 10^{-7}$, and, additionally, a temperature drift of about 4 μK

Table 2.2 The defining fixed points of the International Temperature Scale of 1990

Equilibrium state	T (K)	T (°C)
Boiling point of 4He as a function of pressure	3–5	−270 to −268
Triple point of hydrogen e-2H in equilibrium	13.8033	−259.3467
Boiling point of hydrogen e-2H or 4He as a function of pressure	≈17	About −256
Boiling point of hydrogen e-2H as a function of pressure	≈20.3	About −252.85
Triple point of neon	24.5561	−248.5939
Triple point of oxygen	54.3584	−218.7916
Triple point of argon	83.8058	−189.3442
Triple point of mercury	234.156	−38.8344
Triple point of water	273.16	0.01
Melting point of gallium	302.9146	29.7646
Freezing point of indium	429.7485	156.5985
Freezing point of tin	505.078	231.928
Freezing point of zinc	692.677	419.527
Freezing point of aluminum	933.473	660.323
Freezing point of silver	1234.93	961.78
Freezing point of gold	1337.33	1064.18
Freezing point of copper	1357.77	1084.62

per year [10]. Obviously, changes are even larger in standards of lower rank. The underlying causes of the variations of the thermodynamic temperature of the TPW include slow dissolution of the glass, air pollution and residual air in the water. Dissolution of these substances in water changes its isotopic composition.

As a temporary solution to the problem of water composition, one specific type of water was recommended for use in all the laboratories. Referred to as Vienna Standard Mean Ocean Water (VSMOW), it has a composition defined at a conference in Vienna. However, we still have to do with a base unit of the SI system defined with reference to properties of a chosen substance (water, in this case).

In the meantime, new definitions of the kelvin are being proposed and discussed. Also the unit of temperature is taken into consideration in the attempt to develop a system of measurement based on a set of fundamental physical constants (discussed in Chap. 3). The considered redefinition of the kelvin will relate it to the Boltzmann constant k_B or the molar gas constant $R(R = 8.314472\,\mathrm{J\,mol^{-1}\,K^{-1}})$. The proposed new definitions of the kelvin have been published by the Task Group on the SI of the Consultative Committee of Thermometry at the CIPM [12]. One of the possible new definitions, proposed by the Task Group along with three others, is:

> The kelvin is the change of thermodynamic temperature T that results in a change of thermal energy k_BT by exactly 1.380 65X × 10^{-23} J, where k_B is the Boltzmann constant.

The understanding of the concept of temperature is neither general nor full. The definition of temperature refers to bodies that are themselves in thermodynamic equilibrium. It is assumed that not only the measured body or system is in thermodynamic equilibrium with the thermometer, but also the components of the body or system are in thermodynamic equilibrium with each other during the measurement of temperature. The latter assumption cannot be fulfilled in measurements of very high temperatures, such as plasma temperature, or very low temperatures, below 1 K. At such temperatures the energy of the electrons in the system may be different from the energy of the ions. If the equilibrium condition is not satisfied, controversial results (which does not mean here questionable results) will be obtained in experiments. In measurements of a silver sample cooled to the lowest possible temperature at Helsinki University of Technology the determined temperature of the atomic nuclei of silver proved to be *a negative absolute temperature* [6]; this result, however, concerns a system in the state of internal thermodynamic non-equilibrium. The temperature of the electrons in the silver sample was positive in this measurement.

It is noteworthy that in Poland the lowest temperature, of the order of 10^{-7} K, was achieved by cooling atoms of rubidium ^{87}Rb. The obtained temperature was low enough for a Bose-Einstein condensate to form. The experiment was performed by a team led by W. Gawlik at the University of Torun in 2007. The lowest temperature of an object much heavier than single atoms was attained at the University of Bayreuth, Germany, by cooling a 17 kg copper block to 12 μK [16] in an experiment prepared by a team of physicists led by F. Pobell.

2.7 Standards of Electrical Quantities

The history of measurements of electrical quantities is much shorter than the history of measurements of length, mass and time. Quantitative studies of electrical phenomena have only been conducted since the mid-eighteenth century, at first with electroscopes, which at that time were not yet furnished with a scale. At the end of the eighteenth century Charles Coulomb measured the strength of interaction between electric charges by using a precision torsion balance; these studies provided the basis for the formulation of Coulomb's law in 1785. The foundations of electrodynamics were laid by André-Marie Ampère's measurements and theoretical studies of the interaction between conducting wires carrying electric current in 1820. Note that the laws discovered by Ampère provide the basis for the definition of the ampere, the unit of electric current in the present SI system of units. Also the famous law formulated in 1826 by Georg Simon Ohm, $V = RI$, required many studies, specifically measurements of electrical resistance. Ohm determined the relative values of electrical resistance by measuring the length of the wire used as a resistor in the circuit.

In his paper *Die erdmagnetische Kraft auf ein absolutes Maß zurűckgefűhrt* of 1832 Carl Gauss proposed a system of measurement comprising units of length, mass and time, from which units of electrical and magnetic quantities were derived. Wilhelm Weber developed Gauss' concept by proposing two systems, one with magnetic and the other with electrodynamic units. The magnetic system was based on the magnetic interaction between conducting wires with electric current. The basis of the electrodynamic system was the electrodynamic interaction between two current-carrying wires. The units of electrical and magnetic quantities in these systems depended on the millimeter, gram and second.

The end of the nineteenth century saw a significant increase in the use of electricity. Its applications included, for example, electric lighting and electrically driven trams and subway trains. Electricity has also become the object of trade on an increasing scale. This involved an urgent need for definition and systematization of units of electrical quantities. The systems of electrical and magnetic units proposed by Gauss and Weber were not practical enough to come into common use. Instead, arbitrary units that were easy to realize were adopted in practice.

The currently used names of most units of electrical and magnetic quantities were adopted at the first two International Electrical Congresses (IEC) held in 1881 and 1889. The IEC held in Chicago in 1893 adopted definitions of units of electric current and resistance, the names of which had already been accepted. The IEC in London in 1908 confirmed the definitions of these units and recommended their use. The international acceptance of these definitions of the units of current and electrical resistance was indicated by their names, the international ampere and the international ohm.

> The international ampere of 1908 is that unvarying electric current that, flowing through an aqueous solution of silver nitrate, would cause the deposition of 1.11800 mg of silver per second.
>
> The international ohm of 1908 is the electrical resistance to direct current of a column of mercury with a mass of 14.4521 g, a fixed cross section and a length of 106.300 cm at a temperature of 0 °C.

The ampere was included in the international system of units (the MKSA system) as late as in 1948, as a fourth base unit and a sole representative of electrical units in the set of base units. The definition of the ampere in the MKSA was based on electrodynamics, referring back to the proposals by Gauss and Weber. The name of the unit, the ampere, no longer included the adjective "international":

> The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} N/m of length.

The above definition of the ampere relates an electrical quantity to a mechanical quantity, which is very desirable in the development of a system of measurement. This new definition of the ampere brought metrologists closer to the realization of standards based on the laws of physics, rather than parameters of chosen materials. The standard of the ampere as defined above cannot be reproduced accurately. Obviously, it is impossible (yes!) to strictly meet most of the conditions specified in the definition: that the conductors (wires) must be parallel, straight, infinitely long (!) and have a negligible circular cross section. The latter specification, requiring a cross section small enough to be neglected, does not mean anything from the metrological point of view. Also perfect vacuum is not available, and the quality of the vacuum necessary for the standard is not specified in the definition.

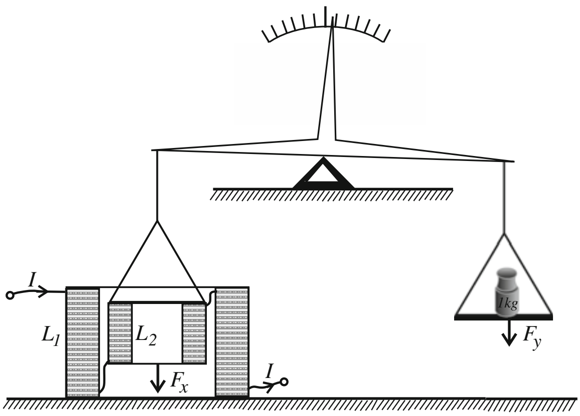


Fig. 2.8 Ampere balance for the realization of the ampere

The standard of the unit of electric current by the above definition is realized with an ampere balance, shown in Fig. 2.8.

The ampere balance, also known as the current balance, compares the force of interaction between coils conducting electric current I with the gravitational force of a weight of known mass. Unrealistic assumptions such as those regarding the infinite length of the wires have been omitted in the design of the ampere balance. The balance has two coils: a fixed coil L_1 and a movable coil L_2, connected in a series electrical circuit. Thus, the same electric current I flows through both coils. The coils interact with a force $F_x = kI^2$ (where k is a structural constant dependent on the configuration of the coils). The force F_x acts on one pan of the balance. A weight of known mass m, with a gravitational force $F_y = gm$ (where g denotes the acceleration of gravity) is put on the other pan as a counterbalance. In the equilibrium $F_x = F_y$. The ampere balance obeys the relation (2.3) below, which allows to determine the current I:

$$\begin{aligned} Ki^2 &= gm \\ I &= \sqrt{\frac{gm}{k}}. \end{aligned} \tag{2.3}$$

The realization of the unit of electric current by means of an ampere balance with a high accuracy and in a manner consistent with the SI is a costly and time-consuming task. For this reason metrology laboratories, including the Polish Central Office of Measures, do not keep a standard of ampere. Instead, they keep two groups of standards of other units of electrical quantities, Weston cells or a matrix of Josephson junctions for the realization of the volt, and a resistance standard for the realization of the ohm. In the SI system:

$$1\,\mathrm{V} = \frac{\mathrm{kg\ m^2}}{\mathrm{s^3 A}}, \quad 1\,\Omega = \frac{\mathrm{kg\ m^2}}{\mathrm{s^3 A^2}}.$$

Single electron tunneling (SET) holds some promise for the construction of a quantum standard of the unit of electric current using an electron pump system or a SET electron turnstile (see Chap. 7). The standard is based on single electron tunneling in SET junctions, in which the electrostatic energy E_C generated by a single electron (from a large population) is much larger than the energy $k_B T$ of thermal fluctuations and the energy hf of quantum fluctuations:

$$E_C = \frac{e^2}{2C_T}, \quad E_C \gg k_B T, E_C \gg hf, \tag{2.4}$$

where C_T is the capacitance of the SET junction (of the order of 10^{-16} F).

The mean current in an electron pump or a SET turnstile is described by the simple formula:

$$\bar{I} = ef, \tag{2.5}$$

where $\bar{I}$ is the mean current, and f denotes the number of carriers per second in the system, equal to the frequency of single electron tunneling.

However, electronic systems such as the SET turnstile or the electron pump only provide weak electric current with a magnitude of 10 nA at the most. Single electron tunneling occurs at very low temperatures (typically $T < 1$ K) and gives very weak signals, which require extremely careful interference suppression. A standard other than both the present one and the SET standard with its current performance and uncertainty levels is needed for practical purposes. A new definition of the ampere was proposed by the CIPM in 2005 [1]:

> The ampere is the electric current in the direction of the flow of exactly 1/(1.602 176 53 × 10^{-19}) elementary charges per second.

This definition of the ampere leads to the conclusion that the direction of the current is the direction of the flow of positive charges. This follows from the definition of the elementary charge as a positive quantity, which is the absolute value of the charge of the electron, or the value of the charge of the proton. The relation (2.6) below allows to determine the value of elementary charge e:

$$I \times T = N \times e, \tag{2.6}$$

where I is the electric current, T the time of accumulation of elementary charges, and N the number of elementary charges.

References

1. BIPM, *The International System of Units (SI) System of Units*, 8th edn (BIPM, Sèvres, 2006)
2. BIPM, Proces-Verbaux de Seances du Comite International des Poids et Mesures CIPM, vol. 57 (CIPM, 1989), pp. 104–105
3. S. Bregni, *Synchronization of digital telecommunications networks* (Wiley, Chichester, 2002)
4. R.H. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 1 (Addison-Wesley, Reading, 1964)
5. G. Girard, The third periodic verification of national propototypes of the kilogram. Metrologia **31**, 317–336 (1994)
6. P.J. Hakonen, S. Yin, O.V. Lounasmaa, Nuclear magnetism in silver at positive and negative absolute temperatures in the low nanokelvin range. Phys. Rev. Lett. **64**, 2707–2710 (1990)
7. P. Kartaschoff, *Frequency and Time* (Elsevier, New York, 1978)
8. D. Kind, *Herausforderung Metrologie* (NW-Verlag, Berlin, 2002)
9. J. Magueijo, *Faster than the Speed of Light* (Arrow Books, London, 2003)
10. A. Merlone, CH. Musacchio, A. Szmyrka-Grzebyk, New definition of the kelvin in terms of the Boltzmann constant. Elektronika **6**, 42–44 (2011)
11. L. Michalski, K. Eckersdorf, J. Kucharski, McGhee, *Temperature Measurement*, 2nd edn (Wiley, Chichester, 2000)
12. M.I. Mills, P.J. Mohr, T.J. Quinn, B.N. Taylor, E.R. Williams, Redefinition of the kilogram, ampere, kelvin and mole: a proposed approach to implementing CIPM recommendation 1. Metrologia **43**, 227–246 (2006)
13. P.J. Mohr, B.N. Taylor, D.B. Newell, CODATA recommended values of the fundamental physical constants: 2010. Rev. Mod. Phys. **84**, 1527–1605 (2012)
14. W. Nawrocki, Revising the SI: the joule to replace the kelvin as a base unit. Metrol. Measur. Syst. **13**, 171–181 (2006)
15. Oscilloquarz, www.oscilloquartz.com
16. F. Pobel, *Matter and Methods at Low Temperatures* (Springer, Heidelberg, 1992)
17. T. Quinn, *Temperature* (Academic Press, London, 1983)
18. T.J. Quinn, Base units of the Systèmeinternatinal d'unités, their accuracy, dissemination and international traceability. Metrologia **31**, 515–527 (1995)
19. M. Stock, The watt balance: determination of the Planck constant and redefinition of the kilogram. Phil. Trans. Roy. Soc. A **369**, 3936–3953 (2011)
20. A.K. Wróblewski, *History of Physics (in Polish)* (Scientific Publisher PWN, Warsaw, 2006)

Chapter 3
The New SI System of Units—The Quantum SI

Abstract This chapter is devoted to the assessment of the quality of the standards and the present SI system of measurement. In this context we present the need for improvement of the system of units, and the international efforts to create a new system. There is unanimous consent in the metrological community that a new system of measurement should be defined on the basis of fundamental physical and atomic constants, as it has been put forward for 180 years, since the first proposals by K. Gauss and J. Clerk Maxwell. We discuss the proposed redefinition of four SI units: kilogram, ampere, kelvin and mole. We present the particular importance of the new definition of the kilogram and the difficulties encountered in the development of a quantum standard of the unit of mass.

3.1 Towards the New System of Units

Classical standards of measurement units are or have been built using the specific properties of material. For example, the electrochemical saturated Weston cell, used as a reference cell of electromotive force is composed of two electrodes: mercury and mercury-cadmium submerged in the electrolyte—cadmium sulfate, $CdSO_4$. The electromotive force of Weston cells is dependent on the technology of its manufacturing (saturated cell or unsaturated cell), the temperature, level of discharge, mechanical shock, and other factors. The quality of the classical standard depends largely on the composition of the materials from which it is made of and on manufacturing technology used (type of metal alloy, heat treatment and others) and on the temperature. Classical standards require periodic comparisons with standards arranged in other metrological laboratories. As stated in Chap. 2, the operations of cleaning, washing and drying, in preparation for comparison, to the International prototype of kilogram take 3 weeks. Transportation classical standards of the volts and of the ohms change their parameters and contributes to increasing uncertainty of realized units of measure. From the above reasons, so important in metrology are standards where accuracy does not affect or material properties, or ambient

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_3

temperature fluctuations, shock or transportation. For many years, have been reported proposals for the implementation of the system units and related standards based on fundamental physical constants.

About 180 years ago, in 1832 Carl Gauss proposed a coherent measurement system of units containing the units of length, mass, and time. Magnetic properties of the Earth were used for defining the units in this system. Already in 1870 J. Clark Maxwell in Cambridge anticipated the use of atomic properties for the realization of natural standards of units. Richard Feynman observed that in principle, only two physical constants are used in quantum electrodynamics, and the majority of the other constants should be deducible from those two [6]. Both of those constants are parameters of the electron: its electric charge, e, and its rest mass, m_e. Standards of electrical units dependent on described constant $e = 1.602 \times 10^{-19}$ C already exist. These are the quantum voltage standard and the quantum standard of electrical resistance. Attempts are made to develop a quantum standard of electric current on the basis of single electron tunneling. A hypothetical development of the system of units based on fundamental physical constants was presented by Petley in 1988 [17] and Quinn in 1995 [19].

The system of units in force today, the International System of Units (SI), is over 50 years old. The SI system was adopted by the 11th CGPM conference in 1960 [1]. The possibility or even necessity of revising definitions of some of the base units of the SI has been discussed over the past 15 years. The discussions concern redefinition of the four base units: the kilogram, the ampere, the kelvin and the mole. Especially, there is a necessity to redefine the kilogram. At present only the kilogram is defined in terms of a material artifact—the international prototype of the kilogram. Taking advantage of recent achievements of physical science in metrology made possible the build of quantum standards of units which use fundamental physical constants or atomic constants. As stated in Chap. 2, the International Committee for Weights and Measures (CIPM) proposes very substantial modification of the International System of Units by redefining four SI base units based on fundamental physical constants [14].

Components of the set of the base units in the SI is discussed as well. Proposals to replace the ampere by the volt and the kelvin by the joule were published. Replacing the ampere by the volt is discussed in paragraph 3.4. The proposal of replacing the kelvin, unit of temperature, by the joule, unit of energy was presented in [16] and discussed in paragraph 3.6.

Redefinitions of the SI base units have been discussed at the 94th Meeting of the CIPM committee (2005) and during the 23rd CGPM conference in 2007. In its 12th Resolution "On the possible redefinition of certain base units of the International System of Units" the 23rd General Conference considered [1]:

- "that, for many years National Metrology Institutes as well as the International Bureau of Weights and Measures (BIPM) have made considerable efforts to advance and improve the SI by extending the frontiers of metrology so that the SI base units could be defined in terms of the invariants of nature—the fundamental physical constants,

- that, of the seven base units of the SI, only the kilogram is still defined in terms of a material artifact—the international prototype of the kilogram (2nd CGPM, 1889, 3rd CGPM, 1901) and that the definitions of the ampere, mole and candela depend on the kilogram, …,
- the many advances, made in recent years, in experiments which relate the mass of the international prototype to the Planck constant h or the Avogadro constant N_A, …,
- initiatives to determine the value of a number of relevant fundamental constants, including work to redetermine the Boltzmann constant k_B,
- that as a result of recent advances, there are significant implications for, and potential benefits from, redefinitions of the kilogram, the ampere, the kelvin and the mole …".

Following the above arguments, the 23rd General Conference recommended in the 12th resolution:

- "pursue the relevant experiments so that the International Committee can come to a view on whether it may be possible to redefine the kilogram, the ampere, the kelvin, and the mole using fixed values of the fundamental constants at the time of the 24th General Conference (2011),
- should, together with the International Committee, its Consultative Committees, and appropriate working groups, work on practical ways of realizing any new definition based on fixed values of the fundamental constants, prepare a *mise en pratique* for each of them, and consider the most appropriate way of explaining the new definitions to users …".

The new SI system will be proposed if new definitions will be ready for all four base units. The new definition of the kilogram, based on the Planck constant, h, is crucial. Therefore, in 2010 the Consultative Committee for Mass at the ICPM committee set out three requirements (confirmed in February 2013), concerning uncertainty in determination of h, which should be fulfilled before the ICPM will send an official proposal to the CGPM conference [1]. Requirements on determination of the Planck constant, h:

- At least three independent results (watt balance and XRCD) with relative uncertainty $u_r < 5 \times 10^{-8}$.
- At least one result with relative uncertainty $u_r \leq 2 \times 10^{-8}$;
- Results consistent.

Both methods of measurements of the Planck constant: the watt balance and the X-ray crystal density technique (XRCD), mentioned above, are described in Chap. 12.

However, the 24th CGPM conference, 2011, considered in its 2nd Resolution:

- "that, although the work (to redefine four base units of the SI) has progressed well, not all the requirements set out by the 23rd General Conference in 2007 have been satisfied and so the CIPM committee *is not yet ready* to make a final proposal."

- The definitions of the metre, kilogram, second, ampere, kelvin, mole and candela will be abrogated.
- The 24th General Conference “invites CODATA to continue provide adjusted values of the fundamental physical constants based on all relevant information available and to make the results known to the International Committee through its Consultative Committee for Units since these CODATA values and uncertainties will be those for the revised SI, …”

Not satisfied requirements, mentioned in the resolution of the 24th CGPM, result from the too large relative uncertainty in determination of the Planck constant, *h,* and the too less consistence of *h* values obtained by different research groups.

The next chance for adopting the new system of units will be in 2018, on the occasion of the 26th General Conferences of Weights and Measures.

3.2 Units of Measure Based on Fundamental Physical Constants

According to the 12th Resolution of the 23rd CGPM conference new definitions based on fixed values of the fundamental constants should be proposed. Now we come back to these ideas considered by Gauss, Maxwell and Feynman. Fundamental physical constants are universal and invariant. Therefore they are good references for units of system of measure. If we decided to create a system of units based on fundamental constants, the problems to be solved are: which of many physical constants should be taken for such system and how units should be redefined?

The simple system of units based on fundamental constants was proposed by Mills et al. [14]. “The International System of Units, the SI, is the system of units scaled so that:

1. Ground state hyperfine splitting transition frequency of the caesium 133 atom $\Delta\nu\ (^{133}\mathrm{Cs})_{\mathrm{hfs}}$ is 9 192 631 770 Hz,
2. Speed of light in vacuum c_0 is 299 792 458 m s^{-1},
3. Planck constant h is 6.626 0693 × 10^{-34} J/s,
4. Elementary charge e is 1.602 176 53 × 10^{-19} C,
5. Boltzmann constant k_B is 1.380 6505 × 10^{-23} J/K,
6. Avogadro constant N_A is 6.022 1415 × 10^{23} per mole,
7. Spectral luminous efficacy of monochromatic radiation of frequency 540 × 10^{12} Hz $K(\lambda_{555})$ is 683 lm/W”.

Accompanying this definition of the SI would be a list of representative units, together with a representative list of the quantities whose values could be expressed in those units”. The list of units includes the metre, the kilogram, the second, the ampere, the kelvin, the mole and the candela as well as the current 22 derived units (from the present SI). One can note that the proposed new SI [14] is rather *a set of physical* *constants*, not a system of units. In the proposed new SI the units are not divided between base units and derived units. "All units are on an equal footing" [14].

In the seven "definitions" presented above physical constants should have fixed values. The values for some fundamental constants, necessary for new definitions of the SI units, must be exactly known—it is the assumption. At present only the speed of light in vacuum, c_0, is fixed as the fundamental constant, $c_0 = 299\ 792\ 458$ m s^{-1}. Two other constants which are not fundamental, namely the spectral luminous efficacy of radiation, $K(\lambda_{555}) = 540 \times 10^{12}$ Hz, and the transition frequency of caesium 133, $\Delta\nu\ (^{133}\mathrm{Cs})_{\mathrm{hfs}} = 9\ 192\ 631\ 770$ Hz, are fixed as well. The values of four remaining fundamental constants must be fixed for the new definition of the SI units. This requirement concerns the Planck constant h, the elementary charge, e, the Boltzmann constant, k_B, and the Avogadro constant, N_A. The contains of the set of the fixed constants is not fixed so far, it can be altered. For example, the Planck constant and the elementary charge could be replaced by the Josephson constant K_J and of the von Klitzing constant R_K (3.1) in the set of fixed constants. The two latter constants appear in equations for quantum effects and are measured very precisely and accurate.

$$K_J = 2e/h, R_K = h/e^2$$

$$e = 2/(K_J \times R_K), h = 2(e/K_J) \tag{3.1}$$

The values to be fixed can be taken from the current CODATA set of recommended values (CODATA, the Committee on Data for Science and Technology). The values of fundamental physical constants (useful for the SI) presented in the latest edition of the CODATA set (2010) are listed below [15].

- $h = 6.626\ 069\ 57(29) \times 10^{-34}$ J s, the relative standard uncertainty $u_r \geq 4.4 \times 10^{-8}$,
- $e = 1.602\ 176\ 565(35) \times 10^{-19}$ C, $u_r \geq 2.2 \times 10^{-8}$,
- $k_B = 1.380\ 6488(8) \times 10^{-23}$ JK^{-1}, $u_r \geq 9.1 \times 10^{-7}$,
- $N_A = 6.022\ 141\ 29(27) \times 10^{23}$ mol^{-1}, $u_r \geq 4.\ 4 \times 10^{-8}$,
- $m_e = 9.109\ 382\ 91(40) \times 10^{-31}$ kg, $u_r \geq 4.4 \times 10^{-8}$.

A relative standard uncertainty u_r of measurements of these five fundamental constants has been improved continuously, e.g. in CODATA 2002 there were the following values: $u_r \geq 1.7 \times 10^{-7}$ for the h, $u_r \geq 8.5 \times 10^{-8}$ for the e, $u_r \geq 1.8 \times 10^{-6}$ for the k_B, $u_r \geq 1.7 \times 10^{-7}$ for the mole, and $u_r \geq 8 \times 10^{-8}$ for the rest electron mass, m_e.

Two physical constants with fixed values in the present SI, the magnetic permeability in vacuum $\mu_0 = 4\pi \times 10^{-7}$ NA^{-2} and the molar mass of carbon ^{12}C, M (^{12}C) = 12 g/mol, will be no more fixed in the new system and no more known exactly. They can be simply measured with uncertainty. At present there are more constants-candidates for new definitions of units than necessary. E.g. the kilogram can be defined using either the Planck constant, h, or the Avogadro constant, N_A, the kelvin can be defined either by the Boltzmann constant, k_B, or by the molar gas constant, R—see Fig. 3.1.

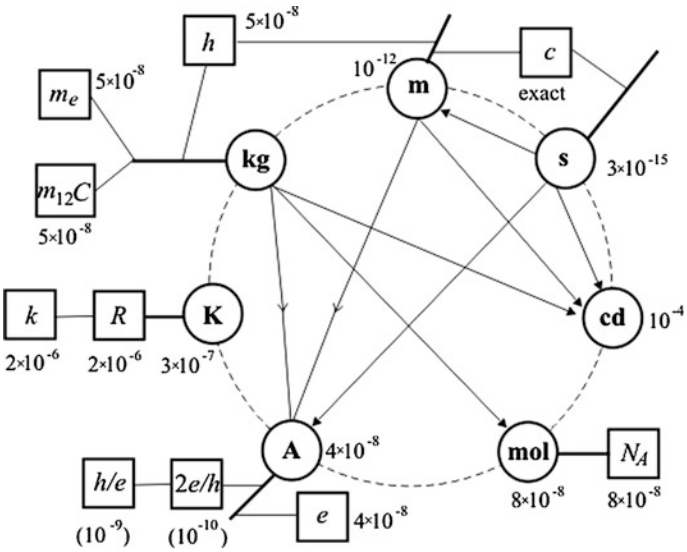


Fig. 3.1 Linking SI base units to fundamental constants [1] (values of relative standard uncertainty according the 2010 CODATA set [15])

3.3 New Definitions of the Kilogram

The kilogram is still defined in terms of a material artifact. The definition in the SI system reads as follows: “The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram”. The international prototype of the kilogram (IPK) is a cylinder with a diameter of 39 mm, it is made of an alloy of platinum 90 % and iridium 10 %. The IPK is kept at the BIPM in Sèvres. Six of the Pt-Ir copies are called official copies of the IPK and kept together with the IPK at the BIPM. Eight other Pt-Ir artifacts are used by the BIPM for calibrations.

The definition of the kilogram is important not only for the unit of mass but for three other base units, namely the ampere, the mole, and the candela as well. Thus, any uncertainty inherent in the definition of the kilogram is transferred on these three units.

There are several difficulties with the present standard of the kilogram. The IPK (standard) can be used in Sèvres only. The IPK can be damaged or even destroyed. The IPK collects contaminants from air, approaching 1 μg per year in mass [3]. However, the most important disadvantage of the international prototype is a long-time drift of its mass. The international prototype of the kilogram is in service since 1889. The second (1939–1953) and third (1989–1992) periodic verifications of national prototypes of the kilogram have shown that the mass of the international prototype changed, with respect to the ensemble copies, by –50 μg after 100 years [8]. It means a relative change of 5×10^{-8} per 100 years. The reason of this effect is unknown. The changes of the masses of Pt-Ir artifacts with respect to fundamental constants has been obvious after the third verification period. Thus, the definition and the standard of mass must be altered. After many proposals and discussions a possible new definition of the kilogram will be related to one of the two fundamental constants, namely: the Planck constant [20], h, or the Avogadro constant, N_A.

In 2007 the Consultative Committee for Mass and Related Quantities at the CIPM (10th meeting of CCM) discussed new definitions of the kilogram and the mole. There were 7 new definitions of the kilogram, each of them relates to a physical constant (to one or many). All seven presented definitions have been still discussed.

1. The kilogram is the mass of a body whose Compton frequency is 1.356392… × 10^{50} Hz exactly [5].
2. The kilogram is the mass of a body whose de Broglie-Compton frequency is equal to exactly [(299 792 458^2/(6.626 0693) × 10^{-34})] Hz [14].
3. The kilogram is the mass of a body whose equivalent energy is equal to that of a number of photons whose frequencies sum to exactly [(299 792 458^2/(66 260 693)] × 10^{41} Hz [14].
4. The kilogram, unit of mass, is such that the Planck constant is exactly h = 6.626 0693 × 10^{-34} J s [14].
5. The kilogram is (6.022 1415 × 10^{23}/0.012) times the rest mass of the ^{12}C atom in the ground state [2].
6. The kilogram is (6.022 1415 × 10^{23}/0.012) times the rest mass of a particle whose creation energy equals that of a photon whose frequency is: [0.012/(6.022 1415 × 10^{23}) × 299 792 458^2/(66 260 693 × 10^{-34})] Hz [2].
7. The kilogram is 1.097 769 24 × 10^{30} times the rest mass of the electron [9].

The (1–3) definitions use formulas of quantum mechanics and two fundamental constants: h and c_0. The (4) definition is very simple but its realization is very far from a practical standard. Physical interpretation of these (1–4) definitions is difficult. At the 10th CCM meeting (2007) E. Williams presented the following explanation of the (4) definition: “The kilogram is the mass of 6.022 1415 × 10^{26} idealized atoms, each of these atoms having the mass such that the Planck constant, the most important constant in quantum mechanics, has the specified value of 6.626 069 57 × 10^{-34} J s” [22]. This explanation is little helpful for understanding of this definition.

The Compton wavelength $\lambda_{C,e}$ and the Compton frequency $\nu_{C,e}$ of the electron, mentioned in the definitions (1, 2), are given by formulas (3.2) and (3.3).

$$\lambda_{\mathrm{C,e}} = h/m_{\mathrm{e}}c_0, \tag{3.2}$$

$$\nu_{\mathrm{C,e}} = c_0/\lambda_{\mathrm{C,e}} = c_0^2 m_e/h. \tag{3.3}$$

Should we put a mass of 1 kg instead of the electron mass into the formula (3.3), we obtain the Campton frequency of ν_C (1 kg) = $c_0^2 m/h$ ≈ 1.356 × 10^{50} Hz which appears in the first definition. Such extremely high frequency of 10^{50} Hz has no practical meaning. The (5–7) definitions are related either to the mass of the carbon atom (5, 6) and the Avogadro constant or to the mass of the electron (7). Each of the (5–7) definitions describes the kilogram as a mass *n*-times larger than the mass of a particle. A standard mass defined this way could be better understood than the mass expressed in frequency units.

The CCM (2007) considered all these seven definitions of the kilogram. Arguments for one from the (5–7) definitions are as follows. "The kilogram mass is a classical, macroscopic quantity, whereas the Compton frequency and the Planck constant describe quantum mechanical effects. How the mass of a macroscopic body (1 kg) is related to quantum mechanics is not clarified and experimentally proved" [9].

3.4 New Definitions of the Ampere, Kelvin and Mole

In the present SI: "The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre of length". The Consultative Committee for Electricity and Magnetism (CCEM) at the CIPM discussed redefinitions of the ampere in 2007 and 2009 [1]. The recommendation E1 (2007) of the CCEM, supported by the meeting of the CCEM in 2009, reads:

> The ampere is the electrical current equivalent to the flow of exactly $1/(1.602\ 176\ 53 \times 10^{-19})$ elementary charges per second.

The very similar definition was proposed by Mills et al. [14] in 2006. However the second definition of the ampere describes the flow of positive charges (protons).

> The ampere is the electric current in the direction of the flow of exactly $1/(1.602\ 176\ 53 \times 10^{-19})$ elementary charges per second.

The first definition is so far the proposal from the Consultative Committee for Electrical measurements (CCEM). The second definition reflects the opinion of the Consultative Committee for Units (I. Mills is the chairman of the CCU). The CCEM considered also the question what quantity should represent all electrical quantities in the set of base units of the SI? It was not obvious that the ampere will be still the base unit. There are two other possibilities: the volt and the ohm. It seems that either the volt or the ohm could be better base units than the ampere because volt standards and ohm standards have been used since 20 years in metrology. Assuming that the values of h and e are fixed (such proposal is generally accepted), so the two other physical constants, then Josephson constant K_J and the von Klitzing constant R_K, will be known exactly. A definition of the volt, an alternative base unit, can read [9]:

> The volt is equal to the difference between two electrical potentials within which the energy of a pair of electrons equals that of photons whose frequency is $4.835\ 978\ 79 \times 10^{14}$ Hz.

In a Josephson voltage standard the device is a matrix with n Josephson junctions. The standard's voltage V depends on a radiation frequency v, which can be measured with a relative standard uncertainty better than 10^{-12}, and the Josephson constant K_J (3.4). The Josephson voltage standards are installed in hundreds laboratories all over the world.

$$\begin{aligned} V &= n \times k \times (h/2e) \times v = n \times k \times KJ^{-1} \times v \\ &= n \times k \times \left[1/\left(483.597879 \times 10^{12}\right] \times v, \end{aligned} \tag{3.4}$$

where V—the voltage of a Josephson voltage standard, n—the number of Josephson junctions in a standard's matrix, k—integer, the step on a voltage-current characteristic of Josephson junction, v—the radiation frequency acting on Josephson junctions.

The ohm is realized by a quantum standard using the quantum Hall effect (QHE). The resistance, R_H, produced by the QHE standard, depends only on the R_K (3.5).

$$R_H = \left(h/e^2\right)/I = R_K/i, \tag{3.5}$$

where R_H—the resistance of a QHE standard, R_K—the von Klitzing constant, i—integer, the step on the resistance-magnetic inductance characteristic of the QHE sample.

The recommendation E1 (2007) of the CCEM reads "that if the concept of base units is retained then the ampere be kept as a base unit for the purposes of historical continuity and SI dimensional analysis although there is no preferential order of traceability within electrical units" [1]. Eventually the CCEM proposes to leave the ampere in the set of base units and to redefine it.

In the present SI "The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water". Proposals of a new definition of the kelvin were discussed during International Symposium on Temperature Measurements in 2007 and after that published in [7, 13]. The authors of the paper [7] are members of the Task Group for SI (TG-SI) of the Consultative Committee of Thermometry at the CIPM. The papers [7, 13] describe results of discussions on the new definition. All four new definitions of the kelvin use the Boltzmann constant, no definition is related to the molar gas constant R = 8.314 472 J/mol^{-1} K^{-1} (the latter is possible, see Fig. 3.1).

1. The kelvin is the change of thermodynamic temperature T that results in a change of thermal energy k_BT by exactly 1.380 65XX × 10^{-23} J, where k_B is the Boltzmann constant.
2. The kelvin is the change of thermodynamic temperature at which the mean translation kinetic energy of atoms in an ideal gas at equilibrium is exactly (3/2) × 1.380 65XX × 10^{-23} J.
3. The kelvin is the change of thermodynamic temperature at which particles have an average energy of exactly (1/2) × 1.38 65XX × 10^{-23} J per accessible degree of freedom.

4. The kelvin, unit of thermodynamic temperature, is such that the Boltzmann constant is exactly 1.380 65XX $\times$ 10^{-23} J/K.

The XX in the above definitions are the appropriate digits of the Boltzmann constant, according to the current CODATA set. What definition will be recommended? "… the TG-SI is recommending the explicit-constant (4) definition because it is sufficiently wide to accommodate future developments and does not favor any special primary thermometer for realizing the kelvin. Should the CCU (Consultative Committee for Units at the CIPM) decide to adopt explicit-unit definitions for the kilogram, the ampere, and the mole, then the second option of the TG-SI would be the (1) formulation for the kelvin in order to be in line with the other new definitions" [7].

However Kilinin and Kononogov [10] see that the new definition of the kelvin using the Boltzmann constant "is not advisable in view of the present-day level of accuracy and reliability of determination of the value of the Boltzmann constant".

An amount of substance (system) is especially important for chemistry. For analyses of chemical processes and reactions being made by chemists, number rations of particles are necessary. Thus the mole, the unit of amount of substance, should be defined without using a unit of the mass. The definition in the present SI reads: "The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12. When the mole is used, elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles". At the 10th meeting (2007) the CCM proposed the following new definition of the mole [3].

> The mole is the unit of amount of substance. It is equal to 6.022 1415 $\times$ 10^{23} mol^{-1} specified identical entities. The entities may be atoms, ions, molecules or other particles.

The definition of the mole proposed in the paper [14] is similar to the previous one. In the definition (2) authors described the particles as "elementary entities" instead of "identical entities" what is in the (1) definition.

> The mole is the amount of substance of a system that contains exactly 6.022 1415 $\times$ 10^{23} mol^{-1} specified elementary entities, which may be atoms, molecules, ions, electrons, other particles or specified groups of such particles [14].

According to the new definitions the amount of substance n does not refer any longer to the unit of mass. The amount of substance can be described by a formula (3.6).

$$n = (\text{number of molecules in a sample})/6.022\,1415 \times 10^{23} \tag{3.6}$$

B.P. Leonard published some comments and critical remarks concerning the definitions of the mole presented above [11]. He introduced his remarks assuming the total amount of substance n(S) and the corresponding number of the entities N (S). The number-specific amount of substance is n(S)/N(S). He has proposed "using the name entity (plural, entities) and symbol ent for this fundamental physical invariant and formally adopting it as an atomic-scale unit in use with SI…" [11].

Because the proposed new unit, entity, is reciprocally related to the Avogadro constant N_A, the amount-specific number of entities, $N_A = N(\mathrm{S})/n(\mathrm{S}) = \mathrm{ent}^{-1}$, this proposal has been not accepted by physicists. The Leonard's comment reads: "The proposals independently redefining the mole (by fixing the Avogadro number, A_N) and the kilogram (by fixing the Planck constant) while keeping the ^{12}C-based definition of an atomic mass unit, the dalton (Da), violate the fundamental compatibility condition stemming from the mole concept. With the mole and kilogram defined independently, if A_N has an exact value, the dalton should be determined by the compatibility condition, Da = $(10^{-3}/A_N)$ kg, exactly. "The correction factor, proposed in [14], for fulfilling the condition for the present definition of the dalton, Da = m_a (^{12}C)/12 "create a degree of complexity that is completely unwarranted, …", where m_a (^{12}C) is the atomic mass of carbon ^{12}C) [11].

3.5 Quantum Metrological Triangle and Pyramid

Three quantum standards of electrical quantities (voltage, resistance and current) can be used for the construction of a compact system that would allow comparison between the measured quantities, thanks to their interrelations. The so-called quantum metrological triangle, corresponding to such system and proposed by Likharev and Zorin in 1985 [12], is shown in Fig. 3.2.

The vertices of the quantum metrological triangle are physical quantities: voltage (V), electric current (I), and frequency (f), coupled by three quantum effects, which represent the sides of the triangle. These quantum effects include Josephson effect, single electron tunnelling (SET) and quantum Hall effect (QHE). Quantization of electric resistance (conductance) in QHE occurs with steps $R_K = h/e^2$. Note that according to Landauer theory of electric conductance, conductance quantization takes place also in a conductor sample of nanometre dimensions, depending on its width and thickness—see Chap. 7. In this quantization effect, quite different from QHE, the conductance step is $G_0 = 2e^2/h = 2/R_K$. The experimental setup for quantum metrological triangle standardization is described by Piquemal [18] and Scherer [21]. The main idea of the experiment was to measure the coherence of values of fundamental physical constants occurring in the Josephson effect, QHE and SET with uncertainty better than 10^{-8}. As a result of the Josephson effect on a junction, placed in an electromagnetic field with a frequency, f, arises constant voltage, V:

$$V = \frac{h}{2e} f_1 = K_J f_1, \tag{3.7}$$

where: h—Planck constant, e—electron charge, f_1—frequency of radiation acting on Josephsona junction, K_J—Josephson constant.

A Josephson voltage standards contains a matrix with many thousands of serially connected Josephson junctions. The voltage at the terminals of the matrix is:

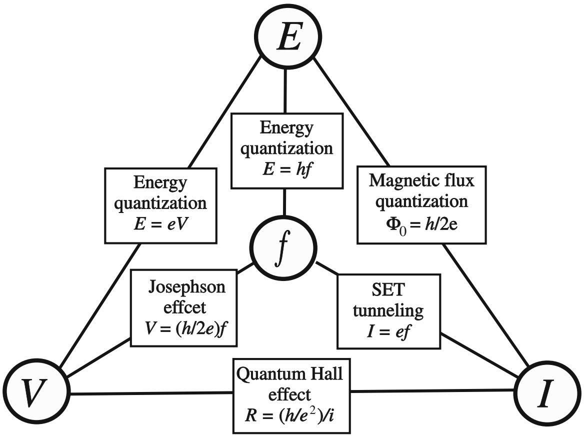


Fig. 3.2 Quantum metrological pyramid

$$E_G = 3.5\, k_B T_c,$$

$$V_J = n_1 \frac{h}{2e} f_1 = n_1 K_J f_1$$

where: n_1—number of Josephson junctions in a matrix. The Josephson effect and voltage standards with Josephson junctions are described in Chap. 4.

The quantum Hall effect (QHE) arises in the Hall sample placed at low temperature ($T < 4.2$ K) and high magnetic field ($B > 1$ T. In the sample volume is formed spontaneously area of the two-dimensional electron gas (2-DEG). The relationship between current and voltage in the Hall sample, i.e. the resistance of the sample is described by the formula:

$$R_H = \frac{V}{I} = \frac{h}{ie^2} \tag{3.8}$$

where: i—integer corresponding to the step on the characteristics of the resistance quantization. The quantum Hall effect and quantum standards of resistance are described in Chap. 6.

The third phenomenon of forming quantum metrological triangle is the tunneling of single electrons. The average value of the current flowing through the tunnel junction is proportional to the frequency f_2 of tunneling of single electrons—formula (3.9)

$$\bar{I} = Qf_2 \tag{3.9}$$

where: Q—single charges passing through the junction during tunneling; it is assumed that charges are elementary.

Sometimes it is considered tunneling of Cooper pairs (with a charge $Q_p = 2e$) via the tunnel junction formed of a superconductor, as a phenomenon equivalent to the single electron tunneling in the tunnel junction made of a conductor.

Note that in the quantum metrological triangle, in addition to realized quantities: voltage, V, current, I, and frequency, f, are only the fundamental physical constants: the Planck constant, h, and the elementary charge, e. Actually coefficients in equations of phenomena in quantum triangle are: the Josephson constant $K_J = 2e/h$, the von Klitzing constant $R_K = h/e^2$ and the electron charge, e. Frequency, forming one of the vertices of the quantum metrological triangle, is a measured with the smallest uncertainty of all physical quantities—in order of 10^{-16}. From the above description it follows that the limits in the uncertainty of voltage standard or electric current standards determine the frequency and an accuracy in the determination of fundamental constants h and e, or constants K_J and R_K.

The experimental set-up to check the quantum metrological triangle is described in paper [18]. The main objective of the experiment was to measure the coherence of the fundamental physical constants occurring in the phenomenon of Josephson effect, QHE and SET with uncertainty of not less than 10^{-8}.The relationship (3.10) was tested.

$$V_{\mathrm{J}} = R_H N I \tag{3.10}$$

Assuming that the physical constants in these phenomena are exactly equivalent to: $2e/h$, h/e^2 and e, respectively. Therefore, measurement results on the quantum metrological triangle would give 1 = 1 from the (3.10). Such an experiment would confirm the initial hypothesis as to the physical constants occurring here, as long as there would not have unlikely case that uncertainty in determining the constant offset each other. Moreover, it would determine whether the SET effect can achieve the uncertainty of 10^{-8} when the SET electronic device is connected to an external circuit. However, in a situation where the of compliance level is only 10^{-7} or less, separate measurements will not say that what phenomenon or what theory contain errors or inaccuracies?

A development of quantum metrological triangle is a system of electrical quantum unis—the quantum metrology pyramid [16], which was formed by the addition of energy unit to the triangle—Fig. 3.3.

Six edges of the pyramid corresponds to six quantum phenomena. Each of the models of quantum electrical quantities (V, I, E) is realized using a frequency standard. All quantum phenomena in this system are dependent only on two fundamental physical constants: e or h. Up to four quantization phenomena occurring in the triangle quantum (quantized resistance occurs in two phenomena) in a quantum metrological pyramid 3 other phenomena are added:

- Quantization of energy transported by electromagnetic radiation; a phenomenon anticipated by Max Planck, $E = hf$;
- Quantization of the magnetic flux, the observed, among others, in SQUID detectors (foe measurement a magnetic flux), $\Phi_0 = h/2e$;
- Energy quantization of electric charge transport process observed in many physical phenomena, $E = eV$. Therefore, in the metrological quantum pyramid is possible to control each of the four physical quantities by 3 others. The

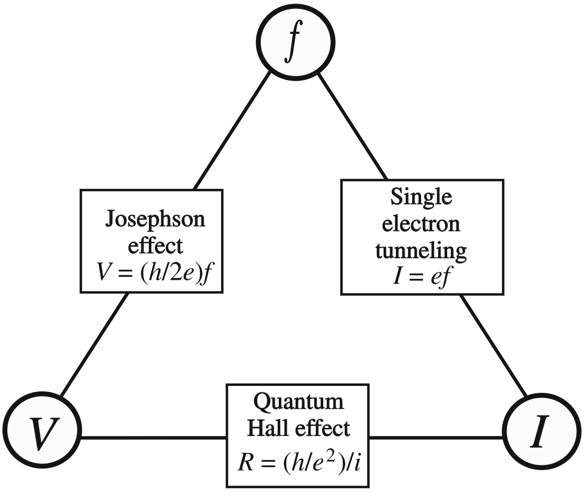


Fig. 3.3 Quantum metrological triangle

frequency is reproduced accurately and in practice the frequency is not controlled, but can be used for independent validation of standards or for measurements: of voltage, current and power.

The current SI system of units, with of seven base units (m, kg, s, A, K, Cd and mol), is not the only possible system. The following is a hypothetical system of units in which the 5 base units of measures may be defined by the fundamental physical constants (*c, e* and *h*) and frequency, *f*, and the other two (candela and mole) using the already defined other base units—Fig. 3.4. In this system the unit of temperature, kelvin, is replaced by the unit of energy, joule [18]. Numerous arguments are presented in favour of such change. The set of base units should comprise the measurement units of physical quantities, which are the most important for human life and activity, e.g. applied in trade. Therefore, among base units there are the units for length (metre), mass (kilogram), and time (second). Because energy is also the subject of common trade exchange on a large scale, the unit of energy (joule) and the accuracy of its standardization have a great effect on trading and commercial activity. Energy is perhaps the most universal physical quantity in nature. One of the formulations of the first thermodynamics principle reads: “The quantity of energy in the universe is always constant”. Different forms of energy: mechanical energy or work, thermal energy, electric energy and nuclear energy, enable mutual comparisons of the standards of mechanical quantities, thermal quantity standards, electric quantity standards and mass standards, as well as comparing them to the standard of energy unit.

$$E = h \times f; E = m \times c^2;\ E = V \times I \times t;\ E = F \times l; E = m \times v^2/2; E = k_B \times T;$$

$$Q = c_h m \times \Delta T$$

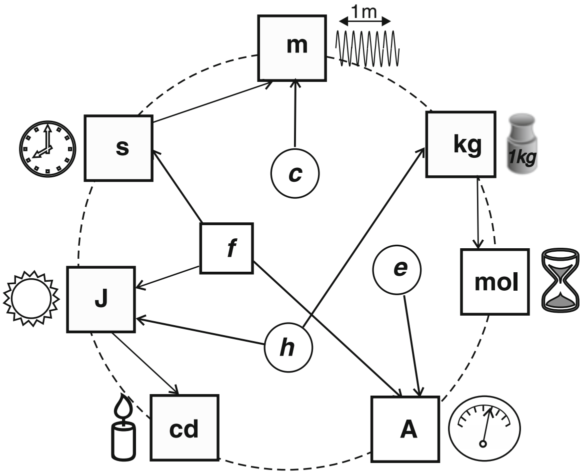


Fig. 3.4 Base units of a hypothetical system of units (with the joule replacing the kelvin) and their mutual relations

where: E—energy, Q—heat, c—speed of light in vacuum, c_h—specific heat, h—Planck's constant, k_B—Boltzmann's constant, f—frequency, F—force, I—electric current, l—length, m—mass, t—time, T—temperature, V—electric voltage, v—linear velocity.

As far as a limit resolution of measurements is concerned, we usually have to take into consideration energy changes affecting (influencing) the measuring sensor. Energy is one of four physical quantities occurring in inequalities describing quantum limits of measurement resolution, according to the Heisenberg uncertainty principle. Three others quantities are: time, length (for position) and momentum: (1.9) and (1.11).

$$\Delta x \times \Delta p \geq \hbar/2; \; \Delta E \times \Delta t \geq \hbar/2$$

So far the idea of replacing the kelvin by the joule has not been supported by many metrologists.

The new definitions of the base units are more sophisticated and much more difficult to understand than old definitions. They can be difficult not only for the public and students but for technical staff as well. In many cases, for the new definition of a unit there is no link to a practical standard of the unit. Laboratories of metrology will use "old standards" for the new defined units. For the kilogram it is the watt balance set (it may be a sphere made of the isotope ^{28}Si in form of a single crystal as well [3]), for the ampere there are the Josephson standard of the volt and the QHE standard of the ohm, and for the kelvin there is The International Temperature Scale of 1990. Some laboratories have established a set-up with the watt balance [21]. Even more results with the watt balance set-up were recently published by NIST and NPL. Unfortunately these results disagree by more than their combined standard uncertainties so far [18].

We can note what kind of arguments were considered at solving the problem of a set of the base units (ampere or volt?). The ampere won "for the purposes of historical continuity" [14]. Cabiati and Bich made some remarks on the proposed quantum SI [4]. "Should the explicit-constant definitions be adopted, the lack of wording could be compensated by moving the focus from SI units to SI quantities, whose definitions might be conveniently refined". The new SI gives "opportunities offered… by reference quantities different from fundamental constants…. such opportunities completely met within the individual metrological system of every quantity, with little implication for the general structure of the SI…" [4].

However, development of the system of units based on fundamental physical constants and quantum standards seems to be a foregone conclusion. We can hope that the new definitions for the base units will be accepted by a community of physics and metrology.

References

1. http://www1.bipm.org/en/
2. P. Becker et al., Considerations on future redefinitions of the kilogram, the mole and other units. Metrologia **2007**(44), 1–14 (2007)
3. ChJ. Borde, Base units of the SI, fundamental constants and a modern quantum physics. Trans. Roy. Soc. A **363**, 2177–2180 (2005)
4. F. Cabiati, W. Bich, Thoughts on a changing SI. Metrologia **46**, 457–466 (2009)
5. R.S. Davies, Redefining the kilogram: how and why? MAPAN J. Metrol. Soc. India **23**, 131–138 (2008)
6. R.H. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 1 (Addison-Wesley, Reading, 1964)
7. J. Fischer et al., Preparative steps towards the new definition of the kelvin in terms of the Boltzmann constant. Intern. J. Thermophys. **28**, 1753–1765 (2007)
8. G. Girard, The third periodic verification of national propototypes of the kilogram. Metrologia **31**, 317–336 (1994)
9. M. Gläser, New definitions of the kilogram and the mole, Presentation on 10th meeting of the CCM (2007), www.bipm.org/cc
10. M.I. Kalinin, S.A. Kononogov, Redefinition of the unit of thermodynamic temperature in the international system of units. High Temp. **48**, 23–28 (2010)
11. B.P. Leonard, The atomic-scale unit, entity: key to a direct and easily understood definition of the SI base unit for amount of substance. Metrologia **44**, 402–406 (2007)
12. K.K. Likharev, A.B. Zorin, Theory of the Bloch-wave oscillations in small Josephson junctions. J. Low Temp. Phys. **59**, 347–382 (1985)
13. L. Lipiński, A. Szmyrka-Grzebyk, Proposals for the new definition of the Kelvin. Metrol. Meas. Syst. **15**, 227–234 (2008)
14. M.I. Mills, P.J. Mohr, T.J. Quinn, B.N. Taylor, E.R. Williams, Redefinition of the kilogram, ampere, Kelvin and mole: a proposed approach to implementing CIPM recommendation 1. Metrologia **43**, 227–246 (2006)
15. P.J. Mohr, B.N. Taylor, D.B. Newell, CODATA recommended values of the fundamental physical constants: 2010. Rev. Mod. Phys. **77**, 1–106 (2005)
16. W. Nawrocki, Revising the SI: the joule to replace the Kelvin as a base unit. Metrol. Meas. Syst. **13**(171–181), 9 (2006)

17. B.W. Petley, *The Fundamental Physical Constants and the Frontier of Measurements* (Adam Hilger, Bristol, 1988)
18. F. Piquemal, G. Geneves, Argument for a direct realization of the quantum metrological triangle. Metrologia **37**, 207–211 (2000)
19. T.J. Quinn, Base units of the SI System, their accuracy, dissemination and international traceability. Metrologia **31**, 515–527 (1995)
20. I.A. Robinson, Toward the redefinition of the kilogram: measurements of Planck's constant using watt balances. IEEE Trans. Instrum. Meas. **58**, 942–948 (2009)
21. H. Scherer, B. Camarota, Quantum metrological triangle experiments: a status review. Meas. Sci. Technol. **23**, 124010 (2012)
22. E. Williams, Proposed changes to the SI, their impact to fundamental constants, and other SI units, Presentation on 10th meeting of the CCM (2007), www.bipm.org/cc

Chapter 4
Quantum Voltage Standards

Abstract In this chapter we focus on superconductivity and superconducting materials. We discuss the discovery of high-temperature (HT_c) superconductors and provide an overview of HT_c superconducting materials. We present the construction of four types of Josephson junction and derive the relations that describe the Josephson effect. We discuss the design of the classical electrochemical standards of direct voltage and standards with Josephson junctions. The parameters of the standards are compared, and the advantages of the quantum standards with Josephson junctions described. Also provided are the results of a comparison of the quantum standard at the Polish Central Office of Measures with the BIPM standard. We present the design of two types of alternating-voltage standards with Josephson junctions: binary divided junction arrays and pulse-driven junction arrays. Also presented are the principles of operation of the memory cell and flip-flop, basic components in digital cryoelectronics.

4.1 Superconductivity

4.1.1 Superconducting Materials

Quantum voltage standards use the effect of voltage quantization in Josephson junctions. A necessary condition for a Josephson junction to operate is the superconducting state of its electrodes. Along with superfluidity, superconductivity is a very peculiar property of matter, which long remained unexplained, and the history of its discovery is worth a brief discussion.

At the turn of 19th and 20th century researchers vied to liquefy helium, discovered as late as in 1868 and remaining the last unliquefied gaseous element. Front-runners in this race, James Dewar at the University of Cambridge, England, and Heike Kamerlingh Onnes at the University of Leiden, Netherlands, had the edge of being more successful than others in attaining low temperatures. Finally, in 1908, helium was first liquefied by Kamerlingh Onnes, at temperature 4.2 K and

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_4

normal atmospheric pressure p_a = 101 kPa. It is worthy of notice that helium is the only element to resist solidification even at the lowest temperatures under normal atmospheric pressure and only solidifies at pressure 2.7 MPa and temperature 1.5 K. After Norway, Poland is Europe's second most important producer of liquid helium. The natural gas resources in Odolanów, Great Poland, include relatively large amounts of helium, which is extracted by liquefaction.

By liquefying helium at 4.2 K Kamerlingh Onnes obtained a cooling agent, which he subsequently used for studying low-temperature physical properties of many substances. These studies led him to discover, in 1911, that the electrical resistance of mercury dropped abruptly at 4.2 K, as illustrated by the plot shown in Fig. 4.1. Although the measurements revealed no resistance of the studied mercury sample at $T < 4.2$ K, in his report of the study Kamerlingh Onnes cautiously stated the electrical resistance decreased to a value too low to be measured. However, later the resistivity of some materials was found to disappear completely at low temperatures, and the phenomenon was named super-conductivity.

The transition to the superconducting phase occurs at a specific temperature referred to as critical temperature T_c. Superconductivity was found to occur in many metals and metallic alloys, among which those which become superconductors at relatively high temperatures are of the most practical importance. Metals with relatively high critical temperatures include niobium, with T_c = 9.3 K, lead, with T_c = 7.2 K, and vanadium, with T_c = 5.4 K (see Table 4.1). Interestingly, among good electrical conductors only aluminum (critical temperature T_c = 1.17 K) has superconducting properties; neither noble metals (gold and silver) nor copper are superconductors. Semiconductors: silicon and germanium require a huge pressure, 12 GPa, to become superconductors at critical temperatures T_c = 7 K (for Si) and T_c = 5.3 K (for Ge) [6].

The pressure dependence is quite different in niobium: the critical temperature of this element decreases with growing pressure from T_c = 9.3 K at normal atmospheric pressure p_a = 101 kPa to T_c = 4.5 K at p = 132 GPa.

In the 8 years' period following the discovery of superconductivity superconducting materials with higher critical temperatures were searched for mainly among substances containing elemental superconductors. Record critical temperatures in such compounds were found in Nb_3Ge (T_c = 23.2 K) and V_3Si (T_c = 17.1 K)

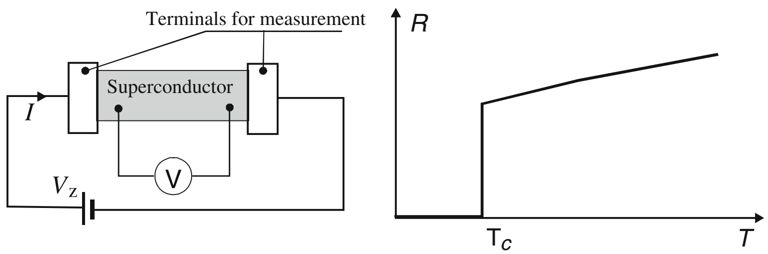


Fig. 4.1 Measurement of resistance R versus temperature T in superconducting materials: measurement system and the resulting R = f(T) plot

Table 4.1 Critical temperature of selected superconducting metals at normal atmospheric pressure

Element	Rhodium	Platinum[a]	Titanium	Aluminum	Indium	Tin	Mercury	Vanadium	Lead	Niobium
T_c (K)	325 μK	1.9 mK	0.40	1.17	3.41	3.72	4.15	5.4	7.20	9.3

[a] Only powdered platinum has superconducting properties

(see Table 4.2). An alloy of titanium and niobium, 60 %Nb40 %Ti, with critical temperature $T_c = 9.8$ K, is commonly used for fabricating superconducting wires [6].

A breakthrough in this search for higher critical temperatures was achieved in 1986, when Karl Alex Müller and Georg Bednorz at the IBM Research Laboratory in Rüschlikon, Switzerland, discovered superconductivity in ceramic materials containing copper oxides CuO_x [1].

The chemical composition of the discovered superconductor was astounding: *none of the constituent elements of the compound is a superconductor itself*! An $La_2Sr_4CuO_4$ sinter studied by Müller and Bednorz showed superconductivity at 36 K. Although announced quite unexpectedly, the finding was so important for science that after its verification thousands of physicists and technologists all over the world became involved in the research on high-temperature superconductors (abbreviated HTS) and their applications. Müller and Bednorz shared the 1987 Nobel Prize, awarded with unprecedented immediateness, as soon as the year following their discovery. Concurrently with the latter, in 1986, Paul Chu (original name Ching-Wu Chu) at the University of Houston published the results of his study of $YBaCu_3O_{7\text{-}x}$, a sinter of yttrium, barium and copper oxide, and announced its critical temperature: $T_c = 91$ K (Table 4.3). The laboratory in Houston became a real mixing plant, producing thousands of combinations of elements and compounds to be studied in terms of their conducting properties, and Chu was very seriously considered for the 1987 Nobel Prize, finally awarded to Müller and Bednorz. The huge practical importance of Chu's discovery consists in the possibility of obtaining the superconducting phase (by cooling the compound to a temperature $T < T_c$) by means of liquid nitrogen (boiling point at 77 K), which is slightly cheaper than petrol, and more than ten times cheaper than liquid helium. Also the cryogenic equipment for liquid nitrogen is much cheaper than the helium equipment because of lower heat insulation requirements.

Among the currently known superconducting materials the highest critical temperature, 138 K, at normal atmospheric pressure (ca. 100 kPa) was found in $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8.33}$ (Table 4.3). At very high pressures, around 30 GPa, this ceramic material achieves superconductivity at even higher critical temperature, ranging from 160 to 165 K. Other superconducting materials include organic

Table 4.2 Critical temperature of selected superconducting compounds and binary alloys

Material	$AuIn_3$	UGe_2	$Nb_{0.6}Ti_{0.4}$	V_3Ga	Ta_3Pb	Nb_3Al	Nb_3Si	Nb_3Ge	MgB_2
T_c (K)	50 μK	≈1	9.8	16.8	17.0	18.0	19.0	23.2	39

Table 4.3 Critical temperature of HTS ceramic materials

Material	$La_2Sr_4CuO_4$	$YBaCu_3O_7$	$Bi_2Ca_2Sr_2Cu_3O_{10}$	$Tl_2Ba_2Ca_2Cu_3O_{10}$	$Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8.33}$
T_c (K)	36	91	110	125	138[a]

[a] The critical temperature of $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8.33}$ at very high pressure ($p_h \approx 30$ GPa) is 165 K, approximately

compounds and fullerenes, a family of carbon molecules including C_{60} (the critical temperature of fullerene superconductors is around 15 K). However, as these two groups of materials are of little importance for electronics and metrology, they will not be discussed further in this book.

4.1.2 Theories of Superconductivity

The theoretical explanation of superconductivity took nearly fifty years after the experimental discovery of the phenomenon. The foundations for the currently adopted theory were laid by London brothers (Germany, 1934) and then by Vitaly Ginzburg and Lev Landau (USSR, 1950), who developed a phenomenological theory formulated within the framework of quantum mechanics. For his contribution to the development of the theory of superconductivity Ginzburg was awarded the Nobel Prize in 2003, more than 50 years after the publication of the theory! Before, in 1962, the Nobel Prize was awarded to Landau, but in recognition of his theory of liquid helium. In 1957 Leon Cooper (USA), a Ph.D. student of John Bardeen (a co-inventor of transistor), proposed to explain the zero electrical resistance in superconductors by the occurrence of electron pairs (referred to as Cooper pairs) in the superconducting phase. Each Cooper pair includes two strongly correlated electrons with opposite orientation of momentum and spin. In a superconducting material Cooper pairs form a coherent quantum condensate (fluid) and are all described by a single wave function fulfilling the Schrödinger equation. The dynamic properties of a quantum condensate differ substantially from the quasi-static properties of an electron gas consisting of single electrons in an electric conductor in its normal (non-superconducting) phase. In the latter case each electron is described by a separate wave function, and their quantum dynamic properties are averaged.

Setting in motion (e.g. that producing an electric current) one electron from a Cooper pair brings about a response of the other electron. As this response involves no energy loss, a number of electrons can move without dissipating any energy to produce a zero-resistance electric current. Developed by Bardeen, Cooper and Schrieffer, the theory based on Cooper pairs is commonly referred to as the BCS theory and represents the currently adopted microscopic explanation of low-temperature superconductivity. One of its consequences is the following relation between the energy gap in the energy spectrum of a superconducting material and its critical temperature [4]:

$$E_G = 3.5\, k_B T_c,$$

where E_G is the energy gap in the superconducting material, k_B is Boltzmann's constant, and T_c is the critical temperature of the material, or the temperature of its transition into the superconducting phase.

4.1.3 Properties of Superconductors

At least two conditions must be fulfilled for a superconductor sample to reach the superconducting state:

- The temperature of the sample must be below the critical temperature T_c,
- The external magnetic field must be below a critical value H_c.

In a sufficiently strong magnetic field the superconductor will quit the superconducting state to return to its normal state (i.e. the state in which the sample has 'ordinary' conducting properties). The minimum value of magnetic field necessary to destroy superconductivity is known as the critical magnetic field H_c. The effect is reversible: reducing the magnetic field will result in the restoration of the superconducting state. The critical magnetic field H_c increases with decreasing temperature of the sample (Fig. 4.2).

Another unusual property of superconductors, discovered in 1933 by German physicists Walter Meissner and Robert Ochsenfeld, is the expulsion of magnetic field from the interior of superconductors, which therefore act as ideal diamagnetic materials (Fig. 4.3). This expulsion of magnetic field by superconductors is referred to as the Meissner effect.

The effect of magnetic field on superconducting materials is a criterion of dividing superconductors into two classes: type I superconductors, in which the effect of magnetic field is as shown Fig. 4.4a, and type II superconductors, with magnetization dependence as in Fig. 4.4b.

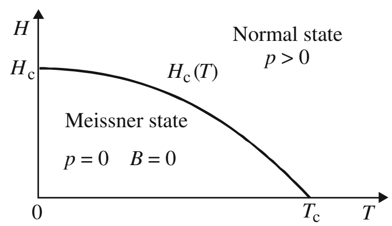


Fig. 4.2 Temperature dependence of critical magnetic field in superconductors

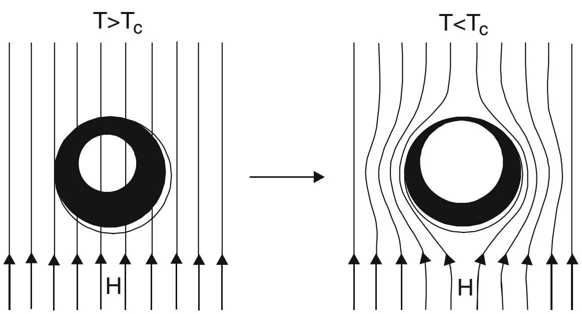


Fig. 4.3 Meissner effect: expulsion of external magnetic field from superconductor interior

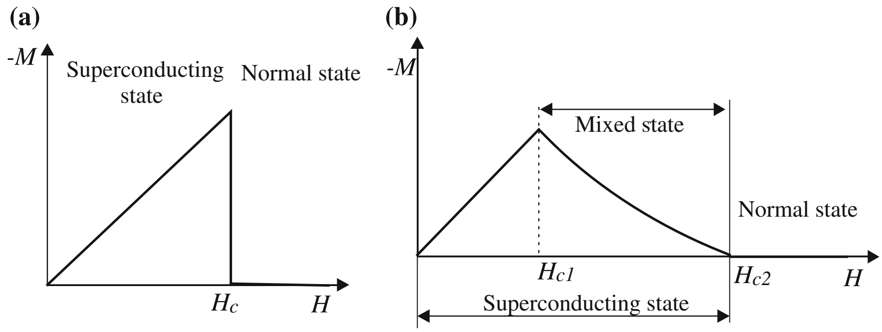


Fig. 4.4 Magnetization M versus magnetic field H **a** in type I superconductors and **b** in type II superconductors

This classification was proposed in 1957 by Alexei Abrikosov, who first discerned these two types of magnetic properties of superconductors and whose finding was recognized by the Nobel Prize in 2003 (after 46 years!). In type II superconductors at temperatures $T < T_c$ complete expulsion of magnetic field from the superconductor interior only occurs in weak magnetic fields, $H < H_{c1}$. In intermediate magnetic fields ($H_{c1} < H < H_{c2}$) a type II superconductor is in a mixed state in which the field expulsion is only partial: vanishing in a part of the sample, in the remainder the magnetic field forms vortices carrying magnetic flux quanta Φ_0. The normal state is restored in magnetic fields $H > H_{c2}$.

4.2 Josephson Effect

In 1962 Brian D. Josephson, 22 years old graduate student in Cambridge, studied a junction consisting of two superconductors separated by an insulating layer with 0.1–1 nm thickness. Josephson predicted the effect described below, which was later given his name [14].

Electrons forming the electric current in a superconductor are bound in pairs, referred to as Cooper pairs. Like other elementary particles, Cooper pairs in a superconductor are described within the framework of quantum mechanics by the wave equation (4.1). The junction under consideration is assumed to represent a one-dimensional space, with coordinate x, for the motion of Cooper pairs. The wave equation reads:

$$\Psi = \Psi_0 \exp\left[\frac{-\mathrm{j}}{\hbar}(Et - \mathbf{p}\mathbf{x})\right], \tag{4.1}$$

where Ψ_0 denotes the amplitude of the wave function, E is the energy of the particle, vectors $\boldsymbol{p}$ and $\boldsymbol{x}$ are its momentum and position, respectively, in the coordinate system, and t is time.

According to the Ginzburg-Landau theory, the product $\Psi \times \Psi^*$ of the wave function Ψ and its conjugate Ψ^* is proportional to the density ρ of particles. This applies also to Cooper pairs, which are regarded as elementary particles of mass $2m$ and electric charge $2e$. All Cooper pairs in a superconductor are described by the same wave function Ψ:

$$\rho = \Psi \times \Psi^* = \Psi_0^2, \tag{4.2}$$

where ρ is the density of Cooper pairs (the number of Cooper pairs per unit of volume) in the superconductor.

Using de Broglie relations: $E = hf$ and $p = h/\lambda$, where f is the frequency of a de Broglie wave and λ denotes its wavelength, (4.1) can be rewritten in the following form, better exposing the phase of the wave function Ψ:

$$\Psi = \Psi_0 \exp[-\mathrm{j}(\omega t - 2\pi \frac{x}{\lambda})], \tag{4.3}$$

where $\omega = 2\pi f$ is the angular frequency.

The wave functions of Cooper pairs in two separated superconductors S1 and S2 are independent. However, if the distance between the superconductors is small (0.1–1 nm), Cooper pairs cross the potential barrier between superconductor S1 and superconductor S2 due to the tunneling effect (Fig. 4.5). When tunneling occurs, the wave functions of Cooper pairs in superconductors S1 and S2 are correlated. If the insulating layer between the superconductors is thin enough, the probability that a Cooper pair will cross the insulator is greater than zero, resulting in a superconducting current that flows across the insulator without any voltage drop in the superconductor-insulator-superconductor junction.

Let Ψ_1 and Ψ_2 be the wave functions of Cooper pairs in superconductors S1 and S2, respectively:

$$\Psi_1 = \Psi_{01} \exp[-\mathrm{j}(\omega t - \varphi_1)] \quad \text{and} \quad \Psi_2 = \Psi_{02} \exp[-\mathrm{j}(\omega t - \varphi_2)], \tag{4.4}$$

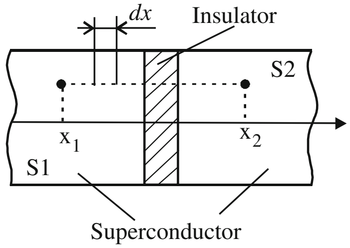


Fig. 4.5 Superconductor-insulator-superconductor junction

where φ_1 and φ_2 denote the respective phases of Ψ_1 and Ψ_2.

The derivative of the wave function (4.1) in each of the separated superconductors fulfils the respective equations:

$$\frac{\partial \Psi_1}{\partial t} = -\mathrm{j}\frac{E_1}{\hbar}\Psi_1, \frac{\partial \Psi_2}{\partial t} = -\mathrm{j}\frac{E_2}{\hbar}\Psi_2. \tag{4.5}$$

The energy difference across the junction is proportional to the potential difference, or the applied voltage V:

$$E_2 - E_1 = 2eV. \tag{4.6}$$

When the distance between the superconductors is small, their wave functions are correlated due to Cooper pair exchange. The correlation of both wave functions can be expressed:

$$\begin{aligned} \frac{\partial \Psi_1}{\partial t} &= -\mathrm{j}\frac{1}{\hbar}(E_1\Psi_1 + K_s\Psi_2), \\ \frac{\partial \Psi_2}{\partial t} &= -\mathrm{j}\frac{1}{\hbar}(E_2\Psi_2 + K_s\Psi_1), \end{aligned} \tag{4.7}$$

where K_s is the superconductor coupling coefficient.

Including (4.3) in the (4.7) leads to:

$$\begin{aligned} &\frac{\partial \Psi_{01}}{\partial t}\exp[-\mathrm{j}(\omega t - \varphi_1)] - \mathrm{j}\Psi_{01}\exp[-\mathrm{j}(\omega t - \varphi_1)](\omega - \frac{\partial \varphi_1}{\partial t}) = \frac{-\mathrm{j}}{\hbar}(E_1\Psi_1 - K_s\Psi_2), \\ &\frac{\partial \Psi_{02}}{\partial t}\exp[-\mathrm{j}(\omega t - \varphi_2)] - \mathrm{j}\Psi_{02}\exp[-\mathrm{j}(\omega t - \varphi_2)](\omega - \frac{\partial \varphi_2}{\partial t}) = \frac{-\mathrm{j}}{\hbar}(E_2\Psi_{21} - K_s\Psi_1). \end{aligned} \tag{4.8}$$

Equation (4.8) can be rewritten in the form:

$$\begin{aligned} \frac{\partial \Psi_{01}}{\partial t} - \mathrm{j}\Psi_{01}(\omega - \frac{\partial \varphi_1}{\partial t}) &= -\frac{\mathrm{j}}{\hbar}\{E_1\Psi_{01} + K_s\Psi_{02}\exp[-\mathrm{j}(\varphi_1 - \varphi_2)]\}, \\ \frac{\partial \Psi_{02}}{\partial t} - \mathrm{j}\Psi_{02}(\omega - \frac{\partial \varphi_2}{\partial t}) &= -\frac{\mathrm{j}}{\hbar}\{E_2\Psi_{02} + K_s\Psi_{01}\exp[-\mathrm{j}(\varphi_2 - \varphi_1)]\}. \end{aligned} \tag{4.9}$$

By using Euler's formula we arrive at:

$$\begin{aligned} \frac{\partial \Psi_{01}}{\partial t} - \mathrm{j}\Psi_{01}(\omega - \frac{\partial \varphi_1}{\partial t}) &= -K_s\frac{\Psi_{02}}{\hbar}\sin(\varphi_1 - \varphi_2) - \frac{\mathrm{j}}{\hbar}[E_1\Psi_{01} + K_s\Psi_{02}\cos(\varphi_1 - \varphi_2)], \\ \frac{\partial \Psi_{02}}{\partial t} - \mathrm{j}\Psi_{02}(\omega - \frac{\partial \varphi_2}{\partial t}) &= -K_s\frac{\Psi_{01}}{\hbar}\sin(\varphi_2 - \varphi_1) - \frac{\mathrm{j}}{\hbar}[E_2\Psi_{02} + K_s\Psi_{02}\cos(\varphi_2 - \varphi_1)]. \end{aligned} \tag{4.10}$$

Equating the real parts of (4.10) leads to:

$$\begin{aligned}\frac{\partial \Psi_{01}}{\partial t} &= K_s \frac{\Psi_{02}}{\hbar} \sin(\varphi_2 - \varphi_1),\\ \frac{\partial \Psi_{02}}{\partial t} &= -K_s \frac{\Psi_{01}}{\hbar} \sin(\varphi_2 - \varphi_1).\end{aligned} \tag{4.11}$$

As the product $\Psi \times \Psi^* = \Psi_0^2$ is the density of particles, the differential $\frac{\partial \Psi_0}{\partial t}$ is proportional to the change in density of Cooper pairs due to tunneling. However, reduced number of Cooper pairs in superconductor S1 implies their increased number in superconductor S2. Thus:

$$\frac{\partial \Psi_{01}}{\partial t} = -\frac{\partial \Psi_{02}}{\partial t}. \tag{4.12}$$

When the superconductors are of the same material, the density of Cooper pairs is equal in both: $\Psi_{01}^2 = \Psi_{02}^2$. A change of density in time corresponds to the superconducting current across the insulator:

$$\frac{\partial \Psi_0}{\partial t} = i_s. \tag{4.13}$$

Thus, (4.11) can be put in the following form, known as the first Josephson equation:

$$i_s = I_C \sin(\varphi_2 - \varphi_1), \tag{4.14}$$

where I_C is the critical current of the junction and $(\varphi_2 - \varphi_1)$ denotes the wave function phase difference across the junction.

By equating the imaginary parts of (4.10) and using (4.6) we obtain the second Josephson equation (4.15), which describes junction oscillations, or alternating current, induced by a potential difference V between the two superconductors. The frequency of these oscillations fulfils the relation:

$$f = \frac{\partial \varphi}{\partial t} = \frac{2e}{h} V, \tag{4.15}$$

where V is the voltage across the junction (between superconductors S1 and S2).

Equation (4.15) implies a 100 μV fixed voltage across a Josephson junction will induce alternating voltage, or oscillations, with frequency 48 GHz. An inverse effect occurs as well: microwave radiation with frequency 48 GHz applied to the junction will induce a constant voltage $V_J \approx 100\ \mu\text{V}$, the exact value of which is determined by the frequency of the radiation and the constant $2e/h$:

$$V_J = \frac{h}{2e} f. \tag{4.16}$$

The value of voltage V_J across a Josephson junction is independent of material (as verified by studies of Nb, Pb, Sn and In [5] junctions). It is also independent of the type and geometry of the junction, the magnetic field (as long as its value is below the critical field H_c) and the power of the microwave radiation [23]. Possible dependence on factors other than $h/2e$ and the frequency of electromagnetic radiation was contradicted by experimental studies. The average difference between V_J values in two Josephson junctions, one made of indium and the other of Nb/Cu/Nb, exposed to electromagnetic radiation from the same source was found to be below 2×10^{-21} V [13]. High-temperature superconductor Josephson junctions have the same properties as those of conventional (low-temperature) superconductors such as niobium, indium or other superconducting materials specified in Tables 4.1 and 4.2. The effect of constant voltage generation in a Josephson junction in external electromagnetic field is used in electronic standards of voltage.

4.3 Josephson Junctions

Governed by (4.14) and (4.15), the Josephson effect manifests itself very distinctly in Josephson junctions. These can be of various types, the most common four of which are shown in Fig. 4.6. First Josephson junctions were based on a point contact between a niobium bolt tip and a flat niobium surface. The current-voltage characteristic of such a point contact shows no hysteresis. Josephson junctions manufactured nowadays include solely thin-film tunnel junctions (Fig. 4.6a, c, d) and microbridge junctions (Fig. 4.6b).

Brian D. Josephson already mentioned in his article [14] that similar properties as the SIS junction (Superconductor-Insulator-Superconductor) shown in Fig. 4.6a should have a junction formed by two superconductors separated by a thin layer of

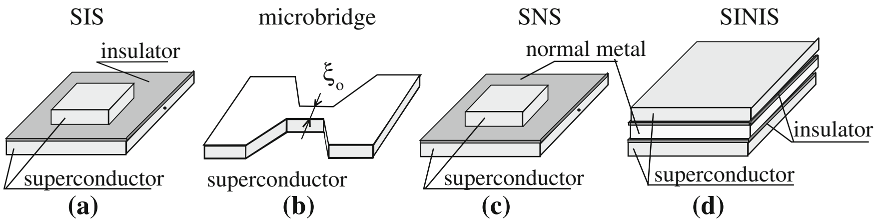


Fig. 4.6 Josephson junctions: **a** point-contact junction, **b** SIS thin-film junction, **c** microbridge, **d** SNS thin-film junction

conductor in the normal state, i.e. SNS junction (Superconductor-Normal metal-Superconductor)—Fig. 4.6c. These suspicions were soon confirmed. SINIS thin-film Josephson junctions (Superconductor-Insulator-Normal metal-Insulator-Superconductor) are used as well. The SNS and SINIS junctions shown a hysteresis voltage-current characteristic. The parameters of Josephson junctions used in quantum voltage standards are presented in Table 4.4.

In a microbridge (Fig. 4.6b) the width of the neck must not exceed the phase coherence length ξ_o in the superconductor. The phase coherence length in HTS materials ($YBa_2Cu_3O_7$) is of the order of a few nanometers, and ca. 1000 times greater in metallic superconductors (e.g. in aluminum ξ_o = 1500 nm). Typically made of niobium, lead or their compounds, Josephson junctions used in metrological devices operate in conditions which require liquid helium cooling (temperature 4.2 K). High-temperature superconductor junctions are manufactured as well, mainly of $YBa_2Cu_3O_7$.

Used in studies of the dynamics of Josephson junctions and electrical circuits of which they are components, the equivalent circuit diagram of a Josephson junction is shown in Fig. 4.7. The equivalent circuit of a Josephson junction includes the

Table 4.4 Parameters of Josephson junctions used in 10 V voltage standards fabricated at PTB, Germany [2] (published with a kind permission of Dr. R. Behr, PTB)

Type of junction	–	SIS	SINIS	SNS
Material	–	Nb-Al/ AlO_x-Nb	Nb-Al/AlO_x/Al/ AlO_x/Al-Nb	Nb-Nb_xSi_{1-x}-Nb with $x \approx 10$ %
Number of Josephson junctions N	–	13 924	69 632	69 632
Number of microstrips	–	4	128	128
Number of JJs in one microstrip N_m	–	3481	136–562	136–562
Junction length l	µm	20	15	6
Junction width w	µm	50	30	20
Current density j	A cm^{-2}	10	750	3000
Critical current I_c	mA	0.1	3.5	3.5
Normal state resistance R_n	Ω	15 at 1.5 mV	0.04 at I_c	0.04 at I_c

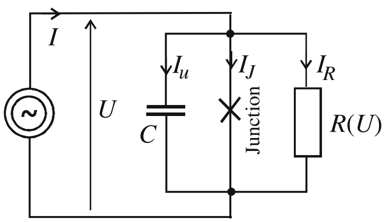


Fig. 4.7 Equivalent circuit diagram of a Josephson junction controlled by current

capacity C of the junction and the nonlinear shunt resistance R dependent on the voltage V across the junction [20]. Including a constant component I_0 and a variable component $I(t)$, the total current I across a Josephson junction and is the sum of the currents in three branches of the circuit depicted in Fig. 4.7:

the Josephson current $I_J = I_C \sin\varphi$, the leakage current $I_R = V/R(V)$,
the capacity current $I_u = C\frac{dV}{dt}$.

$$I = I_0 + I(t) = I_C \sin\varphi + \frac{V}{R(V)} + C\frac{\mathrm{d}V}{\mathrm{d}t}. \tag{4.17}$$

Formula (4.15) implies the following relation between the variable component of the current across the junction and the voltage V:

$$\begin{aligned} \frac{\partial\varphi}{\partial t} &= \frac{2e}{h}V, \\ I &= C\frac{h}{2e}\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} + \frac{h}{2eR}\frac{\mathrm{d}\varphi}{\mathrm{d}t} + I_C \sin\varphi. \end{aligned} \tag{4.18}$$

As the relation (4.18) between the current and the wave function phase difference is strongly nonlinear, numerical methods are used for solving it in practice.

The DC current-voltage characteristic of a Josephson junction has a hysteresis or not, depending on the value of Mc Cumber parameter β_c, defined by parameters of the equivalent circuit of the junction:

$$\beta_c = \frac{2\pi R^2 C I_C}{\Phi_0} = \frac{2\pi L I_C}{\Phi_0}, \tag{4.19}$$

where I_C is the critical current of the junction, Φ_0 is the magnetic flux quantum and L is the inductance of the superconducting circuit incorporating the Josephson junction. Mc Camber parameter describes the damping of Josephson oscillations by the junction shunting resistance R. β_c is equal to the square of the quality factor of a resonant circuit formed by junction parameters shown in Fig. 4.7—$\beta_c = Q^2$.

When the current across a Josephson junction is below the critical current I_C the voltage across the junction is zero (Fig. 4.8a), like in other superconducting sample. This is the DC Josephson effect. A Josephson junction is non-hysteretic if $\beta_c \leq 1$ (overdamped junction) and hysteretic if $\beta_c > 1$ (underdamped junction).

The hysteresis in the current-voltage characteristic of a Josephson junction has a turning point (for decreasing current) at voltage V_{his} determined by the energy gap E_G of the superconducting material and the electron charge e:

$$V_{his} = \frac{E_G}{e}.$$

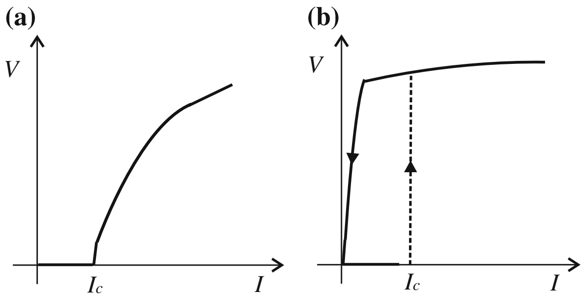


Fig. 4.8 DC current-voltage characteristic of a Josephson junction: **a** without hysteresis, $\beta_c \leq 1$ (over damped junction); **b** with hysteresis, $\beta_c > 1$ (underdamped junction)

When a Josephson junction is exposed to electromagnetic radiation with frequency f_e, internal junction oscillations (referred to as Josephson oscillations) become synchronized with the field frequency f_e. The frequency of non-synchronized Josephson oscillations is called a characteristic frequency of a junction, $f_c = (2e/h) \times R_n I_c$, where R_n is the normal state resistance of a junction and I_c is its critical current. The voltage V across the junction is a stepwise function of the current I (Fig. 4.9a), and the step height ΔV only depends on the frequency f_e and the fundamental physical constants e and h: $\Delta V = (h/2e)f_e$. In particular, ΔV is independent of the material of the junction and its operating temperature (which ensures the superconducting state).

The phenomenon described above is referred to as the AC Josephson effect. As the frequency can be stabilized and measured with an uncertainty of the order of 10^{-15} (offered by the atomic caesium clock), the Josephson effect allows the construction of a quantum voltage standard producing a voltage reference V_{ref} with reproduction uncertainty unavailable to classical standards. The voltage V_{ref} produced by a quantum voltage standard based on the Josephson effect fulfils the relation:

$$V_{ref} = k\frac{h}{2\,e}f_e, \tag{4.20}$$

where k is the step number in the stepwise current-voltage characteristic (Fig. 4.8a) or the line number in the characteristic with hysteresis (Fig. 4.9) ($k = 0, 1, 2, 3,\ldots$) and f_e is the frequency of the applied electromagnetic field.

The value of factor $2e/h$ can be determined with better accuracy than the values of e and h separately! After international consultations the Consultative Committee for Electricity and Magnetism at the International Bureau of Weights and Measures (BIPM) at Sèvres established the following value of $2e/h$, referred to as the Josephson constant and denoted K_J [14],

$$K_J = 2e/h = 483.5979 \times 10^{12}\ \text{Hz/V}. \tag{4.21}$$

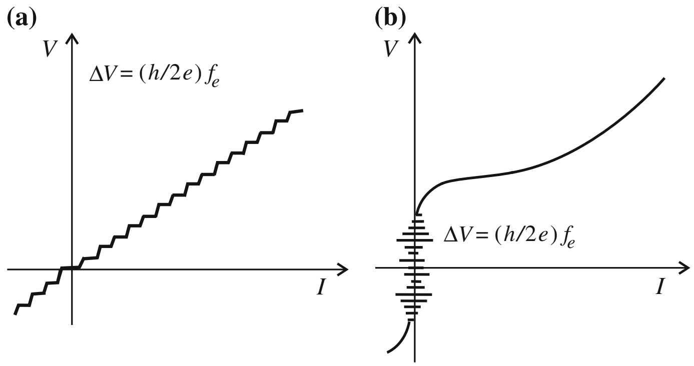


Fig. 4.9 AC Josephson effect caused by electromagnetic irradiation: voltage versus current across a Josephson junction **a** without hysteresis; **b** with hysteresis. The voltage steps on both characteristics are called Shapiro steps

An important parameter of a Josephson junction is its switching timeτ between two states characteristic of the DC Josephson effect: one with junction current greater than the critical current I_c and, consequently, nonzero voltage across the junction, and the other with current $I < I_c$ and zero voltage. The switching time τ fulfils the relation [27, 30]:

$$\tau = \frac{h}{2\pi E_G}, \tag{4.22}$$

where E_G is the energy gap of the material.

The switching time is of crucial importance in Josephson junction-based cryo-electronic circuits including logic gates and counters. Niobium, the most common material of cryoelectronic components, has energy gap E_G = 0.003 eV and switching time τ = 0.23 ps. Calculated as the inverse of τ, the corresponding switching rate of a niobium junction is above 4000 GHz, about 1000 times greater than the clock rate of Intel 4, the fastest PC microprocessor to date (as of 2014), operating at a maximum frequency of 5 GHz.

4.4 Voltage Standards

4.4.1 Voltage Standards with Weston Cells

The first voltage standards were voltaic cells. Selected among the wide variety of these, the saturated Weston cell, producing a very stable electromotive force, which remains constant for many years, is a classical standard realizing the unit of voltage. The saturated Weston cell consists of two electrodes: a high-potential electrode of mercury and a low-potential electrode of an amalgam of mercury and cadmium (with a 13 % proportion of the latter). The electrical connection between the electrodes and the output terminals is realized by platinum wires (Fig. 4.10). The electrodes are immersed in a saturated solution of cadmium sulfate ($CdSO_4$), which acts as the electrolyte and is kept above the electrodes by a separating layer of $CdSO_4$ crystals.

The nominal value of electromotive force (EMF) produced by a saturated Weston cell at temperature 20 °C is $E = 1.01865$ V. The values of EMF produced by different cells only differ in further significant digits. Factors affecting the EMF of a Weston cell include its manufacturing technology, temperature, discharge and the effect of mechanical shocks. Of particular impact on the standard cell are the temperature and mechanical shocks. The temperature-related change E_t in electromotive force produced by a Weston cell fulfils the following empirical formula:

$$E_t = E_{20} - 40.6 \times 10^{-6}(t - 20) - 9.5 \times 10^{-7}(t - 20)^2 + 10^{-8}(t - 20)^3, \quad (4.23)$$

where E_{20} is the electromotive force produced by the cell at temperature 20 °C and t is the temperature in degrees centigrade.

Calculated from (4.23), the electromotive force difference due to a +1 °C temperature change in the temperature range around 20 °C is –41.55 μV. The saturated Weston cell has a relatively high internal resistance, which ranges from 500 to 1000 Ω. The typical internal resistance value is 800 Ω, approximately.

The national voltage standard at the Central Office of Measures (GUM, Warsaw) is a set of selected forty-eight Weston cells immersed in an oil bath and stored in an air-conditioned room. The temperature stability of the immersed Weston cells is ±0.01 °C. The voltage reference is the mean value of the electromotive force produced by the forty-eight cells. As this reference value varies in time (for example, the electromotive force produced by the standard system in 1997 was

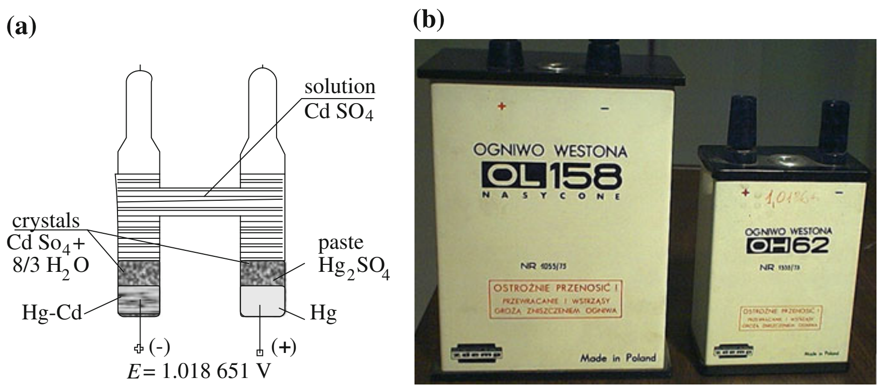


Fig. 4.10 Saturated Weston cell used as classical voltage standard: **a** cell structure, **b** Weston cells manufactured in the Experimental Laboratory of the Silesian University of Technology

E_{GUM} = 1.0186511 V at 20 °C). Both standards: the electrochemical voltage standard and the Josephson quantum voltage standard are used at GUM in Warsaw.

The saturated Weston cell should not be confused with the unsaturated Weston cell, which produces EMF of much worse long-term stability (up to 300 μV per annum), but has lower temperature coefficient (–10 μV/K) and is resistant to shock.

4.4.2 DC Voltage Josephson Standards

Using Josephson voltage standards, i.e., quantum voltage standards based on the Josephson effect, involves some problems due to the low value of the output voltage V_{ref} obtained for typical values of the parameters entering the (4.20). Built in 1968, the first quantum voltage standard using a point-contact Josephson junction [24] produced quantized voltage of up to 5 mV. The low voltage produced by this quantum standard made it difficult to compare with the classical Weston cell standard producing a voltage of 1.018 V. The difficulties consisted in the necessity of using a voltage divider for comparing the two standards, with the quantum standard kept in a helium cryostat, and the Weston cell standard requiring room temperature. The voltage divider contributes to the total uncertainty of the comparison result. On the other hand, using a number of Josephson junctions connected in series and biased by a DC current proved very problematic because of diverse current characteristics of the junctions and different intensity of the electromagnetic field applied to each junction when produced by a single source. Each junction in such a standard system requires individual bias current. Worthy of notice is a voltage standard built in Japan, with twenty individually biased Josephson junctions producing a voltage of 100 mV [8].

A milestone in the construction of voltage standards was achieved by using strongly hysteretic tunneling Josephson junctions in weak electromagnetic fields, a concept proposed by Levinson [16]. The current-voltage characteristic of such a junction (Fig. 4.9b) is split at $I = 0$ to a number of lines referred to as zero-crossing steps (or Shapiro steps). The step height ΔV on the voltage axis of the characteristic fulfils the relation: $\Delta V = (h/2e)f_e$, and is equal for every junction placed in electromagnetic field produced by the same source (typically, a Gunn oscillator). Therefore, in a weak electromagnetic field thousands of Josephson junctions connected in series, without bias current ($I = 0$), act as a quantum voltage standard producing a voltage that fulfils the formula:

$$V_{ref} = Nk\frac{h}{2e}f_e, \tag{4.24}$$

where N is the number of Josephson junctions in the series and k is the number of the line (Shapiro step) in the characteristic of the junction, shown in Fig. 4.9b. In practice k ranges from 0 to 8.

In an integrated circuit (IC) with N Josephson junctions exposed to electromagnetic radiation a part of the junctions will contribute to the total voltage with a voltage $V = k(h/2e)f_e$, a part with $V = (k \pm 1)(h/2e)f_e$, and the remaining junctions (representing 10–20 % of the total) will have zero voltage, i.e., $k = 0$. The number of output voltage steps, defined $N \times k$, in the current-voltage characteristic of a Josephson voltage standard is determined every time the standard system is switched on.

The following necessary conditions must be fulfilled in a voltage standard using a large number of Josephson junctions:

- All the junctions must be connected in series;
- The electromagnetic field (radiation) applied to the junctions must be homogeneous.

The latter requirement is fulfilled by placing a microstrip line or a waveguide inside the integrated circuit with of Josephson junctions, just above the array of thousands junctions. Propagating along the array, a sine microwave signal (usually with a frequency of the range from 70 to 75 GHz), irradiates the junctions, as shown in Fig. 4.11.

Since the Josephson oscillations must be synchronized with the frequency of the microwave radiation, the frequency f_e should be chosen from the range $f_c/2$ to $2f_c$, where f_c is the characteristic frequency of the Josephson junction (see Sect. 4.3).

Built in 1984, the first quantum standard producing a voltage of up to 1.2 V consisted of an array of 1474 Josephson junctions and was the fruit of united efforts of Physikalisch-Technische Bundesanstalt (PTB) in Germany and the National Bureau of Standards (NBS), now National Institute of Standards and Technology (NIST), in the United States [21]. In 1986 J. Niemayer at PTB measured the voltage difference between two arrays of Josephson junctions with a nominal voltage of 1 V each. Both arrays were exposed to a 70 GHz radiation from the same source and biased with equal current. Also the number of quantization steps was equal in both arrays. The voltage difference ΔV_{ref} between the arrays was measured with a SQUID. The measurements indicated zero voltage difference with an uncertainty of 7×10^{-13} [21].

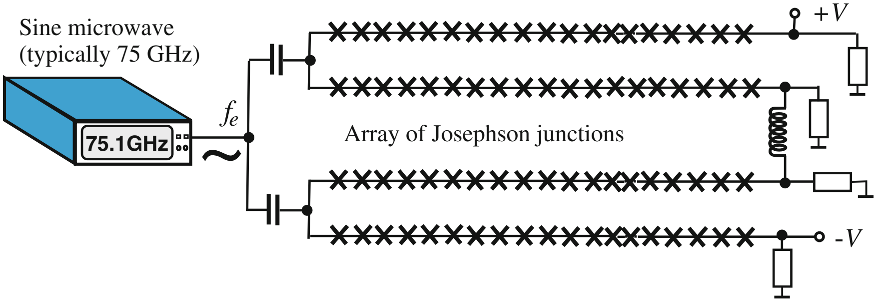


Fig. 4.11 DC Josephson voltage standard (concept borrowed from [3])

In the following year Kautz and Lloyd found that the voltage difference between two 1 V Josephson junction arrays irradiated by the same source was below 2×10^{-17} V [15]. Therefore, it is considered to be experimentally verified that the voltage produced by a standard comprising a fixed number of Josephson junctions is independent of factors other than the frequency of the applied electromagnetic radiation.

The group managed by J. Niemayer at PTB and C.A. Hamilton's team at the NIST were leaders in the design and realization of Josephson voltage standards in the 1990s. Both teams constructed standards consisting of 3000–21 000 $Nb/Al_2O_3/Nb$ Josephson junctions connected in series and irradiated with microwaves with a frequency ranging from 70 to 90 GHz. Each Josephson junction in the array contributed to the total output voltage with about 600–750 μV (at 70 GHz irradiation) because of the activation the 4th or 5th Shapiro step in its voltage-current (V–I) characteristic (see Fig. 4.9b). The controllable voltage produced by these quantum voltage standards ranges from 1 to 10 V [31]. Parameters of Josephson junctions used in both DC and AC voltage standards are presented in Table 4.4.

International voltage comparisons carried out by the BIPM, which sent its quantum voltage standard to different national laboratories, proved voltage coherence with an uncertainty of 10^{-9}, two orders of magnitude better than that obtained in Weston cell standard comparisons [26, 31]. Such measurements are performed as routine procedures in many countries. The measurements of two BIPM 10 V voltage standards in 2002 revealed a voltage difference of 30 pV (3×10^{-12}) with an average standard deviation of 40 pV [25]. Electronic voltage standards with thousands of Josephson junctions connected in series are used in about one hundred metrology laboratories in the world, and since 1998 also at the Central Office of Measures (GUM) in Warsaw.

DC quantum voltage standards operating at a temperature of 4 K are available commercially. Standards with Josephson junction arrays developed in Germany (by PTB and IPHT) are offered by Supracon (Germany), and standards with Josephson junction arrays developed in the United States are offered by Hypres (US). Commercial voltage standards are capable of producing a programmable DC voltage ranging from –10 to +10 V.

4.4.3 AC Voltage Josephson Standards

The successful use of Josephson junctions for producing precise DC voltage was followed by the development of programmable voltage standards and subsequently AC voltage standards. It soon became apparent that underdamped Josephson junctions, with a hysteretic voltage-current (V–I) characteristic, could not be used for this purpose. A disadvantage of hysteretic Josephson junctions is the difficulty of obtaining an exact number of desired Shapiro steps, since the zero-cross step region in their V–I characteristic is unstable. Overdamped junctions, with a non-hysteretic V–I characteristic, are more suitable for programmable voltage standards.

SNS (Fig. 4.6c) or SINIS (Fig. 4.6d) junctions are used in AC voltage standards. Two entirely different systems are used in the construction of AC Josephson voltage standards:

- Binary divided junction arrays (with sections of 1, 2, 4, …, 2^n junctions), $V_{out}(t) = k(t) \times (h/2e) \times f_e$, where $k(t)$ is the number of junctions switched on;
- Pulse-driven arrays, $V_{out}(t) = k \times (h/2e) \times f_e(t)$, where k is the total number of junctions in the array.

The junctions should be maintained in the superconducting phase, which requires a low temperature of 4 K.

4.4.3.1 Voltage Sources with Binary Divided Josephson Junctions

A voltage source with binary divided Josephson junctions is a digital-to-analog converter. Since the junctions are biased with a small current I_b in order to produce the first Shapiro step, the voltage V_J across each Josephson junction is a quantum of the output voltage of the digital-to-analog converter (see Fig. 4.12).

The application of a bias current to a junction section results in the appearance of a Josephson voltage in this section. The desired output voltage is programmed in the form of an n-bit digital word: $a_{n-1}a_{n-2}a_{n-3}\ldots a_2a_1a_0$, where a is 0 or 1, and the binary code number of the word is: $a_{n-1}\,2^{n-1} + a_{n-2}\,2^{n-2} + a_{n-3}\,2^{n-3}\ldots a_2\,2^2 + a_1\,2^1 + a_0\,2^0$.

For example, in a 10-bit digital-to-analog converter the digital control signal 1010101011 will result in the following output voltage V_{out}:

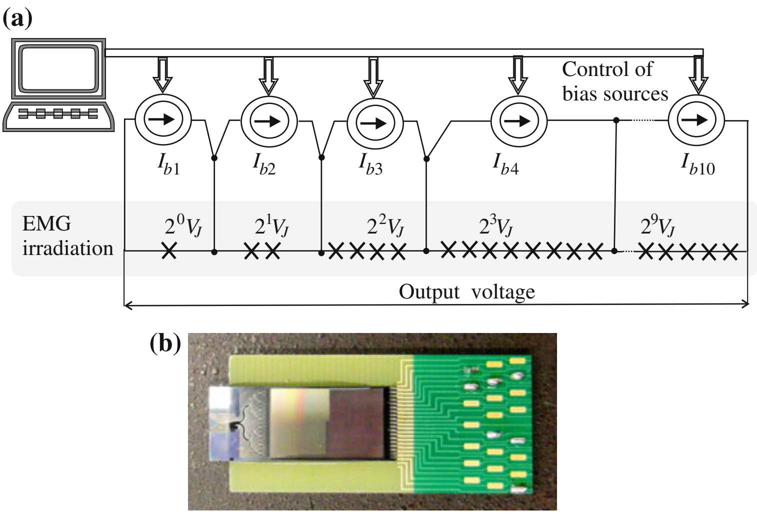


Fig. 4.12 Programmable 10-bit Josephson voltage standard: **a** diagram of a circuit (*top*); **b** a matrix of Josephson junctions (*bottom*)

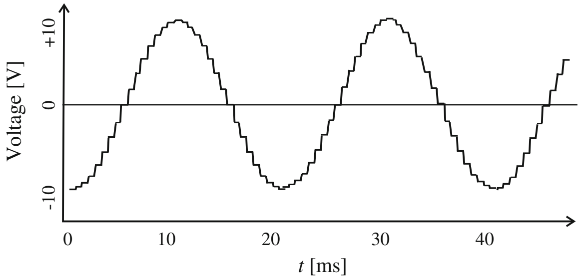


Fig. 4.13 AC signal generated by a programmable Josephson voltage source

$$V_{out} = \left(1 \times 2^9 + 0 \times 2^8 + 1 \times 2^7 + 0 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0\right) \times \mathrm{V}_J = 683V_J..$$

In the case of a non-hysteretic Josephson junction generating voltage at the first Shapiro step (see Fig. 4.9a) and irradiated by microwaves with a frequency of 75 GHz (V_J = 155.2 μV), this means:

$$V_{out} = 683 \times \frac{h}{2e} f_e = 106 \text{ mV}.$$

Important parameters of programmable voltage standards and AC Josephson voltage sources are the voltage V_{rms}, the frequency f_{AC} of the AC voltage, the resolution in the number of bits in control words, and the sampling frequency $f_{samp} = nf_{AC}$, where n is the number of samples per period of the AC signal. Figure 4.13 shows the time dependence of an AC voltage with Shapiro steps that can be seen as voltage jumps with the sampling frequency.

The output voltage is not quantized within each of these transient states (jumps). Moreover, because of the sampling process the output voltage signal is not purely sinusoidal—Fig. 4.13. The frequency spectrum of the output voltage includes numerous harmonics, the largest of which has the sampling frequency f_{samp}. Current AC standards with binary divided junctions are capable of producing voltage with an amplitude of up to 10 V (50 Hz) and a frequency of up to 2,500 Hz.

4.4.3.2 Voltage Standard with Pulse-Driven Josephson Junctions

An entirely different method for the generation of AC voltage with an array of Josephson junctions was proposed and realized by Benz and Hamilton [3, 11], which used an array of non-hysteretic junctions (Fig. 4.9a) generating voltage at the first Shapiro step. Since the voltage V_J across a single Josephson junction depends on the irradiation frequency f_e by the formula $V_J = (h/2e)f_e$, the voltage can be changed also by using a time-varying frequency $f_e(t)$. However, substantial changes in the frequency of a microwave sine signal are ineffective, since they result in breaking the synchronization with the Josephson oscillations within the junction and reduce, even

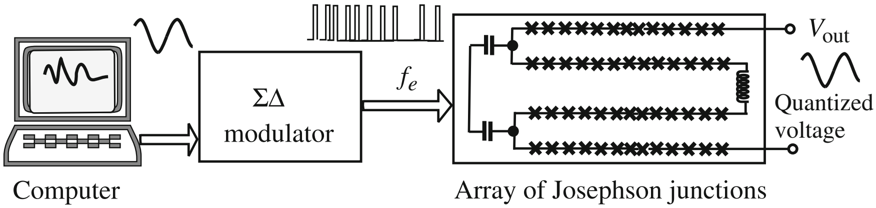


Fig. 4.14 Block diagram of the Josephson arbitrary waveform synthesizer

to zero, the voltage generated by the junction. In their paper [3] Benz and Hamilton proposed irradiation of an array of Josephson junctions with microwaves in the form of a sequence of pulses with a constant width and a variable frequency. The voltage across the Josephson junction array will be higher for pulses with a higher frequency (a denser sequence of pulses), and lower for pulses with a lower frequency. Thus, time-dependent voltage can be generated by varying the frequency of the microwave pulses in the sequence. Previous simulations and subsequent measurements showed that the voltage across a junction irradiated by pulses will not vanish even for the lowest frequencies. This method is the principle of operation of a system referred to as the Josephson arbitrary waveform synthesizer (JAWS). The block diagram of the JAWS is shown in Fig. 4.14.

A generator of an analog signal $v(t)$ with an arbitrary shape of time dependence controls a sigma-delta modulator, which generates a sequence of microwave pulses with a constant width and a variable frequency proportional to the instantaneous voltage: $f(t) \sim v(t)$. The pulses are connected to an integrated circuit containing an array of Josephson junctions. Current pulses result in electromagnetic radiation within the integrated circuit, and the irradiation of the Josephson junctions. The array of Josephson junctions produces an AC output voltage $v_{J\ out}(t)$, the form of which corresponds to the form of the control voltage, and which has definite instantaneous values. Such a source of AC voltage does not generate errors related to sampling or to transient states.

4.4.4 Voltage Standard at GUM

The Josephson voltage standard system in the Laboratory of Voltage and Resistance Standards at the Central Office of Measures (GUM) in Warsaw [28] comprises the following components (Fig. 4.15): an integrated circuit with Josephson junctions, a microwave oscillator with a waveguide and a microwave radiator, an oscilloscope, a frequency meter controlled by a standard atomic clock, a precision digital multimeter, two sets of Zener diode-based secondary standards [7] and a computer with NISTVolt operating software. The integrated circuit with Josephson junctions and the waveguide with a microwave antenna are placed in a Dewar can with liquid helium of temperature 4.2 K. The output voltage V_{ref} produced by the standard is

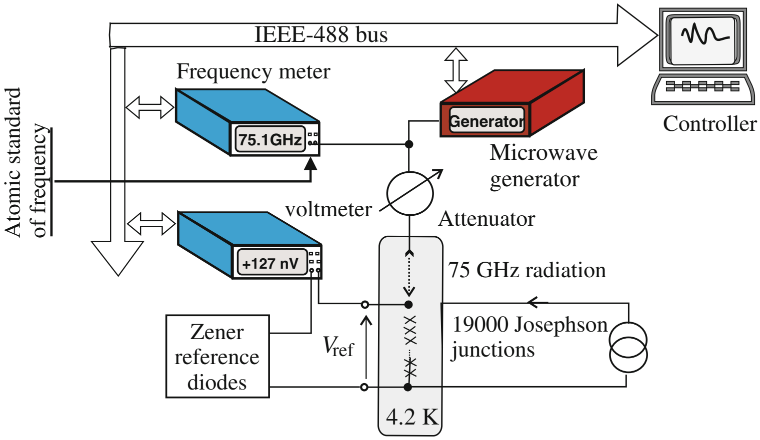


Fig. 4.15 Block diagram of the Josephson voltage standard system at the Central Office of Measures in Warsaw

generated in the integrated circuit, which includes an array with approximately nineteen thousand Josephson junctions made of superconductor $Nb/Al_2O_3/Nb$ and dielectric SiO_2. The Josephson junctions are exposed to electromagnetic field with a frequency of ca. 75 GHz (in the microwave range) generated by the 40 mW microwave radiator.

The high-frequency signal is generated by a microwave oscillator with a Gunn diode. Generated by the atomic frequency standard in GUM Laboratory of Time and Frequency Standards, a reference 10 MHz frequency signal allows precise measurements of the microwave frequency f_e. The frequency standard is a caesium atomic clock with a base frequency of 9192.631770 MHz. The uncertainty of this frequency reference is of the order of 10^{-14}, five orders of magnitude lesser than that of the voltage reference.

The desired value of voltage reference is obtained by setting the radiation power at the damper output and automatic tuning of the microwave frequency. The voltage reference is transferred to the secondary standards, type Fluke 732B, with Zener diodes and internal temperature stabilization controlled by thermistor sensors.

The Josephson junction-based primary standard has nominal output voltage 10 V, and the secondary standards have two nominal output voltages: 10 and 1.018 V. The microwave frequency is set in the range from 74 to 77 GHz. The obtained fixed voltage ranges from –10 to +10 V.

Measurements of the quantum voltage standard carried out at GUM in 1999, with the participation of the author, yielded the following results:

- Voltage reference: V_{ref} = +9.9999928245 V,
- The number of active junctions in the array: 18992,
- Microwave frequency: 76.957 8995 GHz,
- Combined relative standard uncertainty: 6.47×10^{-8}.

A series comparison method is used for transferring the voltage produced by the primary standard (the Josephson junction array) to the secondary standards (Zener diodes in the Fluke 732B set). The difference between the output voltage of the two standards is indicated by an HP 3458A precision digital voltmeter. The relative standard uncertainty of voltage unit realization by the primary standard (without taking into account the uncertainty of the Josephson constant value) is below 2×10^{-9}.

The relative standard uncertainty depends on the uncertainty of microwave frequency and the uncertainty of measurement of leakage current in the circuit with Josephson junctions. As regards the voltage transfer to the secondary standards, the corresponding relative standard uncertainty (calibration uncertainty) does not exceed 10^{-7} and depends on the type A and type B uncertainties. The type A uncertainty stems from the irreproducibility of voltmeter reading, unstable offset voltage and its drift during measurement. The type B uncertainty is caused by imperfect correction of the offset voltage and by the gain error of the voltmeter.

The Josephson voltage standard system for producing and transferring the unit of voltage at the Central Office of Measures in Warsaw is an automated computer-aided measurement system using an IEEE-488 parallel interface and a DACSON-I I/O card (Fig. 4.16).

The cryogenic probe with the Josephson junction array (integrated structure) is placed in a Dewar flask with liquid helium at a temperature of 4.2 K. The NISTVolt software installed in the computer allows to trace metrological characteristics and perform operations such as calibration and linearity measurements of the digital voltmeter, calibration of the secondary standards used as a reference for lower–

Fig. 4.16 Josephson voltage standard system at GUM

order standards, data processing and storage, microwave frequency control and internal check of the measurement system.

Developed at the National Institute of Standards and Technology (USA) and written in QuickBASIC, NISTVolt is designed for high-accuracy voltage measurements with a Josephson voltage standard. Measurement results are displayed in tables or plotted with a possibility of report printout. The main menu gives access to fifteen operations necessary for using the standard. The most important of them include:

- Plotting voltage, frequency or temperature versus time,
- Microwave frequency synchronization,
- Calibration of standards,
- Microwave frequency settings,
- Josephson junction diagnostics,
- Measurements of linearity and accuracy of the digital multimeter,
- Measurements of the multimeter gain error,
- Calibration of the HP 3458A digital voltmeter.

The standard system is placed in a laboratory room screened against radio interferences (screen damping >80 dB) with ambient temperature stabilized at (23 ± 0.3) °C. To minimize the error due to the thermoelectric effect the standard is kept away from other heat-emitting devices, as recommended by the manufacturer. Following the recommendations regarding the operating conditions allows a significant reduction of the type B measurement uncertainty caused by external influences. The block diagram of the Josephson voltage standard system used at GUM is shown in Fig. 4.14.

4.4.5 Comparison GUM Standard with the BIPM Standard

An indirect preliminary comparison of primary voltage standards between GUM and BIPM was carried out by means of a Fluke 732B secondary standard of nominal voltage values 10 and 1.018 V. The secondary standard employed for the comparison was one of the eight standards of the same type used in the Josephson voltage standard system at GUM. The results of the comparison measurements, as reported by BIPM and GUM, are compiled in Table 4.5. The comparison provided a basis for an indirect rough evaluation of the equivalence between the BIPM and GUM primary voltage standards. In the measurements carried out in BIPM and GUM laboratories the secondary standard was referred directly to the primary standard with Josephson junctions [28].

Both indirect and direct bilateral comparisons of primary voltage standards between GUM and BIPM were performed in 2001 by the procedure required in key comparisons. Direct bilateral comparisons are carried out by BIPM and those laboratories which have their own primary standard. The uncertainty values specified in Table 3.4 refer to the standard uncertainty and take into account the

Table 4.5 Results of the indirect bilateral comparison of quantum voltage standards between GUM and BIPM

Measurement data from BIPM and GUM	BIMP		GUM	
	10 V	1.018 V	10 V	1.018 V
Standard voltage (V)[a]	9.9999878	1.0181772	9.9999893	1.0181779
Drift coefficient (nV/day)	– 38.1 ± 1.6	1 ± 0.4	– 38.8 ± 17.8	8.2 ± 3.8
Pressure coefficient (nV/hPa)	19.4 ± 0.4	1.924 ± 0.06	–	–
Atmospheric pressure (hPa)	1013.25	1013.25	994.2	994.2
Drift correction (nV)	–	–	–2.286	60
Temperature correction (nV)	–	–	236.8	–2.4
Pressure correction (nV)	–	–	369.6	36.6
Corrected GUM result (V)	–	–	9.9999877	1.0181779
Difference between BIPM and GUM results (μV)	0.1	−0.7	–	–
Combined uncertainty (nV)	141	14.1	484	99
Comparison uncertainty (nV)	505	100	–	–
Deviation index	0.28	7.2	–	–

[a] Unofficial comparison, measurement data as reported by GUM, 99-01-10, and BIPM, 99-03-10

corrections made. A measure of coherence between the results obtained in the two laboratories, the deviation index is defined as the ratio of result difference to comparison uncertainty. Deviation index value below one (corresponding to a result difference within the comparison uncertainty) means sufficient coherence of the results.

The Josephson voltage standard system at GUM allows to reproduce the unit of voltage with uncertainty 2×10^{-9}, two orders of magnitude better than that of the classical Weston cell standard at GUM. The quantum voltage standard at GUM has been adopted as the legitimate Polish national standard.

4.4.6 Precision Comparator Circuits

Besides being used as a voltage reference, Josephson junctions in high-frequency electromagnetic fields can be employed as a source of voltage reference in comparator circuits for measuring small voltage changes. One of the first applications of this type was a Josephson junction-based comparator circuit used for measuring changes in the electromotive force produced by a Weston cell.

The indirect comparison showed a good equivalence between the 10 V quantum voltage standard at GUM and the international 10 V voltage standard at BIPM. The required corrections for drift, temperature difference and pressure difference during the measurements at GUM and BIPM were taken into account. The result of the

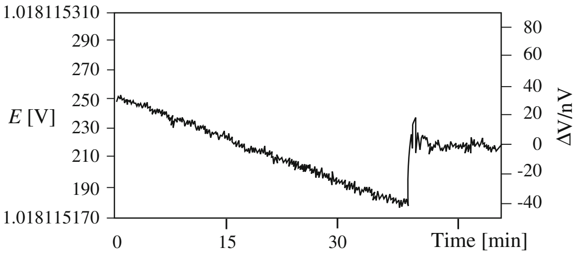


Fig. 4.17 Measured with a quantum voltage standard, a change in electromotive force of a Weston cell caused by a temperature change $\Delta T = 1$ mK [21]

comparison is similar to those obtained in comparisons of other countries' national standards to the international BIPM standard.

Niemayer at PTB recorded a change in EMF caused by a 1 mK stepwise temperature change [21] (Fig. 4.17). The EMF rose by 70 nV, approximately, 40 min after the temperature change. Measurements of this type allow to select the best Weston cells, i.e. those producing the most stable EMF, from a group of manufactured cells.

4.5 Superconductor Digital Circuits

4.5.1 Prospective Development of Semiconductor Digital Circuits

The evolution of semiconductor digital circuits is hampered by two major limitations: limited switching frequency, and limited performance of heat sinks absorbing heat from large-scale integration (LSI) circuits. The first crisis in the development of semiconductor digital circuits occurred in years 1994–1995, when forced cooling proved necessary in the fastest processor at that time, Alpha 21164 (manufactured by DEC), running at 500 MHz with a 60 W power input. The 1994 National Roadmap for Semiconductors (ITRS) [29] forecast the clock rate of digital systems to reach 1.1 GHz in 2010. However, systems with this clock frequency appeared as soon as in 2000, and the typical clock rate of currently available multicore processors is 3 GHz (as of 2014). Regardless of the missed forecast (the progress in integrated circuit technologies is faster than predicted), the increase in clock rate of both silicon and gallium arsenide semiconductor circuits is constrained by the growing quantities of heat to be absorbed from integrated circuits operating with still higher frequencies. Table 7.1 in Chap. 7 presents data from the 2013 ITRS [29] with the current forecast of the semiconductor component evolution. Note the predicted increase in clock rate of semiconductor circuits assumes a reduction of their supply voltage, which implies reduced power consumption and heat dissipation.

4.5.2 Digital Circuits with Josephson Junctions

Because of the above-discussed constraints of the evolution of semiconductor digital circuits, superconductor digital circuits are considered for prospective use in digital systems, particularly in ultrafast computers. Superconductor circuits have two major advantages:

- High operating frequency (thousands of gigahertz), demonstrated theoretically and partly verified experimentally,
- No ohmic resistance of circuit components and intercircuit connections (paths), implying a substantial reduction of dissipated heat; moreover, the signal voltage levels range from 1 to 10 mV.

A closed superconducting circuit can be used as a memory of a superconductor digital system.

The idea of data storage based on the effect of magnetic flux quantization in superconducting loops was proposed by T. Clark in 1967. The total magnetic flux Φ passing through a circuit is a multiple of the magnetic flux quantum Φ_0: $\Phi = n\Phi_0$ (see formula (5.6) in Chap. 5), and the integer n can be used for encoding the counted quanta. Although the time necessary for the magnetic flux in a superconducting material to change by a single quantum, or the flux switching time, is of the order of hundreds of picoseconds, an analysis performed by Konstantin Likharev proved the switching process can be much faster if at least one Josephson junction is included in the superconductor loop (e.g. a superconducting ring) [17] (Fig. 4.18a).

When a magnetic field is applied to the ring, the incorporated Josephson junction allows fast single flux quanta penetration into or expulsion from the ring. In a niobium ring the process only takes a fraction of a picosecond. This means the digital data encoded by the number n of magnetic flux quanta can be stored and processed at very high rates.

Unlike the applied external magnetic flux Φ_e, which can take on any value, the flux Φ inside the ring is quantized. The difference between Φ_e and Φ is related to the superconducting current I throughout the ring:

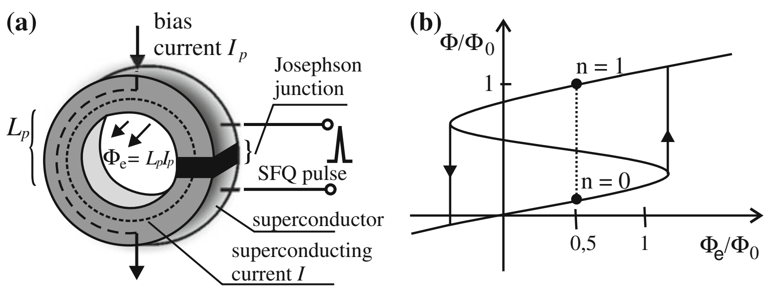


Fig. 4.18 SFQ dynamic memory cell: **a** superconducting circuit with Josephson junction; **b** quantization of internal magnetic flux Φ in the circuit versus external flux Φ_e (*dots* represent flip-flop stable states)

$$\Phi_e = \Phi + LI, \tag{4.25}$$

$$\Phi = n\Phi_0, \tag{4.26}$$

where Φ_e is the external magnetic flux applied to the closed superconducting circuit (ring), Φ is the magnetic flux inside the superconducting ring, L is the self-inductance of the ring, I is the current flowing around the ring, and consequently through the Josephson junction (superconducting current), and n is the number of magnetic flux quanta Φ_0 inside the ring.

In digital circuits the external magnetic flux Φ_e is usually generated by the bias current I_b in a part of the ring: $\Phi_e = L_b I_b$ (L_b being the inductance of the part of the ring in which the bias current flows). Obviously, the flux Φ_e can also be produced by an external source galvanically separated from the ring under consideration. The current I through the Josephson junction is described by the first Josephson equation (4.14):

$$I = I_C \sin\varphi.$$

The relation between the magnetic flux inside the superconducting ring and the wave function phase difference is expressed by the basic formula (5.12) (see Chap. 5):

$$\varphi = 2\pi \frac{\Phi}{\Phi_0}, \tag{4.27}$$

$$\varphi + \frac{2\pi L I_C}{\Phi_0} \sin\varphi = 2\pi \frac{\Phi_e}{\Phi_0}. \tag{4.28}$$

The above equation can be written as:

$$\varphi + \lambda \sin\varphi = 2\pi \frac{\Phi_e}{\Phi_0}, \tag{4.29}$$

where $\lambda = \frac{2\pi L I_C}{\Phi_0}$ is a parameter of the superconducting ring.

Equations (4.27) and (4.29) describe the switching properties of a circuit (ring) with a Josephson junction: the circuit has such properties if $\lambda > 1$. The external magnetic flux Φ_e causes different wave function phases, and consequently, different stable states of the internal magnetic flux, according to (4.26). For example, for $\Phi_e = 0.5\Phi_0$ the internal magnetic flux in a ring with $\lambda = 2\pi$ fulfils two relations: $\Phi/\Phi_0 = 0.25$ and $\Phi/\Phi_0 = 0.93$, as indicated in Fig. 4.18b. Calculated from (4.29), the number of stable states of the ring, or the number of magnetic flux quanta carried by the internal magnetic flux Φ, is λ/π. In a ring with $\lambda = 2\pi$ only two stable states occur, and the ring can be used as a binary element.

In superconductor digital circuits the logic zero corresponds to a state with no magnetic flux quanta inside the ring, $n = 0$. In the case referred to by Fig. 4.18b this state occurs at $\Phi = 0.25\Phi_0$ for the specific value of external flux $\Phi_e = 0.5\Phi_0$. For a different value of Φ_e a different value of Φ corresponds to $n = 0$. The logic one is a state with $n = 1$ (in Fig. 4.18b the corresponding internal magnetic flux value is $\Phi = 0.93\Phi_0$). Switching between two stable states occurs as a result of external magnetic flux changes due to changing the bias current I_b.

Faraday's law of electromagnetic induction is employed for data transfer between digital circuits. A change in magnetic flux in a closed circuit generates a voltage $v = -\frac{\mathrm{d}\Phi}{\mathrm{d}t}$, which is held by the Josephson junction as the superconducting loop cannot support the electric field. In particular, a change in magnetic flux by a single quantum Φ_0 generates a voltage pulse $v = -\frac{\mathrm{d}\Phi_0}{\mathrm{d}t} = 2\,[\mathrm{mV/ps}]$. The data transfer is dynamic. A single flux quantum (SFQ) can be transferred over a transmission line to other circuits in a multipart system even if the line does not have the superconducting properties.

A family of super-conductor combinatory and sequential logic circuits with SFQ elements were proposed by Likharev et al. [18]. K. Likharev has considerable achievements in the construction of superconductor digital circuits as well as in the research on the dynamics of Josephson junctions, in the field of single electron tunneling and in the theory of noise in SQUIDs. Digital circuits with SFQ elements are referred to as Rapid Single Flux Quantum (RSFQ) circuits. The simplest RSFQ circuit is a reset-set (RS) bistable flip-flop (shown in Fig. 4.19), consisting of a superconducting circuit (e.g. a ring) with two incorporated Josephson junctions. An external magnetic flux $\Phi_e = \Phi_0/2$, generated by the bias current, provides equal conditions for the occurrence of two possible stable states of the flip-flop, corresponding to $n = 0$ and $n = 1$. The flip-flop is switched from state 0 to state 1 by a set pulse, which transmits a single flux quantum through Josephson junction JJ1. A reset SFQ pulse restores the initial state by expelling the flux quantum from inside the ring. While the flip-flop passes from state $n = 1$ to $n = 0$, the other Josephson junction, JJ2, generates an output SFQ pulse, which can be used for switching other RSFQ elements [9].

The internal speed of an RSFQ system is determined by the time τ_o of duration of an SFQ pulse [18]:

$$\tau_o = \frac{\Phi_0}{2\pi I_c R_n}, \tag{4.30}$$

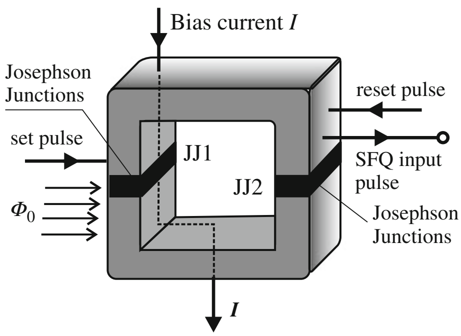


Fig. 4.19 RS bistable flip-flop, a super-conductor digital circuit of a family of RSFQ circuits with two Josephson junctions (JJ1 and JJ2)

Table 4.6 Parameters of RSFQ digital integrated circuits made of Nb/AlAlO$_x$/Nb low-temperature superconductors

Technology	–	Hypres	PARTS	VLSI
Junction size a (the smallest dimension)	μm	3	1.5	0.4
Critical current density j_c	kA/cm^2	1	6.5	100
Voltage I_cR_N	mV	0.3	0.6	1.5
Minimum power P per logic gate	nW	30	60	150
SFQ pulse duration time τ_o	ps	1	0.5	0.2
Maximum operating frequency f	GHz	20–100	40–80	100–770

where I_c is the critical current of the Josephson junction and R_n is the shunt resistance (an extra resistor or the normal resistance of the junction).

The approximate minimum time of duration of an SFQ pulse can be evaluated from the formula:

$$\tau_{\min} \cong \frac{0.18\hbar}{k_B T_c}, \tag{4.31}$$

where $\hbar = h/2\pi$.

Estimated from (4.31), the approximate minimum pulse duration time in niobium (T_c = 9.3 K) SFQ circuits is $\tau_{min} \cong 0.15$ ps, and in high-temperature superconductor (T_c = 90 K) SFQ circuits $\tau_{min} \cong 0.015$ ps.

An RSFQ element of the simplest type can operate with frequency up to $1/2\tau_o$; the operating frequency of an IC circuitry will be 3 to 6 times lower. Parameters of commercial and prototype low-temperature superconductor RSFQ integrated circuits are presented in Table 4.6.

Hypres offers Nb/AlAlO$_x$/Nb superconductor digital circuits (parameters as specified in Table 4.6) manufactured by a HYPRES niobium trilayer technology [12]. Measured by the junction size a, the workmanship precision of HYPRES circuits is relatively low (a = 3 μm). An on-chip oscillator developed by HYPRES has a quality factor of the order of 10^6 and operating frequency ranging from 10 to 100 GHz [10]. Higher-precision PARTS and VLSI prototype technologies, developed in collaboration with IBM and Bell Laboratories and enhanced at the Stony Brook University, USA, offer better circuit packing and higher switching frequencies [19]. For example, an RSFQ frequency divider discussed in [4] runs at up to 770 GHz.

4.6 Other Applications of Josephson Junctions

4.6.1 Voltage-to-Frequency Converter

Josephson junctions are also used in many other cryoelectronic systems. These include SQUID detectors, voltage-to-frequency converters and, prospectively, sources of terahertz radiation.

Josephson junction-based voltage-to-frequency converters use the AC Josephson effect, which has two forms:

- A Josephson junction exposed to electromagnetic radiation generates a constant voltage V, which only depends on the radiation frequency and the physical constants e and h; the Josephson effect in this form is used in quantum DC voltage standards;
- A constant voltage V applied to a Josephson junction generates oscillations with frequency f, which only depends on the voltage V and the physical constants e and h:

$$f = \frac{2e}{h}V = K_J \times V = 483597.9\ (\text{GHz/V}) \times V. \tag{4.32}$$

Thus, the Josephson junction acts as a voltage-to-frequency converter with high conversion coefficient K_J, which can be determined very precisely (a 1 μV DC voltage, applied to a Josephson junction, induces oscillations with frequency of 484 MHz). The high values of K_J and the precision of its determination make the Josephson junction a good voltage-to-frequency converter for processing the weak signal in low-temperature systems. Because of the low power of the oscillating signal in the junction, a cryoelectronic amplifier is indispensable in systems with a Josephson junction converter. The amplifier can be an rf-SQUID quantum magnetometer, which uses a detector incorporating one Josephson junction. Also the detector junction can be used as a voltage-to-frequency converter. The RF-SQUID detector and magnetometer are discussed in this chapter. In cryogenics the Josephson junction can be used e.g. for measuring the thermal electromotive force (emf), or the voltage V_T created by a temperature difference (Fig. 4.20):

$$V_T = k_T(T_x - T_1),$$

$$f_x = \frac{2e}{h}V_T = k_2(T_x - T_1).$$

The thermal EMF, V_T, is proportional to the temperature difference. Applied to a Josephson junction, the voltage V_T generates oscillations with frequency f_x. The

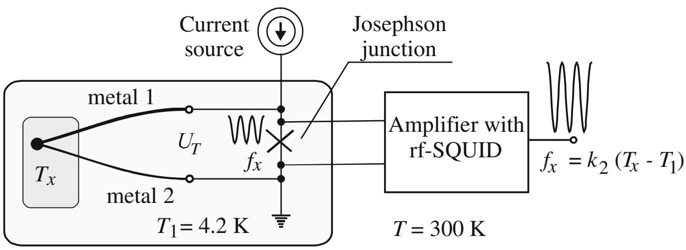


Fig. 4.20 Josephson junction-based voltage-to-frequency converter for measurements of the thermal electromotive force

Josephson junction-based voltage-to-frequency converter can also be used in noise thermometers for low-temperature applications. This application is discussed in this chapter, which also explains the operation of the SQUID detector used as a superconductor analog-to-digital converter.

4.6.2 Source of Terahertz Radiation

Terahertz (THz) radiation, or T-rays, are electromagnetic waves with frequency ranging from 300 GHz to 3 THz (corresponding to the wavelength range from 100 μm to 1 mm). Propagating in a straight line, terahertz waves do not penetrate conducting materials such as metals or water and are strongly absorbed by the Earth's atmosphere, which contains water vapor. At the same time, terahertz radiation can penetrate both soft and hard human body tissues as well as materials such as leather, plastics, paper, cardboard, wood, ceramics or stones. As non-ionizing radiation, T-rays are not detrimental to health, or at least much less detrimental than, for example, X-rays. Because of its strong absorption by the atmosphere, terahertz radiation is of little use as a medium of wireless transmission. Still, it is of much interest for potential applications in medicine (e.g. for three-dimensional imaging of teeth in dentistry) as well as for security applications, such as luggage and body scanning at airports and other places where security screening is used. These applications, however, require efficient sources of terahertz radiation.

Though T-rays are present in the spectrum of black body radiation at temperatures above 10 K, their thermal emission is very weak. Stronger emission is offered by the far infrared laser (FIR), the free electron laser (FEL), the backward wave oscillator (BWO) and some other techniques requiring costly specialist equipment. These remained the only efficient sources of terahertz radiation available until 2007, when an international group of scientists led by U. Welp at the U.S. Department of Energy's Argonne National Laboratory developed a device that allows the construction of a portable source of terahertz radiation emitted by Josephson junctions made of a high-temperature superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ [22]. The optimum operating temperature of the Josephson junctions ranges from 24 to 45 K and the maximum power of T-ray emission is 5 μW (at efficiency of 3×10^{-4}). Already high enough for some applications, this power of emission is considered augmentable to 1 mW by increasing the efficiency of the source to 6 %.

References

1. G. Bednorz, K.A. Müller, Supraleitung in LaBaCuO in 36 K. Z. Phys. **B64**, 189–193 (1986)
2. R. Behr et al., Development and metrological applications of Josephson arrays at PTB. Meas. Sci. Technol. **23**, 124002 (2012)
3. S.P. Benz, C.A. Hamilton, A pule-driven programmable Josephson voltage standard. Appl. Phys. Lett. **68**, 3171–3173 (1996)

4. W. Chen et al., Rapid single flux quantum T-flip-flop operating at 770 GHz. IEEE Trans. Appl. Supercond. **8**, 3212–3215 (1998)
5. J. Clarke, Experimental comparison on the Josephson voltage-frequency relation in different superconductors. Phys. Rev. Lett. **21**, 1566–1569 (1968)
6. M. Cyrot, D. Pavuna, *Introduction to Superconductivity and High Tc Materials* (World Scientific, Singapore, 1992)
7. Documentation of the RMC Josephson standard, 1998
8. T. Endo, M. Koyanagi, A. Nakamura, High accuracy Josephson potentiometer. IEEE Trans. Instrum. Meas. **IM-32**, 267–271 (1983)
9. D. Gupta, Y. Zhang, On-chip clock technology for ultrafast digital superconducting electronics. J. Appl. Phys. **76**, 3819–3821 (2000)
10. C.A. Hamilton et al., A 24-GHz Josephson array voltage standard. IEEE Trans. Instrum. Meas. **IM-40**, 301–304 (1991)
11. C.A. Hamilton, Josephson voltage standards. Rev. Sci Instrum. **71**, 3611–3623 (2000)
12. HYPRES Design Rules, Internet resources: www.HYPRES.com
13. T. Jaw-Shen, A.K. Jain, J.E. Lukens, High-precision test of the universality of the Josephson voltage-frequency relation. Phys. Rev. Lett. **51**, 316–319 (1983)
14. B.D. Josephson, Possible new effects in superconducting tunneling. Phys. Lett. **1**, 251–263 (1962)
15. R.L. Kautz, L. Lloyd, Precision of series-array Josephson voltage standards. Appl. Phys. Lett. **51**, 2043–2045 (1987)
16. M.T. Levinson et al., An inverse ac Josephson effect voltage standard. Appl. Phys. Lett. **31**, 776–778 (1977)
17. K.K. Likharev, Superconductors speed up computation. Phys. World **10**, 39 (1997)
18. K.K. Likharev, O.A. Mukhanov, V.K. Semenov, Ultimate performance of RSFQ logic circuits. IEEE Trans. Magn. **23**, 759–762 (1987)
19. K.K. Likharev, Superconductor devices for ultrafast computing, Section 5 in *Applications of Superconductivity*, ed. by H. Weinstock, NATO ASI Series (Kluwer, Dordrecht, 2000), pp. 247–293
20. D.E. McCumber, Effect of ac impedance on dc-voltage-current characteristics of Josephson junctions. J. Appl. Phys. **39** 3113–3118 (1968)
21. J. Niemayer, L. Grimm, C.A. Hamilton, R.L. Steiner, High-precision measurement of a possible resistive slope of Josephson array voltage steps. IEEE Electron. Dev. Lett. **7**, 44–46 (1986)
22. L. Ozyuzer et al., Emission of THz coherent radiation from superconductors. Science **318**, 1291–1293 (2007)
23. W.H. Parker et al., Determination of e/h using macroscopic quantum phase coherence in superconductors. Phys. Rev. **177**, 639–664 (1969)
24. B.W. Petley, *Quantum Metrology and Fundamental Constants*, ed. by P.H. Cutler, A.A. Lucas (Plenum, New York, 1983)
25. T. Quinn, New from BIPM. Metrologia **39**, 117 (2002)
26. D. Reymann, T.J. Witt, J. Balmisa, P. Castejon, S. Perez, Comparisons of the Josephson voltage standards of the CEM and the BIPM. Metrologia **36**, 59–62 (1999)
27. M. Shukrinv Yu et al., Phase dynamics of two parallel stacks of coupled Josephson junctions. Supercond. Sci. Technol. **27**, 124007 (2014)
28. D. Sochocka, W. Nawrocki, Quantum voltage standard at central office of measures (in Polish). Elektronika **11/42**, 15–18 (2001)
29. The International Roadmap for Semiconductors 2013, Internet resources: http://public.itrs.net/Files
30. T. van Dutzer, G. Lee, Digital signal processing, in *Superconducting Devices*, ed. by S.T. Ruggiero, D.A. Rudman (Academic Press, New York, 1990)
31. C.M. Wang, C.A. Hamilton, The fourth interlaboratory comparison of 10 V Josephson voltage standards in North America. Metrologia **35**, 33–40 (1998)

Chapter 5
SQUID Detectors of Magnetic Flux

Abstract This chapter presents the principles of operation and design of superconducting quantum interference devices, known under the acronym SQUID. RF-SQUIDs, which have a single Josephson junction in a closed magnetic circuit and are biased with radio-frequency signal, are discussed with their parameters provided. At much greater length we discuss DC-SQUIDs, which have two Josephson junctions and are biased with direct current. The equation describing the conversion of the magnetic flux to the output voltage is derived for both the RF- and DC-SQUIDs. We analyze the energy resolution of SQUIDs, citing the record resolution of the DC-SQUID, equal to $0.5h$. We provide many examples of SQUID-based measurement systems, including systems for biomagnetic studies, non-destructive evaluation (NDE) systems and noise thermometers for the measurement of very low temperatures.

5.1 Quantization of Magnetic Flux

A quantum interference effect in a superconducting loop with two Josephson junctions was first observed in 1964 [15]. It is also then that the critical current in a circuit with two Josephson junctions was shown to be a periodic function of the magnetic flux threading the loop, with a period equal to the flux quantum Φ_0. Conclusions drawn from these studies were soon used for the development of a magnetic flux sensor referred to as SQUID (acronym for Superconducting QUantum Interference Device). The SQUID is a magnetic flux to voltage converter. There are two types of these superconductor sensors, the DC-SQUID, biased by direct current and containing two Josephson junctions, and the RF-SQUID, biased by a high-frequency signal and containing a single Josephson junction. The operation of the SQUID sensor is based on two effects characteristic of superconductors: the quantization of magnetic flux in a closed magnetic circuit, and the Josephson effect, in which characteristics of superconductivity appear in Josephson junction. SQUIDs have the *highest sensitivity* of all known sensors of *all physical quantities*.

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_5

The high sensitivity of the SQUID allows, for example, to detect a car at a distance of 1 km.

SQUID sensors are used in the following fields of medicine, science and technology:

- In medicine, for non-invasive diagnostics of human organs based on the measurement of their magnetic field (biomagnetic studies). Since the strongest magnetic field is generated by the heart, magnetocardiography is the most advanced of these techniques;
- In technology, for non-destructive evaluation of materials, a technique that allows to detect flaws in the tested material by measuring the inhomogeneity of distribution of a current-induced magnetic field. SQUIDs are used for non-destructive testing of vital parts of aircrafts and spacecrafts in periodic controls in the US Army;
- In geology, for the localization of mineral deposits by the measurement of magnetic field inhomogeneity at the surface of the Earth;
- In electrical metrology, for the measurement of weak current and voltage signals;
- In thermometry, for the measurement of absolute temperature based on thermal noise. The noise thermometer is a primary thermometer, using fundamental laws of physics and thus not requiring calibration;
- In the army, for coast surveillance against foreign objects, mainly submarines.

The RF-SQUID has only one Josephson junction; probably for this reason its fabrication technology was developed before the DC-SQUID technology. The improvement of the RF-SQUID in the 1970s allowed to obtain the energy resolution as high as 7×10^{-32} J/Hz [22]. Little attention was paid to the DC-SQUID at that time. However, the DC-SQUID was shown theoretically to have a potentially higher sensitivity than the RF-SQUID. Moreover, the development of the thin-film technology allowed the fabrication of two Josephson junctions at the same expenditure as a single junction. By the end of the 1980s the resolution of the DC-SQUID increased by a few orders of magnitude to ca. 3×10^{-34} J/Hz = $0.5h$! [19], whereas the sensitivity of the RF-SQUID improved only slightly.

The above data refer to SQUIDs of niobium or its compounds cooled with liquid helium. These are referred to as low-temperature (LTS) SQUIDs. The discovery of high-temperature superconductors (HTS) in 1986 soon resulted in their application for the fabrication of SQUIDs. Presently the energy resolution of HTS SQUIDs (3×10^{-31} J/Hz) is about one hundred times inferior to that of LTS SQUIDs [18]. Therefore, only LTS SQUIDs have been used in metrology laboratories so far.

The conduction of electricity in a superconductor is based on the motion of electrons bound in pairs known as Cooper pairs. Cooper pairs form a coherent quantum condensate, which can be described by a *single wave function* fulfilling the Schrödinger equation. In the normal conductivity phase single electrons form an electron gas in the material, and each electron in described by a separate wave function. The motion of Cooper pairs in a superconductor is described by a wave function Ψ [23]:

$$\Psi = \Psi_0 \exp[-\mathrm{j}(\omega t - 2\pi \frac{x}{\lambda})], \tag{5.1}$$

where λ and ω are the wavelength and the angular frequency of the wave, respectively, and x is the distance along the direction of propagation.

When current flows in the superconductor, the wave function at points P and Q has the phase difference:

$$\varphi_P - \varphi_Q = \frac{4\pi m}{h e n_s} \int_P^Q \vec{J_s}\, \mathrm{d}\,\vec{l} + \frac{4\pi e}{h} \int_P^Q \vec{A}\, \mathrm{d}\,\vec{l}, \tag{5.2}$$

where m and e denote the electron mass and charge, respectively, h is the Planck constant, n_s the concentration of non-superconducting electrons, $\vec{J}_s$ the current density, $\vec{A}$ the vector potential of the magnetic field, and $d\vec{l}$ an infinitesimal segment of the line between P and Q.

The vector potential $\vec{A}$ and the magnetic induction $\vec{B}$ are related as follows:

$$rot\vec{A} = \vec{B}.$$

From the above relation and (5.2) it follows that the phase difference of the wave function depends on the current density between P and Q, as expressed by the first term at the right-hand side of (5.2), and on the magnetic induction through the second term.

Very important in the theory of the SQUID is the case, illustrated in Fig. 5.1, in which the superconducting current flows in a closed contour in a superconductor S surrounding a normal (non-superconducting) region N. Placed in a magnetic field of induction B, the superconductor exhibits perfect diamagnetism (Meissner effect), while in the N region the magnetic induction is equal to B. The magnetic induction in N induces superconducting current in S.

Let us consider a contour, or closed path, represented by the dashed line in Fig. 5.1. The phase difference at any two points of this path will be as described by (5.2).

For a closed path, e.g. P→Q→P, the phase difference is:

$$(\varphi_P - \varphi_Q) + (\varphi_Q - \varphi_P) = \frac{4\pi m}{hen_s} \oint \vec{J}_s \mathrm{d}\,\vec{l} + \frac{4\pi m}{h} \oint \vec{A}\, \mathrm{d}\vec{l}. \tag{5.3}$$

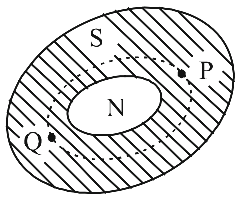


Fig. 5.1 Superconductor *S* around a normal (non-superconducting) region *N*

The Stokes theorem states that:

$$\oint \overrightarrow{A}\,\mathrm{d}\,\overrightarrow{l} = \int\int\limits_{s} \mathrm{rot}\overrightarrow{A}\,\mathrm{d}\,\overrightarrow{S},$$

where $\mathrm{d}\,\overrightarrow{S}$ is the area delimited by the closed path.

By *the Stokes theorem* the term $\oint \overrightarrow{A}\,\mathrm{d}\,\overrightarrow{l}$ can be replaced with the double integral:

$$\oint \vec{A}\,\mathrm{d}\vec{l} = \int\int\limits_{S} \vec{B}\,\mathrm{d}\vec{S}.$$

Thus, (5.3) becomes:

$$\left(\varphi_P - \varphi_Q\right) + \left(\varphi_Q - \varphi_P\right) = \frac{4\pi m}{hen_s}\oint \overrightarrow{J}_s\,\mathrm{d}\,\overrightarrow{l} + \frac{4\pi m}{h}\int\int\limits_{s} \overrightarrow{B}\,\mathrm{d}\,\overrightarrow{S}. \tag{5.4}$$

The phase condition for the wave function describing the motion of the electrons requires that its change in phase on a closed path must equal $2\pi n$, where n is an integer:

$$\frac{4\pi m}{hen_s}\oint \overrightarrow{J}_s\,\mathrm{d}\,\overrightarrow{l} + \frac{4\pi m}{h}\int\int\limits_{s} \overrightarrow{B}\,\mathrm{d}\,\overrightarrow{S} = 2\pi n \tag{5.5}$$

$$\frac{m}{n_s e^2}\oint \overrightarrow{J}_s\mathrm{d}\,\overrightarrow{l} + \int\int\limits_{s} \overrightarrow{B}\,\mathrm{d}\,\overrightarrow{S} = n\frac{h}{2e}.$$

The term $\frac{h}{2e} = \Phi_0 = 2.07 \times 10^{-15}$Vs is the magnetic flux quantum, referred to as fluxon.

The (5.6) below equates, on the right-hand side, the internal magnetic flux Φ_i within the superconductor, representing an integer multiple of the magnetic flux quantum Φ_0, and, on the left-hand side, the sum of the external flux Φ_e and the flux induced by the superconducting current of density J_s:

$$\Phi_i = n\Phi_o, \quad \Phi_e = \int\int \overrightarrow{B}\,\mathrm{d}\,\overrightarrow{S}$$

$$\frac{m}{n_s e^2}\oint \overrightarrow{J}_s\,\mathrm{d}\,\overrightarrow{l} + \Phi_e = \Phi_i \tag{5.6}$$

$$\frac{m}{n_s e^2}\oint \overrightarrow{J}_s\,\mathrm{d}\,\overrightarrow{l} + \Phi_e = n\Phi_0$$

$$\frac{m}{n_s e^2} \oint \vec{J}_s \, \mathrm{d}\vec{l} + \Phi_e = \Phi_i.$$

The (5.6) explains the phenomenon of *quantization of magnetic flux* within a closed superconducting circuit. When the external flux Φ_e acting on the superconductor grows from zero to Φ_0, the internal flux Φ_i remains zero, Φ_e being compensated by the magnetic flux that comes from the superconducting current J_s induced in the circuit. At $\Phi_e = \Phi_0$ the internal flux Φ_i changes in a stepwise manner from zero to Φ_0. After this stepwise change, further increase in Φ_e is again compensated by the flux from the superconducting current until $\Phi_e = 2\Phi_0$, and so forth. Thus, the $\Phi_i(\Phi_e)$ dependence is a step function. Observed in many experiments, the quantization of magnetic flux in a superconducting loop is also apparent in the operation of the SQUID.

The *Josephson effect* (discussed in detail in Chap. 4) takes place in a Josephson junction consisting of two superconductors in the superconducting phase separated by a thin film (1–5 nm thick) of a dielectric material. Besides single electrons, also Cooper pairs tunnel across the dielectric in such a junction. The current I through the junction is the sum of two components, I_e and I_p, resulting from the flow of single electrons and Cooper pairs, respectively:

$$I = I_e + I_p. \tag{5.7}$$

Electrons in Cooper pairs tunnel across the dielectric thin film coherently in terms of the phase of the wave function describing the motion of Cooper pairs.

When a bias voltage lower than E_G/e is applied to the Josephson junction, $I_e = 0$; E_G denotes the bandgap width, i.e. the energy difference between the bottom of the conduction band and the Cooper pair level. The component I_p does not cause any voltage drop across the junction, which behaves as a superconductor. Voltage appears across the junction above a certain critical current I_c of the direct current of density j, known as the Josephson current, flowing through the junction:

$$j = j_c \sin \varphi, \tag{5.8}$$

where j_c is a constant measured in current density units, and $\varphi = (\varphi_2 - \varphi_1)$ is the phase difference of the wave function of Cooper pairs on opposite sides of the junction.

If the junction is placed in a magnetic field of vector potential $\vec{A}$, the density of the Josephson current depends on the magnitude of the magnetic field through the vector potential dependence of the phase:

$$j = j_c \sin\left(\varphi - \frac{4\pi e}{h} \int \vec{A} \, \mathrm{d}\vec{l}\right). \tag{5.9}$$

5.2 RF-SQUID

5.2.1 RF-SQUID Equation

Application studies in superconductivity resulted in the development of a very sensitive measurement device, the superconductor magnetic flux sensor known as the RF-SQUID [23]. According to (5.6), the quantization of magnetic flux in a superconducting closed electrical circuit without a Josephson junction can be expressed as:

$$\Phi_e + LI_i = \Phi_i = n\,\Phi_0, \tag{5.10}$$

where Φ_e is the external magnetic flux acting on the superconducting circuit, Φ_i the internal magnetic flux within the superconductor, L the inductance of the magnetic circuit, and I_i the superconducting current induced in the circuit.

The RF-SQUID is created by including one Josephson junction in the magnetic circuit. The (5.10) still applies to this case, but the current I_i must fulfill the Josephson relation (5.8):

$$I_i = I_c \sin\varphi, \tag{5.11}$$

where I_c is the critical current of the junction.

In the case considered the phase difference φ in (5.11) can be expressed as:

$$\varphi = \frac{2\pi\,\Phi_i}{\Phi_0}. \tag{5.12}$$

The substitution of (5.11) and (5.12) in (5.10) leads to:

$$\Phi_i = \Phi_e + LI_c \sin\frac{2\pi\,\Phi_i}{\Phi_0}, \tag{5.13}$$

where L is the self-inductance of the SQUID.

The (5.13) is the basic equation of the RF-SQUID. The plot in Fig. 5.2 shows the $\Phi_i = \mathrm{f}(\Phi_e)$ dependence as a curve oscillating around the straight line $\Phi_i = \Phi_e$ with a period Φ_0. The slope of this magnetization curve can be positive, negative or infinite.

Depending on the McCumber parameter β_c (the hysteresis coefficient of the Josephson junction):

$$\beta_c = \frac{2\pi L I_C}{\Phi_0},$$

the internal flux Φ_i is either a one-to-one function or a multifunction of the external flux Φ_e (see Chap. 4, (4.19)).

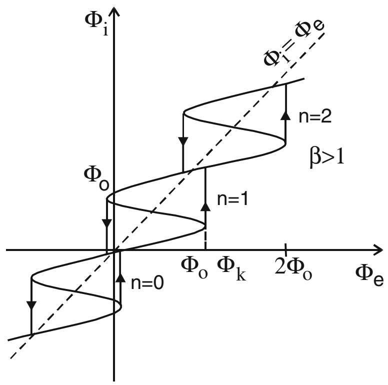


Fig. 5.2 Quantization of magnetic flux in an RF-SQUID

The operation of the SQUID is hysteretic if $\beta_c > 1$. The fabrication of a non-hysteretic ($\beta_c < 1$) RF-SQUID with a high inductance and a low critical current I_c is difficult. High inductance is desirable because of the necessity of coupling the SQUID with the external circuit. For this reason, the SQUIDs in use have $\beta_c > 1$, typically ranging from 3 to 5. Jump points in the $\Phi_i = f(\Phi_e)$ characteristic plotted in Fig. 5.2 correspond to states in which the current I_i reaches the critical value and the SQUID loop passes from the superconducting to normal phase. Consecutive Φ_i values in states in which such transitions are possible differ by Φ_0. Thus, the process is analogous to the quantization of magnetic flux, and Φ_0 is the magnetic flux quantum in the described quantization process. It should not be understood as an indivisible portion of magnetic flux, in contrast to, e.g., the electric charge quantum e, which *is* an indivisible portion of electric charge.

When the SQUID is exposed to an external magnetic field with a high-frequency component of amplitude $\Phi_{er} > \Phi_k$ and a constant component $\Phi_{er} \gg \Phi_0$, the magnetization follows the $n = 0$ line and is reversible and lossless. For higher values of the high-frequency magnetic flux, $\Phi_{er} > \Phi_k$, the magnetization curve includes at least one hysteresis loop. The energy ΔE lost in one hysteresis loop is:

$$\Delta E = -\frac{1}{L}(2\Phi_k - \Phi_0) \times (\Phi_0 - \Phi_c). \tag{5.14}$$

For $\Phi_{er} \ll \Phi_0$: $\Delta E = \frac{\Phi_0^2}{L}$, if $\Phi_c \ll \Phi_0$.

5.2.2 Measurement System with an RF-SQUID

Figure 5.3 shows the block diagram of a measurement system with an RF-SQUID. A measurement signal (electric current) brought to the input coil generates the measured magnetic flux in this coil. Made of a superconducting block or thick-walled sleeve with a Josephson junction formed by a weak link of the magnetic circuit, the

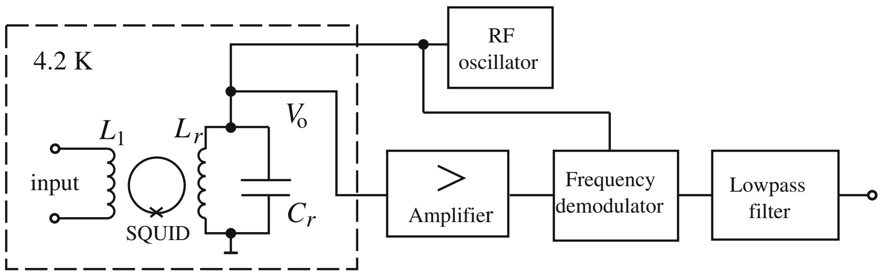


Fig. 5.3 Measurement system with an RF-SQUID

RF-SQUID plays the role of a magnetic flux sensor in the system. Niobium is the typical superconducting material for sensors of this type. The first RF-SQUIDs, known as Zimmerman SQUIDs, were made of a niobium block (e.g. a rod) with a self-inductance L, with two holes for the input coil L_i and the resonance circuit coil L_r [35].

Currently SQUIDs are fabricated in the form of an integrated circuit containing a superconducting loop with a Josephson junction, and L_i and L_r coils coupled with the loop. Thus, the magnetic circuit of the SQUID is coupled inductively with the input coil and a parallel L_rC_r resonance circuit powered by a high-frequency (radio frequency—RF) oscillator. The amplitude of the RF current in the resonance circuit is adjusted so that the amplitude of the flux Φ_{er} induced by this current in the magnetic circuit is slightly lower than Φ_k.

Constant or low-frequency external magnetic flux Φ_{e0} can be determined by the measurement of the voltage V_0 at the terminals of the resonance circuit. For $\Phi_{e0} = 0$ the voltage V_0 is the highest, since the circuit of the sensor is non-hysteretic. If the flux Φ_{e0} increases so that the total flux $(\Phi_{e0} + \Phi_{er})$ exceeds Φ_k, an energy ΔE is taken up from the resonance circle in the hysteresis loop, as a result of which the voltage in the resonance circle will drop by ΔV_0:

$$\Delta V_0 = \frac{2\pi L f_r \Phi_0}{2M}, \tag{5.15}$$

where L is the inductance of the magnetic circuit of the SQUID, f_r denotes the resonance frequency of the L_rC_r circuit, and M is the mutual inductance between L and L_r.

The voltage V_0 depends on Φ_{e0} and has a minimum at $\Phi_{e0} = \Phi_k/2 = \Phi_0/2$. As Φ_{e0} grows from $\Phi_0/2$ to Φ_0, V_0 increases, but decreases again to reach a minimum at $3\Phi_0/2$ with further increase in Φ_{e0}. Thus, V_0 is a periodic function, with a period Φ_0, of the external magnetic flux Φ_{e0}. The measurement of V_0 allows to determine the external flux Φ_{e0} in the range $\pm\, \Phi_0$ or the much wider range $\Phi_{e0} \gg \Phi_0$. In the latter case the quantum magnetometer is a counter of magnetic flux quanta.

In measurement systems the magnetic flux Φ_{e0} is generated by the current in the input coil L_i. A parameter commonly used in practice, the flux sensitivity at input, denoted S_i, is defined as the change ΔI_i in the current in the input coil that causes the magnetic flux acting on the SQUID to change by one fluxon, $\Delta\Phi = \Phi_0$. The inverse of S_i is the mutual inductance M_i between the input coil L_i and the SQUID of self-inductance L:

$$S_i = \frac{\Delta I_i}{\Phi_0}, \quad M_i = \frac{1}{S_i}.$$

The quantum magnetometer is used for the measurement of many nonmagnetic physical quantities and in that case often referred to as the "SQUID measurement system".

5.3 DC-SQUID

5.3.1 DC-SQUID Equation

A quantum interference effect in a superconducting loop with two Josephson junctions was first observed in 1964 [15]. The critical current of a circuit with two junctions is a periodic function of the magnetic flux threading the loop, with a period equal to the flux quantum Φ_0. Studies in this field soon led to the development of the DC-SQUID.

Schematically depicted in Fig. 5.4, the direct-current (DC) SQUID consists of a superconducting loop with two Josephson junctions (XY and ZW) and two terminals for connection to an external circuit. We assume that the superconducting loop is placed in a magnetic field of induction $\vec{B}$ and the magnetic field lines are perpendicular to the plane of the loop. The superconducting current in the loop of the DC-SQUID is determined by the parameters of the Josephson junctions, which represent weak links in the loop. Caused by the motion of Cooper pairs (see Paragraph 4.1.2), the superconducting current I_j in each of the two junctions fulfills the relation:

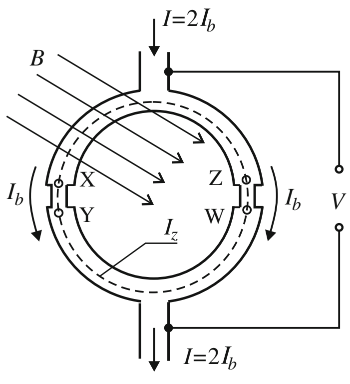


Fig. 5.4 Schematic representation of the DC-SQUID

$$I_j = I_{jc} \sin \varphi, \tag{5.16}$$

where I_{jc} is the critical current of the junction, and φ is the phase difference of the electron wave of Cooper pairs across the junction.

The superconducting current in the loop causes a substantial difference in the phase of the electron wave across the junctions XY and ZW, and a negligible phase difference elsewhere in the loop, i.e. in its thick parts. In contrast, the applied magnetic field results in a difference in the phase of the electron wave between X and Z and between W and Y, i.e. in the thick parts of the loop. Since the length of the junctions is insignificant compared to the circumference of the loop, the phase difference across the junctions due to the magnetic field is negligible.

We assume that no current I from the external circuit flows through the SQUID, and the maximal current I_{jc} induced by the magnetic field $\vec{B}$ generates a flux much smaller than one fluxon, $\Phi = LI_{jc} \ll \Phi_0$, where L is the inductance of the SQUID loop. The dashed line in Fig. 5.4 is the line of integration of the phase shift in the loop. Except for the Josephson junctions, the line runs beyond the penetration depth of the magnetic field. The current density on this line, with the exception of the Josephson junction regions, is zero. The total change in phase due to the magnetic field and the current along the line XYWZX is:

$$\varphi = \varphi_{ZX} + \varphi_{XY} + \varphi_{YW} + \varphi_{WZ}. \tag{5.17}$$

Let the phase differences across the junctions be denoted as α_j and β_j. Thus:

$$\varphi = \alpha_j + \beta_j + \varphi_{YW} + \varphi_{WZ}. \tag{5.18}$$

Using (5.4) we get:

$$\varphi = \alpha_j + \beta_j + \frac{4\pi e}{h} \int\int_S \vec{B}\, d\vec{S}. \tag{5.19}$$

The double integral in (5.19) is equal to the flux Φ_e of the magnetic induction $\boldsymbol{B}$ acting on the SQUID:

$$\Phi_e = B \times S + LI_j. \tag{5.20}$$

Having assumed that $LI_{jc} \ll \Phi_0$, we can neglect the element LI_j in (5.20), and transform the factor in front of the integral:

$$\frac{4\pi e}{h} = 2\pi \frac{2e}{h} = \frac{2\pi}{\Phi_0}.$$

The (5.20) becomes:

$$\varphi = \alpha_j + \beta_j + 2\pi \frac{\Phi_e}{\Phi_0}. \tag{5.21}$$

The wave function of a Cooper pair must fulfill the phase condition: the total change in the phase on a closed path must be an integer multiple of a full angle, $2\pi n$, where n = 1, 2, 3, …. Thus:

$$2\pi n = \alpha_j + \beta_j + 2\pi \frac{\Phi_e}{\Phi_0}. \tag{5.22}$$

The phase shifts α_j and β_j depend not only on the current flowing across the junctions. When no current flows through the SQUID and the junctions are identical, the phase difference across each junction is:

$$\alpha_j = \beta_j = \pi\left(n - \frac{\Phi_e}{\Phi_0}\right). \tag{5.23}$$

Now, let us give up the previous assumption and assume that a current $I = 2I_b$ from the external circuit flows through the SQUID depicted in Fig. 5.4 and splits symmetrically in the two branches of the loop. When the SQUID is driven with a direct current I, the phase shifts across the junctions are different because of different currents flowing through the junctions. The current across the XY junction is $(I_b + I_j)$, and the current across the WZ junction $(I_b - I_j)$. However, the phase condition must be fulfilled, and the total phase shift $\alpha_j + \beta_j$ remains unchanged:

$$\alpha_j = \pi\left(n - \frac{\Phi_e}{\Phi_0}\right) - \upsilon \tag{5.24}$$

$$\beta_j = \pi\left(n - \frac{\Phi_\varepsilon}{\Phi_0}\right) + \upsilon, \tag{5.25}$$

where υ is the phase difference dependent on the direct current I.

Using (5.16), (5.24) and (5.25), the currents across the two junctions can be written as:

$$I_j - I_b = I_{jc}\, sin\left[\pi\left(n - \frac{\Phi_e}{\Phi_0}\right) - \upsilon\right] \tag{5.26}$$

$$I_j + I_b = I_{jc} \sin\left[\pi\left(n - \frac{\Phi_e}{\Phi_0}\right) + \upsilon\right]. \tag{5.27}$$

Subtraction of these equations and further transformations lead to the expression (5.28) of the direct current flowing through the DC-SQUID:

$$I = 2I_b = 2I_{jc}\left|cos\left[\pi\left(n - \frac{\Phi_e}{\Phi_0}\right)\right]\right|sin\upsilon. \tag{5.28}$$

Since $|\sin\upsilon|$ cannot be greater than one, (5.28) can only be fulfilled for the current I:

$$I \leq 2I_{jc}\left|cos\left[\pi\left(n - \frac{\Phi_e}{\Phi_0}\right)\right]\right|. \tag{5.29}$$

The critical current I_c of the SQUID, i.e. the maximal current for which the superconductivity of the junctions is maintained, is:

$$I_c = 2I_{jc}\left|\cos\left[\pi\left(n - \frac{\Phi_e}{\Phi_0}\right)\right]\right|, \text{ or}$$

$$I_c = 2I_{jc}\left|\cos\pi\left(n - \frac{\Phi_e}{\Phi_0}\right)\right|. \tag{5.30}$$

Thus, the critical current I_c of the DC-SQUID is a periodic function of the magnetic flux Φ_e acting on the DC-SQUID, as shown in Fig. 5.5.

The relation (5.30), illustrated by Fig. 5.5a, refers to the case in which the maximal superconducting current I_{jc} circulating in the loop meets the assumption $LI_{jc} \ll \Phi_0$. Without this assumption the dependence of the critical current I_c on the flux Φ_e is as shown in Fig. 5.5b.

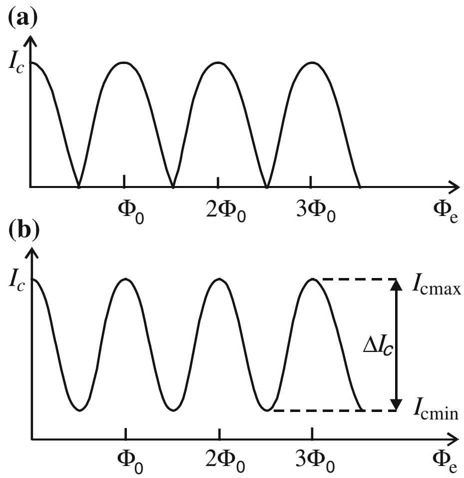


Fig. 5.5 Critical current I_c of the DC-SQUID versus magnetic flux Φ_e: **a** for $LI_{jc} \ll \Phi_0$, **b** general dependence

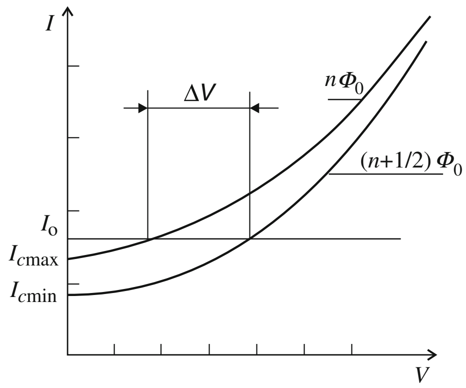


Fig. 5.6 Current-voltage characteristic of the DC-SQUID

The current-voltage (I–V) characteristic of the DC-SQUID is contained between the curve starting from $I = I_{cmax}$, for the flux $\Phi_e = n\Phi_0$, and the curve starting from $I = I_{cmin}$, corresponding to $\Phi_e = (n + ½)\Phi_0$—see Fig. 5.6.

There are two possible ways of biasing a DC-SQUID:

- Current (I_0) biasing and the measurement of the voltage across the sensor;
- Voltage (V_0) biasing and the measurement of the current flowing through the sensor.

In the first, more commonly used method, the DC-SQUID is biased by a direct current I_0 greater than the maximal critical current I_{cmax}. For a constant value of I_0 the voltage V at the terminals of the SQUID pulsates with increasing flux Φ_e acting on the SQUID. The SQUID has a nonlinear dynamic resistance $R_{dyn} = \frac{dV}{dI}$, which depends on the bias current I_0. An important parameter of the sensor, the conversion coefficient k_Φ measures the conversion of the magnetic flux acting on the sensor to its terminal voltage. Measured in the units of $\mu V/\Phi_0 = 5 \times 10^8$ Hz, the conversion coefficient is equal to the ratio of ΔV to $\Phi_0/2$; its value depends on the current bias point I_0:

$$k_\Phi = \left(\frac{\Delta V}{\Phi_0/2}\right)_{I_0}. \tag{5.31}$$

A voltage-biased SQUID is characterized with the current sensitivity M_{dyn}, dependent on the bias voltage V_0 and measured in inductance units:

$$M_{dyn} = -\left(\frac{d\Phi}{dI}\right)_{V_0} = \frac{R_{dyn}}{k_\Phi}. \tag{5.32}$$

In almost all measurement system applications the DC-SQUID plays the role of a zero magnetic flux sensor operating in a flux-locked loop. A modulation flux $\Phi(t) = 0.5\Phi_0 \sin(2\pi f_m t)$ with a frequency f_m ranging from 100 to 500 kHz is applied

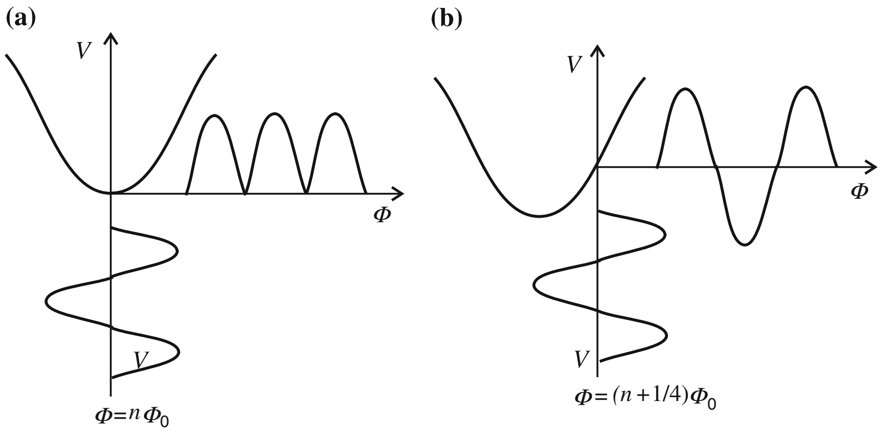


Fig. 5.7 Flux-to-voltage transfer characteristic of the DC-SQUID with flux modulation and bias: **a** $\Phi = n\Phi_0$, **b** $\Phi = (n + ¼)\Phi_0$

to the SQUID. When the bias flux is exactly equal to $\Phi = n\Phi_0$, the form of the sensor terminal voltage is as after full-wave rectification (see Fig. 5.7a). The flux-to-voltage transfer function of the sensor is strongly nonlinear in that case. However, with the bias flux value shifted by $\Phi_0/4$, i.e. equal to $\Phi = (n + ¼)\Phi_0$, the DC-SQUID is a linear flux-to-voltage converter (Fig. 5.7b). Typically, the conversion coefficient ranges from 20 to 100 μV/Φ_0.

In quantum magnetometer circuits used in practice the measured magnetic field does not act on the SQUID directly, but through a superconducting magnetic flux transformer (Fig. 5.8).

In a simple magnetometer the flux transformer consists of a pick-up coil (also known as antenna) connected with an input coil L_i coupled inductively with the SQUID. The measured magnetic flux Φ_{sig} induces a current I_i in the pick-up coil, and I_i generates a magnetic flux Φ_x that acts on the SQUID (Fig. 5.8a).

First-order (Fig. 5.8b) or higher-order gradiometers are used to reduce the effect of interference Φ_{int} on the measurement result. A first-order gradiometer consists of two coaxial windings, an antenna and a compensating coil, separated by a distance *d*. The antenna and the compensating coil are wound in opposite directions. Acting on both coils, the interference flux induces opposite currents in them. If the two coils have perfectly equal areas, the current in the input circuit is unaffected by the interference flux and only depends on the measured flux Φ_{sig} that acts solely on the antenna.

Modern integrated SQUID sensors contain a DC-SQUID and a modulation coil (feedback coil) in a single integrated circuit (Fig. 5.9). The DC-SQUID and the input coil are made of a superconducting material, and the modulation coil usually has the form of a track made of copper foil. To ensure a strong effect of the signal brought to the terminals of the input coil on the DC-SQUID, strong magnetic coupling is necessary between the input coil of inductance L_i and the DC-SQUID of

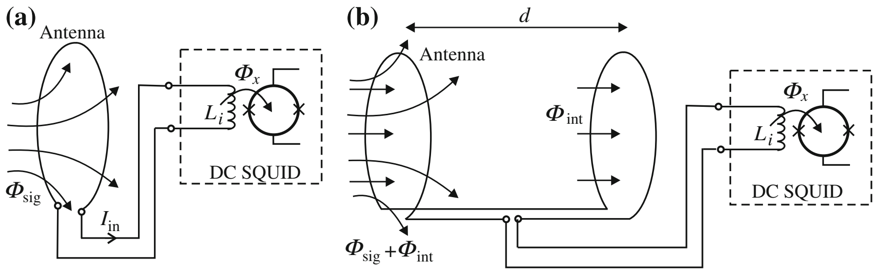


Fig. 5.8 Magnetic flux transformer in a SQUID magnetometer: **a** simple magnetometer, **b** first-order gradiometer

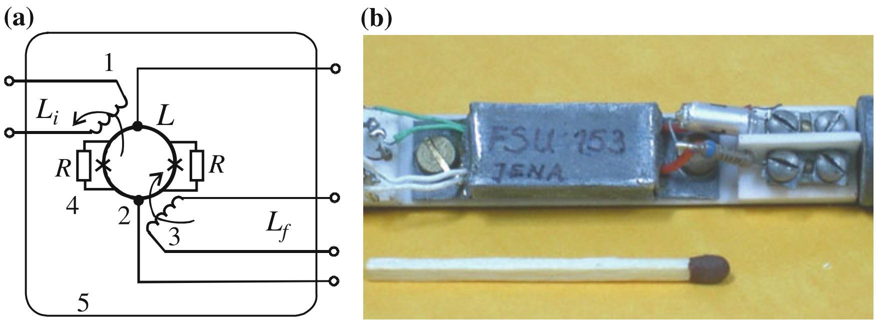


Fig. 5.9 Integrated DC-SQUID sensor: **a** electrical diagram: *1* input coil, *2* DC-SQUID, *3* modulation coil, *4* shunt resistance R of each junction, *5* superconducting shielding; **b** SQUID sensor in a lead shielding, with niobium terminals (*to the right*) for connection of the input coil to an external measurement circuit

inductance L. A measure of this coupling is the mutual inductance M_i between the input coil L_i and the SQUID loop of self-inductance L: $M_i = k_i\sqrt{L_i L}$. Other important parameters are the mutual inductance $M_f = k_f\sqrt{L_f L}$ between the modulation coil L_f and the SQUID loop, and the mutual inductance $M_t = k_t\sqrt{L_i L_f}$ between L_i and L_f. Each of the two Josephson junctions is shunted by an external resistor or, in the case of integrated sensors, the resistance of the substrate.

5.3.2 Energy Resolution and Noise of the DC-SQUID

Usually SQUIDs are optimized for minimal noise. Two noise criteria in common use, the energy resolution $\varepsilon(f)$ and the energy resolution at input coil $\varepsilon_c(f)$, are the minimum magnetic energy per band unit that can be measured by a plain SQUID or a SQUID with an input coil, respectively. It is important to carefully distinguish between these two parameters to avoid possible confusion between the literature data concerning the plain SQUID energy resolution ε, which has reached the level of $0.5h$ [19] (where h is the Planck constant), and the much worse energy resolution at input coil ε_c. The energy resolution $\varepsilon(f)$ and the energy resolution at input coil $\varepsilon_c(f)$ fulfill the relations:

$$\varepsilon(f) = \frac{\Phi_r^2}{2L}, \quad \varepsilon_c(f) = \frac{\varepsilon(f)}{k_i^2},$$

where Φ_r is the spectral density of magnetic flux noise, L is the inductance of the SQUID, and k_i denotes the coupling coefficient between the input coil and the SQUID.

The minimal change ΔI_i in the input coil current that can be detected by the magnetometer is a function of its noise level Φ_r:

$$\Delta I_i = \frac{\Phi_r}{M_i}, \tag{5.33}$$

where ΔI_i is the minimal change ΔI_i in the current in the input coil, Φ_r the noise level of the DC-SQUID in the units of magnetic flux per $\sqrt{\text{Hz}}$, the mutual inductance M_i between the input coil L_i and the SQUID loop of self-inductance L, and k_i the coupling coefficient.

The energy resolution ε of the DC-SQUID is:

$$\varepsilon = \frac{1}{2} L_i (\Delta I_i)^2 = \frac{L_i \Phi_r^2}{2M_i^2} = \frac{\Phi_r^2}{2k_i^2 L}. \tag{5.34}$$

The parameter ε_c characterizes a SQUID with an input coil. The lower the noise level and the higher the coupling coefficient k_i between the SQUID and the input coil, the better the resolution of the SQUID.

The noise level of the SQUID has a minimum for frequencies above the corner frequency of the $1/f$ noise (ranging from a 1 Hz to tens of kHz) and below the Josephson oscillation. In this frequency range the noise of the SQUID is a white noise, mainly originating in the thermal noise of the shunt resistors of the Josephson junctions. Each of the two Josephson junctions is shunted by a resistor R in order to obtain an I–V characteristic without hysteresis. In DC-SQUID sensors made of high-temperature superconductors the shunt resistance is included in the structure of the material. The perfect I–V characteristic is given by:

$$V = R\sqrt{\left(I_b^2 - I_{jc}^2\right)}, \tag{5.35}$$

where V is the voltage across the terminals of the SQUID, I_b and I_{jc} denote the current in a single junction and its critical current, and R is the resistance of a single shunt.

The *I*–*V* characteristic is nonhysteretic when the McCumber parameter β_c does not exceed one:

$$\beta_c = \frac{2\pi R^2 I_{jc} C}{\Phi_0} \leq 1, \tag{5.36}$$

where β_c is the McCumber parameter and *C* the capacitance of the junction.

Likharev and Semenov [21] derived an analytical formula for the power spectral density S_v of the white noise of the DC-SQUID for a voltage signal at its terminals:

$$S_v = 4k_B TR \frac{1 + \frac{1}{2}\left(\frac{I_{jc}}{I_b}\right)^2}{1 + \left(\frac{I_{jc}}{I_b}\right)^2}. \tag{5.37}$$

Usually parameters of the whole sensor, I_c, I_0 and the shunt resistance $R_s = R/2$ of the whole SQUID, are used instead of the single-junction parameters I_{jc}, I_b and *R*. The power spectral density at the operating point $I = I_0$ is:

$$S_v = 8k_B TR_s \frac{1 + \frac{1}{2}\left(\frac{I_c}{I_0}\right)^2}{1 + \left(\frac{I_c}{I_0}\right)^2}. \tag{5.38}$$

The power spectral density S_v of the voltage signal can be converted to the power spectral density Φ_r^2 of the magnetic flux noise figuring in (5.34). The (5.39) provides the optimal low-noise relation between the change in the voltage and the change in the flux acting on the SQUID [7]:

$$\frac{dV}{d\Phi} = \frac{R_s}{L}. \tag{5.39}$$

From (5.38) and (5.39) it follows that the power spectral density of the magnetic flux, i.e. the noise level of the DC-SQUID, can be written as:

$$\Phi_r^2 = \frac{S_v}{\left(\frac{dV}{d\Phi}\right)^2} = \frac{S_v}{k_\Phi^2} = \frac{8k_B TL^2}{R_s} \frac{1 + \frac{1}{2}\left(\frac{I_c}{I_0}\right)^2}{1 + \left(\frac{I_c}{I_0}\right)^2}. \tag{5.40}$$

Given by (5.34), the energy resolution ε can be also expressed as:

$$\varepsilon = \frac{\Phi_r^2}{2k_i^2 L} = \frac{4k_B TL^2}{k_i^2 R_s} \frac{1 + \frac{1}{2}\left(\frac{I_c}{I_0}\right)^2}{1 + (I_c/I_0)^2}. \tag{5.41}$$

The formula (5.41) provides the conditions to be fulfilled to obtain a good resolution of the DC-SQUID. The operating temperature of the SQUID is very often fixed; its typical values are 4.2 K (the temperature of liquid helium) for low-temperature SQUIDs and 77 K (the temperature of liquid nitrogen) for high-temperature SQUIDs. Good energy resolution of the SQUID (low value of ε) requires a low self-inductance L, a high coefficient k_i of the coupling between the inductance L of the SQUID loop and the inductance L_i of the input coil, and a high shunt resistance R_s of the Josephson junctions. The energy resolution is determined indirectly by measuring the noise level Φ_r and the current sensitivity $\Delta I_i/\Phi_0$ of the DC-SQUID:

$$\varepsilon = \frac{1}{2} L_i \Phi_r^2 \left(\frac{\Delta I_i}{\Phi_0}\right)^2. \tag{5.42}$$

The energy resolution ε determined from (5.42) strongly depends on the conditions, such as the frequency band or the external interference level, in which the noise level Φ_r is measured.

5.3.3 Parameters of a DC-SQUID

Low-temperature SQUIDs play a more important role in metrology than high-temperature SQUIDs. This is due to the lower noise level of LTS sensors (the temperature is an important factor determining the noise, an its low value is desirable) as well as to their higher durability. A high-temperature superconductor will change its chemical composition after a few months' time as a result of chemical interaction with the environment, and finally lose its superconducting properties. Thus, the electrical parameters of an HTS DC-SQUID will change in time to degenerate completely at the end.

The noise levels of the best DC-SQUIDs are: $22 \times 10^{-7}\ \Phi_0/\sqrt{\text{Hz}}$ for Nb/Al-AlO$_x$ sensors [11] at the temperature of 4.2 K, and $42 \times 10^{-5}\ \Phi_0/\sqrt{\text{Hz}}$ for $YBa_2Cu_3O_{7\text{-}x}$ HTS SQUIDs at 77 K [34]. The best energy resolution of a DC-SQUID with an input coil, $\varepsilon = 2.1 \times 10^{-32}$ J/Hz = 32 h, was obtained for a low-temperature sensor with the noise level $6 \times 10^{-7}\ \Phi_0/\sqrt{\text{Hz}}$ for the frequency $f = 1$ kHz [4]. Due to the $1/f$ noise, at the point $f = 10$ Hz on the frequency scale the noise level and the energy resolution of this sensor were three times worse, $2 \times 10^{-6}\ \Phi_0/\sqrt{\text{Hz}}$ and $\varepsilon = 100h$, respectively. The best HTS DC-SQUID energy resolution, $\varepsilon = 3 \times 10^{-31}$ J/Hz = $450h$, was measured at 77 K for the frequency 70 kHz [18]. For the frequency $f = 1$ Hz the resolution ε of HTS sensors is three orders of magnitude worse. Let us recall that the record energy resolution of a plain DC-SQUID is as good as $\varepsilon = 0.5h$ [19] at 290 mK in the frequency range above 500 Hz. E.g. the integrated UJ111 DC-SQUID

(Fig. 5.9b) made at the Friedrich Schiller University Jena, Germany. Its parameters are:

- Maximal critical current of the SQUID: I_{cmax} = 14 μA;
- Difference between the maximal and minimal critical currents of the SQUID: ΔI_c = 11 μA;
- Self-inductance of the SQUID loop: L = 50 pH;
- Output voltage at the optimal operating point (Fig. 5.6): ΔV = 16 μV, or conversion coefficient k_Φ = 32 μV/Φ_0;
- Inductance of the niobium input coil: L_i = 770 nH;
- Mutual inductance M_i between the input coil and the SQUID: M_i = 4.6 nH;
- The inverse of the mutual inductance M_i, current sensitivity of the input coil: $\Delta I_x/\Phi_0 = 1/M_i$ = 0.45 μA/Φ_0;
- Mutual inductance M_f between the modulation coil and the SQUID: M_f = 110 pH;
- The inverse of the mutual inductance M_f, current sensitivity of the modulation coil: $\Delta I_f/\Phi_0 = 1/M_f$ = 19.5 μA/Φ_0;
- Shunt resistance of the junctions: R_s = 1.2 Ω;
- Noise level: $\Phi_n = 2 \times 10^{-5}\, \Phi_0/\sqrt{\text{Hz}}$ (the noise level of the best UJ111 sensor was ten times lower);
- Nominal operating temperature: T = 4.2 K.

The DC-SQUID specified above operated in a measurement system at Poznań University of Technology. It should be noted that currently fabricated commercial DC-SQUIDs have noise levels ten times lower than that of the presented UJ111 sensor. The noise level of present-day commercial DC-SQUIDs cooled with liquid helium is of the order of $\Phi_n \approx 10^{-6} \Phi_0/\sqrt{\text{Hz}}$.

5.4 Measurement System with a DC-SQUID

5.4.1 Operation of the Measurement System

Figure 5.10 presents the block diagram of a measurement system with the DC-SQUID. Measuring the current at the terminals of the input coil, the system can be used for high-sensitivity measurements of any physical quantity convertible to electric current in the input coil. Since the DC-SQUID is a magnetic flux sensor, a measurement system with a plain DC-SQUID (without input coil) will be a quantum magnetometer.

In real measurement systems the SQUID is very carefully shielded from the external magnetic field. Even the Earth magnetic field is the source of very strong interference for an unshielded SQUID. Because of the high interference level, measurements with an unshielded SQUID are very difficult and sometimes practically impossible.

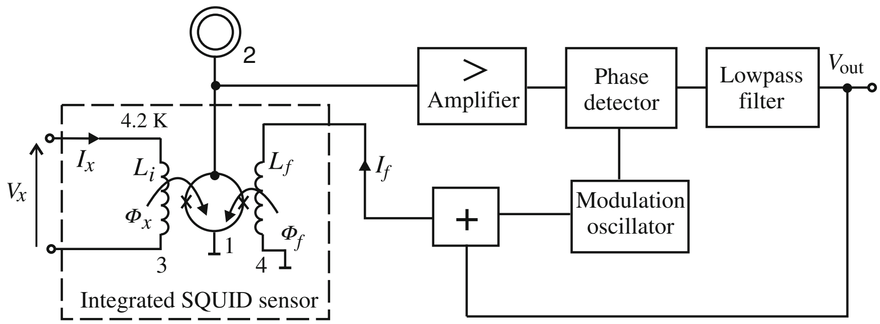


Fig. 5.10 Measurement system with a DC-SQUID: *1* DC-SQUID, *2* source of direct current, *3* input coil, *4* modulation coil

A shielding of lead or niobium, both of which are superconducting metals, is used for interference reduction. The superconducting shielding plays the role of a screen, the external magnetic field being repulsed from the superconductor due to its perfect diamagnetism. The DC-SQUID is only exposed to the magnetic flux generated by the current I_x in the input coil integrated with the SQUID loop (Fig. 5.9). The system measures the signal I_x at the terminals of the input coil. The measurement system operates in a magnetic feedback loop with magnetic flux modulation [10]. A source of direct current (2) provides a bias current to the DC-SQUID (1). The total magnetic flux acting on the magnetic field sensor includes:

- The magnetic flux Φ_x from the input coil (3) of inductance L_i, to which the measurement signal is brought;
- The negative feedback flux Φ_f induced by the modulation coil (4), proportional to the modulated signal generated by the modulation signal oscillator.

The signal is processed by amplitude modulation. The modulation takes place in the SQUID. The AC voltage generated at the terminals of the sensor is amplified by the preamplifier. After the phase detection in the phase detector the measurement signal is filtered by the low-pass filter. The current in the modulation coil is proportional to the voltage V_{out} at the output of the low-pass filter. The measured magnetic flux Φ_x is automatically compensated by the feedback flux Φ_f. The stronger the input signal I_x, the greater the current I_f in the modulation coil generating the feedback flux. Since the current I_f is caused by the voltage V_{out}, also the latter increases with the input signal. The voltage V_{out} at the output of the low-pass filter is the output voltage of the measurement system.

The system allows to measure the current I_x in the input coil or any physical quantity that can be converted to it. The main parameters of a measurement system with a SQUID are the noise level and the dynamic characteristics, such as the passband or the slew rate. Other important parameters include linearity (harmonic distortion), feedback range and crosstalks in multichannel systems. Usually the internal noise of the DC-SQUID defines the total noise of the system, and the dynamic properties are limited by the electronic units involved in the magnetic feedback.

5.4.2 Input Circuit

Signal analysis in the input circuit is of much use for the evaluation of the parameters of a measurement system with a DC-SQUID. The equivalent circuit diagram of the input circuit is shown in Fig. 5.11. The input circuit consists of a source of electromotive force E_x with internal resistance R_{int}. The source of E_x is connected to the terminals of the input coil of the DC-SQUID. The input coil has a self-inductance L_i, and the DC-SQUID a self-inductance L. The self-inductance L of the sensor is involved in the mutual inductance M_i between the sensor and the input coil, and the mutual inductance M_f between the sensor and the feedback coil of inductance L_f. The mutual inductance between L_i and L_f is denoted as M_t.

The electromotive force E_x drives the current I_x that flows through the input coil; a feedback current I_f flows through the feedback coil. As a result of negative magnetic feedback the magnetic flux $\Phi_x = M_i I_x$ acting on the DC-SQUID and generated by the input current I_x is compensated by the magnetic flux $\Phi_f = M_f I_f$ generated by the feedback current I_f in the modulation coil. On the assumption that the magnetic flux Φ_r of the internal noise of the DC-SQUID is negligible compared to Φ_x and Φ_f:

$$M_i I_x + M_f I_f = 0. \tag{5.43}$$

The following equation, resulting from Kirchhoff's voltage law, is fulfilled in the input circuit:

$$E_x(t) = R_{int} I_x + \mathrm{j}2\pi f L_i I_x + \mathrm{j}2\pi f M_t I_f. \tag{5.44}$$

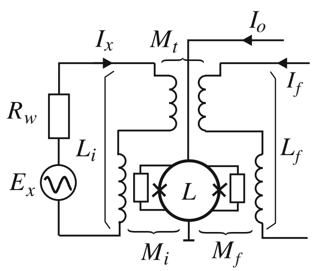


Fig. 5.11 Input circuit of a measurement system with a DC-SQUID (equivalent circuit diagram)

Using (5.43) and (5.44), the current I_x in the input circuit can be written as:

$$I_x^2 = \frac{E_x^2}{R_{int}^2 + (2\pi f)^2 [L_i(1 - \frac{M_i M_t}{L_i M_f})]^2}. \tag{5.45}$$

The (5.45) implies that for the input current I_x the input circuit plays the role of a low-pass filter with a cutoff frequency f_c:

$$f_c = \frac{R_{int}}{2\pi L_i(1 - \frac{M_i M_t}{L_i M_f})} = \frac{R_{int}}{2\pi L_{i.e.}}, \tag{5.46}$$

where L_{ie} is the effective inductance of the input coil.

Introduced by Clarke [5], the effective inductance L_{ie} of the input coil of the SQUID is the effective value to which the self-inductance L_i of the input coil is reduced as a result of the negative feedback in the measurement system:

$$L_{ie} = L_i(1 - \frac{M_i M_t}{L_i M_f}). \tag{5.47}$$

The introduction of the cut-off frequency f_c of the input circuit simplifies the expression (5.45) of the frequency dependence of the input current I_x to:

$$I_x = \frac{E_x}{R_{int}\sqrt{1 + (f/f_c)^2}}. \tag{5.48}$$

The parameters L_i, M_i, M_f of the integrated DC-SQUID can be easily measured in an operating system (L_i directly, M_i and M_f by the measurement of the current sensitivity of the input and modulation coils in the units of $\mu A/\Phi_0$). However, it is hard to determine the effective inductance L_{ie} of the input coil on the basis of (5.47), since the low value of the mutual inductance M_t involved in the calculation is difficult to measure directly [5]. The effective inductance L_{ie} is necessary for the evaluation of the dynamics of the measurement system, specifically, the filtering properties of the input circuit (low-pass filter).

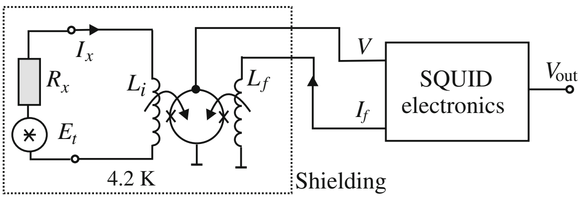


Fig. 5.12 Setup for the measurement of resistor thermal noise and the determination of the effective inductance of the input coil (the resistor and the DC-SQUID are kept the same ambient temperature of 4.2 K)

The paper [26] proposes a new method for the measurement of the effective inductance L_{ie}. In the reported study L_{ie} was determined by the measurement of the thermal noise current of a resistor connected to the terminals of the input coil of the SQUID sensor in a measurement system, the block diagram of which is shown in Fig. 5.12.

Being made of a superconducting material, the winding of the input coil has zero resistance. The mean square of the electromotive force E_t of the thermal noise of the resistor R_x at the temperature T is given by the Nyquist formula:

$$d\langle E_t^2 \rangle = 4k_B T R_x df. \tag{5.49}$$

The thermal noise generated by the resistor is a white noise, i.e. has a constant power spectral density, in a wide frequency range up to GHz frequencies. The mean square of the thermal noise current in the input circuit is:

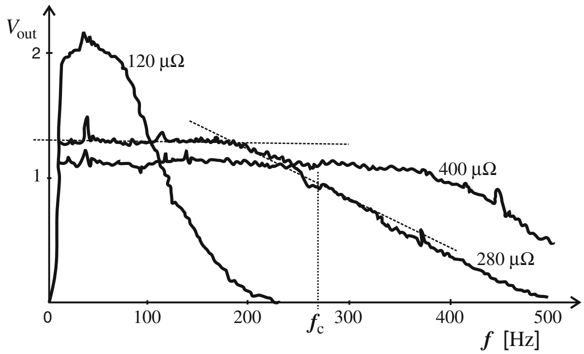
$$d\langle I_n^2 \rangle = \frac{4k_B T R_x\, df}{R_x^2 + (2\pi f L_{ie})^2} = \frac{4k_B T}{R_x} \times \frac{df}{(f/f_c)^2} \tag{5.50}$$

$$L_{ie} = \frac{R_x}{2\pi f_c}.$$

Figure 5.13 presents the plot of the frequency dependence of the output voltage in the measurement system for different resistors. The frequency band of the SQUID system used for the measurements ranged from 0 to 2 kHz. Thus, the low-pass filtering of the measurement signal shown by the plot is only due to the filtering by the input circuit of the sensor.

The shape of the spectral characteristic allows to determine graphically the cutoff frequency f_c for the calculation of the effective inductance of the input coil of the SQUID. In the presented case the negative feedback reduces the inductance of the input coil from L_i = 770 nH to the effective value L_{ie} = 170 nH [25, 26].

Fig. 5.13 Spectral characteristic of the output voltage of a DC-SQUID measurement system with a resistor connected to the input coil



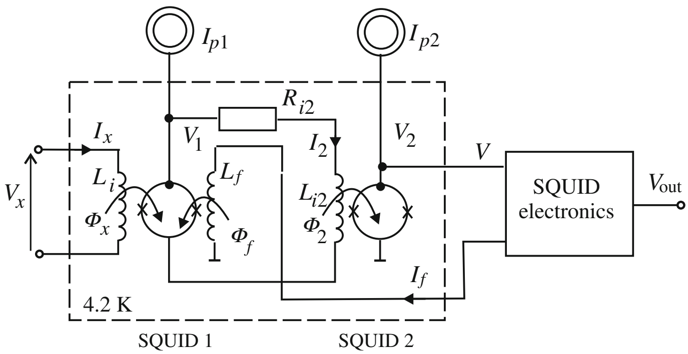


Fig. 5.14 Two-SQUID measurement system

5.4.3 Two-SQUID Measurement System

SQUID measurement systems have a very high sensitivity, described, among others, by the ratio $k_p = V_{out}/\Phi_x$ of the output voltage to the magnetic flux acting on the SQUID. The sensitivity can be further increased by incorporating another SQUID to act as a preamplifier in the system [7], as shown in Fig. 5.14. The two SQUID sensors in the measurement system, SQUID 1 and SQUID 2, are biased with direct currents I_{p1} and I_{p2}, respectively. The voltage V_1 across the terminals of SQUID 1 is supplied to a series consisting of the resistor R_2 and the input coil of the other sensor. The current I_2 flowing through the coil L_{i2} generates in it a magnetic flux Φ_2 that acts on SQUID 2. Since the negative magnetic feedback involves both SQUIDs in the system, the current in the modulation coil of SQUID 1 is proportional to the output voltage V_{out} of the system. Let $k_{\Phi 1} = \frac{V_1}{\Phi_x}$ and $k_{\Phi 2} = \frac{V_2}{\Phi_2}$ denote the conversion coefficients of SQUID 1 and SQUID 2, respectively, and M_{i2} be the mutual inductance between the input coil L_{i2} and the self-inductance of SQUID 2. The magnetic flux acting on SQUID 2 is $\Phi_2 = M_{i2}I_2$.

After some transformations, the formula for the conversion coefficient of the two-SQUID system becomes:

$$k_\Phi = \frac{V_2}{\Phi_x} = \frac{k_{\Phi 1} V_2}{V_1} = \frac{k_{\Phi 1} V_2}{\sqrt{R_2^2 + (2\pi f L_{i2})^2}\, I_2} = \frac{k_{\Phi 1} M_{i2} V_2}{\sqrt{R_2^2 + (2\pi f L_{i2})^2}\, \Phi_2}$$
$$k_\Phi = k_{\Phi 1} k_{\Phi 2} \frac{M_{i2}}{\sqrt{R_2^2 + (2\pi f L_{i2})^2}}. \tag{5.51}$$

Thus, the conversion coefficient of the two-SQUID system is the product of the conversion coefficients of the two SQUIDs. It is also proportional to the mutual inductance M_{i2}, and inversely proportional to the modulus of the impedance of the $(R_2 + L_{i2})$ series.

Let $k_{\Phi 1} = k_{\Phi 2} = 100\ \mu\text{V}/\Phi_0 \cong 5 \times 10^{10}$ Hz be the conversion coefficient of each sensor in the discussed system, with the mutual inductance $M_{i2} = 5$ nH and the resistance $R_2 = 10\ \Omega$. For these parameters the conversion coefficient of the two-SQUID system for the DC component will be $k_\Phi = 2500\ \mu\text{V}/\Phi_0 \cong 12 \times 10^{11}$ Hz, a value ten times higher than that of the conversion coefficient of a single sensor. An additional advantage of the effective sensitivity of the whole measurement system is that the impedance matching between the low-resistance output of the sensor (its terminals) and the low-resistance ($R_2 + L_{i2}$) series is better than that between the sensor and the electronic preamplifier.

The increased sensitivity of the system with two or more SQUIDs involves a narrower frequency band for signals in the circuit. This is due to the additional ($R_2 + L_{i2}$) series, which represents an additional low-pass filter with a cutoff frequency given by: $f_{c2} = R_2/(2\pi L_{i2})$.

5.4.4 SQUID Measurement System with Additional Positive Feedback

D. Drung of Physikalisch-Technische Bundesanstalt, Berlin, proposed and first realized an interesting idea for increasing the conversion coefficient k_Φ of the SQUID by the introduction of additional positive feedback (APF). Let us recall that $k_\Phi = V_S/\Phi_x$, where V_S is the voltage at the terminals of the SQUID, and Φ_x is the measurement signal, i.e. the magnetic flux acting on the SQUID. Figure 5.15 shows the circuit of the DC-SQUID with an additional $R_{ad}L_{ad}$ series circuit connected to the terminals of the SQUID [7, 8].

Let us assume that the sensor is biased by magnetic flux so that its operating point **W** lies on the positive slope of the V–Φ characteristic, as shown in Fig. 5.15b. An increase $\delta\Phi$ in the magnetic flux will result in an increase δV_S in the terminal voltage. The consequent increase in the current in the APF coil will induce an additional positive flux Φ_{ad} acting on the SQUID by the mutual inductance M_{ad}. This additional flux will further increase the voltage V_S across the terminals of the sensor and consequently augment also its conversion coefficient k_Φ. In a similar way k_Φ will decrease if the operating point **W** lies on the negative slope of the V–Φ characteristic. Thus, as a result of introducing the APF circuit the V–Φ characteristic becomes asymmetric. Additional positive feedback slightly reduces the peak-to-peak voltage range V_{pp}, since the $R_{ad}L_{ad}$ series circuit acts as a low-pass filter.

The properties of the SQUID with APF can be explained with the use of a low-frequency equivalent circuit. Shown in Fig. 5.15c, the equivalent circuit describes the properties of the system for small changes in the flux around the operating point **W**. The SQUID with APF behaves as a SQUID without the APF circuit, but connected to a voltage amplifier with a gain G_{APF}:

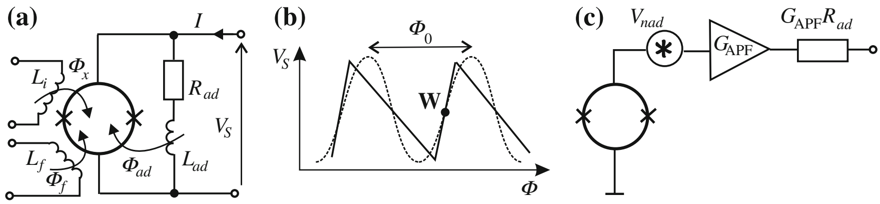


Fig. 5.15 SQUID with additional positive feedback: **a** schematic diagram, **b** V–Φ characteristic with the operating point **W**, **c** equivalent circuit diagram

$$G_{APF} = \frac{1}{1 - \beta_{APF}} \tag{5.52}$$

$$\beta_{APF} = \frac{k_\Phi (M_{ad} - M_{dyn})}{R_{dyn}} \leq 1, \tag{5.53}$$

where G_{APF} is the increase in the flux-to-voltage conversion coefficient; G_{APF} is infinite for $\beta_{\mathrm{APF}} \cong 1$; β_{APF} denotes the positive feedback coefficient, k_Φ is the conversion coefficient of the SQUID without APF, $R_{dyn} = \frac{dV}{dI}$ is the dynamic resistance of the SQUID, and $M_{dyn} = R_{dyn} / k_\Phi$ is the current sensitivity, defined by (5.32).

The output resistance of the amplifier in Fig. 5.15c, $G_{\mathrm{APF}}R_{dyn}$, is the dynamic resistance of the SQUID with APF. The dynamic resistance and the conversion coefficient increase equally. The contribution V_{nad} of the resistor R_{ad} to the total noise is represented by a source of noise voltage at the input of the amplifier, with the spectral density:

$$S_{APF} = \frac{V_{nad}^2}{\Delta f} = 4k_B T R_d \beta_{APF}^2 \approx 8k_B TR, \tag{5.54}$$

where R is the resistance of the SQUID in the normal (non-superconducting) state.

For minimal noise the value of R_{ad} should be as low as possible. However, high peak-to-peak voltage V_{pp} requires $R_{ad} \gg R/2$. Thus, R_d cannot be too small. The approximate value on the right-hand side of (5.54) has been obtained with $G_{\mathrm{APF}} = 10$ and $R_{ad} = 2.5R$, values seen in practice. For a low-temperature SQUID the noise related to the APF circuit, around $8k_BTR$ [8], is sensibly lower than the noise at the terminals of the SQUID, ca. $16k_BTR$ [5].

According to the analysis performed by Drung, a SQUID system with APF provides a much better signal-to-noise ratio than the system without APF [7].

5.4.5 Digital SQUID Measurement System

A conventional SQUID measurement system produces an analog output signal. However, in many applications the output signal of the SQUID is further processed (by filtering, averaging or arithmetic operations) in digital systems. This requires previous high-resolution analog-to-digital conversion. The need for the digital signal grows if the system contains many SQUIDs (as, e.g., multichannel SQUID systems for biomedical studies). The requirements regarding the analog-to-digital converter are lower if the analog-to-digital conversion takes place at the input, rather than at the output, of the electronic circuit of the SQUID system. After the analog-to-digital conversion at the input the digital signal is averaged by a digital signal processor, converted to an analog signal by a digital-to-analog converter, and applied to the SQUID by feedback. The digital signal at the input of the digital-to-analog converter is the digital result of the measurement performed by the SQUID system [8, 9]. In this pattern the operation of the SQUID sensor is still analog.

A SQUID system with a DC-SQUID sensor for digital operation was proposed and realized by Drung [9]. The functioning of the digital SQUID system is based on high-frequency detection of the critical current. A single hysteretic Josephson junction can operate as a latching current comparator. In this case a high-frequency clock current signal I_{clock} and an input current signal I_{sig} flow across the junction, which switches between two states:

- The normal state (junction voltage $V_J > 0$), when the total current across the junction is greater than its critical current, $I_{clock} + I_{sig} \geq I_c$;
- The superconducting state ($V_J = 0$), if the total current drops below the critical current value $I_{clock} + I_{sig} < I_c$.

Thus, the Josephson junction produces a series of voltage pulses generated upon switching between the superconducting and normal states. To improve the sensitivity of the system the single Josephson junction was replaced with a hysteretic DC-SQUID. The DC-SQUID converts the magnetic flux Φ_x generated by the current in the input coil or by the applied magnetic field into a series of pulses with given probability. The pulses are more frequent with higher magnetic flux Φ_x acting of the SQUID.

Figure 5.16 shows a digital SQUID system with a digital flux-locked loop and the SQUID operating as a comparator. Current from a high-frequency clock generator is passed through the SQUID. The resulting voltage pulses at the terminals of the sensor, with an amplitude of a few mV, are supplied to the input of an up/down counter kept at room temperature. The counter plays the role of a digital integrator circuit. Its reading increases when the state of the SQUID changes from superconducting to normal, and decreases if the opposite. The counter reading, which is the digital result of the measurement, is converted to current in the feedback coil by a digital-to-analog converter.

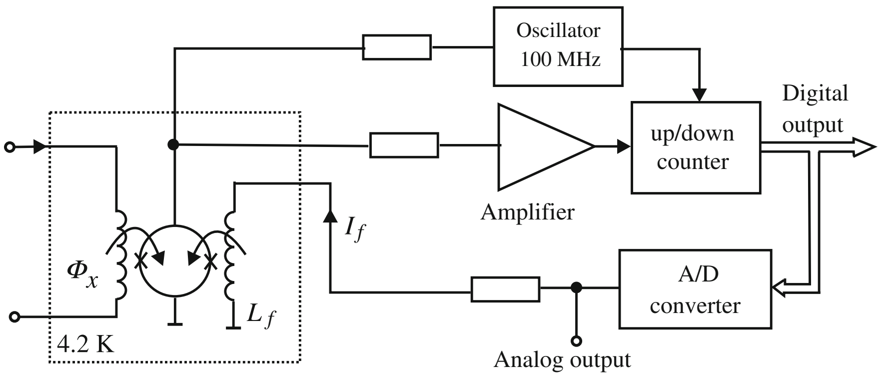


Fig. 5.16 Digital SQUID measurement system

The feedback loop automatically assumes the switching probability $p = 50\ \%$ if the amplitude of the clock pulses is between the maximal and minimal critical currents (Fig. 5.5) of the SQUID. The effect of the amplifier noise voltage is eliminated, as the system only detects and records pulses regardless of their amplitude or phase. The amplifier noise current, however, is of some importance, since the current flows across the SQUID and affects its bias. The energy resolution ε of the SQUID and its magnetic flux noise decrease with increasing clock frequency f_{clock}. Simulations and measurements indicate that in the frequency range from 1 to 700 MHz the energy resolution ε of the SQUID and its switching frequency f_{clock} obey the relation:

$$\frac{\varepsilon}{h} = \frac{20\,\mathrm{GHz}}{f_{clock}}. \tag{5.55}$$

The first digital measurement system constructed by Drung had a DC-SQUID made of Nb-Nb_2O_5-PbInAu. The measured noise level of the sensor operating in this system was $\Phi_n = 4.5 \times 10^{-6}\ \Phi_0/\sqrt{\mathrm{Hz}}$, and its energy resolution $\varepsilon = 4300h$ for the optimal frequency of 20 MHz of the clock generator. Better parameters were obtained in a two-SQUID digital system with a coil used as a gradiometer, integrated with the sensor in a single structure: the noise level was $\Phi_n \leq 10^{-6}\ \Phi_0/\sqrt{\mathrm{Hz}}$ and the energy resolution $\varepsilon \leq 70h$. The best dynamics of this digital system was $2 \times 10^6\ \Phi_0$/s [9].

5.5 Magnetic Measurements with SQUID Systems

5.5.1 Magnetic Signals and Interference

SQUID magnetic flux sensors are used for the measurement of magnetic quantities: magnetic flux, magnetic induction and magnetic field. Because of their excellent energy resolution (sensitivity), better than that of any sensor of any physical

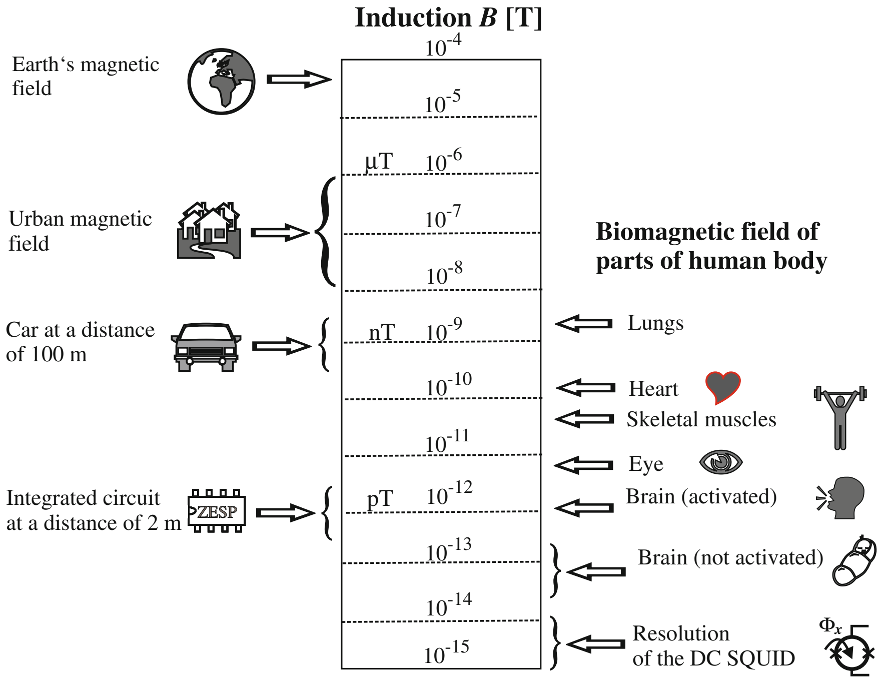


Fig. 5.17 Magnetic induction generated by technological devices and biological objects

quantity, SQUID measurement systems allow measurements that otherwise would be impossible. SQUID measurement systems cleared the way for the development and improvement of new test and diagnostic methods in technology and medicine. Figure 5.17 presents the levels of magnetic induction related to biological and technological processes compared with the measurement capacities of SQUIDs [1].

The specified data indicate that the SQUID requires very efficient shielding from the magnetic field generated by the Earth and various technological devices. Commercial SQUIDs are mostly installed in systems for biomagnetic measurements.

A sensitive SQUID measurement system for biomagnetic studies is very susceptible to interference. Two methods are used for interference reduction: the measurements are carried out in a shielded room, or the interference is compensated in the measurement system. Especially biomagnetic studies require a shielded room in the form of a Faraday cage, ensuring particularly efficient shielding for interference in the frequency range from 0.1 Hz to ca. 100 Hz. Shielded rooms are also necessary for the installation of the quantum voltage standard. Usually the shield consists of a few layers of mu-metal (an alloy of 77 % Ni + 15 % Fe + 5 % Cu + 3 % Mo), which has a high magnetic permeability, and one layer of a good electric conductor (copper or aluminum) for electric-field screening. Probably the best-shielded rooms are those constructed in PTB in Berlin. One of them has 6 layers of mu-metal and one layer of

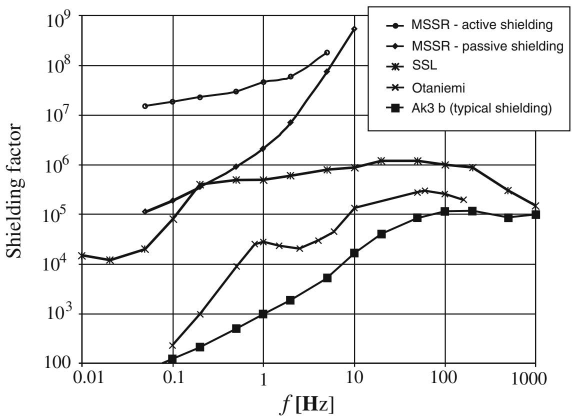


Fig. 5.18 Shielding factors of some shielded rooms (borrowed from [3], courtesy of Dr. M. Burghoff)

copper sheet. Another, new room, known as the Magnetically Super Shielded Room (MSSR), is a $(2.9\ \text{m})^3$ cube shielded with 7 layers of mu-metal and one 10 mm thick layer of aluminum [3]. This multilayer shielding is additionally supported by an active shield, with coils installed on the outer walls of the shielded room. The coils play the role of antennas for receiving interference and transmitting anti-interference at the same time. Anti-interference signals are generated by electronic control circuits of the active shield.

The plot in Fig. 5.18, borrowed from [3], compares the frequency dependences of the shielding factors of different shielded rooms. The shielding factor is defined as the ratio of the outside magnetic field H_{out} to the magnetic field H_{in} in the shielded room. The shielding factor depends on the frequency f; measured in the MSSR in PTB, its value ranged from 10^7 for $f = 0.03$ to 2×10^8 for $f = 5$ Hz. A two-layer mu-metal/Al shielding had a shielding factor of 20 for $f = 0.01$ Hz and 15×10^3 for $f = 1$ kHz [1, 2].

5.5.2 *Biomagnetic Studies*

A good example of magnetic measurements with SQUIDs, biomagnetic studies in medicine have led to the development of a new diagnostic method based on the measurement of the distribution of the magnetic field generated by human organs. A map of magnetic field distribution allows to draw conclusions regarding those parts of the studied organs in which the field is distorted, and find the cause of the distortion.

Studies of the electrical signals of an organ, the magnetic field distribution and the tissue density are believed to be necessary for the fullest image of the internal structure and functioning of the studied organ available by noninvasive methods. For example, such noninvasive methods of heart studies include electrocardiography (ECG or EKG), magnetocardiography (MCG) and magnetic resonance imaging

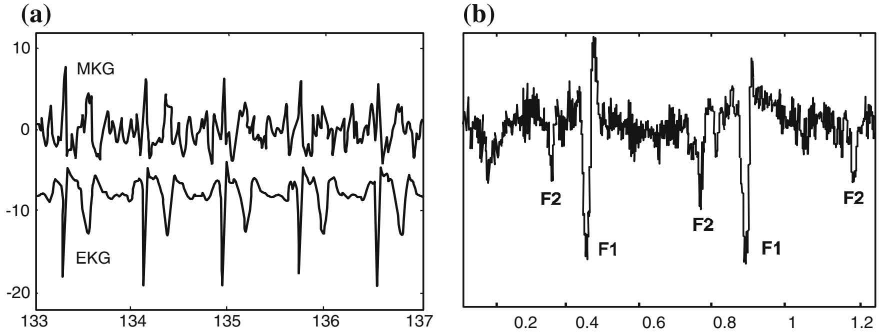


Fig. 5.19 Magnetocardiography signal traces: **a** MCG signal measured with an HTS DC-SQUID, against the ECG signal [33], **b** magnetic signals F1 and F2 generated by the hearts of twins in the mother's womb [27]. Both plots have time in seconds on the abscissa axis

(MRI) tomography, a technique using nuclear magnetic resonance. Medical biomagnetic studies are mainly used in five fields:

- Magnetocardiography (MCG), complementary to ECG (Fig. 5.19);
- Fetal magnetocardiography (FMCG), possible in conditions in which ECG cannot be performed;
- Magnetoencephalography (MEG), the study of the magnetic field of the brain for diagnostic purposes (e.g. epilepsy diagnostics);
- Magnetoneurography (MNG), for the identification of the regions of the brain responsible for the coordination of distinct functions, such as seeing, hearing or extremity movement;
- Magnetic marking, used to identify the path and velocity of micromagnets introduced to the digestive or respiratory tract or to the blood circulatory system for diagnostic purposes.

Used in biomagnetic studies, multichannel SQUID measurement systems with up to 250 channels with one sensor each, allow simultaneous measurement of the magnetic field in the whole studied area. Low-noise amplifiers, efficient low-pass filters (with the attenuation of up to 96 db/octave) and analog-to-digital converters with a resolution ranging from 12 to 16 bits are used in the measurement channels. Although the processes in the human body are relatively slow (e.g., the approximate breathing frequency is 0.2 Hz and the heart rate ranges from 1 to 2 Hz), simultaneous processing of hundreds of signals in the system imposes high demands regarding the dynamics of a single channel.

Since the strongest magnetic field is generated by the heart, magnetic measurements of the heart with SQUID systems are the most developed among the many MCG techniques. Only the use of SQUID sensors made MCG studies useful for clinical diagnosis, the other methods of measurement of the magnetic field of living organisms not being sensitive enough. Fetal magnetocardiography is of special practical importance. Many successful attempts to measure the magnetic

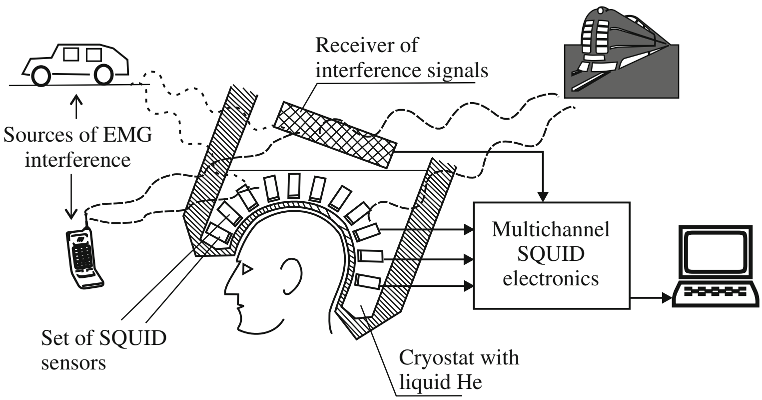


Fig. 5.20 Multichannel SQUID measurement system for MEG brain magnetic field studies

field of the heart of a child in the mother's womb have been made to date, as reported e.g. in [27].

Of more importance in medicine than MCG, magnetoencephalography, the study of the magnetic field of the brain, inspires both admiration for the advanced technology behind it, and fear of possible abuse. Magnetoencephalography allows to identify the regions of the brain responsible for disorders such as epilepsy. Figure 5.20 provides a schematic representation of a multichannel SQUID system for the measurement of the magnetic field of the brain in MEG studies.

The operation of the system shown in Fig. 5.20 includes interference compensation. The SQUIDs are placed in a helmet in the form of a nonmagnetic Dewar flask filled with liquid helium, shaped to fit the human head. All interference acts both on the SQUID and a sensor in the reference signal receiver. In the SQUID electronics the interference signal V_{int} is subtracted from the output signal $V_{meas} + V_{int}$ of the SQUID that has sensed both the measurement and interference signals. Measurements of the weakest magnetic signals, e.g. MEG measurements, require the simultaneous use of both techniques of interference shielding.

Magnetoneurography is used for precise localization of the regions of the brain responsible for the coordination of distinct functions, such as seeing, hearing or extremity movement. The magnetic field of the brain is measured after its optical, acoustic, tactile or other activation. Magnetoneurography studies are of help before brain surgery, as they indicate the regions of the brain to be avoided by the surgeon to maintain important functions of the body after the surgery.

Biomagnetic studies of the brain include also measurements of the brain magnetic field and its changes in various emotional states. This direction of research arouses ethical concerns, though. If the distribution of the brain magnetic field can be measured, magnetic field of the same distribution can as well be produced, allowing electronic emotional or even behavioral manipulation.

The directions of development of biomagnetic measurements include:

- Construction of still more sensitive (SQUID + gradiometer) sensors and low-noise amplifier circuits;
- Construction of systems with more and more measurement channels;
- Search for new, more efficient methods of interference shielding;
- Development of the system software, especially for magnetic field image processing.

Multichannel SQUID systems for biomedical studies have been constructed and implemented in clinical practice in many countries. The most technologically advanced are:

- The 256-channel system in the Electrotechnical Laboratory ETL and the Superconducting Sensor Laboratory in Japan [16] (national project);
- The 143-channel system (plus 26 reference channels) in Canada [32], designed and constructed by CTF Systems Inc. of Port Coquitlam;
- The 162-channel system in Italy, IECC-CNR, Rome [28];
- The 122-channel system in Finland, Neuromag Ltd., Helsinki [20];
- The 83-channel system in PTB, Berlin [10]; a 304-channel system is nearing completion;
- The 64-channel system in Biomagnetisches Zentrum at the Jena University, Germany, made by Philips [1].

Most commercial SQUIDs are installed in measurement systems for biomagnetic studies.

5.5.3 Nondestructive Evaluation of Materials

Used in the diagnostics of elements of mechanical constructions, nondestructive evaluation (NDE) is the detection and localization of inhomogeneities in the material structure (defectoscopy) without causing damage. The detected inhomogeneities include clefts, bubbles and impurities inside the structure, as well as cracks, notches and fissures at the surface (sometimes covered with a protective coat of paint). Major defectoscopy methods include X-ray examination, ultrasonic testing and measurement of the magnetic field generated by electric current in the studied object. However, metals, representing the most important class of tested materials, cannot be tested with X-rays, and multilayers (e.g. materials consisting of a few layers of metal sheet) are unfit for ultrasonic testing. Surface studies of metalic machine parts can be performed by scanning their magnetic field with various magnetic field sensors, such as common electrical coils or conventional measurement systems; these methods have their limitations, though. The skin effect must be taken into account in the induction of electric current in metal machine parts: the alternating current only flows at the surface and does not allow the localization of material flaws below the skin depth. Table 5.1 specifies the skin depth $d(f)$ for different current frequencies [6].

Table 5.1 Skin depth of different materials [6]

Frequency	Skin depth d [mm]			
	Material			
	Aluminum	Stainless steel	Plain steel	Graphite
1 Hz	83	330	5.6	4000
100 Hz	8.3	33	0.56	400
10 kHz	0.83	3.3	0.056	40
1 MHz	0.083	0.33	0.0056	4
Resistivity ρ, $\times\ 10^{-8}$ (Ωm)	2.7	43	10	6200
Relative permeability μ	1	1	800	1

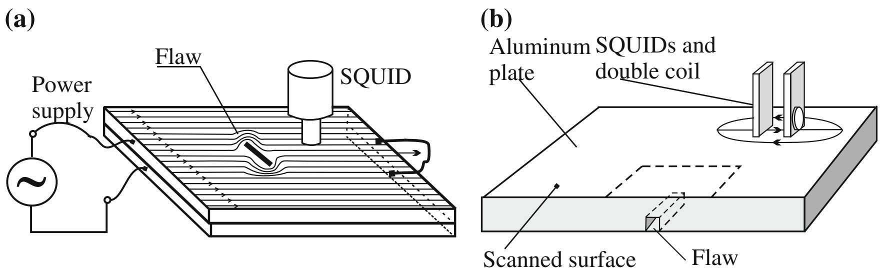


Fig. 5.21 System for nondestructive evaluation of machine parts using a SQUID: **a** with external power supply, **b** with integrated power supply and the SQUID

The data in Table 5.1 indicate that in major construction materials, plain steel and aluminum, the applied alternating current with a frequency above 100 Hz will flow at the surface of the metal with a maximum penetration depth of a few millimeters. Note that the application of low-frequency current requires large coils and reduces the scanning resolution.

In some of the developed NDE systems current is passed through the studied material to allow the measurement of the surface magnetic field distribution, as shown in Fig. 5.21a. The tested surface is scanned with a SQUID sensor placed in a cryostat. Other NDE systems include a measurement probe with a SQUID and coils inducing eddy currents at the studied surface, as depicted in Fig. 5.21b. Satisfactory sensitivity has been achieved in such systems with HTS DC-SQUID sensors cooled with liquid nitrogen [6, 29].

Independently of other NDE tests, SQUIDs are used for nondestructive evaluation of vital parts of machines and devices, such as nuclear reactor parts, lifting parts of aircraft (controlled periodically in the US Army) or lifting parts and covering of spacecraft. The space shuttle Columbia disaster in 2003 shows that such tests, far from being excessive precaution, are even insufficient.

5.6 SQUID Noise Thermometers

5.6.1 R-SQUID Noise Thermometer

In a noise thermometer the role of a temperature sensor is played by a resistor; the measurement signal is the thermal noise voltage (or electromotive force) across the resistor or its thermal noise current. Noise thermometers with an R-SQUID or a DC-SQUID have applications in low-temperature physics and metrology.

The resistive SQUID, or R-SQUID, is a high-frequency RF-SQUID with a resistively shunted Josephson junction. Direct current can flow through the shunt resistor to produce a DC voltage V_0 across the resistor. If a DC voltage V_0 and an AC voltage V_n sin ωt (the measurement signal) are applied to the junction, a signal with a frequency f will appear at the junction as a result of the internal AC Josephson effect:

$$f = \frac{2e}{h}(V_0 + V_n \, sin \, \omega t). \tag{5.56}$$

This periodic-frequency signal modulates the amplitude of the high-frequency bias signal. The modulated signal is amplified and then demodulated. The signal at the output of the amplitude demodulator has a changing frequency f proportional to the total voltage applied to the junction. The range Δf of the frequency f is proportional to the peak-to-peak voltage V_{pp} of the measurement signal. Thus, a weak voltage signal V_n is converted to a frequency signal Δf.

In noise thermometry the shunt resistor R in the R-SQUID, playing the role of a temperature sensor, is placed together with the RF-SQUID in the medium the temperature of which is being measured. The resistance R of the sensor is optimized to meet conflicting requirements. On the one hand, R should be as high as possible, since the thermal noise voltage increases with R. On the other hand, the resistor is connected in parallel with the Josephson junction, the correct functioning of which requires the shunt resistance to be many times lower than the resistance of the junction in its normal (non-superconducting) state.

Figure 5.22 shows the block diagram of the first R-SQUID noise thermometer [17], constructed in a National Bureau of Standards (at present NIST) laboratory in the USA by R. Kamper and J. Zimmerman.

The system uses an RF-SQUID with a bias frequency of 30 MHz. The Josephson junction in this noise thermometer is shunted by a 10 μΩ resistor. A direct current of the order of 1 μA produces a 10 pV voltage drop across the resistor; this voltage causes oscillations with a frequency of 5 kHz in the junction, according to (5.56). The fundamental frequency varies in the range 5 kHz ± $\Delta f/2$ as a result of the fluctuations of the electromotive force E_t of the thermal noise of the resistor, described by the Nyquist formula:

$$d\langle E_t^2 \rangle = 4k_B T R_x df. \tag{5.57}$$

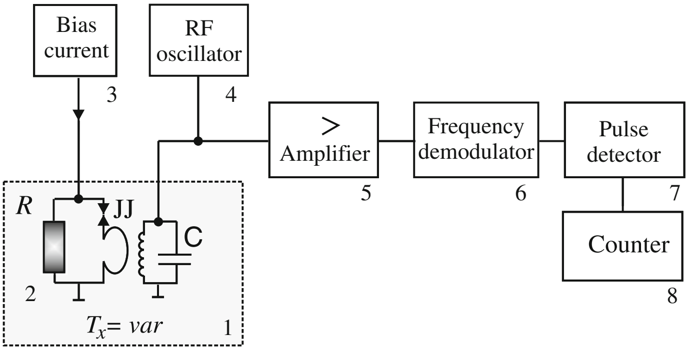


Fig. 5.22 R-SQUID noise thermometer: *1* measured medium, *2* sensor, *3* source of direct current, *4* high-frequency generator, *5* amplifier, *6* frequency demodulator, *7* pulse detector, *8* pulse counter

These changes in the frequency of the measurement signal can be measured with a spectrum analyzer or a frequency counter. The spectrum analyzer determines the fluctuation bandwidth Δf, which allows to calculate the temperature. The frequency counter measures the output frequency f. The temperature can be calculated from the variance of frequency in a series of measurements.

The Nyquist formula can be transformed to the formulas describing the discussed quantities:

$$\delta f = \frac{4\pi k_B R T_x}{\Phi_0^2} = 4.06 \cdot 10^7 \cdot R T_x \tag{5.58}$$

$$\sigma^2 = \frac{1}{N}\sum_i^N (f_i - f_m)^2 = \frac{2k_B R T_x}{\Phi_0^2 \tau}, \tag{5.59}$$

where δf is the bandwidth of the output signal of the SQUID, σ is the variance of frequency in a series of measurements, N the number of frequency measurements, f_i the frequency value obtained in the i-th measurement, f_m the mean value of the frequency measured in the series, and τ the gate time of the frequency counter.

Only fundamental physical constants and easily measurable quantities figure in the formulas (5.58) and (5.59), which therefore provide the basis for the measurement of absolute temperature. The R-SQUID noise thermometer is a *primary thermometer*. The resistance R of the sensor in the measured temperature can be determined from a series of frequency measurements:

$$R = \frac{hf_m}{2eI_0}, \tag{5.60}$$

where I_0 is the direct current flowing through the resistor R.

The statistical error of temperature measurement by this method depends on the number N of frequency measurements and is given by:

$$\delta_s = \sqrt{2/N}. \quad (5.61)$$

R. Soulen at NBS (presently the NIST) in Washington used an R-SQUID noise thermometer for the measurement of temperature in the range from 10 to 50 mK [30, 31]. The thermometer sensor was made of silicon copper. The bias frequency of the RF-SQUID in this device was 22 MHz. The readings of the noise thermometer were compared with those of a nuclear orientation thermometer with the cobalt isotope ^{60}Co used as the radiation source, and two radiation detectors, NaI and Ge(Li). In the range from 10 to 50 mK the reading of these two thermometers are in agreement with an error below 0.5 %. Soulen calibrated the resistance of the sensor of the noise thermometer in the temperature range from 10 to 520 mK. The resistance of the sensor was 17.2 μΩ at the ends of this range.

An improved noise thermometer for the range from 6 to 740 mK was developed at NIST by Soulen and Marshak [30]. The 350 MHz high-frequency signal used in this system allowed the measurement of f with better resolution and lower uncertainty. This noise thermometer was used, among others, for the measurement of the transition temperature of superconducting metals. Over its 10 years' operating period (1982–1992) the uncertainty of this thermometer was found to range from 0.08 to 0.27 % at different points of the scale.

An R-SQUID noise thermometer for the temperature range from 0.005 to 4.2 K was constructed by Hoffmann at PTB, Berlin [13]. In this system the SQUID was biased with a 320 MHz high-frequency signal. The frequency measurement was improved by the elimination of the dead time of the frequency counter. The time interval in which the measured pulses were counted started with the coming of the first pulse and ended once the last pulse had passed through the gate. The system used a PdPtAu alloy resistor, with a resistance ranging from 10 to 20 μΩ in the measurement range of the thermometer. In two measurements, lasting 400 h each, the readings of the noise thermometer were found to agree with the critical temperature T_c of indium with an uncertainty of 0.01 % [13].

More than 10 years later a new noise thermometer system with a DC-R-SQUID was developed by Menkel at PTB [24]. The circuit of the DC-R-SQUID consists of two Josephson junctions biased with direct current, as in the DC-SQUID, and a resistor R connected to them. Also the resistor is biased with direct current, which, however, is supplied by a difference source. The constant voltage across the resistor R is modulated by the thermal noise voltage, see Fig. 5.23. The frequency of the signal V_{out} at the terminals of each junction depends on the voltage across the resistor. The frequency of V_{out} oscillates, and the frequency changes δf are proportional to the temperature of the resistor sensor, according to the relation (5.58).

Advanced electronic technologies allowed to integrate the superconducting Josephson junctions and the resistor playing the role of the temperature sensor in a single chip. The resistor sensor was made of 99.9999 % high-purity silver in the

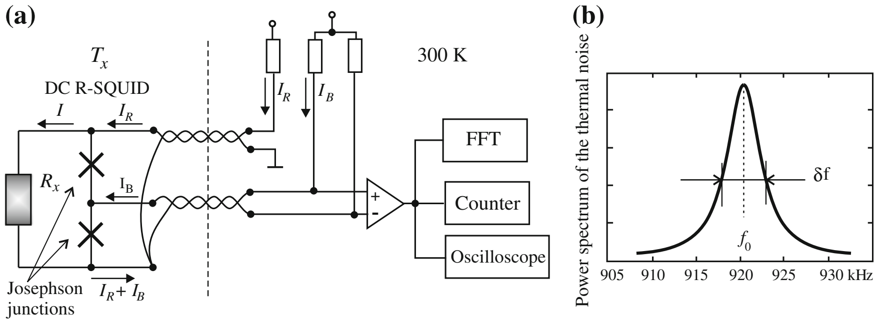


Fig. 5.23 DC R-SQUID noise thermometer [24]: **a** schematic diagram, **b** voltage power spectrum of the thermal noise in the temperature sensor

form of a 18 μΩ track. Good operation of this noise thermometer was confirmed in the range from 0.04 to 8.4 K, but uncertainty below 1 % was only attained in the narrower temperature range from 0.14 to 5.9 K. In the best experiment in the latter range the relative uncertainty was 0.53 % and the absolute uncertainty 1.1 mK [24].

Some disadvantage of the R-SQUID noise thermometer is the necessity to place the resistor sensor together with the SQUID in the measured medium (with a temperature T_x), as the resistor sensor is connected both mechanically and electrically with the SQUID. Variable temperature T_x can affect the physical properties of the SQUID.

5.6.2 DC-SQUID Noise Thermometer

In the DC-SQUID noise thermometer the SQUID is used as a high-sensitivity cryoelectronic amplifier. Because of the potential high sensitivity of the DC-SQUID, the DC-SQUID noise thermometer can provide better resolution than the noise thermometer with an RF-SQUID. In the DC-SQUID noise thermometer the temperature sensor and the SQUID can be spatially separated. The block diagram of the DC-SQUID noise thermometer is shown in Fig. 5.24. The thermal noise in the temperature sensor R_x produces an electromotive force E_t:

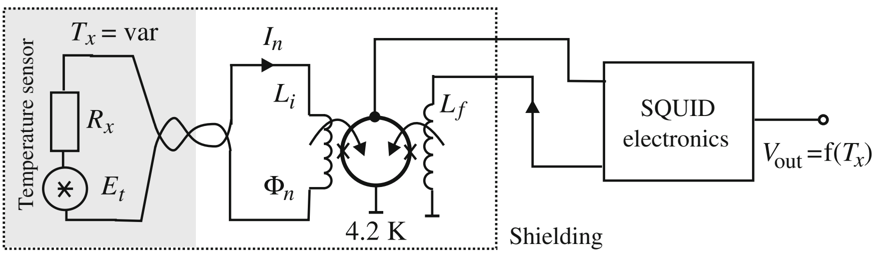


Fig. 5.24 DC-SQUID noise thermometer

$$d\langle E_t^2 \rangle = 4k_B T_x R_x df. \tag{5.62}$$

In the noise thermometer system presented in Fig. 5.24 the temperature sensor R_x is connected to an input coil L_i by means of superconducting wires, typically of niobium. Also the input coil is made of niobium. The thermal noise current I_n in the input circuit is a function of the absolute temperature T_x. Thus, also the magnetic flux Φ_n and the output voltage V_{out} are functions of T_x [26].

After a series of measurements of the output voltage, provided that the frequency band $f_{\max}$ of the SQUID system is much wider than the cutoff frequency f_c of the input circuit, $f_m \gg f_c$ (formula (5.50)), the measured temperature T_x is determined from the relation:

$$T_x = \frac{L_{ie}}{k_B k_p^2 M_i^2} \langle V_{out}^2 \rangle, \tag{5.63}$$

where $\langle V_{out}^2 \rangle$ is the mean square of the output voltage of the system, and k_p is its conversion coefficient, $k_p = V_{out}/\Phi_n$.

In the presented noise thermometer the measurement error ΔT_x due to the internal noise Φ_r of the SQUID is given by:

$$\Delta T_x = \frac{R_x \Phi_r^2}{4k_B M_i^2}. \tag{5.64}$$

The measurement error calculated for a noise thermometer with a sensor of resistance R_x = 1 mΩ and a DC-SQUID with noise level $2 \times 10^{-6}\, \Phi_0/\sqrt{Hz}$ and mutual inductance M_i = 5.2 nH is ΔT_x = 12 μK [26]. Obviously, also other sources of measurement error, mainly interference induced in the input circuits, can appear in the noise thermometer system.

A major advantage of the noise thermometer systems discussed so far is their ability to operate in low temperature conditions. A primary thermometer operating in a low temperature range is of much more use than e.g. a thermometer for the measurement of room temperature. The LTS DC-SQUID noise thermometer measures the temperature in the range T_x < 5 K. However, the superconducting sensor cannot operate above the critical temperature T_c, which imposes an upper bound on the measurement range. The measurement range of the noise thermometer can be extended to higher temperatures by using HTS SQUIDs (T_c > 80 K) [12].

Another solution, presented in Fig. 5.25, is galvanic isolation of the temperature sensor from the input circuit of the SQUID [14].

Experiments have been reported with a noise thermometer with an RF-SQUID circuit with an input coil or a DC-SQUID with an input coil [26] galvanically isolated from the temperature sensor R_x. The temperature sensor in the form of a solenoid of inductance L_x is placed in the measured medium of temperature T_x. In the noise thermometer the thermal noise signal from the sensor is transformed to the input circuit of the RF-SQUID, and the output signal V_{out} of the system is

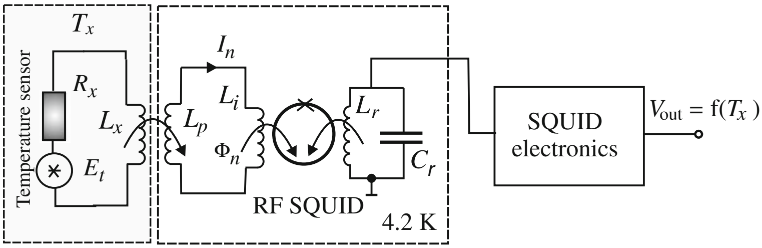


Fig. 5.25 Noise thermometer with a temperature sensor coupled inductively with the input circuit of the SQUID

proportional to the measured temperature. The noise thermometer shown in Fig. 5.25 operated with a commercial SHE 330 RF-SQUID system [14].

Tested in temperature measurements from 1.5 to 90 K, the system showed an uncertainty ranging from 0.2 to 1 %. A noise thermometer with a DC-SQUID coupled inductively with the temperature sensor was used for temperature measurements in the range from 4 to 290 K [26]. The measurement uncertainty was of a few percent, with a major contribution of interference. Along with the thermal noise signal from the temperature sensor, the pick-up coil, L_p, in the input circuit of the SQUID receives interference as well. Because of the substantially increased effect of interference, noise thermometers with inductive coupling cannot be recommended for operation in unshielded rooms.

5.6.3 Other Applications of SQUIDs

Apart from the applications discussed above, SQUID measurement systems are used also in the following fields of science and technology:

1. In geological studies conducted in the US for the localization of mineral deposits by the measurement of inhomogeneities of the magnetic field of the Earth on its surface. The high sensitivity of SQUID systems allows to detect a magnetic field gradient on the surface of the Earth, inferring a change in the structure of its crust. This method can be also used in attempts to localize and characterize mineral deposits under the bottom of the sea or ocean.
2. In electrical metrology, for the measurement of weak current and voltage signals, particularly in low-temperature systems. For example, the SQUID is used as a null detector in the cryogenic current comparator (CCC) discussed in Chap. 6.
3. For the detection of gravitational waves in terrestrial and space experiments.
4. In the army, for coast surveillance against foreign objects, mainly submarines, and for detecting hidden metal objects, e.g. mines, from aircraft equipped with a SQUID system.

References

1. W. Andrä, H. Nowak, *Magnetism in Medicine* (Wiley-VCH, Londyn, 1998)
2. J. Bork, H.-D. Hahlbohm, R. Klein, A. Schnabel, The 8-layer magnetically shielded room of the PTB. In *Proceedings of 12th International Conference on Biomagnetism*, ed. by J. Nenonen, R. Ilmoniemi, T. Katila. pp. 970–973, 2000
3. M. Burghoff, Noise reduction in biomagnetic recordings. Referat na 13th International Conference on Biomagnetism (Jena (Niemcy), 2002)
4. R. Cantor, T. Ryhänen, H. Seppä, in *Superconducting Devices and Their Applications*, ed. by H. Koch, H. Lübbig. A Compact Very Low Noise DC-SQUID Magnetometer (Springer, Heidelberg, 1992), pp. 276–280
5. J. Clarke, W.M. Goubau, M.B. Ketchen, Tunnel junction DC-SQUID: fabrication, operation and performance. J. Low Temp. Phys. **25**, 99–144 (1976)
6. G.B. Donaldson, A. Cochran, McKirdy D. McA., The Use of SQUIDs for Nondestructive Evaluation, Chapter XXX, ed. by H. Weinstock, NATO ASI Series. SQUID Sensors (Kluwer, New York, 1996), pp. 599–628
7. D. Drung, Advanced SQUID Read-Out Electronics, [ibid 6], Chapter 2, pp. 63–116
8. D. Drung, M. Mück, in *SQUID Electronics*, Chapter 4, ed. by J. Clarke, A. Braginski. SQUID Handbook (Wiley, New York, 2003)
9. D. Drung, Digital feedback loops for DC-SQUIDs. Cryogenics **26**, 623–627 (1986)
10. D. Drung, The PTB 83-SQUID system for biomagnetic applications in a clinic. IEEE Trans. Appl. Supercond. **5**, 2112–2117 (1995)
11. L. Grönberg, et al., A Low Noise DC-SQUID Based on Nb/Al-AlO_x/Nb Josephson Junctions, ibid [3], pp. 281–285
12. L. Hao, et al., Simulations and experiments on HTS resistive SQUIDs, in *Proceedings of the 4th European Conference on Applied Superconductivity,* Sitges (Hiszpania) Institute of Physics Conference Series, 1999, No 167, pp. 469–472
13. A. Hoffmann, B. Buchholz, UHF resistive SQUID noise thermometer at temperatures between 0.005 and 4.2 K, J. Phys. E: Scien. Instrum. **17**, 1035–1037 (1984)
14. M. Itoh et al., Measurements of Nyquist noise spectra in passive electric circuits by using a SQUID. Japan. J. Appl. Phys. **25**, 1097–1105 (1986)
15. R.C. Jaklevic, J. Lambe, A.H. Silver, J.E. Mercereau, Quantum interference effects in Josephson tunneling. Phys. Rev. Lett. **12**, 159–160 (1964)
16. K. Kado, G. Uehara, Multichannel SQUIID system. FED J. **5**, 12–19 (1994)
17. R.A. Kamper, J.E. Ziemmerman, Noise thermometer with Josephson effect. J. Appl. Phys. **42**, 132–136 (1971)
18. M. Kawasaki, P. Chaudhari, T. Newman, A. Gupta, Sub-micron YBaCuO Grain Boundary Junction DC-SQUID, ibid [5], pp. 150–154
19. M.B. Ketchen, Design and Fabrication Considerations for Extending Integrated DC-SQUIDs to the Deep Sub-micron Regime, ibid [5], pp. 256–264
20. J. Knuutila et al., A 122-channel whole-cortex SQUID system for measuring brain's magnetic fields. IEEE Trans. Magn. **29**, 3315–3320 (1993)
21. K. Likharev, V. Semenov, Fluctuation spectrum in superconducting point junctions. Pisma Zhurnal Eksperim. i Teoret. Fizyki **15**, 625–629 (1972)
22. A.P. Long, T.D. Clark, R.J. Prance, M.G. Richards, High performance UHF SQUID magnetometer. Rev. Sci. Instrum. **50**, 1376–1381 (1979)
23. O.V. Lounasmaa, *Experimental Principles and Methods Below* 1 K (Academic Press, New York, 1973)
24. S. Menkel, Integrierte Dünnschicht-DC-RSQUIDs für die Rauschthermometrie, Ph.D. thesis, Faculty of Physics (Jena University, Germany, 2001)
25. W. Nawrocki, K.-H. Berthel, T. Döhler, H. Koch, Measurements of thermal noise by DC-SQUID. Cryogenics **28**, 394–397 (1988)

26. W. Nawrocki, *Noise Thermometry* (in Polish) (Publishing House of Poznan University of Technology, Poznan, 1995)
27. H.W.P. Quartero, J.G. Stinstra, H.J.G. Krooshoop, M.J. Peters, in *Abstracts of Fetal Biomagnetism Satellite Symposium of 4th Hans Berger Conference*. Clinical applications for fetal magnetocardiography (Jena 1999) pp. 14–20
28. G.L. Romani, C. Del Gratta, V. Pizzella, *Neuromagnetism and Its Clinical Applications*, ibid [7], pp. 445–490
29. T. Schurig et al, NDE of semiconductor samples and photovoltaic devices with high resolution utilizing SQUID photoscanning, IEICE Trans. Electron. (Japan), **E85-C**, 665–669 (2002)
30. R.J. Soulen, H. Marshak, The establishment of a temperature scale from 0.01 to 0.05 K using noise and gamma-ray anisotropy thermometers. Cryogenics **20**, 408–411 (1980)
31. R.J. Soulen, W.E. Fogle, J.H. Colwell, Measurements of absolute temperature below 0.75 K using a Josephson-junction noise thermometer. J. Low Temp Phys. **94**, 385–487 (1994)
32. J. Vrba et al., Whole cortex, 64 channel SQUID biomagnetometer sensor. IEEE Trans. Appl. Supercond. **3**, 1878–1882 (1993)
33. R. Weidl, Realisierung eines biomagnetischen Meßsystems fűr die Intensiv-Kardiologie auf Basis von Hochtemperatursupraleitersensoren, PhD thesis, Faculty Physics, Jena University (Germany) (Shaker Verlag, Aachen, 2000)
34. V. Zakosarenko, F. Schmidl, L. Dörrer, P. Seidel, Thin-film DC-SQUID gradiometer using single YBaCuO layer. Appl. Phys. Lett. **65**, 779–780 (1994)
35. J.E. Zimmerman, A.H. Silver, Macroscopic quantum interference effects through superconducting point contacts. Phys. Rev. **141**, 367–375 (1966)

Chapter 6
Quantum Hall Effect and the Resistance Standard

Abstract This chapter opens with a presentation of the classical Drude theory of electrical conduction and a theory proposed by Landauer. Based on the assumption that electrical conduction can be modeled as transfer of electrons between two electron reservoirs, the Landauer theory proves to describe particularly well the electrical resistance in nanoscale conductors, i.e., in nanostructures. Surprisingly, this theory implies that the conductance (and resistance) of a nanostructure is independent of its material and temperature, and only depends on the dimensions of the sample, changing in a stepwise manner with a step of $h/2e^2$ representing the conductance quantum. The quantization of electrical and thermal conductance in nanostructures has been verified experimentally. Conductance quantization in nanostructures is used in the analysis of large-scale integration circuits, as required by the currently used 14 nm technology and future technologies.

6.1 Hall Effect

The Hall effect was discovered in 1879 by Edwin Hall, at that time a student at the University of Baltimore in the United States. Investigating the distribution of electric potential in a rectangular golden foil conducting electric current, depicted in Fig. 6.1, Hall observed a potential difference not only along the sample (the V_x connection), but also across the sample, when a current I flowed along the foil in a magnetic field of induction B.

The value of the transverse voltage V_H was proportional to the density j of the electric current and to the induction B of the applied magnetic field. This phenomenon is known as the classical Hall effect and the potential difference V_H is referred to as the Hall voltage.

The production of the Hall voltage is related to the effect of the magnetic field on the electron charges flowing along the sample (a plate with a width a and a thickness b). The electrons experience a Lorentz force F, which curves their trajectory in the direction perpendicular to the current density vector $\boldsymbol{J}$ and the magnetic induction vector $\boldsymbol{B}$:

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_6

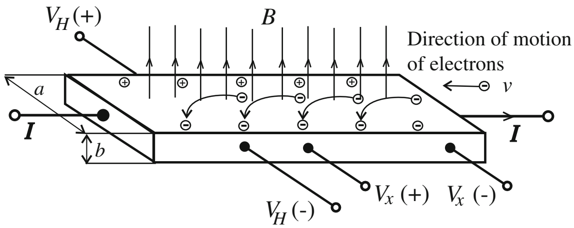


Fig. 6.1 Distribution of electric potential in a conductive sample and the production of the Hall voltage

$$\boldsymbol{F} = e\boldsymbol{v}\boldsymbol{B}, \tag{6.1}$$

where F is the force acting on the electron, e is the electron charge, v the velocity of the electron (oriented in the direction opposite to that of the current density vector), and B the magnetic induction.

There are more electrons along one side of the sample than along the other side. Negative electric charge accumulates on one side of the sample, and positive charge (positively charged ions) accumulates on the other side, as shown in Fig. 6.1. The electric charge produces an electric field of magnitude $\boldsymbol{E_p}$ across the sample. The force exerted by this electric field on the electron counterbalances the Lorentz force:

$$\begin{aligned} e\boldsymbol{E}_p &= e\boldsymbol{v}\boldsymbol{B} \\ \boldsymbol{E}_p &= \boldsymbol{v}\boldsymbol{B}. \end{aligned} \tag{6.2}$$

The potential difference across the sample, i.e. the Hall voltage V_H, depends on the magnitude of the electric field and the width a of the sample:

$$V_H = a\boldsymbol{E}_p = a\boldsymbol{v}\boldsymbol{B}. \tag{6.3}$$

The current density j in the conductor depends on the concentration N of free electrons (in a metal: 10^{22} cm^{-3} < N < 10^{23} cm^{-3}) and the velocity v of the electrons:

$$j = \frac{I}{ab} = Nev,$$

where j is the current density, b the thickness of the sample, and N the concentration (in m^{-3}) of free electrons in the material of the sample.

From this relation it follows that:

$$v = \frac{j}{Ne} \tag{6.4}$$

$$V_H = \frac{a}{Ne} jB = \frac{a}{Ne}\frac{I}{ab} B = \frac{B}{Neb} I = R_H I. \tag{6.5}$$

The coefficient R_H is known as the Hall resistance. Sometimes the density of free electrons is defined per unit surface of the conductor, as a surface concentration $N_s = Nb$ measured in units of m^{-2}. Expressed by the surface concentration, the formula for the Hall resistance becomes:

$$R_H = \frac{B}{Neb} = \frac{B}{N_s e}. \tag{6.6}$$

The (6.6) implies that the Hall resistance only depends on the magnetic induction and two constants: the electron charge e and the surface concentration N_s of free electrons, which is a material constant. Representing the ratio of the Hall voltage V_H to the current I along the sample, the Hall resistance R_H can be neither described in terms of the theory of electrical conduction, nor measured with an ohmmeter.

6.2 Quantum Hall Effect

6.2.1 Electronic Devices with 2-DEG

The quantum Hall effect (QHE) is observed as quantized (stepwise) change in the resistance of a sample as a function of the applied magnetic induction at a temperature of a few kelvins or lower. The quantum Hall effect was discovered in 1980 by Klaus von Klitzing [11], born in Sroda, Poland, in 1943. For this discovery von Klitzing received the Nobel Prize in 1985. The QHE is observed in samples with two-dimensional electron gas (2-DEG). Figure 6.2 shows schematically depicted samples with a decreasing number of degrees of freedom for the motion of electrons:

- A macroscopic sample in which electrons move freely in the three spatial directions, forming a 3-dimensional electron gas (3-DEG). There is no size-related quantization in a 3-DEG sample.
- A sample in which the motion of electrons is only possible in two dimensions. Thus, the electrons in the sample form a 2-dimensional electron gas (2-DEG). In a 2-DEG sample quantization only occurs in one direction.
- A sample in which electron transport is only possible in one dimension, along the sample. The electrons in the sample form a 1-dimensional electron gas (1-DEG). Examples of 1-DEG samples are nanowires (discussed in Chap. 7 or chains of atoms. In a 1-DEG sample quantization occurs in two directions.
- A quantum dot, which is a zero-dimensional system (a 0-DEG sample). A quantum dot is a space bounded by potential barriers in three dimensions. A particle with a wavelength comparable to the size of the dot is locked within this

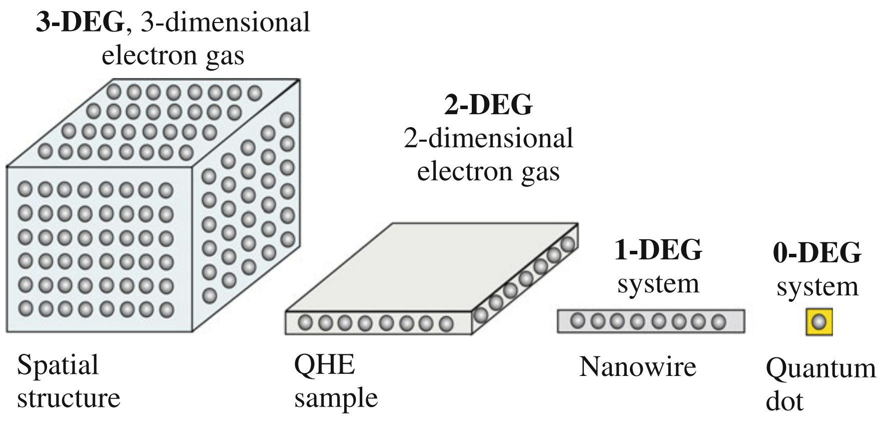


Fig. 6.2 Samples with a decreasing number of degrees of freedom for the motion of electrons

space. The states of the particle locked in a quantum dot are described in the formalism of quantum mechanics. Only discrete states (determined from the Schrödinger equation) of the particle are possible in each of the three directions.

Electrons in a 2-DEG electron gas are frequently compared to balls on a billiard table. Pressed to the surface of the table by the force of gravity, the balls can only move in the plane of the surface of the billiard table. A similar behavior of electrons must be imposed in a 2-DEG sample, in which the motion of electrons should only be affected by their interaction and by the applied magnetic field. Particularly undesirable is scattering of electrons by donors, impurities and ions in thermal oscillation. The latter can be reduced by cooling the sample.

6.2.2 *Physical Grounds of the Quantum Hall Effect*

Quantization of the Hall resistance takes place when a strong magnetic field with a magnetic induction of a few Tesla is applied in the direction perpendicular to the surface of a heterostructure junction within the sample. At a low temperature T and with a low surface concentration N_s of charge carriers conduction electrons are only present near the surface of the junction. In the direction perpendicular to the surface of the junction the motion of electrons is disallowed. According to the laws of classical physics, free electrons in a magnetic field move along a circular path as a result of the Lorentz force. On the other hand, by the rules of quantum mechanics only some trajectories (orbits) are allowed to the motion of electrons, just as only some trajectories are available to an electron traveling on a circular orbit around the nucleus of an atom. The energy levels available to the motion of free electrons are known as the Landau levels. The spacing ΔE_L between the Landau levels in an energy band is equal throughout the band and depends on the magnetic induction B:

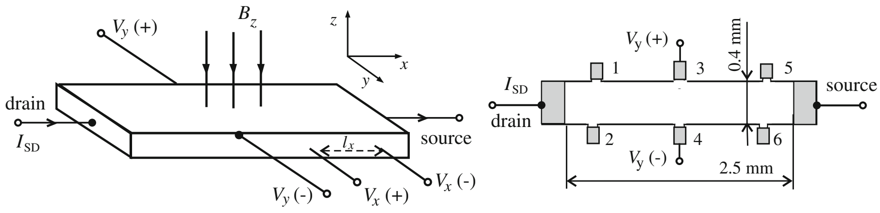


Fig. 6.3 Sample with two-dimensional electron gas (2-DEG) in which the quantum Hall effect is produced

$$\Delta E_L = \frac{heB}{4\pi m}, \tag{6.7}$$

where ΔE_L is the spacing between Landau levels, e denotes the electron charge, and m is the electron mass.

If the magnetic induction B_z is low, the electrons occupy states in a continuous manner up to the Fermi level, and move freely in the xy plane, as shown in Fig. 6.3. Under these conditions, the Hall resistance R_H, given by the formula (6.8) below, is independent of the size of the sample:

$$R_H = \frac{V_{y1} - V_{y2}}{I_{SD}} = \frac{B_z}{N_s e}, \tag{6.8}$$

where R_H is the Hall resistance, B_z the magnetic induction, and N_s the surface concentration of electrons.

A characteristic parameter of a device with two-dimensional electron gas, the degeneracy d is the maximum number of electrons that can occupy each of the Landau levels. The degeneracy depends on the magnetic induction B and the physical constants e and h:

$$d = \frac{eB}{h}, \tag{6.9}$$

where d is the degeneracy of the device (sample) with two-dimensional electron gas.

Let us consider a 2-DEG sample with a surface concentration N_s of free electrons. Elementary particles strive to occupy states at the lowest available energy levels. If the magnetic induction $B_1 = hN_s/e$ is large enough, all the electrons occupy the lowest energy level. The Hall resistance in such a state is determined by the relation (6.6) and the formula for the induction B_1:

$$R_H = \frac{B_1}{N_s e} = \frac{hN_s}{e}\frac{1}{N_s e} = \frac{h}{e^2} = R_K, \tag{6.10}$$

where $R_K = 25812.807 \times (1 \pm 2 \times 10^{-7})\ \Omega$ is the von Klitzing constant value recommended by the Consultative Committee for Electricity at the General Conference on Weights and Measures.

If the magnetic induction is reduced to $B_2 < B_1$, by the formula (6.8) the degeneracy d will decrease as well. Consequently, the number of states available to electrons at every energy level, including the lowest one, will be lower. A part of the electrons will have to occupy the neighboring higher energy level. Further reduction of the magnetic induction will result in further decrease in the degeneracy d; thus, successive Landau levels will become available to free electrons. For a magnetic induction corresponding to i energy levels available to electrons the Hall resistance is:

$$R_H(i) = \frac{h}{ie^2} = \frac{R_K}{i}. \tag{6.11}$$

Figure 6.4a shows the Hall resistance plotted versus the magnetic induction in the quantum Hall effect. In real systems the Landau levels are broadened as a result of splitting in the magnetic field; however, they will remain well separated from each other on the condition that $\mu B_z \gg 1$, where μ is the electron mobility. For example, in a semiconductor GaAs sample with an electron mobility $\mu \approx 25\ \mathrm{T}^{-1}$ at a temperature $T = 1$ K the Landau levels are separated when $B_z \geq 10$ T. Under the conditions of separation of the Landau levels the state distribution consists of states at separate split Landau levels, each with $N_s = eB_Z/h$ electrons per unit area.

Surprising results were obtained in studies of the quantum Hall effect in samples cooled to a temperature below 1 K when the magnetic induction was increased to exceed substantially the value corresponding to the first quantization step ($i = 1$) of the Hall resistance. The fractional quantum Hall effect [6, 10], illustrated by the plot in Fig. 6.4b, was found to occur in these conditions. The explanation of the

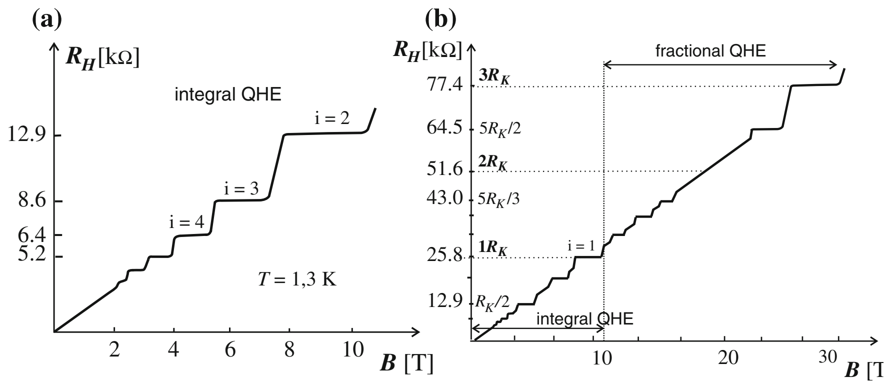


Fig. 6.4 Two types of quantum Hall effect: **a** integral QHE, **b** fractional QHE ($R_K = h/e^2 \approx 25.8$ kΩ)

fractional quantum Hall effect is based on the formation of a new phase, referred to as a quantum liquid. The quantum liquid forms as a result of the interaction between electrons and vortices of the magnetic field. Correlated electrons in the quantum liquid exhibit an electric charge that represents a fraction of the elementary charge e, e.g. $e/3$ or $e/5$. For the discovery of this phenomenon Robert B. Laughlin, Horst L. Störmer and Daniel C. Tsui received the Nobel Prize in 1998. A detailed description of the fractional Hall effect would go beyond the scope of this book, but can be found in an excellent paper by Störmer [10].

Robert Laughlin proposed an interesting physical interpretation of the Hall resistance in both the integral and fractional quantum Hall effects [10]. His interpretation is inferred from the physical conditions necessary for the occurrence of the quantum Hall effect, specifically the interaction of free electrons with the magnetic field (in the vortex form). Laughlin transformed the relationship $R_H = h/ie^2$ to obtain the formula:

$$R_H = \frac{h}{ie^2} = \frac{2(h/2e)}{ie} = \frac{2\Phi_0}{ie},$$

where h is Planck's constant, e the electron charge, and $\Phi_0 = h/2e \approx 2.07 \times 10^{-15}$ Vs is the quantum of magnetic flux, or fluxon.

Laughlin pointed out that the coefficient R_H represents the ratio of the magnetic flux quantum $\Phi_0 = h/2e$ to the electron charge e. This interpretation is acceptable at least as an alternative for the theory of electrical conduction, which does not explain the quantum Hall resistance. The Hall resistance is simply a coefficient representing the ratio of the transverse voltage to the current along the sample, and only for this reason its value is expressed in ohms. In the quantum Hall effect the magnetic flux interacts with fractions of the elementary charge, e.g. $R_H = 3h/ie^2 = \Phi_0/(e/3) \approx 77.4$ kΩ.

6.2.3 QHE Samples

Electronic components with two-dimensional electron gas include metal-oxide semiconductor field-effect transistors (MOSFET) and heterostructures. Two-dimensional electron gas is also present in one-atom-thick carbon sheets known as graphene. A system that exhibits the quantum Hall effect is referred to as a QHE sample. In a silicon MOSFET transistor electron gas forms in silicon at its interface with an insulator layer of silicon dioxide (SiO_2). Attracted by a positive voltage V_{GS} applied to the metal gate of the transistor, electrons accumulate at the surface of the silicon/insulator junction, as shown in Fig. 6.5a. The voltage V_{GS} produces an electric field which, acting on the electrons, keeps them in the plane of the silicon/insulator junction. MOSFET samples were used in the first measurements of the quantum Hall effect performed by von Klitzing.

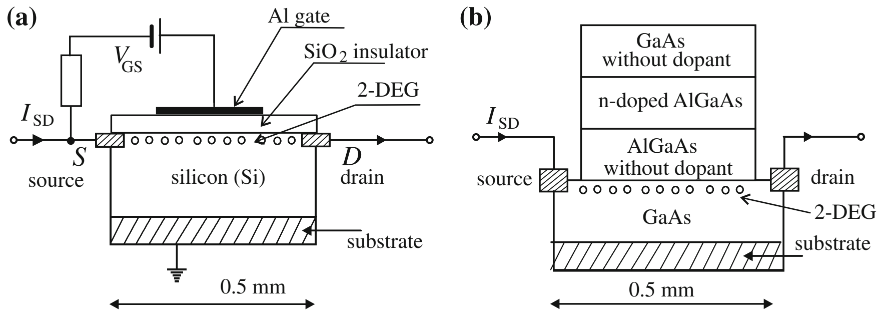


Fig. 6.5 Quantum Hall effect samples: **a** MOSFET transistor, **b** GaAs/AlGaAs heterostructure [2]

Better quality samples with two-dimensional electron gas can be obtained by using semiconductor multilayers with one type of charge carriers, i.e. either n-semiconductors, with a majority of electrons, or p-semiconductors, with a majority of holes, in all the layers. Such semiconductor structures are known as heterostructures. An important technical parameter of a heterostructure is the electron mobility μ, expressed in $m^2/Vs = T^{-1}$. The higher the electron mobility, the better the heterostructure. The electron mobility is defined as the ratio of the drift velocity v of the electric charge to the magnitude of the electric field E: $\mu = v/E$. Figuratively speaking, the electron mobility is a measure of the degree of freedom for the motion of electrons. Figure 6.5b shows the cross section of a heterostructure. Of crucial importance in this heterostructure is the junction between the 500 μm thick layer of gallium arsenide (GaAs) and the 0.5 μm thick layer of undoped aluminum gallium arsenide (AlGaAs). Since these two semiconductor materials have the same lattice constant (the spacing between atoms in the crystal lattice), the interface (junction) between the layers is free of stress. The two semiconductors differ slightly in electron affinity. Silicon atoms are added to the upper AlGaAs layer as a donor dopant (Fig. 6.3b), and each silicon atom releases one electron. Free electrons diffuse to the GaAs layer, where they can occupy states in the conduction band at an energy level lower than that available in the AlGaAs layer. Electrons are attracted to the plane of the junction by positively charged silicon atoms, and therefore can only move in two dimensions. Thus, the conditions for the formation of 2-DEG are fulfilled: electrons can only move in a plane, and their motion is not disturbed by collisions with silicon atoms from which they originate.

Figure 6.6a explains the formation of two-dimensional electron gas in a heterostructure.

The n-semiconductor in layer 1 is undoped or doped only slightly. The band gap, or forbidden energy band, in layer 1 is the energy gap E_{S1} between the bottom E_{C1} of the conduction band and the top E_{V1} of the valence band in this layer, $E_{S1} = E_{C1} - E_{V1}$. In layer 2 the band gap $E_{S2} = E_{C2} - E_{V2}$, is wider than E_{S1}, as the semiconductor 2 (AlGaAs, n-semiconductor as well) is heavily doped. At the interface between layers 1 and 2 the band gaps must meet, hence the band gap

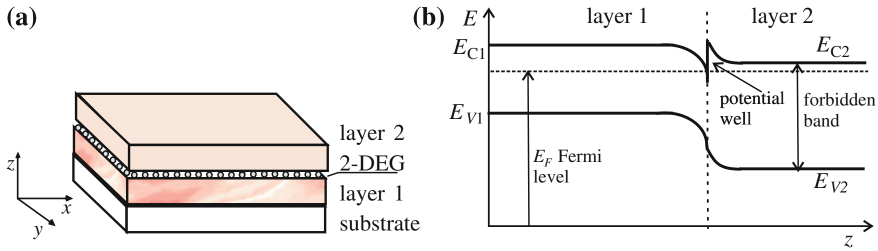


Fig. 6.6 Formation of 2-DEG at a GaAs/AlGaAs heterostructure junction: **a** schematic depiction of the sample; **b** energy levels in the semiconductor layers in the heterostructure

bending shown in Fig. 6.6b. Free electrons in the junction region are those electrons that have quit the dopant atoms in the semiconductor material constituting layer 2 and moved to layer 1, since the Fermi level in layer 1 was lower than that in layer 2, $E_{F1} < E_{F2}$, before the two semiconductors were joined. In the joint layers, however, the Fermi level is equal. In Fig. 6.6b this Fermi level is denoted E_F. As a result of the bending of the band gap at the junction between the two layers a narrow potential well forms in the z direction (the triangular segment of the line representing the conduction band bottom E_C in the plot). Free electrons in the junction region can only move in the plane parallel to the junction, which is the xy plane. Thus, these electrons form a two-dimensional electron gas.

Typically, heterostructures used as quantum Hall effect samples are 2.5 mm × 0.5 mm samples made of GaAs/($Al_{0.33}$/$Ga_{0.67}$)As [4, 10]. The n-doped AlGaAs layer boosts the creation of electron gas by the basic GaAs/AlGaAs junction, while the upper GaAs layer protects from oxidation the AlGaAs layer beneath it. Here are some examples and parameters of QHE samples used in quantum Hall effect standards:

- The resistance standard in the BIPM laboratory in Sèvres: GaAs heterostructure, charge carrier mobility μ = 30 T^{-1}, surface carrier concentration $N_s = 5.1 \times 10^{15}$ m^{-2}, parameter $i = 2$, operation temperature $T = 1.3$ K, magnetic induction $B = 10.5$ T, measurement current of 40 μA [1];
- The resistance standard in the Swiss Federal Office of Metrology (METAS): GaAs heterostructure, charge carrier mobility $\mu = 42$ T^{-1}, surface carrier concentration $N_s = 4.8 \times 10^{15}$ m^{-2}, parameter $i = 2$, operation temperature $T = 0.3$ K, magnetic induction $B = 9.9$ T, measurement current of 40 μA [1]—see Fig. 6.7b.
- The resistance standard in the American Institute of Standards and Technology (NIST): GaAs heterostructure, charge carrier mobility $\mu = 11$ T^{-1}, surface carrier concentration $N_s = 5.6 \times 10^{15}$ m^{-2}, parameter $i = 2$, operation temperature $T = 0.3$ K, magnetic induction $B = 10.5$ T, measurement current of 40 μA [2];
- The resistance standard in the Central Office of Measures (GUM) in Warsaw, Poland: GaAs heterostructure, charge carrier mobility $\mu = 40$ T^{-1}, surface carrier

concentration $N_s = 3.9 \times 10^{15}$ m^{-2}, parameter $i = 2$, operating temperature $T = 0.3$ K, magnetic induction $B = 8.35$ T, measurement current of 10 or 100 μA.

Metrologists test arrays with numerous QHE samples in order to obtain not only the values of quantum resistance equal to R_K or $R_K/2$ but values $N \times R_K$ without of the fractional quantum Hall effect [9].

6.2.4 Quantum Hall Effect in Graphene

A new material with excellent physical properties, graphene was first produced in 2004 by Andre Geim and Konstantin Novoselov at the University of Manchester [7]. Graphene is a new allotrope of carbon with amazing characteristics (the other allotropes of carbon include graphite, diamond and fullerenes). It is a structure of carbon atoms in the form of a flat one-atom-thick sheet. Atoms in graphene form a hexagonal lattice, shown in Fig. 6.7a, with a lattice constant of 1.42 Å. For their discovery of graphene Geim and Novoselov received the Nobel Prize in 2010.

Geim and Novoselov were also the first to study the physical properties of graphene, and observed the quantum Hall effect in graphene samples at room temperature [8]. Due to the two-dimensional structure of graphene the motion of electrons in a graphene sample is two-dimensional as well. As electrons in graphene form a two-dimensional electron gas, a graphene sheet is a 2-DEG sample. It can be used as a QHE sample without additional application of electric field.

The physical properties of graphene include, for example:

- A very high electron mobility, $\mu = 25$ T^{-1} at a temperature of 300 K, and $\mu = 100$ T^{-1} at 4 K, with phonon scattering assumed; for comparison, at 300 K

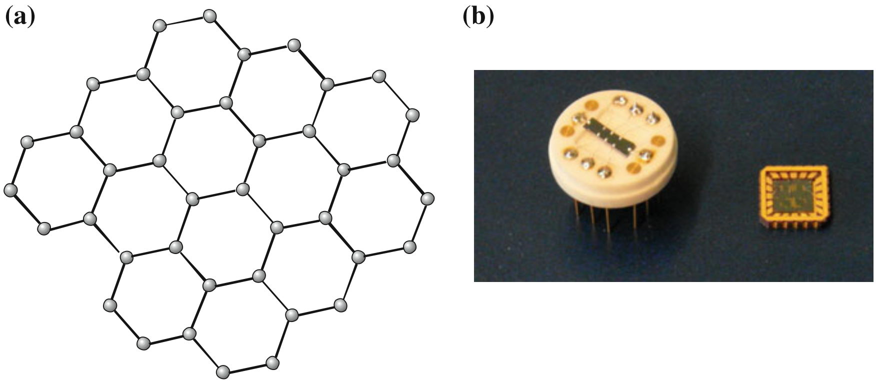


Fig. 6.7 The structure of graphene (a one-atom-thick two-dimensional lattice of carbon atoms) and QHE samples in PTB laboratory, Germany: a heterostructure sample fabricated by METAS (*left*) and a graphene sample (*right*)

the electron mobility in silicon (in MOSFET QHE samples) is $\mu = 0.15\ \mathrm{T}^{-1}$, and in gallium arsenide (in heterostructures used as QHE samples) $\mu = 0.85\ \mathrm{T}^{-1}$.

- A very high thermal conductivity, of the order of 5000 W/m K. For comparison, the conductivity of diamond, a very good conductor of heat, ranges from 900–2,300 W/m K, and the conductivity of silver is 429 W/m K.
- A low resistivity of the order of $10^{-8}\ \Omega$ m; (lower than the resistivity of copper, which is $1.72 \times 10^{-8}\ \Omega$ m.
- Very high mechanical strength and stability.

The QHE in graphene samples has been confirmed experimentally at 300 K and in a very strong magnetic field (with a magnetic induction of 45 T) [8]. However, there is a long way ahead until graphene might be used in quantum standards of electrical resistance. The technology of manufacturing graphene samples by micromechanical methods is extremely expensive and makes graphene the most expensive material in the world: the price of a plate with a diameter of 100 μm and a thickness of 0.154 nm (the diameter of the carbon atom) ranges from $500–$1000. Graphene sheets are used as QHE samples in metrology. Although the quantum Hall effect is observed in graphene at room temperature, the observation of this phenomenon requires extremely high magnetic induction.

Research is in progress for the potential use of graphene, instead of semiconductors, in electronic devices. Controlled switching between two states, conducting and non-conducting, is required for applications in electronic digital circuits. These two states correspond to two logic states of the digital system. As mentioned above, graphene is a good electrical conductor. However, when treated with hydrogen, it becomes an insulator [3]. Hydrogenated graphene is known as graphane. The first papers on potential applications of graphene in electronics appeared 2 years after its discovery. Researchers at the Massachusetts Institute of Technology (MIT) in Massachusetts, USA, have built a frequency multiplier with a graphene field effect transistor (G-FET) [12]. The potential upper operating frequency of the multiplier, as well as that of other graphene-based digital circuits, is as high as 1000 GHz.

6.3 Measurement Setup of the Classical Electrical Resistance Standard at the GUM

As an example of classical standard of the unit of electrical resistance, in this section we shall discuss the national standard of electrical resistance kept at the Central Office of Measures (GUM) in Warsaw. The national standard of the unit of electrical resistance at the GUM is a group standard consisting of six calibrated resistors with a nominal resistance of 1 Ω each. The standard has been legally introduced at the GUM in December 1980. The resistive part of each calibrated resistor has the form of a manganin strip. Each resistor is enclosed in a metal casing

Fig. 6.8 Laboratory resistors of the Polish national standard of electrical resistance: a 9330 resistor manufactured by Guildline (on the *left*) and a P321 resistor manufactured by ZIP (on the *right*)

with four contacts atop; two of these are current terminals and the other two are voltage terminals.

Three of the resistors that constitute the national standard were manufactured by Guildline and the other three made in ZIP factories in the former Soviet Union—Fig. 6.8. The resistance of each resistor is compared with the resistance of the other resistors in the group by means of a bridge current comparator. The mean value of resistance of the six resistors is determined as a result of this comparison. The mean resistance of the resistors constituting the standard is the correct resistance value of the national standard. Four of the six resistors constituting the national standard (always the same four) are periodically compared with the international standard at the BIPM. A correction related to the long-term stability (drift), and a correction of the effect of temperature and pressure are taken into account in the measurements of all the resistors calibrated at the BIPM.

Table 6.1 Basic data of components of the Polish national standard of electrical resistance

	Component	Technical characteristic	Type	Manufacturer
1	Standard resistor	Nominal resistance of 1 Ω	P321	ZIP
2	Standard resistor	Nominal resistance of 1 Ω	P321	ZIP
3	Standard resistor	Nominal resistance of 1 Ω	P321	ZIP
4	Standard resistor	Nominal resistance of 1 Ω	9330	Guildline
5	Standard resistor	Nominal resistance of 1 Ω	9330	Guildline
6	Standard resistor	Nominal resistance of 1 Ω	9330	Guildline
7	Current comparator	Bridge direct-current comparator	9975	Guildline
8	Oil tub	Oil temperature of (23.0 ± 0.1) °C	9730 CR	Guildline
9	Resistance bridge	For measurements of oil temperature	8640	Tinsley
10	Platinum thermometer	For measurements of oil temperature	5187 SA	Tinsley

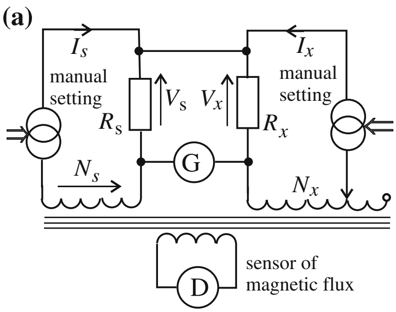


Fig. 6.9 DC current comparator: **a** schematic diagram of the comparator, **b** Kusters comparator with feedback

The technical characteristics and identification features of parts and components of the Polish national standard of electrical resistance are specified in Table 6.1.

Because of their relatively large temperature coefficient the resistors are stored and measured in stable temperature conditions ensured by placing them in an oil bath with a temperature stabilized at 23 °C, in accordance with BIPM recommendations. The temperature dependent uncertainty of measurement and the change in temperature during the measurement should not exceed 0.01 °C. The resistance of the standard resistors is also sensitive to changes in the atmospheric pressure. This effect is taken into account by measuring the atmospheric pressure and correcting the measured resistance values of the standard resistors with a known pressure factor. A crucial part of the setup of the national resistance standard at the GUM is a system for precise measurements of resistance. The system consists of a direct-current comparator in the bridge configuration, conceived by N. L. Kusters and manufactured by Guildline Instruments of Canada. The principle of operation of the Kusters direct-current comparator is illustrated in Fig. 6.9a.

The comparator is a toroidal transformer in which direct current flows in the primary and secondary windings. A detector winding and a modulator winding are used for detecting magnetic flux in the magnetic core of the comparator. The procedure of measurement of the resistance R_x consists in adjusting to zero the reading of two instruments, the galvanometer G and the zero magnetic flux detector D.

The currents I_s and I_x in the comparator circuit, shown in Fig. 6.7a, are adjusted manually so that the voltage V_x across the measured resistor R_x is equal to the voltage V_s across the standard resistance R_s.

The state in which V_x equals V_s is indicated by the galvanometer G:

$$\begin{gathered} V_x = V_s \\ R_x I_x = R_s I_s \\ \frac{R_x}{R_s} = \frac{I_s}{I_x}. \end{gathered} \tag{6.12}$$

Equal values of ampere-turns, $N_x I_x = N_s I_s$, in opposite directions should be obtained by changing the number of turns N_x in the I_x circuit. In that case the resultant magnetic flux in the transformer core is zero. The state with zero magnetic flux is indicated by the zero magnetic flux detector D:

$$\begin{gathered} N_x I_x = N_s I_s \\ \frac{N_x}{N_s} = \frac{I_s}{I_x}. \end{gathered} \tag{6.13}$$

From the formulas (6.12) and (6.13) it follows that:

$$\begin{gathered} \frac{R_x}{R_s} = \frac{I_s}{I_x} = \frac{N_x}{N_s} \\ R_x = \frac{N_x}{N_s} R_s. \end{gathered} \tag{6.14}$$

Thus, the resistance value of the resistor R_x is determined from the fixed reference resistance value R_s, the number of turns N_s in the reference circuit, and the adjustable number of turns N_x. A disadvantage of the comparator shown in Fig. 6.9a is that it requires voltage balance $V_x = V_s$ and ampere-turn balance $N_S I_S = N_x I_x$ *at the same time*. In practice the balance of ampere-turns means zero magnetic flux. Because of the necessity to balance simultaneously the voltage and the ampere-turns the DC power sources in the system should have a very high stability. The measured value of resistance is given by the relation (6.14).

The Kusters comparator has an improved design, presented in Fig. 6.9b. In this system the manually controlled source of current I_s is replaced with an automatically controlled source of current. The automatic control is based on the reading of the magnetic flux detector. In the Kusters comparator the simultaneous balance of

Fig. 6.10 Kusters bridge comparator in the Central Office of Measures (GUM), and one of the standard 1 Ω resistors

voltage and ampere-turns is ensured by the system of automatic control of the current I_s. Figure 6.10 shows the 9975 direct-current Kusters comparator manufactured by Guideline, the model used at the GUM in Warsaw. The comparator allows to compare resistors with a ratio of 1:10 and measure resistance in the range from 0.1 Ω to 1 MΩ. In comparisons of 1 Ω resistors the limit relative uncertainty of the Kusters bridge comparator does not exceed $\pm 2 \times 10^{-7}$.

The correct resistance value of the Polish national standard of the unit of electrical resistance is the mean resistance of the six resistors which constitute the national standard. The mean resistance is determined in comparisons between the six resistors and corrected in the calibration of selected four resistors carried out periodically at the BIPM. These four resistors constitute a transfer reference between the Polish national standard and the primary standard, based on the quantum Hall effect, of the unit of electrical resistance at the BIPM.

The resistance standard system at the GUM is kept in a laboratory room with temperature stabilization. The temperature is stabilized at the oil-bath value, i.e. at (23.0 ± 0.3) °C. The room is shielded from radio interference by a screen providing attenuation better than 80 dB. Such laboratory conditions allow a significant reduction of type B uncertainty components related to unstable temperature and electromagnetic interference during the measurements.

A quantum standard of the unit of resistance based on the quantum Hall effect was installed in the Laboratory for Voltage and Resistance Standards at the Central Office of Measures in Warsaw in 2005. Manufactured by Cryogenics, the standard system can realize the unit of electrical resistance with a relative uncertainly of the order of 10^{-9} (estimated value). The parameters of the QHE sample used by the quantum standard are: the charge carrier mobility $\mu = 40\ \mathrm{T}^{-1}$, the surface carrier concentration $N_s = 3.9 \times 10^{15}\ \mathrm{m}^{-2}$, and the operating temperature $T = 0.3$ K.

The QHE sample has two resistance plateaus: $R = R_K/i \approx 6453\ \Omega$ for $i = 4$ and $B = 3.74$ T, and $R = R_K/i \approx 12\,906\ \Omega$ for $i = 2$ and $B = 7.88$ T [5]. A SQUID sensor is used as a zero flux detector in the cryogenic comparator in the system. The potential relative uncertainty of the quantum standard of the unit of resistance at the GUM is of the order of 10^{-9}. The system is on standby, but has not yet participated in international comparisons, and its expected good uncertainty remains to be confirmed experimentally.

6.4 Quantum Standard Measurement Systems

Two different measurement systems, a potentiometric bridge or a cryogenic current comparator (CCC), are used for comparing the resistance values $R_H(i)$ of QHE samples with the resistance of room-temperature standard resistors. An advantage of the potentiometric method is its simplicity and the easy construction of the system, but the CCC bridge has an edge in many aspects, including a low type A uncertainty (the random uncertainty component) and a reduced type B uncertainty (the systematic component) in comparisons with low-resistance reference resistors.

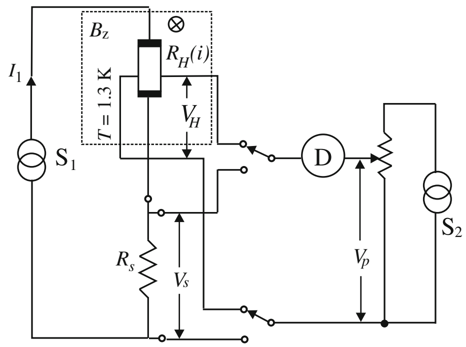


Fig. 6.11 Measurement system with the potentiometric bridge

Figure 6.11 shows a simplified diagram of the measurement system with the potentiometer bridge. The current I_1 from an electronic current source S_1 flows through the source-drain terminals of the QHE sample and through the reference resistor R_S kept at controlled temperature. The nominal resistance of the resistor R_S is equal to $R_H(i)$. The Hall voltage V_H is compared with a voltage V_P of similar value generated by an independent current source S_2. The current from S_2 flows through the potentiometer resistor with a much lower resistance value.

The voltage difference $\Delta V_{HP} = V_H - V_P$ is detected by a zero flux detector D and amplified by an instrumentation amplifier (not shown in Fig. 6.9). The voltage V_P and the amplifier are switched by mechanical keys in order to amplify and measure the voltage difference ΔV_{SP}:

$$\Delta V_{SP} = V_S - V_P, \tag{6.15}$$

where $V_S = I_1 R_S$ is the voltage across the reference resistor;

$$V_{HP} = V_H - V_P, \quad V_{SP} = V_S - V_P. \tag{6.16}$$

A measurement technique in which the direction of the current I_1 is switched must be used to eliminate the thermoelectric power (TEP), its drift in time and the time drift of the current I_1. For one direction of I_1 the measured voltage is ΔV_{HP} (+), and for the opposite direction ΔV_{HP} (−). The correct value of V_{HP} is the mean value of these two:

$$V_{HP} = 0.5[\Delta V_{HP}(+) + \Delta V_{HP}(-)].$$

The potentiometer is calibrated by short-circuiting the resistor R_S with another resistor of a known and sufficiently large value of resistance. As a result of the short-circuiting R_S is replaced by a resistance R_{S1} in the circuit. Thus, we first obtain the voltage difference V_{SP} across the resistor R_S, and then the voltage difference V_{SP1} across the resistor R_{S1}. The operation of switching (short-circuiting) is controlled by a computer. The voltage ΔV_{HP} is measured with a digital voltmeter connected to the output of the amplifier.

For R_S:

$$\Delta V = V_{HP} - V_{SP} = (R_H - R_S)I_1. \tag{6.17}$$

For R_{S1}:

$$\Delta V_1 = V_{HP} - V_{SP1} = (R_H - R_{S1})I_1; \tag{6.18}$$

$$\frac{V_{HP} - V_{SP}}{V_{HP} - V_{SP1}} = \frac{\Delta V}{\Delta V_1} = \frac{R_H - R_S}{R_H - R_{S1}}. \tag{6.19}$$

The (6.19) leads to the formula for the resistance of the QHE sample:

$$R_H = \frac{R_S\Delta V - R_{S1}\Delta V_1}{V_{SP} - V_{SP1}} = \frac{(R_S - R_{S1})V_{HP} - R_S V_{SP1} + R_{S1}V_{SP}}{V_{SP} - V_{SP1}}. \tag{6.20}$$

Since the voltage values V_{SP} and V_{SP1} can vary widely from the order of millivolts to volts, the amplifier used in the measurement system must have a good linearity. Measurements of the current I_1 (e.g. $I_1 \approx 50$ μA) last a few hours, the time necessary to reduce by averaging the type A uncertainty (1 σ) to the order of 10^{-9}.

The other measurement system, widely used for comparisons of QHE resistance standards, is the CCC, schematically depicted in Fig. 6.12.

A current I_1 from a direct-current source S_1 flows through the QHE sample connected in series to the primary winding of the CCC comparator. The primary winding has N_1 turns. A controlled source S_2 generates a current $I_2 \approx (N_1/N_2)I_1$, which flows through the reference resistor R_S and through the secondary winding of the CCC. The number of turns in the secondary winding is N_2. The intensity of the current I_2 depends on the voltage V across the resistor R: $V = RI_1$. The resultant magnetic flux generated by the two windings, with N_1 and N_2 turns, should be close to zero. The zero-flux detector used in this system is an RF-SQUID magnetic flux sensor (for a detailed discussion of SQUID sensors, see Chap. 5).

The RF-SQUID is strongly coupled to the CCC by a flux transformer. The relative accuracy of the turns ratio N_1/N_2 can be better than 10^{-10} [4]. Any deviation of the current ratio I_1/I_2 from the turns ratio N_1/N_2 is corrected by feedback via the SQUID and the current source S_2. The $R_H(i)/R$ ratio can be determined on the basis of the reading of the zero-flux detector D. The voltage output of the detector D may be also used for generating a balance current I_b in the secondary winding N_2 of the CCC comparator. The current I_b is used for balancing the RF-SQUID. Consequently, the gain stability and the linearity of the RF-SQUID are of no importance. From the combination of the formulas for balanced zero-flux detector, $R_H(i) I_1 - R_s I_2 = 0$, and for balanced comparator:

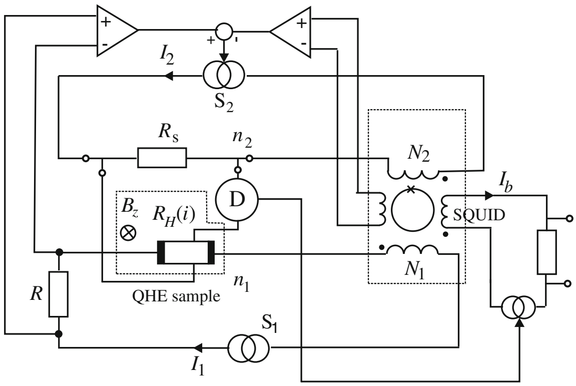


Fig. 6.12 Cryogenic current comparator (CCC)

$$(N_1/N_2)I_1 + I_b - I_2 = 0 \tag{6.21}$$

it follows that, when the current I_b is low:

$$\frac{R_H(i)}{R_s} = \frac{N_1}{N_2}\left[1 + \frac{I_b}{I_2}\right]. \tag{6.22}$$

The current I_b is measured with a digital voltmeter and a resistor. A 100 MΩ shunt resistor R_s is used for calibration. Full accuracy of the CCC bridge can only be achieved on the condition that the undesirable leakage currents are carefully eliminated. It is particularly important that the current flowing through $R_H\,(i)$ should flow also through the winding N_1, and the current flowing through R_s must flow through the winding N_2. Also of importance is a low intensity of the bias currents in the separating amplifier (which must have a high impedance). The CCC bridge can be connected to a computer to provide control of the switching of direction of the currents I_1 and I_2, and to register the readings of the digital voltmeter. Optical fiber is a good solution for connection between the CCC and a computer. The random uncertainty can be reduced to 2×10^{-9} in a measurement time of ca. 5 min if a power of 1 mW is dissipated in the resistor R_S.

6.5 Quantum Standard of Electrical Resistance in the SI System

In the SI system each unit is related to the base units. The Thompson-Lampard cross capacitor, shown in Fig. 6.13, is an impedance standard providing a value, calculated from its size, in SI units. It is the only standard accurate enough for comparison with the von Klitzing constant R_K (or with another impedance). The electrostatic Thompson-Lampard theorem states that the cross-capacitances C_1 and

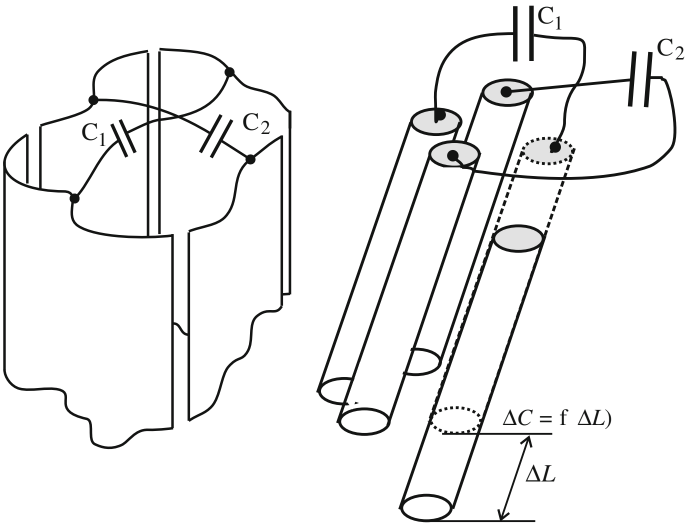


Fig. 6.13 Calculable Thompson-Lampard capacitor with cross capacitances C_1 and C_2

C_2 per unit length L of an infinitely long system in vacuum are given by the formula:

$$\exp\left(-\frac{\pi C_1}{\varepsilon_0}\right) + \exp\left(-\frac{\pi C_2}{\varepsilon_0}\right) = 1. \tag{6.23}$$

In many calculable capacitors the electrodes are made in the form of cylinders as symmetric as possible, so that:

$$\frac{C_1 - C_2}{C_1} < 10^{-4}. \tag{6.24}$$

Under such conditions, with an accuracy of 1×10^{-9}, the mean value of the capacitances C_1 and C_2 is changed in function of the displacement ΔL:

$$\frac{\Delta C}{\Delta L} = \left(\frac{\varepsilon_0 \ln 2}{\pi}\right)\frac{\mathrm{F}}{\mathrm{m}} = 1.953549043\,\frac{\mathrm{pF}}{\mathrm{m}} \tag{6.25}$$

for the defined value of vacuum permeability, $\mu_0 = 4\pi \times 10^{-7}$ H/m, and the speed of light c = 299 792 458 m/s.

The capacitances C_1 and C_2 are changed by shifting the position l of a concentric tubular screen, which effectively eliminates the effect of the finite size of the capacitor. The shifting end of the electrode system is realized in a similar way by an inserted solid cylinder. The "exact" expected change in capacitance C is 0.4 pF when the central cylinder is shifted by 0.2 m (which corresponds to 2 pF/m); the shift is

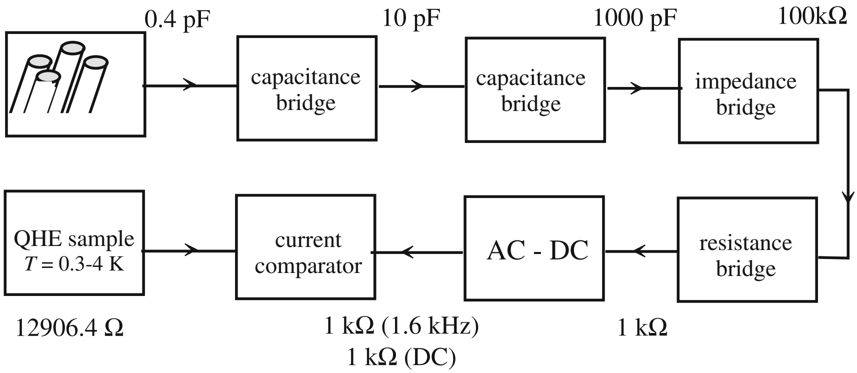


Fig. 6.14 Comparison of a capacitance standard with a quantum standard of electrical resistance

measured by a laser interferometer. However, the achievement of the expected accuracy of 1×10^{-8} requires many improvements in the construction of the capacitor. A chain of AC bridges is used for measuring the R_K/C ratio in the NPL standard system. The AC bridges are followed by two steps of direct-current CCC comparators, as shown in Fig. 6.14. Experiments have shown that the R_H value is independent of the experimental conditions and on the parameters of the device within the experimental uncertainty, which is 3.5×10^{-10} at best. By measuring R_K in their laboratories metrological institutions are able to determine the drift in time of resistance standards with a resolution of the order of 10^{-8}–10^{-9} per year.

As in the case of the Josephson effect, those national laboratories which have a quantum resistance standard no longer need to rely on reference resistors for verification of the resistance standard. Thus, the quantum standard of resistance allows to avoid the negative effect of transportation, which reduces the accuracy of comparison with reference resistors.

The BIPM has established a portable QHE standard, which is transported to national laboratories. Comparative measurements of a 100 Ω reference resistor with BIPM and OFMET quantum standards yielded equal results with an uncertainty of $(9 \pm 17) \times 10^{-10}$ [1].

References

1. F. Delahaye, T.J. Witt, B. Jeckelmann, B. Jeanneret, Comparison of quantum Hall effect resistance standards of the OFMET and BIPM. Metrologia **32**, 385–388 (1996)
2. F. Delahaye, T.J. Witt, R.E. Elmquist, R.F. Dziuba, Comparison of quantum Hall effect resistance standards of the NIST and the BIPM. Metrologia **37**, 173–176 (2000)
3. D.C. Elias et al., Control of grapheme`s properties by reversible hydrogenation. Science **323**, 610–613 (2009)
4. A. Hartland, The quantum Hall effect and resistance standards. Metrologia **29**, 175–190 (1992)

5. H. Ibach, H. Lüth, *Solid-State Physics. An Introduction to Principles of Materials Science* (Springer, Heidelberg, 1995)
6. R.B. Laughlin, Ułamkowe kwantowanie, Wykład Noblowski, Postępy Fizyki. **51**, 68–84 (2000)
7. K.S. Novoselov et al., Electric field effect in atomically thin carbon films. Science **306**, 666–669 (2004)
8. K.S. Novoselov, Z. Jiang, Y. Zhang, S.V. Morozov, H.L. Stormer, U. Zeitler, J.C. Maan, G.S. Boebinger, P. Kim, A.K. Geim, Room-temperature quantum Hall effect in graphene, Science **315**, 1379 (2007)
9. T. Oe et al., Fabrication of the 10 kΩ QHR Array device, Elektronika, No 6/2011, 47–49
10. H.L. Störmer, Ułamkowy kwantowy efekt Halla. Wykład Noblowski, Postępy Fizyki **51**, 113–133 (2000)
11. K. von Klitzing, G. Dorda, M. Pepper, New method for high-accuracy determination of fine-structure contact based on quantized Hall resistance. Phys. Rev. Lett. **45**, 494–497 (1980)
12. H. Wang, D. Nezich, J. Kong, T. Palacios, Graphene frequency multipliers. IEEE Electron Device Lett. **30**, 547–549 (2009)

Chapter 7
Quantization of Electrical and Thermal Conductance in Nanostructures

Abstract This chapter opens with a presentation of the classical Drude theory of electrical conduction and a theory proposed by Landauer. Based on the assumption that electrical conduction can be modeled as transfer of electrons between two electron reservoirs, the Landauer theory proves to describe particularly well the electrical resistance in nanoscale conductors, i.e., in nanostructures. Surprisingly, this theory implies that the conductance (and resistance) of a nanostructure is independent of its material and temperature, and only depends on the dimensions of the sample, changing in a stepwise manner with a step of $h/2e^2$ representing the conductance quantum. The quantization of electrical and thermal conductance in nanostructures has been verified experimentally. Conductance quantization in nanostructures is used in the analysis of large-scale integration circuits, as required by the currently used 14 nm technology and future technologies.

7.1 Theories of Electrical Conduction

Quantization of electrical conductance $G(G = 1/R)$ in a metal or semiconductor sample of nanometer (atomic) size is an effect of major importance for the understanding of the process of electrical conduction and for applications of the theory of conduction phenomena. Conductance quantization depends on the dimensions of the nanosized sample. In nanostructures the effect takes place at room temperature and does not necessitate the application of a strong magnetic field. Thus, the physical conditions necessary for conductance quantization in nanostructures are much easier to create than those required for the quantum Hall effect. In fact, conditions such as room temperature, atmospheric pressure and lack of strong magnetic field are available naturally and do not require additional efforts.

The first theory of electrical and thermal conduction in metals was proposed by Paul Drude in 1900. It was based on the assumption that the electrons in the conductor form an ideal electron gas. According to the Drude theory all the free electrons in the metal participate in the current. In a static electric field E_p applied to the sample the current density j_p is described by the formula:

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_7

$$j_p = en\mu E_p = e^2 n\tau E_p/m, \tag{7.1}$$

where E_p is the applied electric field, n the concentration of charge carriers, μ the mobility of charge carriers, τ the relaxation time, and m the electron mass.

In the Drude model the conductivity σ of the material is given by:

$$\sigma = j_p/E_p = e^2 n\tau. \tag{7.2}$$

Using a statistical approach to the motion of electrons, the Drude theory describes well the electrical conduction in macroscopic samples, in which the number of atoms and free electrons is large enough to allow a statistical description of their motion. From the model of band structure of metals (and conductors in general) it can be concluded that only those electrons which lie near the Fermi level are involved in the flow of current. In contrast, electrons deep below the Fermi level remain unaffected by the electric field [8]. The Drude theory does not describe the phenomenon of superconductivity.

The model of electrical conduction proposed by Drude is contrary to the Pauli exclusion principle, formulated thirty years later in the framework of quantum mechanics. The Pauli exclusion principle prohibits electrons which occupy energy states much below the Fermi level E_F from shifting to a slightly higher energy level under the influence of the electric field, since all the energy states nearby are already occupied. In 1928 the theory of electrical conduction was provided with a quantum mechanical formalism by Felix Bloch, then working in Leipzig in Heisenberg's team.

Electrical conduction in nanometer-sized samples is described much better by a model developed by Rolf Landauer at IBM (New Jersey, USA). His theory of electrical conduction, applicable also to nanostructures and predicting conductance quantization, was first presented by Landauer in 1957 [10]. Thirty years later Landauer published its advanced version [11]. The 1980s saw the publication of the first results of conductance quantization studies in nanostructures [5, 18]. Developed further by Markus Büttiker, the theory is now commonly used and known as the Landauer-Büttiker formalism.

Let us consider a constriction of length L and width W between two wide metal terminals (Fig. 7.1a), referred to as electron reservoirs in the Landauer theory [11]. Let the mean free path in the conductor be denoted as λ, and the Fermi level as E_F. As a result of the application of a potential difference V to the ends electrons flow through the constriction. The transport of electrons is ballistic, i.e. without collision, if the length of the constriction is shorter than the mean free path of the electrons, $L < \lambda$. Under this condition the scattering of electrons can be neglected. Electron scattering is caused by the presence of impurities and donors as well as irregularities of the surface.

The following five assumptions of the Landauer theory allow to present the above description in a synthetic manner:

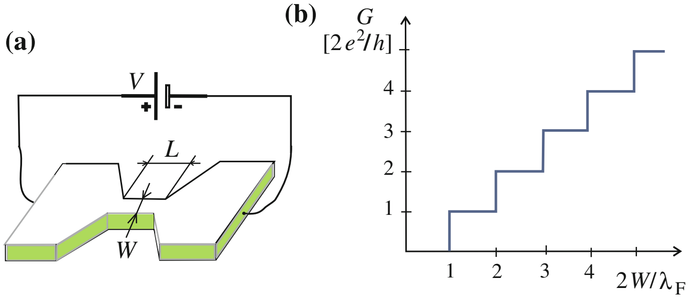


Fig. 7.1 Conductance quantization in a nanowire (conductor with a length $L < \lambda$ and a width W comparable with the Fermi wavelength λ_F): **a** schematic depiction of the nanowire (the third dimension is not considered); **b** quantization of conductance G versus the nanowire width W

- Electrons are transferred between two electron reservoirs with different chemical potentials;
- The transport of charge carriers (electrical conduction) results from a charge carrier density gradient between the reservoirs of electrons;
- Only elastic scattering of electrons, caused by their reflection from the potential barriers and the edges of the conductor, takes place in the perfect conductor;
- The observed conduction is related to the transmission probability in the conduction channels formed in the scattering;
- Power is dissipated during the transport of electrons in the reservoirs, but not in the interior of the perfect guide.

Another characteristic parameter of the system is the Fermi wavelength $\lambda_F = 2\pi/k_F$, where k_F is the Fermi wavevector. Metals such as copper or gold have a Fermi wavelength $\lambda_F \approx 0.5$ nm, much shorter than the electron mean free path $\Lambda(\Lambda_{Au} = 14$ nm). If the dimensions of the system are smaller than the electron mean free path, the scattering due to the presence of impurities can be neglected, and, consequently, the electron transport can be regarded as ballistic. A metal wire with an external diameter W comparable with the Fermi wavelength λ_F and a length L shorter than Λ can be treated as a one-dimensional (1D) waveguide, with electrons regarded as waves, and quantum effects can be expected to occur. An object of a width comparable to the Fermi wavelength λ_F can be treated as a quasi-one-dimensional waveguide. Also in this case electrons are regarded as waves, and can be expected to manifest their quantum nature.

Let us consider a perfect conductor with a diameter W and a length L (Fig. 7.1) connecting two wide contacts (electron reservoirs), and the electrical conduction between them. On the assumption that the electron reservoirs are infinitely large, the electrons are in thermodynamic equilibrium described by Fermi-Dirac statistics. When electrons enter the 1-D conductor, nonequilibrium states with negative and positive velocities appear. If a resultant current flows through the conductor, states with positive velocities have higher energy [1]. According to the Büttiker [8] model the Hamiltonian of the perfect conductor can be expressed as follows:

$$H = \frac{1}{2m^*}\left(\hbar^2 k_x^2 + \hbar^2 k_y^2\right) + V(x), \tag{7.3}$$

where y is the coordinate along the wire, x is the coordinate in the transverse direction, m^* is the effective mass, $V(x)$ denotes the potential well of the width W, k_y is the wave vector component along the y axis, and k_x the wave vector component in the x direction. Because of the narrowness of the potential well $V(x)$ the energy in the transverse propagation is quantized:

$$E_{Tj} = \frac{\hbar^2 k_x^2}{2m^*} = \frac{\hbar^2}{2m^*}\left(\frac{j\pi}{W}\right)^2. \tag{7.4}$$

The formula (7.4) holds if the potential energy tends to infinity at the boundary of the quantum well. For the Fermi level $E_F = E_j$ there are $N \sim 2W/\lambda_F$ states E_{Tj} below the Fermi level. Let us assume that the thermal energy k_BT is much lower than the energy gap between levels, and that the wide contacts have chemical potentials μ_1 and μ_2, with $\mu_1 > \mu_2$. Then, the current of electrons in the j-th state is:

$$I_j = ev_j\left(\frac{dn}{dE}\right)_j \Delta\mu, \tag{7.5}$$

where v_j is the velocity along the y axis and $(dn/dE)_j$ is the density of states at the Fermi level for the j-th state. In a 1-D conductor the density of states is:

$$\frac{dn}{dk} = \frac{1}{2\pi} \quad \text{and} \quad \left(\frac{dn}{dE}\right)_j = \left(\frac{dn}{dk}\frac{dk}{dE}\right)_j = \frac{2}{hv_j}. \tag{7.6}$$

The factor 2 in the (7.6) results from spin degeneracy. Hence, the current related to the j-th state, $I_j = \frac{2e^2}{h}V$ (where the voltage difference $V = \Delta\mu/e$), does not depend on j. The total current is $I = \sum_{j=1}^{N} I_j$. Consequently, the conductivity can be expressed as:

$$G = \frac{2e^2}{h}N, \tag{7.7}$$

where N depends on the width of the wire (Fig. 7.1).

However, defects, impurities and irregularities in the shape of the conductor can cause scattering, which is taken into account in the Landauer formula:

$$G = \frac{2e^2}{h}\sum_{i,j=1}^{N} t_{ij}, \tag{7.8}$$

where t_{ij} denotes the probability of transition from the j-th to i-th state. In the case of lack of scattering $t_{ij} = \delta_{ij}$ and the (7.8) reduces to (7.7).

Measurements of electrical resistance (or conductance) of samples of a size close to Λ (mesoscopic range) show that the Landauer theory describes the electrical conduction in such samples better than the Drude model. The conductance G_0 for a single channel, with $T_j = 1$, is:

$$G_0 = \frac{2e^2}{h} \cong 77.2 \times 10^{-6}\text{A/V} = (12.9\ \text{k}\Omega)^{-1}.$$

For steps of $2e^2/h$ to be observed in the conductance characteristic the probability T_j should be equal to 1, which means the lack of electron scattering by impurities in the crystal or by surface irregularities, since this would lead to a reduction in the conductance steps. The total number of modes of the electron wave, i.e., the number of conductance channels, is equal to $2W/\lambda_F$ in a one-dimensional constriction, and $(2W/\lambda_F)^2$, approximately, in a two-dimensional constriction (examples of which are dynamically formed nanowires). *Assuming a transmission* $T_j = 1$, the conductance of the constriction (sample) is a function of its width W. In real systems, the transmission coefficient is usually close to 1 for the first transmission channel. For other channels the values of T_j are lower, and can even fall below 0.5. The conductance of a one-dimensional constriction is given by the (7.9) below, and that of a two-dimensional constriction by the (7.10):

$$G = \frac{2W}{\lambda_F}\frac{e^2}{h} = NG_0 \tag{7.9}$$

$$G \approx \left(\frac{2W}{\lambda_F}\right)^2 \frac{e^2}{h} = NG_0, \tag{7.10}$$

where W is the width of the sample, λ_F the Fermi wavelength, N the number of channels for electron transport, equal to the number of modes of the electron wave, and G_0 is the conductance quantum, $G_0 = e^2/h = 77.2 \times 10^{-6}$ A/V.

When the width of the constriction changes, each time the energy exceeds the allowed Fermi level (which is equivalent to a new mode entering or exiting the waveguide) the conductance increases by one or two quanta, depending on the degree of degeneracy. This results in a stepwise dependence, shown in Fig. 7.1b, of the conductance on the width W of the sample.

It is noteworthy that the Landauer model is not the only theory of electrical conduction in nanostructures. An identical formula for conductance has been obtained by Kamenec and Kohn without the first two of the Landauer assumptions [9], and by Das and Green without any of them, as reported in their paper *Landauer Formula without Landauer's Assumpions* [3].

7.2 Macroscopic and Nanoscale Structures

Metallic nanocontacts, semiconductor nanocontacts (including heterostructures), nanowires, nanotubes (especially carbon nanotubes), nanorods, quantum dots and other nanostructures are the subject of ongoing research and studies for potential applications.

According to the Landauer theory the conductance G of a nanostructure (and its resistance $R = 1/G$) *does not depend* on either the type of conductive material of the sample, or the temperature, or any material constant. By the formula (7.8) the conductance only depends on the dimensions of the 1-DEG sample. This is an entirely new and rather unexpected feature of nanostructures, not observed in samples of macroscopic size. However, careful consideration of the conductance of nanostructures made of different metals (or, generally, different conductive materials) leads to the observation that the metals differ in the values of the Fermi wavelength λ_F and the mean free path λ (e.g. for Au λ = 14 nm, and for Cu λ = 30 nm at T = 295 K). In addition, the mean free path increases steeply with decreasing temperature. Thus, nanostructures of the same dimensions, but made of different materials or in different temperature conditions may differ in conductance.

Let us consider the resistance of a cubic sample of a conductor with a side length a. Resistance is analyzed in four size ranges:

- Macroscopic samples, with the size a larger than the coherence length of the electron wave in the conductive material, $a > L_\varphi$ (where L_φ is the coherence length of the electron wave); these are considered to contain three-dimensional electron gas (3-DEG) (see Fig. 6.2 in Chap. 6);
- Mesoscopic samples, with the size a smaller than the coherence length of the electron wave, but larger than the mean free path of electrons in the conductive material, $\lambda \geq a > L_\varphi$ (where λ is the mean free path);
- Samples with ballistic transport of electrons, with the size a smaller than the mean free path of electrons in the conductive material, but larger than the Fermi wavelength, $N\lambda_F \geq a > \lambda$ (where λ_F is the Fermi wavelength, λ denotes the mean free path of electrons, and N is a natural number);
- Nanometer-sized samples, in which the smallest dimension a is comparable to the Fermi wavelength λ_F and can be expressed as its multiple, $a = N\lambda_F$ (where $N = 1, 2, 3, \ldots$), and the length is smaller than the mean free path λ of electrons. Nanowires are an example of such samples.

The resistance R of a cubic sample calculated in accordance with the classical theory of electrical conductance, the Drude theory, is given by the formula:

$$R = \frac{\rho\, a}{a^2} = \frac{\rho}{a}, \tag{7.11}$$

where R is the resistance of the cubic sample, ρ the resistivity of the material of the sample, and a the side length of the cube.

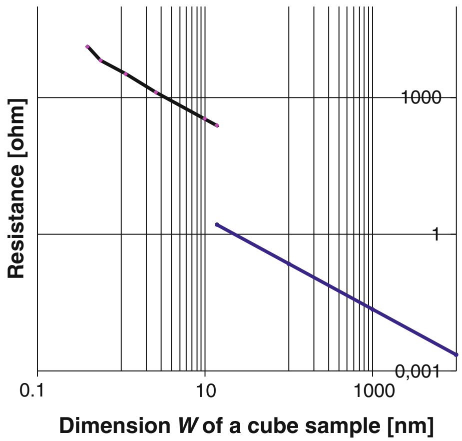


Fig. 7.2 Resistance of a cubic gold sample as a function of its width W, calculated by the Drude theory (*right line* in the plot), and in the Landauer model (*left line*) for $T = 295$ K

In cubic samples the resistance R is first calculated by the Drude theory and then by the Landauer model of electrical conduction, with the conversion $R = 1/G$. Calculations by these two theories can be expected to lead to different resistance (or conductance) values. For example, let us consider a gold cube with a side length a, and calculate its resistance by the Drude theory, (7.11), and in the Landauer model, (7.10). The material parameters of gold are: the atom diameter of 0.408 nm, the lattice constant (the spacing between atoms in the crystal lattice) of 0.288 nm, the Fermi wavelength $\lambda_F = 0.52$ nm, the electron mean free path $\lambda = 14$ nm (at a temperature $T = 295$ K), and the resistivity $\rho = 22.4 \times 10^{-9}\ \Omega$ m (at $T = 295$ K).

The plot in Fig. 7.2 shows the resistance of a cubic gold sample with a side length a versus the width W of the sample, calculated by the Drude theory for $W = a = 14$ nm to 10 μm, and in the Landauer model for $W = a = 0.4$ nm (the diameter of an atom) to 14 nm (the electron mean free path). The plot shows (Fig. 7.2) a substantial difference in the resistance data calculated by the two theories for the width $a = \lambda = 14$ nm, equal to the electron mean free path in gold at a temperature of 295 K: $R_D = 1.6\ \Omega$ (Drude's theory) and $R_L = 240\ \Omega$ (Landauer's theory). Without the simplifying assumptions, including the assumption that the transmission coefficient $T_j = 1$, the discrepancy between the calculated resistance values would be even larger. The actual transmission coefficient values in nanostructures are usually less than 1; thus, the conductance of the sample is lower than calculated, and the actual resistance larger than R_L.

7.3 Studies of Conductance Quantization in Nanostructures

7.3.1 Formation of Nanostructures

The effect of quantization of electrical conductance was first observed and measured by Gimzewski and Möller at IBM in Zurich in 1987 with a scanning tunneling microscope [5] (STM, discussed in detail in Chap. 11). When the iridium tip of the STM was about to detach from the smooth silver surface of the studied sample, with which it was still in contact, the measured electrical conductance of the contact between the sample and the tip proved to be a step function of its size, as in the plot shown in Fig. 7.1b. Similar steps in the conductance characteristic were observed by van Wees in a two-dimensional electron gas with modulated voltage applied to the sample in order to obtain a narrow path within the 2–DEG [19]. After this experimental confirmation of the Landauer theory of electrical conduction also other research teams undertook studies of electron transport and conductance quantization in nanostructures.

Nanostructures used in conductance studies can be formed by a number of methods. The formed nanostructures tend to be unstable, their lifetime ranging from microseconds to one hour at the most. The following techniques are used for the dynamic formation of nanostructures:

- A method consisting in creating and breaking a contact between an STM tip and a flat surface of a conductor or semiconductor sample. Involving the use of an STM, this is a rather expensive technique, and frequently leads to the destruction of the tip [5].
- A method consisting in creating and breaking a contact between macroscopic samples (e.g. wires with a diameter ranging from 0.1 to 0.5 mm) of conductive material. The relative position of the macroscopic samples is either controlled by a precision actuator or subject to change in spontaneous oscillations. In the first stage the samples are brought in contact, which is broken in the last stage. As a result, a nanocontact, or nanoscale junction, is formed due to surface roughness of the macroscopic samples. The formation of nanowires by this technique is shown in Fig. 7.3 [11, 14].

A voltage-controlled piezoelectric actuator with a conversion coefficient of ca. 1 μm/1000 V, or a cheaper magnetostrictive transducer driven by DC current are usually used for positioning the samples. Figure 7.4a shows the surface of a copper

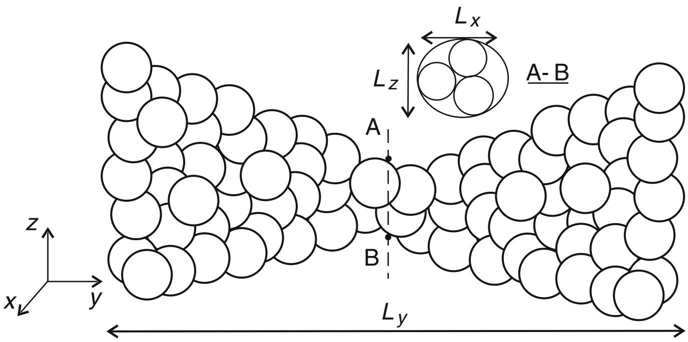


Fig. 7.3 Formation of a nanowire between two macrowires in contact. The *A–B* cross section of the nanowire includes as few as three atoms [13]

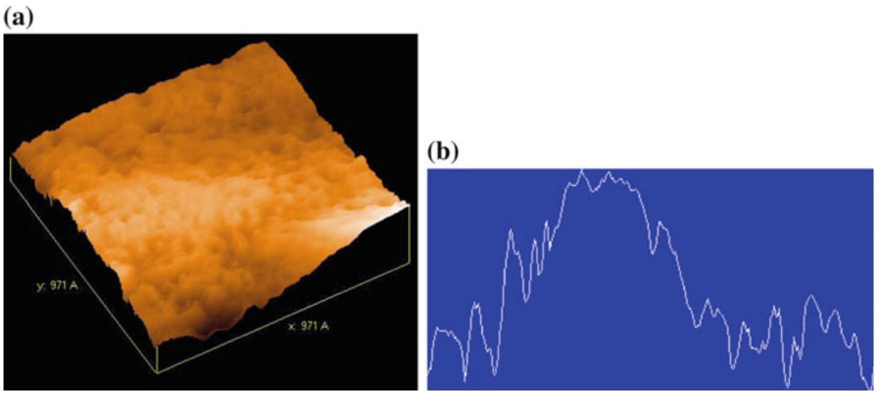


Fig. 7.4 STM image showing the surface of a copper wire with a diameter of 0.5 mm: **a** surface of a 97 nm × 97 nm area, h = 13.5 nm; **b** cross section of the surface shown in (**a**) (measurements by S. Szuba)

wire with a diameter of 0.5 mm (determined in STM measurements). Presented in Fig. 7.4b, the cross section of the scanned surface of the copper wire shows surface irregularities of the macroscopic sample. The extent of surface roughness is the reason for using the above-described dynamic method for the formation of nanowires. Both actuators—the piezoelectric actuator operating with the reverse piezoelectric effect, and the magnetostrictive actuator—have been used in conductance quantization studies at Poznan University of Technology (PUT) in Poland. In the magnetostrictive actuator the length of a magnetic core changes with the induced magnetic field. The voltage to elongation conversion ratio of the tube-shaped piezoelectric actuator used in the studies carried out in PUT was $\Delta l/\Delta V = 1$ μm/750 V. A magnetostrictive actuator with a current to elongation conversion ratio $\Delta l/\Delta I = 1$ μm/150 mA was proposed and manufactured by Wawrzyniak [20].

- A simple and cheap method for the formation of nanowires was proposed by Garcia and Costa-Krämer from Madrid [2]. They used two macroscopic wires (with a diameter of 0.5 mm and a length of 10 cm), which were put in spontaneous vibrations and in this way brought in both mechanical and electrical contact, as shown in Fig. 7.5.

Garcia and Costa-Krämer predicted that an electrical nanocontact would form between the wires as a result of creating and breaking a contact due to surface irregularities. This original idea for the dynamic formation of nanowires between macroscopic wires connecting and disconnecting in vibration has been verified experimentally, and after its publication used by many research teams, as reported e.g. in [7, 12, 14].

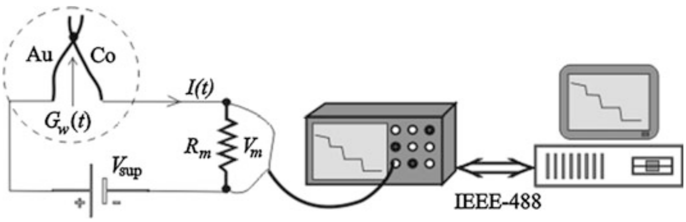


Fig. 7.5 System for forming nanowires between two vibrating macrowires

7.3.2 Measurements of Dynamically Formed Nanowires

The electrical conductance of dynamically formed nanowires is measured with the experimental setup shown in Fig. 7.5.

The system includes a measurement circuit and a digital oscilloscope connected to a PC computer by the IEEE-488 interface. The serial measurement circuit comprises two macrowires between which a contact is created spontaneously, a measuring resistor R_m (typically with a resistance of ca. 1 kΩ) and a source of DC voltage V_{sup}. The measurement signal in the system is the current I in the circuit, proportional to the conductance $G_w(t)$ (for constant power supply).

No current flows ($I = 0$) in the state of disconnected wires, while in a stable state of contacting wires a maximum current ($I = V_{sup}/R_m$) flows in the circuit. The sequence of transient states that appear in the processes of creating and breaking the contact results in a time-dependent current $I(t)$. The voltage V_m across the measuring resistor is proportional to the current in the circuit. The voltage V_m is measured and recorded with a digital oscilloscope. The signals in the measurement circuit are:

$$I(t) = \frac{V_{sup}}{R_m + 1/G_w}; \quad V_m = R_m I(t) = \frac{V_{sup} R_m}{R_m + 1/G_w}. \tag{7.12}$$

The digital voltage signal $V_m(t)$ is sent from the digital oscilloscope to the PC, which additionally performs statistical data processing. The conductance G_w of the nanowire is calculated by the formula (7.13):

$$G_w(t) = \frac{V_m}{R_m(V_{sup} - V_m)}. \tag{7.13}$$

Figure 7.6 presents two conductance traces showing conductance quantization in metal nanowires studied with the measurement system described above.

The above-presented method for the formation of nanowires by repeated connection and disconnection of macroscopic wires (or other pieces of conductive material) has been improved by the addition of precise control of the movement of the macroscopic wires.

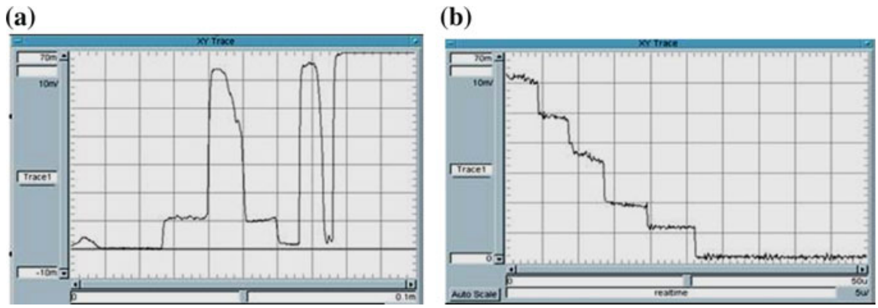


Fig. 7.6 Conductance traces showing conductance quantization in metal nanowires formed by creating and breaking a contact at 295 K: **a** conductance trace with a $1G_0$ step corresponding to the creation and breaking of the contact; **b** stepwise conductance trace indicating conductance quantization. The charts have 10 μs/div on the abscissa (time) axis and $1G_0$/div on the ordinate (voltage) axis, where $G_0 \approx (12.9\ \text{k}\Omega)^{-1}$ is the conductance quantum

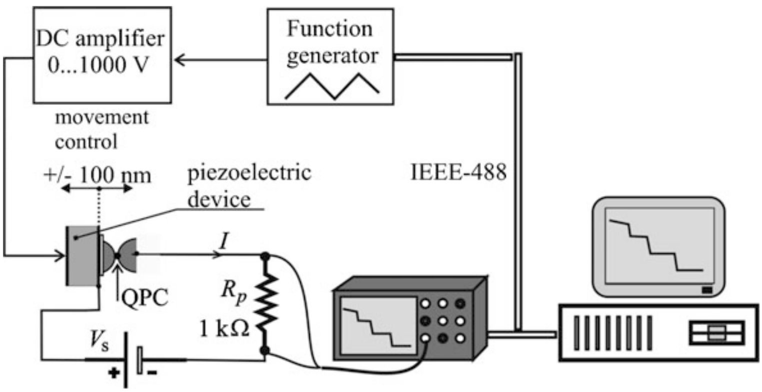


Fig. 7.7 System with a piezoelectric actuator for the formation of nanowires between two vibrating macrowires

A piezoelectric device has been used for controlling the backward and forward movement of the macroscopic wires between which nanowires, or quantum point contacts (QPCs), are formed (Fig. 7.7).

7.4 Quantization of Thermal Conductance in Nanostructures

Electron transport in nanostructures is described by electrical conductance G_E and thermal conductance G_T. There are a number of analogies between these two physical quantities. By analogy with the quantization of electrical conductance, thermal conductance can be expected to be quantized as well. Electron transport in a nanowire has two effects: one is the electric current $I = G_E \Delta V$, and the other the heat flux density $Q_D = G_T \Delta T$, where G_E is the electrical conductance of the sample, ΔV the electric potential difference, G_T the thermal conductance of the sample, and ΔT the temperature difference. The electrical conductance G_E and the thermal conductance G_T are described by the equations:

$$G_E = \sigma A/l, \quad G_T = \lambda A/l, \tag{7.14}$$

where σ denotes the electrical conductivity, λ the thermal conductivity, l the length of the sample (e.g. a nanowire), and A the cross-section area of the sample.

However, thermal conduction is more complex than electrical conduction, since both electrons and phonons can be involved in heat transfer. The quantization of thermal conductance in one-dimensional systems was predicted theoretically by Greiner [6] for ballistic transport of electrons and by Rego [15] for ballistic transport of phonons. Thermal conductance is described in terms similar to those used in the analysis of electrical conductance. Heat conduction channels are considered to form in one-dimensional systems, and each channel to contribute to the total thermal conductance with a thermal conductance quantum G_{T0}.

The quantization of thermal conductance (for the ballistic transport of electrons) was confirmed experimentally by Schwab [16]. The universal thermal conductance quantum G_{T0} depends on the temperature:

$$G_{T0}[\mathrm{W/K}] = (\pi^2 k_B^2/3h)T = 9.5 \times 10^{-13} T. \tag{7.15}$$

At T = 300 K the thermal conductance quantum is $G_{T0} = 2.8 \times 10^{-10}$ W/K. However, this value is determined on the assumption that electron transport in the nanowire is ballistic (without scattering), with the transmission coefficients t_{ij} = 100 %. This means that in real systems (with t_{ij} < 100 %) the actual thermal conductance is lower than the limit value defined by the formula (7.15).

A single metal or semiconductor nanowire should be considered together with its terminals (Fig. 7.8), referred to as reservoirs of electrons. The assumed ballistic character of electron transport in the nanowire means that there is no scattering of electrons and no energy dissipation in electron transport. The energy is partially dissipated in the terminals, though. Because of the energy dissipation the local temperature T_{term} of the terminals is higher than the temperature T_{wire} of the nanowire (Fig. 7.8). The temperature distribution in the terminals of the nanostructure should be taken into account in the analysis of thermal conduction.

In small structures the amount of dissipated energy is quite large. In the first thermal conductance step, $G_E = G_{E0} = 7.75 \times 10^{-5}$ A/V, for a supply voltage V_{sup} = 1.4 V the current in the circuit is I = 100 μA (I = 190 μA in the second step). The power dissipated in the terminals of the nanowire is $P = I^2/G_{E0}$ = 130 μW for the first step and P = 230 μW for the second step. It is worthy of notice that the current density in nanowires is extremely high. Assuming the estimate diameter D = 0.4 nm of a gold nanowire in the first conductance step, for I = 100 μA we obtain a current density as high as $J \cong 8 \times 10^{10}$ A/cm^2.

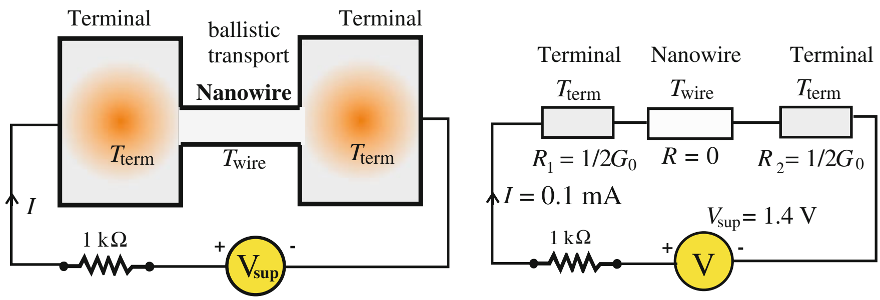


Fig. 7.8 Temperature and resistance distribution in a nanowire in the ballistic regime

7.5 Scientific and Technological Impacts of Conductance Quantization in Nanostructures

Research in electrical conductance quantization in nanostructures is of importance for at least five areas of science and technology, namely:

- Fundamental research in which the properties of matter are considered on the smallest, atomic scale; in particular, research in electron transport in conductors with different atomic structures. The finite values of the electric and thermal conductivities of materials result from the fact that in reality electron transport is non-ideal.
- The technology of large-scale integration (LSI) circuits. The increasingly large scale of integration involves the need to scale down to nanometers integrated circuit components such as transistors, resistors, capacitors or conductive tracks.
- Electronic circuits with relays, especially supersensitive amplifiers with a relay (chopper) in the input circuit. The process of short-circuiting of the relay contacts involves the formation of nanowires in the circuit and the occurrence of short-lived transient states. In high-gain amplifiers this can lead to oscillations in circuits with relay contacts (see Fig. 7.9).
- Conductance studies of nanoscale samples.
- A quantum standard of electrical conductance and resistance operating at room temperature (as opposed to the QHE resistance standard).

According to the forecasts of the development of semiconductor devices, published periodically in the International Technology Roadmap for Semiconductors (ITRS), already in 2023 the size of a transistor gate in a semiconductor chip will be smaller than 10 nm, and will further decrease to 5 nm by 2028 [17]. Table 7.1 shows collected data on the expected performance of selected integrated circuits (MPU means Microprocessor Unit).

The full version of the ITRS 2013 documentation has several hundred pages. Besides involving the need for precision lithographic technologies and system diagnostics, production of integrated circuits containing nanoscale elements

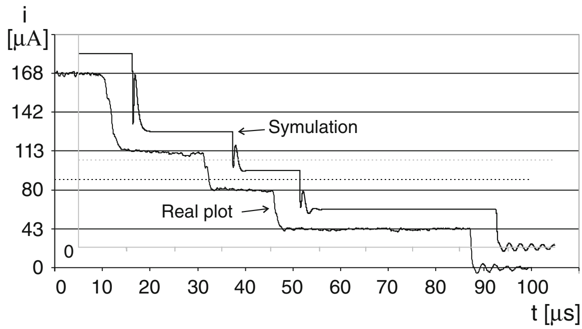


Fig. 7.9 Transient states of the electric current in a circuit with a dynamically formed nanowire

Table 7.1 Forecasts of the 2013 international technology roadmap for semiconductors

Year		2013	2017	2021	2025	2028
On-chip local clock frequency for MPU	GHz	5.50	6.44	7.53	8.8	9.9
Flash generations label (per chip)	Gbit	64	256	512	2000	4000/8000
Supply voltage	V	0.88	0.80	0.74	0.68	0.64
Maximal power (for cooled IC)	W	180	198	198	198	198
Printed gate length in a high-performance MPU integrated circuit	nm	28	18	11	7	5

requires in-depth understanding of the physical phenomena that occur in structures of such dimensions. Estimated on the basis of the Landauer theory, the resistance of IC components and a 6 nm-width conductive tracks in integrated circuits may be so large (of the order of hundreds of ohms) that it will have to be taken into account in the balance of power dissipation in the chip, as well as in the calculation of the system dynamics.

Research in nanostructures led to the discovery of giant magnetoresistance by Peter Grünberg (Germany) and Albert Fert (France) in 1988. The effect of giant magnetoresistance is observed as a major change in the resistance of the sample upon the application of a magnetic field. Giant magnetoresistance is used in the fabrication of magnetic disks for computers. Its discovery brought Grünberg and Fert the Nobel Prize in 2007.

Chances are that conductance quantization will provide the basis for a quantum standard of electrical resistance and conductance operating at room temperature. Strikingly, conductance quantization in metal nanowires and the QHE are very similar despite their completely different physical grounds. It is only due to the spin degeneracy of the electrons that the conductance step in the conductance quantization in nanostructures is $G_0 = 2e^2/h = 2/R_{K,}$ and not $1/R_K$, as in the QHE. If the spin degeneracy is lifted by the application of a strong magnetic field, the conductance steps in the two effects become equal, $G_M = e^2/h = 1/R_K$. The effect of removing the spin degeneracy was confirmed experimentally in a study of conductance quantization in a GaAs/AlGaAs semiconductor heterostructure in a 2.5 T magnetic field in low-temperature conditions (at 0.6 K) [19]. The estimated magnetic induction necessary to lift the spin degeneracy in metals at room temperature is as large as 1000 T, which makes the experimental verification technically infeasible.

Presently it cannot be determined whether the quantization of electrical conductance in nanostructures at room temperature could be used for the development of a conductance or resistance standard. If so, it would overcome the two major difficulties that restrict the use of QHE standards: the high cost of the sample and the extreme physical conditions, i.e. a very low temperature and a very strong magnetic field. To be used as a conductance or resistance standard, a nanostructure would have to be stable in time. It would have to be made of a high-purity conductive material, in which the impurity scattering of electrons would be negligible. Also, the design of the nanostructure, especially the connections between the electrodes (terminals) and the constriction, should provide a transmission T of electrons close to 100 %. At present there are no nanostructures that could meet these requirements.

In voltage amplifiers and high-sensitivity low-frequency amplifiers the input signal is keyed with electromechanical relays. The repeated closing and opening of macrocontacts in each circuit with an electromechanical relay results in the spontaneous formation of nanowires. When nanowires form between the relay contacts, transient current oscillations resulting from conductance quantization are observed in the circuit, as shown in Fig. 7.9. An analysis of this effect and relevant measurement data are provided in [13].

In systems with high-gain amplifiers such oscillations can cause instability of the whole multi-stage amplifying circuit.

The phenomenon of conductance quantization in nanostructures, especially in nanowires, can be used for the determination of the geometric dimensions of 1–DEG conductors (nanowires) on the basis of measurements of their electrical conductance.

References

1. N. Agrait, G. Rubio, S. Vieira, Plastic deformation of nanometer-scale gold connective necks. Phys. Rev. Lett. **74**, 3995–3998 (1995)
2. J.L. Costa-Krämer et al., Nanowire formation in macroscopic metallic contacts: quantum mechanical conductance tapping a table top. Surf. Sci. **342**, L1144–L1149 (1995)
3. M.P. Das, F. Geen, Landauer formula without Landauer`s assumptions. J. Phys.: Condens. Matter **14**, L687 (2003)
4. T.S. Fisher, *Thermal Energy at the Nanoscale* (World Scientific, New Jersey–London, 2014)
5. J.K. Gimzewski, R. Möller, Transition from the tunneling regime to point contact studied using STM. Phys. Rev. B **36**, 1284–1287 (1987)
6. A. Greiner, L. Reggiani, T. Kuhn, L. Varani, Thermal conductivity and lorenz number for one-dimensional ballistic transport. Phys. Rev. Lett. **78**, 1114–1117 (1997)
7. K. Hansen et al., Quantized conductance in relays. Phys. Rev. B **56**, 2208–2220 (1997)

8. H. Ibach, H. Lüth, Solid-State Physics. An Introduction to Principles of Materials Science (Springer, Heidelberg, 1995)
9. A. Kamenec, W. Kohn, Landauer conductance without two chemical potentials. Phys. Rev. B **63**, 155304 (2001)
10. R. Landauer, Spatial variation of currents and fields due to localized scatters in metallic conduction. IBM J. Res. Dev. **1**, 223–231 (1957)
11. R. Landauer, Conductance determined by transmission: probes and quantised constriction resistance. J. Phys. Condens Matter **1**, 8099–8110 (1989)
12. C.J. Muller et al., Qunatization effects in the conductance of metallic contacts at room temperature. Phys. Rev. B **53**, 1022–1025 (1996)
13. W. Nawrocki, M. Wawrzyniak, J. Pajakowski, Transient states in electrical circuits with a nanowire. J. Nanosci Nanotechnol. **9**, 1350–1353 (2009)
14. F. Ott, J. Lunney, Quantum conduction: a step-by-step guide, Europhysics News. **29**, January/February, 13–15 (1998)
15. L.G.C. Rego, G. Kirczenow, Qunatized thermal conductance of dielectric quantum wire. Phys. Rev. Lett. **81**, 232–235 (1998)
16. K. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes, Measurement of the quantum of thermal conductance. Nature **404**, 974–977 (2000)
17. The International Technology Roadmap for Semiconductors. Internet site (2013)
18. B.J. van Wees et al., Quantized conductance of point contacts in a two-dimensional electron gas. Phys. Rev. Lett. **60**, 848–850 (1988)
19. B.J. van Wees et al., Quantum ballistic and adiabatic electron transport studied with quantum point contacts. Phys. Rev. B **43**, 12431–12453 (1991)
20. M. Wawrzyniak, Measurements of electric nanocontacts (in Polish). Serie: Dissertations, Publishing House of Poznan University of Technology (2012)

Chapter 8
Single Electron Tunneling

Abstract We begin this chapter with a brief presentation of the theory of single electron tunneling (SET) through a potential barrier. The conditions necessary for the observation and registration of this effect include a low temperature of a few kelvins and a small-sized sample. Low temperature will reduce the thermal energy that could interfere with the tunneling process. A small size of the tunneling junction is necessary to ensure its low electric capacitance. Further in this chapter we describe the basic electronic systems using SET: the SETT transistor, the electron pump, and the turnstile device. We discuss the attempts to construct a direct-current standard with SET junctions and explain why the results are unsatisfactory. We also describe two SET systems of more practical importance: the electron-counting capacitance standard (ECCS) and the Coulomb blockade thermometer (CBT).

8.1 Electron Tunneling

8.1.1 Phenomenon of Tunneling

Charge carriers tunneling through the potential barrier is the phenomenon of those explained on the basis of quantum mechanics. It generally refers to the transition of electrons from the conductor to the conductor by a thin insulating layer (1–3 nm) that forms the barrier. The two metal plates separated by the insulating layer form a thin tunnel junction, which is a system in which there electron tunneling occurs—Fig. 8.1a. To simplify the analysis, we assume that the plates are made of the same metal as is usually in manufactured junctions. Current flow, i.e. transport of electrons between two plates (electrodes), requires the performance of work E_G electrons (electron work function of metal) and overcome the potential barrier (energy gap) formed by the insulating layer. There is possible electron transport through the junction between the plate 1, in which energy levels in the conduction band are fully occupied, and the plate 2, in which are allowed but unfilled energy levels in this band. In order to shift the Fermi level E_F each to other, it means E_{F1} related to

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_8

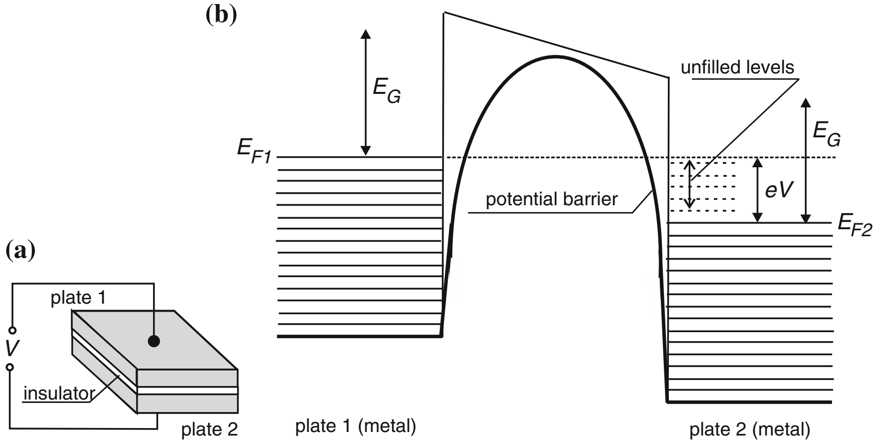


Fig. 8.1 Electron tunneling: **a** the electrodes for tunneling separated by insulation, **b** the distribution of energy in the tunneling electrodes (E_G—electron work function, E_{F1} and E_{F2}—fermi levels)

E_{F2}, the electrical voltage V should be applied between the two plates. This voltage causes a shift in the energy levels of $eV = (E_{F1} - E_{F2})$ [6].

A voltage bias V and the shift of energy levels allows to pass of electrons from the plate 1–2 by the potential barrier, i.e. the flow of a tunneling current through the junction. A width of the potential barrier is a function of the thickness of the insulating layer between the plates forming a junction. Generally, the width is a function of the distance between the electrodes. The phenomenon of tunneling is important in explaining the structure of matter and for the construction of highly sensitive measuring instruments. Among other things, it is the basis of operation of the scanning tunneling microscope, as described in the Chap. 10. Tunneling in semiconductors was discovered by Leo Esaki [1], and in superconductors by Ivar Giaever [2], for which both received the Nobel Prize in 1973.

8.1.2 Theory of Single Electron Tunneling

A variation of the above-described phenomenon of tunneling is the tunneling of single electrons—SET (Single Electron Tunneling). Possibility of manufacturing of devices in which we can track the movement of single electrons or control them is high interested for many potential users. The technology roadmap for semiconductor industry ITRS lists SET devices and systems as a possible technology that will replace the CMOS technology after you-reap its capabilities. The ability to monitor or control the movement of single electrons in a conductor is associated with defeat by each electron Coulomb blockade. Energy to charge a sample by one additional electron receives a significant obstacle to the transfer—just the Coulomb blockade energy.

We are considering single electron box, a SET junction, consisting of two metal plates. Capacity of a SET junction is labeled C_T, and its resistance as R_T—Fig. 8.2. The junction is biased with a source of a current I_{bias}. The internal conductance of the source is G_S. V_T voltage on the junction is measured using an oscilloscope with an input impedance Z_{in}. A Coulomb blockade energy of such a SET junction is described by the formula:

$$E_C = \frac{e^2}{2C_T} \tag{8.1}$$

where: E_C—Coulomb blockade energy, C_T—capacity of the tunnel junction.

Generally, E_C represents the tunneling energy scale. It is important to compare E_C with other types of energy in the system: heat, radiation, and so on. E_C has to be much greater than the thermal energy if we want to observe or to measure it:

$$E_C \gg k_B T \tag{8.2}$$

where: k_B—Boltzmann constant, T—temperature of a SET junction.

SET junctions are usually made of aluminum plates isolated by AlO_x oxide layer. So the structure of a junction is $Al/AlO_x/Al$. The technology of junctions provides oxidation of a surface of Al plates. The formed AlO_x oxide layer has the required insulating properties. The typical junction capacity C_T is of the order 0.05–1 fF (5×10^{-17}–10^{-15} F). Junctions have dimensions nanometer range. Knowledge of the value of the capacity of the junctions allows you to specify the Coulomb blockade energy, for example for $C_T = 0.1$ fF, the energy is $E_C = 1.3 \times 10^{-22}$. For the condition $E_C = 3\ k_B T$ the highest operating temperature of junction is 3.1 K.

Insulation separating the metal plates in the SET junction should ensure a sufficiently high resistance R_T of the junction. R_T value is compared with the von Klitzing constant $R_K = h/e^2 \approx 25.8$ kΩ; $R_T > R_K$. High resistance of the junction is necessary for tunneling of electrons ($1e$) through it. If the resistance of the junction is not greater than R_K, that electrons are not well localized to the junction

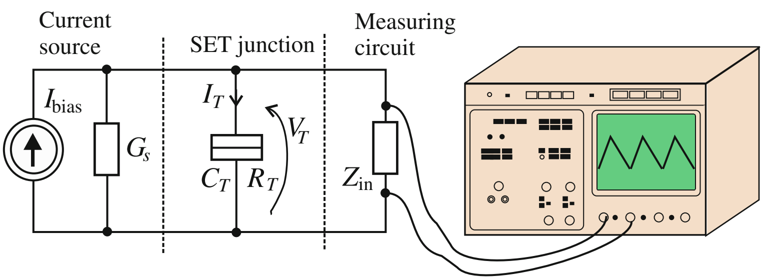


Fig. 8.2 SET junction for single electron tunneling in a measuring circuit

electrodes, and consequently the quantum fluctuations of electrical charge suppress the Coulomb blockade [11]. However, the resistance of the junction should not be too large. A SET junction is the source of the measurement signal and the internal resistance of the source, R_T, should be significantly smaller than the input impedance of the measuring circuit Z_{in}, for example an input circuit of oscilloscope. Resistance R_T of SET junctions should be in the from 100 kΩ to 1 MΩ, which together with the capacitance values of C_T (from 0.1 to 1 fF) yields a time constant: R_TC_T from 10 ps to 1 ns. The addition of one electron or loss of one electron on the SET junction changes the voltage ΔV_T:

$$\Delta V_T = e/C_T \tag{8.3}$$

You can calculate the change in ΔV_T voltage of a SET junction of a capacity 10^{-16} F (for example) after the addition or loss of one electron, ΔV_T = 1.6 mV. Changing the voltage of such a value is easily measured.

It is worth mentioning that both the single SET junction and SET electronic devices are very sensitive to interference from microwaves included in the electromagnetic field. Quantum of energy transported by microwave radiation should be less than the Coulomb blockade energy:

$$h \times f < E_C, \tag{8.4}$$

where h—the Planck constant.

On the basis of the Coulomb blockade energy E_C (e.g. $E_C = 1.3 \times 10^{-22}$ $C_T = 0.1$ fF) it is possible to determine the frequency limit of harmful interferences, e.g. $f = E_C/h = 1.3 \times 10^{-22}/6.62 \times 10^{-34} \approx 200$ GHz. The most important devices and systems using single-electron tunneling phenomenon are: the SETT transistor, electrons pump, electron counting capacitance standard ECCS and CBT thermometer.

8.2 Electronic Circuits with SET Junctions

8.2.1 SETT Transistor

Two tunnel junctions create SETT transistor (Single Electron Tunneling Transistor). Two SET junctions in a SETT transistor are created of two metallic plates and a single metal speck, which is a common component of the two junctions (Fig. 8.3).

The speck is called an island. By analogy to the FET transistor the electrodes in SETT are called: drain, source and gate-island. SETT transistors comprise junctions of an $Al/AlO_x/Al$ material structured, as a single SET junction. If we apply a voltage of 1 mV between the drain and source, the current I_T will flow through the transistor as a stream of sequential tunneling electrons. At any time, only one electron can tunnel to the island and only one electron must tunnel from the island

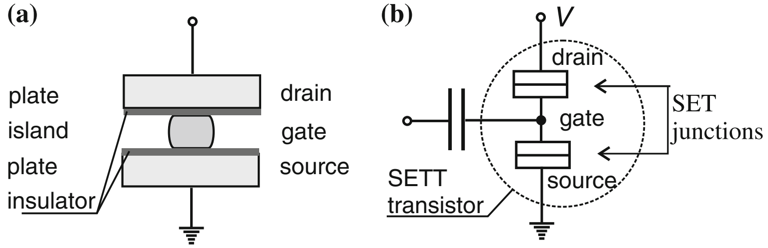


Fig. 8.3 SETT transistor using single electron tunneling phenomenon: **a** SETT structure, **b** SETT in an electronic circuit

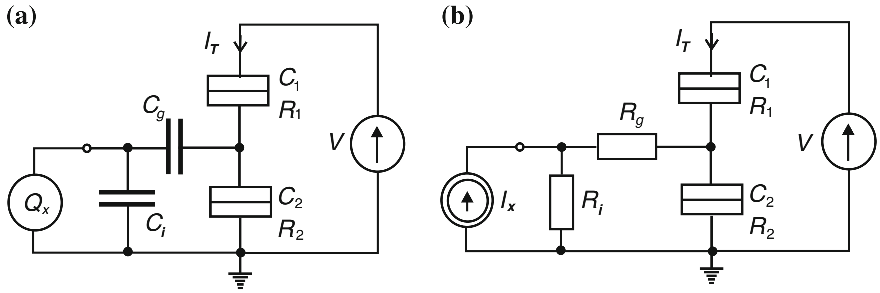


Fig. 8.4 SETT transistor: **a** in a circuit with capacitive coupling, **b** in a circuit with resistive coupling

before the next electron can tunnel to the island. Current I_T is controllable by the charge applied to the gate or by the voltage connected to the gate. Accordingly, we distinguish between the type of transistors controlled by capacitive coupling (Fig. 8.4a, used much more often), and transistors controlled by resistive coupling (Fig. 8.4b).

Sensitivity of a SETT transistor (Fig. 8.4a), $\Delta I_T/\Delta Q$, can reach 10^{-4} e/s, which is possibly 10^{-4} e in band of 1 Hz. These features make the transistor SETT ultra sensitive electrometer [7]. A very small value of the gate capacitance C_g causes that scattered (parasitic) capacities, of small values of electric capacity, in the system are particularly harmful, and thus the resolution of the electrometer may be reduced by several orders of magnitude! This leads to a typical value of the current $1e/RC \approx 10$ nA.

It should be mentioned that all important metrological applications of SETT transistors use SET junctions made in the technology of $Al/AlO_x/Al$. Kinds of materials, in which Coulomb blockade occurs, is, however, much more, including carbon nanotubes and granular Si coating. However, research is underway to find new materials. The main objective of this research is to develop a SET junction able to operate at higher temperature than that of the $Al/AlO_x/Al$. For best performance and the best control of these devices is found in the structure of CMOS with Si-on-insulator SiO, working at temperature up to 100 K, under the best conditions [10].

8.2.2 Electron Pump and Turnstile Device

Figure 8.5a shows the simplest version of the electron pump. It is an electronic circuit with three junctions to the tunneling of single electrons, including two islands. SET junction matrix may comprise more than three of these junctions. Compared to the SETT transistor—the electron pump has one more island. The addition of this second island allows trapping of electric charge. In this system the islands act as gates, using a voltage to control of tunneling of electrons to/from the island. The supply voltage V_{sup}, symmetrical to the ground, is applied to electrodes "drain" and "source". The voltage $V_{g1}(t)$ and $V_{g2}(t)$ controls gate 1 and gate 2, respectively. The voltages are attached to the gates using coupling capacitors (C_{g1} and C_{g2}) of less capacity than the capacity of the SET junction itself.

Like in the SETT transistor—only one electron can tunnel from the source to the first island. After reaching the island the electron is closed in an electrostatic trap there. Until we apply the voltage V_{g2} to the second gate and we will not lower the energy barrier to allow tunneling, as long electron can't tunnel from the island of 1–2. Attaching voltage pulses V_{g1} and V_{g2} sequentially on subsequent gates in a junction matrix can be sequentially lower the energy barrier for them and bring to the flow of electrons through the matrix according to the principle: tunneling through one junction in one period. The process of opening the gates must be done synchronously, so the signals V_{g1} and V_{g2} should have the same frequency and be out of phase. In this way, the "pumping" an electron from island to island the electron can be passed by all junctions, it means by a matrix composed of any number of SET junctions.

It should be underlined that the pump works properly for a zero drain-source voltage (i.e., $V_{\mathrm{sup}} = 0$) as well. Current I_P in the circuit, flowing as result of electron pumping, may flow in both directions, depending on the order in which the gates are controlled by voltages. For the description as in Fig. 8.5a, the $+I_P$ current requires the pulses sequence: V_{g1} first and V_{g2} second and the $-I_P$ current for the sequence V_{g2} first and V_{g1} second.

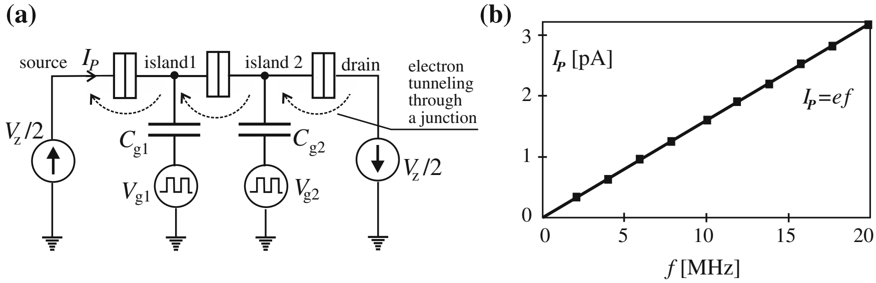


Fig. 8.5 Electron pump: **a** electronic circuit, **b** mean current I_P in the electron pump as a function of clock frequency [10]

In the electron pump circuit one and only one electron can pass in one cycle: $T_{rep} = 1/f_{rep}$. It can be seen that the intrusive application of an electron pump is a electric current standard: the timing of the pump frequency f in order to achieve the current by the basic relationship:

$$\bar{I}_P = e \times f \tag{8.5}$$

where: $\bar{I}_P$—average current (with averaging period τ), e—electron charge, f—clock frequency of a pump control (in description above $f = f_{rep}$).

Figure 8.5b shows a graph of the average current $\bar{I}_P$ flowing through the pump as a function of pulse frequency in the range 0–20 MHz. Measurement results of the current are consistent with the theoretical values predicted by the formula (8.5).

Tunneling phenomenon occurs at very weak signals, which are susceptible to interference despite the use of suppression and protections against interference. The best uncertainty of realization of the quantum current source by the means of the electron pump was 9×10^{-9} in the averaging period of 900 s [3, 5]. Electron pump can be used as a very precise standard of electric current, acting according to the rules of quantum mechanics. But there are many reasons why the pump does not perform a SET standard function in metrology yet.

- Simple observation shows that, in contrast to standards of voltage, resistance, electrical charge, or capacitance, the quantum SET current standards (reproducing current non-changeable) are more difficult to implement. Thus, as a practical representation of the size of the electric capacity standard is more convenient than the current standard.
- The voltage standard (using Josephson effect) and the resistance standard (using the quantum Hall effect) can now be used to define the current (using Ohm's law: $I = V/R$), and therefore the SET standard of current will only be attractive if it will be simpler and less expensive than the quantum voltage standard or the quantum resistance standard. Currently, it is not.
- The main disadvantage of the current standard using SET is the low achievable current value. A typical time constant RC of a SET junction leads to a maximum value of current of 10 pA [4] with the theoretical relative uncertainty of 10^{-8}. In view of the need for standards of current from 1 mA to 1 A a SET current standard offers too low current. Some research groups have decided to close the quantum metrological triangle with a current of this magnitude (10 pA), others are trying to significantly increase the value of the current from SET standards.
- Comparison of two weak quantum phenomena: the Josephson to the SET allows to estimate the difficulty of realization of a current using the SET. The difficulty is much higher than when realizing of the voltage by the Josephson effect. Ratio between the frequency and voltage in the Josephson formula (4.20) is small and it is 2.07×10^{-15} Vs, while the ratio between the frequency and the current in the SET effect is even smaller and is 1.6×10^{-19} As, both factors in basic units.

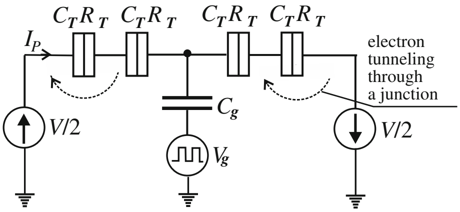


Fig. 8.6 Turnstile device for transfer of tunneling electrons

$$V_{\mathrm{ref}} = k\frac{h}{2e}f_e$$
$$\bar{I}_P = ef$$

In SETs, there are two main sources of error. Firstly, the permitted time T_{rep} of tunneling of electrons is short, due to the intended increasing the frequency of f_{rep} tunneling, thereby increasing the current I_p. If the time T_{rep} is too short, some electrons do not tunnel. Secondly, some electrons can tunnel through more junctions than the one during the open gate. Theoretical calculations showed that for a 5-junction pump producing the current of 10 pA, the generation of electric current can be done with the uncertainty of 10^{-8}. This corresponds to resolution 1 electron per second, this is about 1000 times better the resolution of the measurement systems used at present [3]. Therefore, many laboratories strives to achieve the ability of the measurements required by quantum SET standards.

A *SET turnstile device* is similar to the electron pump. A turnstile device is composed of four SET junctions. It contains only one control gate [8, 11]—Fig. 8.6.

In the turnstile device is only one control voltage V_g applied to the central gate. Each oscillation of the V_g voltage stimulates one electron to tunnel into the adjacent junction and causes in the next junction the same process, because the junctions are polarized symmetrical by supply voltage. A value of current in the turnstile device is similar to this in the electron pump.

The turnstile device is able to generate a current of intensity proportional to the frequency of tunneling, according to the formula (8.5). The direction of current flow in the circuit, unlike the electron pump, is only one.

8.3 Capacitance Standard Based on Counting Electrons

Limitation of current values in SET electron pumps, estimated numerically long time ago, reduced the possibility of building a practical standard of current. However, the NIST proposed a different metrological use of SET a pump [12]. Instead of using for the current standard construction—the SET pump is used to build a standard of electrical capacitance. It was made the following calculation of electric values for quantum capacitance standard: if you are pumping a current of 1 pA to 1 pF capacitor for 1 s, then the voltage on capacitor will be 1 V. For a known electric charge Q supplied to the capacitor—the capacity C is determined by measuring the voltage:

$$C = \frac{Q}{V} = \frac{N\,e}{V} \tag{8.6}$$

where: C—capacitance of the capacitor, N—number of electrons delivered by electron pump to the capacitor, V—voltage on the capacitor.

Schematic diagram of the NIST electric capacitance counting electrons (ECCS) standard is shown in Fig. 8.7. Three essential parts of the system are: an electron pump, a SETT transistor and a cryogenic capacitor C_{ref} of about 1 pF, cooled together with SET circuits.

The ECCS system operates in the two phases selected by switching the switches P1 and P2. In the first phase the switch P1 is closed and the switch P2 open. Then the pump is pumping electrons on the inner plate of a capacitor C_{ref}. During this process, the voltage along the pump must be kept close to zero, or practically must be zero voltage at the point of "virtual zero". The zero detector (transistor SETT) and the feedback system provide keeping the zero voltage. In order to force such voltage on the outer plate of the capacitor C_{ref} the frequency of pumping should be set up precisely. With this condition it ensured that the electric charge carried by the pump is accumulated at C_{ref}, and does not flow to ground through the parasitic capacitances. After pumping about 10^8 electrons in one direction, the pump stops and the voltage V is measured—it should be approximately +10 V. Then the same number of electrons is pumped in the opposite direction and then the voltage is measured it should be approximately −10 V. This process of reload of C_{ref} capacitor

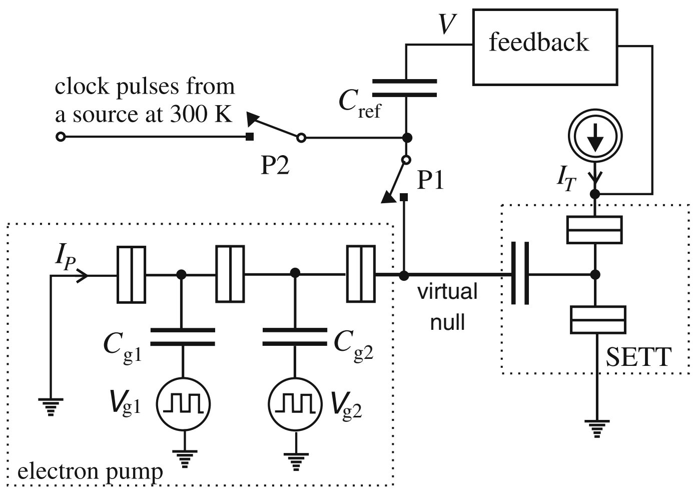


Fig. 8.7 Capacitance standard based on counting electrons—ECCS

is repeated 10–100 times (each reloading series takes about 100 s) to obtain a one, average value C_{ref}.

In the second phase of the operation cycle of the ECCS the switch P1 is open and P2 is closed. Then, using an alternating current bridge, the capacitance of the capacitor C_{ref} is compared to another capacitor placed at room temperature. This allows to determinate the electric charge (and further the capacity) of the last capacitor in an elementary charge units. After that this capacitor can be used for calibrating other capacitors.

The components of ECCS system must meet high requirements. In the system described in [11] temperature of the cryogenic part of the circuit was of 0.1 K. The electron pump was clocked at 10 MHz. It is required that in one clock cycle pump conveyed exactly one electron (not 0 and not 2). The quality of the pump was estimated at several electrons transferred too much or too little when pumping 10^9 electrons. Also estimated that the leakage of the reference capacitor C_{ref} should not cause the loss of stored charge more than 10^{-8} Q for 100 s. This means that the insulation resistance R_{ins} of the capacitor must in order of 10^{22} Ω. The system uses vacuum capacitor with plates of copper and sapphire bearing structure of the insulation resistance $R_{ins} \geq 10^{21}$ Ω. Presented in the paper [11] the relative standard uncertainty of the capacity is 10^{-6}, but the aim of the study was to further improve its value to 10^{-8}. The authors [9] proposes the use of capacitance standard described above to complete a quantum metrology triangle in an alternative manner to that shown in Chap. 3.

8.4 Thermometer with the Coulomb Blockade

Junctions for single electron tunneling, the SET junctions, are used in the primary thermometer at low temperature range—the Coulomb blockade thermometer CBT. The author of the concept of CBT is Pekola from Finland [9]. It is noted that temperature measurements in low temperature physics and cryogenics are often carried out with the primary thermometer. The primary thermometer is an instrument, whose operation is based on the fundamental laws of physics and not on a temperature dependence of the parameter of the temperature medium. At low or very low temperature parameters of materials can change significantly under the influence of other factors not only of temperature. Therefore indications of parametric thermometers can't be fully trusted. Examples of phenomena with a spectacular change in the parameters of materials at low temperatures are superconductivity and superfluidity. For these reasons, the physics community well accepted the news of a new kind of original primary thermometer—the CBT. Other primary thermometers for low temperature range are: a gas thermometer, a magnetic thermometer (measurement of magnetic susceptibility), a nuclear orientation thermometer (changing the spatial distribution of nuclear radiation), a ^{3}He melting curve thermometer and a noise thermometer (see Chap. 4) [8].

For two SET junctions, connected serially, the theoretical dependence of the dynamic resistance $R = \mathrm{f}(V) = \mathrm{d}V/\mathrm{d}I$ in the function of bias voltage V was obtained [9]. Characteristic of $R = \mathrm{f}(V)$ has the shape of a bell and is dependent on the absolute temperature T of SET junctions—Fig. 8.8.

A sensor of the CBT thermometer is a matrix containing total number N_{tot} SET junctions instead of two SET junctions, considered above. The junction matrix consists of several branches parallel connected, containing $N(N \geq 10)$ junctions connected in series in each branch—see Fig. 8.8a. As a result of serial connection of N SET junctions we obtain the $N/2$ times better sensitivity of the characteristics of the dynamic resistance $R = \mathrm{f}(V)$ relative to the two junctions in series.

While the parallel connection of serial branches is arranged in order to reduce the dynamic resistance R to the value of 100 kΩ, so that the R value is easily measurable. Important are three points on the characteristics of dynamic resistance $R = \mathrm{d}V/\mathrm{d}I$: $R = R_{\mathrm{max}}$ for $V = 0$, $R = R_T$ for $V \to \infty$ (in practice for: -10 mV > V > +10 mV) and $R_{1/2} = (R_{\mathrm{max}} + R_T)/2$. On the basis of the measured characteristic $R = \mathrm{d}V/\mathrm{d}I$ the half-bias voltage $V_{1/2}$ is calculated. The half-bias voltage $V_{1/2}$ is equal to the difference (positive and negative) for which the dynamic resistance of the junctions matrix is equal to the $R_{1/2}$. The absolute temperature T of the sensor (i.e. matrix of SET junctions) is a function only of the half bias voltage $V_{1/2}$, the number N of SET junctions in the matrix connected in series and the fundamental physical constants: k_B and e [9]:

$$T = \frac{e}{5.439\, k_B N} V_{1/2} \tag{8.7}$$

where: e—electron charge, k_B—the Boltzmann constant, N—number of SET junctions connected in series in a matrix-sensor, $V_{1/2}$ half bias voltage of a sensor.

The Coulomb blockade energy in a CBT sensor should be much larger than the thermal energy k_BT:

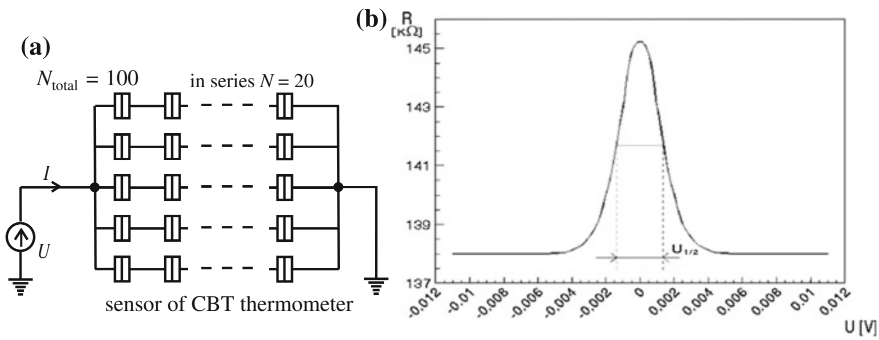


Fig. 8.8 Coulomb blockade thermometer CBT: **a** matrix of 100 SET junctions forming a temperature sensor, **b** resistance of a junction matrix in function of a bias voltage

$$\frac{e^2}{2N C_T} \gg k_B T \tag{8.8}$$

For exemplary characteristics of $R = \mathrm{f}\,(V) = \mathrm{d}V/\mathrm{d}I$ shown in Fig. 8.8: $N = 10$, $R_{\max} = 146$ kΩ, $R_T = 138$ kΩ, $R_{1/2} = 142$ kΩ and $V_{1/2} = 2.6$ mV. For these values of R we obtain the measured temperature of 0.55 K.

A CBT thermometer made by a team from the University of Jyväskylä has been tested in the measuring range of 8 mK to 30 K. SET junctions in the sensor matrix waere made of $Al/Al_2O_3/Al$ layers. The recommended measuring range of the CBT is narrower: from 20 mK to 30 K. The relative error of the thermometer in the range is of 0.3 % [3]. The upper limit of the measuring range is determined by the condition expressed by the formula (8.8). The lower limit is associated with the breaking of the coupling between the electron system in SET and a grid of integrated circuit substrate at too lower temperature.

A CBT thermometer is insensitive to strong magnetic fields (it was tested in the field of induction to 27 T), which is a valuable advantage when conducting research at low temperature and in high field. Indications of CBT was compared with measurements using a ^{3}He melting curve thermometer in a temperature range from 8 mK to 1 K at realization of the Provisional Low Temperature Scale of 2000 from 0.9 mK to 1 K. In addition to the use of CBT as a primary thermometer, CBT system can also be used as secondary thermometer [4]. It uses the dependence of the dynamic conductance of the temperature sensor ($G = 1/R = \mathrm{d}I/\mathrm{d}V$):

$$\frac{\Delta G}{G_T} = \frac{N-1}{N} \times \frac{e}{k_B\, C\, T} \tag{8.9}$$

where: G—dynamic conductance of a sensor (matrix of SET junctions), N—number of SET junctions connected in series in a matrix-sensor, C—capacitance of a sensor.

A secondary CBT thermometer requires calibration due to unknown exactly sensor capacitance C occurring in formula (8.9). The advantage of the secondary CBT thermometer is shorter measurement time (approximately 1 s) than at those at the primary CBT thermometer.

References

1. L. Esaki, New phenomenon in narrow germanium p-n junctions. Phys. Rev. **109**, 603–604 (1958)
2. I. Giaever, Energy gap in superconductors measured by electron tunneling. Phys. Rev. Lett. 5, 147–150 (1960)
3. H.D. Jensen, J.M. Martinis, Accuracy of the electron pump. Phys. Rev. B **46**, 13407–13427 (1992)
4. J.P. Kauppinen et al., Coulomb blockade thermometer: tests and instrumentation. Rev. Sci. Instrum. **69**, 4166–4175 (1998)

5. M.W. Keller et al., Accuracy of electron counting using a 7-junction electron pump. Appl. Phys. Lett. **69**, 1804–1806 (1996)
6. Ch. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996)
7. A.N. Korotkov, D.V. Averin, K.K. Likharev, S.A. Vasenko, *Single-Electron Transistors as Ultrasensitive Electrometers*. *Springer Series in Electronics and Photonics*, vol. 31 (Springer, New York, 1992), pp. 45–59
8. W. Nawrocki, *Introduction to Quantum Metrology (in Polish)* (Publishing House of Poznan University of Technology, Poznan, 2007)
9. J.P. Pekola, K.P. Hirvi, J.P. Kauppinen, M.A. Paalanen, Thermometry by arrays of tunnel junctions. Phys. Rev. Lett. **73**, 2903–2906 (1994)
10. Y. Takahashi et al., Silicon single electron devices. Int. J. Electron. **86**, 605–639 (1999)
11. C. Urbina, et al., *Manipulating Electrons One by One, Single-Electronics—Recent Developments. Springer Series in Electronics and Photonics*, vol. 31 (Springer, Berlin, 1992), pp 23–44
12. N.M. Zimmerman, M.W. Keller, Electrical metrology with single electrons. Meas. Sci. Technol. **14**, 1237–1242 (2003)

Chapter 9
Atomic Clocks and Time Scales

Abstract This chapter presents atomic frequency standards with high-stability frequency of oscillations. Also discussed is the Allan variance used for measuring the frequency fluctuations in such standards. We present the design of the 9.2 GHz cesium standard, which is presently the most important atomic standard. We discuss the directions of development and the metrological parameters of cesium atomic standards: more complicated and expensive, but providing better accuracy cesium fountain standards, and much cheaper miniature cesium standards with the size of a matchbox. Also discussed are other frequency standards: hydrogen masers, rubidium standards and the currently implemented standards with a visible wavelength. Optical frequency standards are potentially 10^5 times more stable than the cesium standard, though at present their actual stability is only 10 times higher. Together with the optical frequency standards we discuss the optical comb, used for mapping frequencies of the order of 10^{14} Hz to MHz frequencies. Also discussed are the major time scales, the International Atomic Time (TAI) and the Coordinated Universal Time (UTC), as well as the role of satellite navigation systems (GPS, GLONASS, BeiDou) in the dissemination of the standard time signal.

9.1 Theoretical Principles

9.1.1 Introduction

Periods of astronomical phenomena, like day and night, or year, are obvious examples of natural periodic processes. Centuries-old investigation on the stability of periodic phenomena as well as searching for other physical periodic processes of great stability lead to build *atomic clocks*, in which the most important part is atomic frequency standard. Sometimes "atomic frequency standard" and "atomic clock" are considered to be synonyms and used in an interchangeable way.

It was possible to build the atomic clock only at a certain stage of physics development, when quantum mechanics was already developed. According to one

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_9

of its principles, an energy level of atom can take only discrete, quantified values: E_1, E_2, E_3, E_4, etc., where $E_1 < E_2 < E_3 < E_4$. Transition of an atom from a higher level E_2 to a lower level E_1 is accompanied with emitting an energy quantum, i.e. a photon, with value:

$$E_2 - E_1 = h\nu_0 \tag{9.1}$$

where: h is the Planck constant, ν_0—electromagnetic radiation frequency.

The transition of atom from a lower energy level, e.g. E_2 onto an adjacent, higher level E_3, requires the absorption of an electromagnetic radiation photon with frequency ν_1: $E_3 - E_2 = h\nu_1$. Atomic clocks apply transitions "down" and "up" only between two quantum levels in an atom, e.g. E_1 and E_2. Because atomic energy levels are unchangeable (according to state-of-the-art), frequency ν_0 of transitions between them is connected exclusively with the Planck constant, a basic physical constant.

Energy levels E_i in atoms are determined by hyperfine interactions between the magnetic moment of unpaired electron and the magnetic moment of atomic nucleus in an atom being in ground (fundamental) state [9]:

$$E_i(F, m_F, H) = \frac{-h\nu_0}{2(2I+1)} \pm \frac{h\nu_0}{2}\sqrt{1 + \frac{4m_F}{2(2I+1)}x + x^2} \tag{9.2}$$

where: x—magnetic field intensity function H_0 (9.5), I—quantum number of nuclear spin, F—quantum number of total angular momentum of an atom, m_F—quantum number of the projection of the vector of angular momentum F onto the direction of the vector of magnetic field H_0.

$$F = I \pm \frac{1}{2} \tag{9.3}$$

$$m_F = 0,\ \pm 1,\ \pm 2,\ \ldots,\ \pm F \tag{9.4}$$

$$x = \sqrt{\frac{2C_H}{\nu_0}}H_0 \tag{9.5}$$

where C_H is the quadratic Zeeman effect coefficient, describing dependence $\nu_0 = f(H_0)$.

The Heisenberg relation determines the uncertainty limits of energy measurement ΔE in time interval Δt:

$$\Delta E \times \Delta t \geq \hbar/2 \tag{9.6}$$

where Δt is the measurement duration, or interaction time.

From dependences (9.1) and (9.6) it results that the uncertainty of determination (or measurement) of frequency Δv_0 depends on the measurement time according to formula:

$$\Delta v \times \Delta t \geq 1/4\pi \tag{9.7}$$

If the band of frequency uncertainty Δv is to remain narrow, the time Δt of measurement or interaction should be possibly long—Fig. 9.1.

Hyperfine interactions in the atom are weak, which causes a narrow energy gap $\Delta E = E_2 - E_1$ between the atomic energy levels. Small value of ΔE causes a resonant frequency v_0 to be in the range of microwaves, which enables to build an electronic circuit dedicated to control the atomic frequency standard. Narrow energy gap $E_2 - E_1$ causes, however, problems with separating atoms located either at energy level E_1 or E_2. Such separation is indispensable to obtain in the system a sufficient number of atoms capable of absorbing an energy quantum during atomic transition (change of energy level) $E_1 \rightarrow E_2$ or, in another system, to emit an energy quantum during transition $E_2 \rightarrow E_1$. Two methods are applied to separate the atoms with labelled energy levels. The first method is based on the properties of an atom that at energy level E_1 has an opposite magnetic moment than at level E_2. Therefore, atoms with energy E_1 can be separated from those with energy E_2 by means of deviation, with the use of heterogeneous magnetic field. This method is applied in caesium resonators and hydrogen masers.

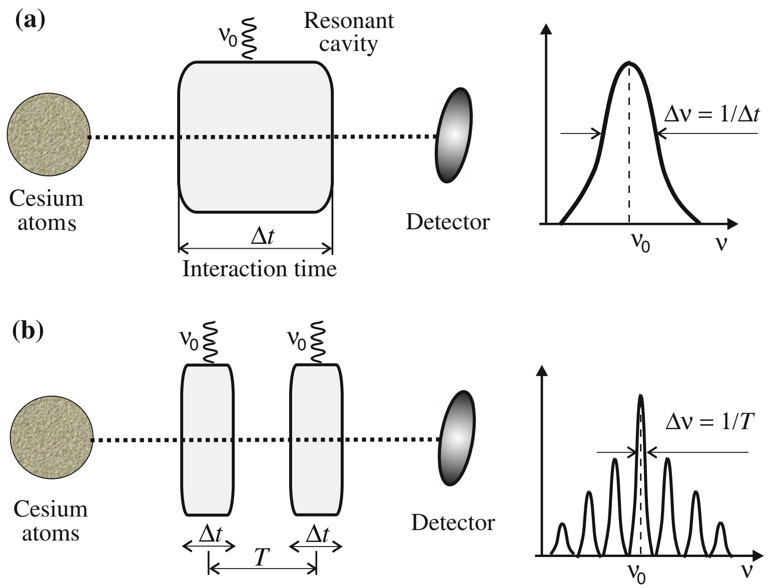


Fig. 9.1 Interaction time Δt of atoms in atomic generator and deviation band Δv_0 of resonant frequency: **a** single resonant cavity, **b** double resonator system (Ramsay chamber)

The other method applies laser optical radiation. Laser light with a selected wavelength is used to activate (stimulate) atoms to transit e.g. from level E_2 to higher-energy level E_3. Spontaneous transitions from level E_3 to level E_2 or E_1 are very fast, and their result is an increase in the number of atoms per one of the two levels, E_2 or E_1. This principle is applied in rubidium standards and caesium resonators with optical pumping (Cs atoms fountain).

9.1.2 Allan Variance

Atomic standard frequency generators are the most accurate standards known in science. A basic metrological problem concerning atomic generators is the stability measurement of standard frequency ν_0. Let us pay attention to the relation between the standard frequency value and the inaccuracy of its reproduction. For the caesium atomic generator, standard frequency ν_0 is of the order of 10^{10} Hz, and relative frequency $\Delta\nu/\nu_0$ of its reproduction is about 10^{-15}. It means that absolute inaccuracy $\Delta\nu = \nu_x - \nu_0$ is of the order of 10^{-5} Hz. In order to measure the difference of frequencies $\Delta\nu$ corresponding to *a single period* of a signal with investigated frequency ν_x, a counting period $T = 1/\Delta\nu = 10^5$ s $\cong$ 28 h is needed. If the differences $\Delta\nu$ were to be processed statistically, the counting time would have to be several times longer, to obtain a sufficient measurement resolution $\Delta\nu$. For example, a counting series of a hundred pulses with frequency 10^{10} Hz, in order to average the results, would require constant measurements during 2800 h, i.e. during around 4 months. So, it is quite useless to express the generator accuracy by the standard deviation of a series of such frequency measurements. The period of signal with an exemplary frequency 10^{10} Hz is equal to 100 ps. It corresponds to the full angle of phase shift 360°. In order to investigate stability, it is better to measure the phase shift time of a signal measured after N periods, instead of counting the pulses of that signal.

Allan variance was used to define the stability [2], in which the phase time $x(t)$ is processed, i.e. the phase shift converted into time. For a given signal with period 100 ps, the phase time 1 ps corresponds to phase angle of 3.6°. In order to define the Allan variance, we assume that there is a sinusoidal voltage at the generator output [1]:

$$u(t) = U_0 \sin\left[2\pi\nu_0 t + \Phi(t)\right] \tag{9.8}$$

where: $u(t)$—instantaneous voltage, U_0—constant voltage amplitude, ν_0—nominal (resonant) frequency, $\Phi(t)$—signal instantaneous phase shift. Signal instantaneous frequency is the sum of constant component ν_0 and constant variable ν_Z:

$$\nu(t) = \nu_0 + \nu_Z = \nu_0 + \frac{1}{2\pi} \times \frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{9.9}$$

In a further description of stability the following notions are applied: relative deviation $y(t)$ of instantaneous frequency (9.10) and phase time $x(t)$, which is the integral of relative deviation (9.11), it has a time dimension, and its value is proportional to the instantaneous phase shift $\Phi(t)$:

$$y(t) = \frac{\nu_Z}{\nu_0} = \frac{1}{2\pi\nu_0} \times \frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{9.10}$$

$$x(t) = \int_0^t y(t')\,\mathrm{d}t' = \frac{\Phi(t)}{2\pi\nu_0} \tag{9.11}$$

When analysing the generator stability, both deduced values, $y(t)$ and $x(t)$, are treated as random variables. The measurement of frequency deviation $\Delta\nu$ comprises the process of averaging the value $y(t)$ in time range τ. The beginning of averaging process is instant t_k on the time scale.

$$\bar{y}_k(t_k, \tau) = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t)\mathrm{d}t = \frac{1}{\tau}[x(t_k+\tau) - x(t_k)] \tag{9.12}$$

where $\bar{y}_k(t_k, \tau)$—k-th sample of the relative deviation of instantaneous frequency, at a determined averaging period (time) τ.

Further on, we are considering a series of N frequency measurements $\bar{y}_k$ conducted in intervals T, with averaging period τ. The results of those N measurements, expressed with relative frequency deviation, is a series: $\bar{y}_k(t_k, \tau) = \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N$. Figure 9.2 illustrates the waveform of random variable y(t) in the function of time as well as the results of a series of frequency measurements expressed in the values of relative deviation $\bar{y}_k$.

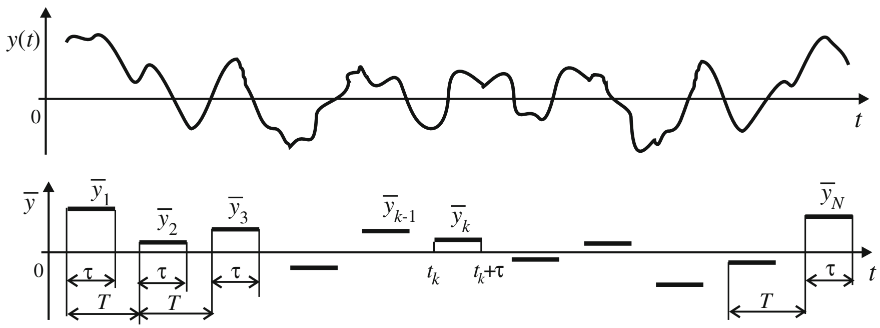


Fig. 9.2 Relative deviation $y(t)$ of the instantaneous frequency value from the nominal value ν_0 for a series of N measurements and averaging period τ

The mean value of a series of N measurements $\langle \bar{y}_k \rangle_N$ is given with formula (9.13), and variance $\sigma_y^2(N, T, \tau)$ of that series—with formula (9.14):

$$\langle \bar{y}_k \rangle_N = \frac{1}{N} \sum_{k=1}^{N} \bar{y}_k \tag{9.13}$$

$$\sigma_y^2(N, T, \tau) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\bar{y}_i - \langle \bar{y}_k \rangle_N \right)^2 \tag{9.14}$$

where i—a successive sample of relative deviation $\bar{y}_i$ of frequency in the series of N averaging processes.

It was demonstrated [1] that for finite values of N, T and τ there is a limit for the number of M processes of averaging of variance (9.14), where $M \to \infty$.

$$\left\langle \sigma_y^2(N, T, \tau) \right\rangle = \lim_{M \to \infty} \frac{1}{M} \sum_{j=1}^{M} \sigma_{yj}^2(N, T, \tau) \tag{9.15}$$

where j—a successive sample (value) of variance $\sigma_y^2(N, T, \tau)$ in a series of M calculations.

Expression (9.15) is Allan variance. For the Allan variance calculated with averaging of only two adjacent values $\bar{y}_k$, i.e. for $N = 2$, the calculation method it is very simple:

$$\left\langle \sigma_y^2(2, T, \tau) \right\rangle = \frac{1}{2} \left\langle \left(\bar{y}_2 - \bar{y}_1 \right)^2 \right\rangle \tag{9.16}$$

According to formula (9.16), Allan variance for $N = 2$ and $T = \tau$ is recommended as stability indicator of standard generators in metrology and telecommunications [1]. Formulas (9.15) and (9.16) describe the "real" value of the variance, i.e. its value for an infinite number of averaging M. The Allan variance estimator is determined according to formula (9.17) for a limited number of M averagings of function $\sigma_y^2(2, \tau, \tau)$. Telecommunication recommendations limit the selection of averaging time τ of a single sample of interval 0.1 s $\leq \tau \leq 10^5$ s.

$$\sigma_y^2(\tau, M) = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} \left(\bar{y}_{k+1} - \bar{y}_k \right)^2 \tag{9.17}$$

In atomic clocks deviation $\sigma_y(\tau)$, which is the square root of Allan variance, is estimated with formula:

$$\sigma_y(\tau) \cong \frac{\Delta \nu}{\pi \nu_0} \sqrt{\frac{T}{\tau N}} \tag{9.18}$$

where: $\Delta\nu$—frequency bandwidth around resonant frequency ν_0 (Fig. 9.1), i.e. the so called transition width, τ—averaging time during frequency deviation measurement, T—interaction time of atoms in resonant cavity, N—number of atoms participating in atomic transition.

9.1.3 Structure and Types of Atomic Standards

An atomic frequency standard contains three main components—Fig. 9.3:

- Quantum resonator with high quality factor Q, in which resonant atomic transitions occur,
- Voltage-Controlled Crystal Oscillator VCXO with high short-term stability,
- Feedback system (circuit).

A signal with resonant frequency ν_0 is closed in a feedback loop, which ensures a maximum number of atomic transitions between the labelled quantum energy levels. A feedback system receives the signal from the cavity resonator, it tunes the crystal oscillator VCXO with the signal, and synthesizes the frequency of VCXO signal.

In passive atomic standard generators (e.g. in frequency standard with a beam of caesium atoms) a microwave signal excites atomic transitions—energy level changes. In order to induce a maximum number of atomic transitions (i.e. to induce the state of resonance), we use a feedback system that adjusts or tunes the synchronized feedback system that is responsible for adjusting or retuning the frequency of synchronized generator VCXO. For the maximum number of transitions, corresponding to microwave radiation with the greatest intensity, resonance is achieved, and then the frequency of stimulating microwave signal is precisely equal to ν_0.

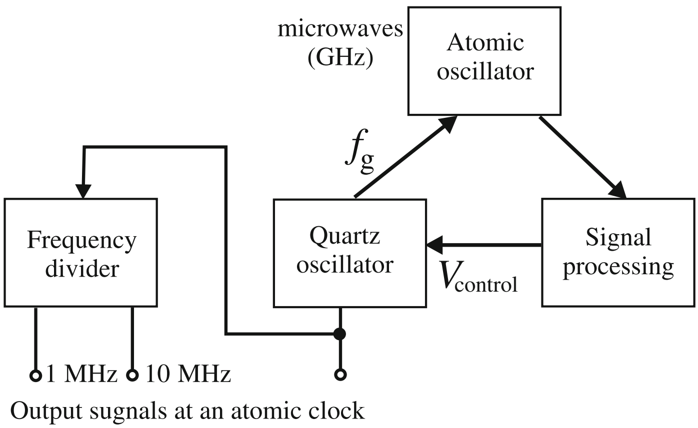


Fig. 9.3 Block diagram of atomic or molecular frequency standard

In active atomic standard generators (e.g. hydrogen maser chambers) a device generates self-sustaining oscillations. An electronic system keeps the frequency of VCXO oscillations thanks to a phase loop. Because VCXO operates in a frequency range between 1 and 100 MHz, a signal mixer and heterodyne detection are applied to synchronize the signal of VCXO with the microwave signal.

To build atomic frequency standards in microwave range we use quantum transitions in the following substances:

- Ammonia NH_3 (hydride nitrogen), resonant frequency $\nu_0 \cong 23.870140$ GHz;
- Thallium ^{205}Tl, resonant frequency $\nu_0 = 21.310833945$ GHz;
- Caesium ^{133}Cs, resonant frequency $\nu_0 = 9.192631770$ GHz—primary standard;
- Rubidium ^{87}Rb, resonant frequency $\nu_0 = 6.834682610904$ GHz—recommended as secondary standard (Recommendation CCTF 2, 2006);
- Hydrogen ^{1}H, resonant frequency $\nu_0 = 1.420405751$ GHz.

Currently, atomic standards and hydrogen masers are applied as primary frequency standards. Rubidium standards are also applied, although they are slightly less accurate. The source of standard frequency and time signal in atomic standards is the crystal generator with frequency ν_g, synchronized with atomic resonator by means of microwave signal.

Ammonia frequency standard was the source of signal in a molecular clock set working in NBS (USA) in 1949. That frequency standard used selective energy quanta absorption with frequency 23.87 GHz. This frequency corresponds to transitions between quantum energy levels in the ammonia particle. Extensive descriptions of physical properties of ammonia particle and its energy states are given in [5]. The frequency instability of ammonia standard, with values $2–10 \times 10^{-8}$ [9], depends on a number of factors, like gas (ammonia) pressure in the chamber. Microwave signal amplification became for the first time possible in an ammonia maser (*Microwave Amplification by Stimulated Emission of Radiation*) in which ammonia particles concentrated in a beam were stimulated with microwave radiation. Microwave stimulation forced the emission of radiation with resonant frequency 23.870 140 GHz. The frequency bandwidth of signal *emitted* in ammonia maser is much less than that of signal *absorbed* in ammonia absorption frequency standard. It means that ammonia maser can be a much better standard frequency source for clocks.

First ammonia masers were built independently in USA and USRR in 1954. The appearance of masers is considered as the beginning of *quantum electronics*. The authors of maser: Charles H. Townes, Nikolay G. Basov and Aleksandr M. Prochorov were awarded the Nobel Prize in 1964 "for fundamental work in the area of quantum electronics, which lead to build oscillators and amplifiers based on the working principle of laser and maser". The first caesium frequency standard (with a beam of Cs atoms) was also built in NBS in 1954, and the first caesium atomic clock was built in NPL in England in 1955 (by L. Essen and J.V.L. Parry). In Poland the first ammonia maser was built in the Institute of Physics, Polish Academy of Sciences in Poznań in 1964 (by A. Piekara et al.) [11], and 1 year later —the first caesium frequency standard in the Institute of Fundamental Technological Research, Polish Academy of Sciences in Warsaw.

9.2 Caesium Atomic Frequency Standards

9.2.1 Caesium-Beam Frequency Standard

Caesium atomic clock uses the transition between energy levels in caesium atoms, and more precisely—the frequency ν_0 of radiation emitted during such transition. The distribution of energy levels in caesium atoms, in shell 6, shells s and p, was shown in Fig. 9.4 [2]. The caesium-beam frequency standard uses the transition of ^{133}Cs atoms in ground state between hyperfine levels characterised with the quantum number of total angular momentum of atom $F = 4$ (level 1 with energy E_1) and momentum of atom $F = 3$ (level 2 with energy E_2).

A pictorial diagram of caesium atomic clock was shown in Fig. 9.5.

A heater with a small quantity of caesium (a few grams of ^{133}Cs isotope) emits a beam of caesium atoms in the vacuum lamp (chamber). The probability that a Cs atom is found in one of the states of Zeeman structure of one of the two levels in the hyperfine structure of ground level is approximately equal for each of those energy states. When the microwave radiation frequency corresponds to energy difference $E_2 - E_1 = h\nu_0$ ($\nu_0 = 9.192631770$ GHz), the radiation induces magnetic-dipole transitions between levels E_2 and E_1. Thus level E_1 is completed again (at the cost of completing level E_2) in caesium atoms moving in the resonant cavity.

In magnetic separator 1, atoms with energy E_2 are deviated. Only those atoms get to the resonant cavity radiated with microwaves. Then, the heterogeneous magnetic field in magnetic separator 2 deviates only the atoms with energy E_1 and directs them to an ionisation detector. The detector output current is proportional to the inflowing stream of ^{133}Cs atoms, and, at the same time, to the intensity of transition from level E_2 to E_1. The frequency of electromagnetic field, interacting with the microwave cavity, is swept in a range around ν_0, from ($\nu_0 - \Delta\nu$) to ($\nu_0 + \Delta\nu$), and the resonance state is indicated with the maximum output current achieved during the operation. A determined maximum value of detector current

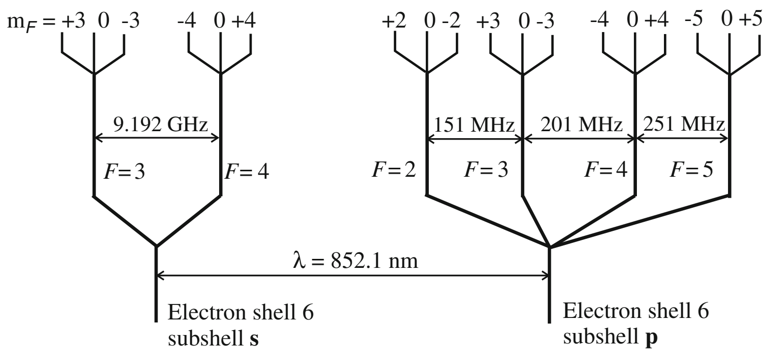


Fig. 9.4 Distribution of energy levels in ^{133}Cs atom—sub shells *s* and *p* in shell 6—were shown (*F*—quantum number of total angular momentum of atom)

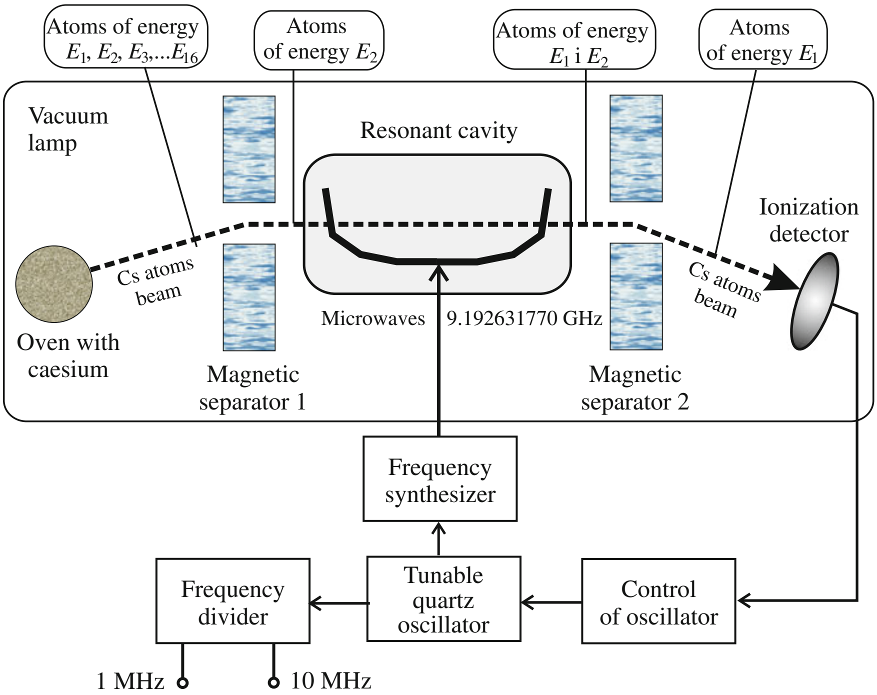


Fig. 9.5 Caesium-beam atomic frequency standard

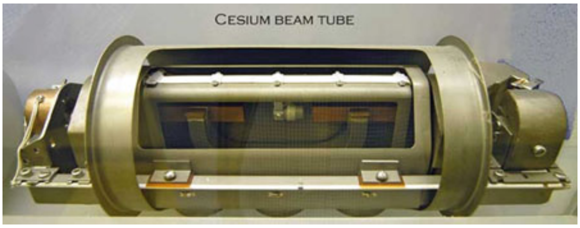


Fig. 9.6 Vacuum tube of caesium-beam atomic clock [6] (published upon consent of Dr. Lee Breakiron of US Navy)

(lack of changes in current intensity) indicates the state of oscillator tuning to the resonance. The spectrum line width around ν_0 is typically equal to about 100 Hz, which is caused by the time of atomic flow through the cavity. The vacuum tube of caesium atomic clock was shown in Fig. 9.6.

Resonant frequency $\nu_0 \cong 9.2$ GHz is not very useful in a majority of devices requiring an accurate time scale because it is too high. Therefore the signal from crystal generator, synchronized with frequency ν_0, is divided in frequency divider up to the value: 5 or 10 MHz (sometimes 1 MHz) and those signals are available for the user. Output signal from the detector is applied to control the crystal generator VCXO, from which microwave radiation and the "timing" output signal (time marker) are synthesized, in order to maximize the transition number. So, a very good short-term stability of crystal generator is combined with good long-duration stability of controlled atomic generator [2, 3].

Transition frequency ν_0 in not disturbed ^{133}Cs atoms is used to define the second in the SI system as the number of 9 192 631 770 periods of resonant oscillations. The best uncertainty of caesium-beam atomic clock amounted to the value of 5×10^{-15}—NIST-7 atomic clock in an American metrological institute NIST (National Institute of Standards and Technology) [15]. Caesium atomic standards are manufactured by some companies, from which the instruments by Agilent Technologies (former Hewlett-Packard) of type HP5071A are most common in metrological laboratories. Other manufacturers of caesium atomic clocks are: Datum, Frequency Electronics, Kernco and Oscilloquartz. Commercial caesium atomic clocks have the size and form similar to the desktop personal computer. The biggest laboratories, e.g. PTB in Braunschweig, use exclusively atomic clocks self-manufactured.

Development of caesium atomic frequency standards went in two directions. The first is the expansion of caesium generators in order to improve the stability of their frequency, which resulted in atomic standards with fountain of caesium atoms, described in the next section. A Cs fountain frequency standard has 10 times better stability (around 10^{-16}), but it costs about 100 times more than the classical, commercial caesium standard, without the fountain. The second direction of development was initiated by physicists at NIST (USA) in 2004, when they presented a miniature atomic caesium generator.

The principle of miniature clocks based on Coherent Population Trapping (CPT), which occurs in a compact and very small sealed vacuum cell. The cell contains alkali vapors (^{133}Cs or ^{87}Rb). It is radiated by a modulated laser beam. The frequency stability of atomic clocks depends on electron transitions between the ground state hyperfine levels of alkali atoms. The frequency stability is in order of magnitude of 10^{-10}. In miniature CPT clocks a microwave cavity to probe the atomic resonance is necessary. Then, it is possible to create the compact micro-clock. The cell, containing caesium atoms in this miniature generator developed at NIST, was the size of a rice grain (diameter 1.5 mm, length 4 mm). The entire generator system comprising a laser diode, photodiode, electronic sources of bias currents, had a volume of 1 cm^3. At present many companies offer cesium or rubidium miniature atomic clocks with better performances than quartz oscillators (more about rubidium atomic clocks—in Sect. 9.3). For instance, the rubidium miniature atomic clock of the SA.45s CSAC type is commercially offered by Microsemi company (CSAC—Chip Scale Atomic Clock)—Fig. 9.7.

The SA.45s rubidium atomic clock provides 10 MHz-pulse output signal and one Pulse Per Second (PPS) signal. It means the frequency of 1 Hz of the PPS signal. Then, the frequency of the quantum transition in ^{87}Rb gas, $\nu_0 \cong 6.834$ GHz, is transferred to the frequency of 10 MHz and 1 Hz, much more easy to process and to use in electronic circuits than 6.8 or 9 GHz signals. The PPS signal is a

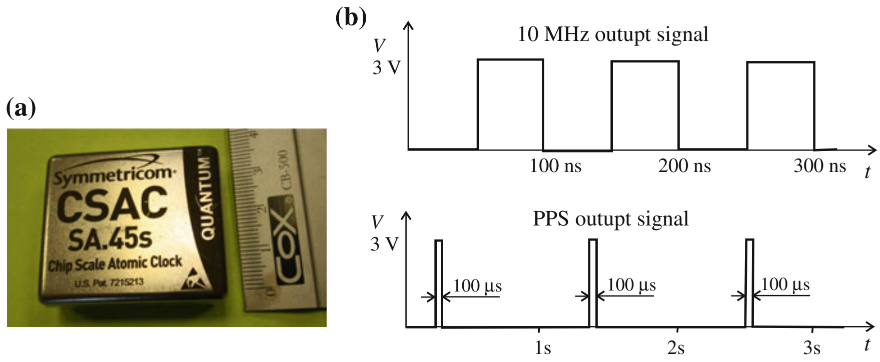


Fig. 9.7 Miniature rubidium atomic clock: **a** picture; **b** output signals

100 μs-width pulse, of a fall/rise time shorter than 10 ns, at CMOS digital levels: $V_{\max} = 0.3$ V for logic 0, $V_{\min} = 2.8$ V for logic 1 (at supply voltage of 3.3 V). The accuracy of the SA.45s clock is of 5×10^{-11} (Allan deviation). The short term stability is 2.5×10^{-11} per 100 s, the long term stability (aging) is about 20 times less, it is 5×10^{-10} per 1 year. Energy consumption of the SA.45s clock is very low, $P < 125$ mW.

Other miniature atomic clock, the E10-MRX rubidium clock manufactured by Quartzlock, provides an output sin signal of reference frequency of 10 MHz. The accuracy of the E10-MRX clock is of 5×10^{-11}, the short term stability of 8×10^{-12} per 100 s, the long term stability of 5×10^{-10} per 1 year and power consumption of 6 W.

9.2.2 Caesium Fountain Frequency Standard

Atomic clock with laser cooling of Cs atoms, called also caesium fountain atomic clock, is an improved form of caesium-beam atomic clock. Cooling of atoms is understood as slowing down, i.e. a decrease in an average kinetic energy of atoms, converted to temperature. This slowing down means a decrease in atomic speed from 1000 km/s to about 10 cm/s, i.e. by 7 orders of magnitude. The kinetic energy of atoms with random movements in optical molasses, described further in this chapter, is expressed in temperature units. Some atoms achieve a temperature of the order of single μK, and for the Cs atoms the lowest achievable temperature is 2 μK, which corresponds to the atomic speed of 11 mm/s [2]. The author of laser method for cooling atoms by generating the so-called optical molasses is Steven Chu (USA). Atomic transitions in cold Cs atoms, moving similarly as water in a fountain, was used in 1995 in an atomic fountain standard generator, Fig. 9.8, according to the theory of Claude Cohen-Tannoudji (France). Both scientists (together with W. Phillips, who also measured the temperature of cold atoms), were

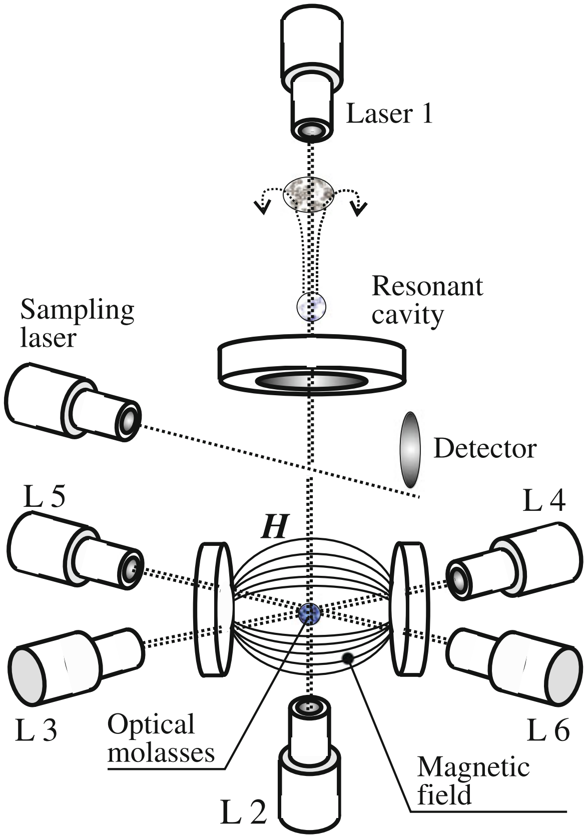


Fig. 9.8 Caesium fountain atomic frequency standard

awarded the Nobel Prize in 1997 "for the development of methods for cooling and trapping atoms with laser".

Basic component of a fountain clock is the magneto-optical trap [4]. It consists of 6 lasers (from Laser 1 to Laser 6) and a coil generating heterogeneous magnetic field *H*—Fig. 9.8. The laser cooling method applies the properties of interaction between the light wave and the particle slowed down by transmitting the momentum of light photons to its atoms. The laser light wavelength must be matched to the type of atoms, so that a photon wave could be absorbed during the atomic transition. For example, the caesium atoms are slowed down with infrared radiation IR (852 nm), on account of atomic transition [6 $^2S_{1/2}$ ($F = 4$) → 6 $^2P_{3/2}$ ($F = 5$)], and the sodium atoms—with yellow light. A laser radius is a concentrated beam of light or IR radiation, with a big number of photons and big power of interaction.

The task of 6 laser radii (3 pairs of parallel backward radii, oriented perpendicular pair-to-pair), crossing in one point, is to moderate free atoms (atomic gas) in three dimensions of space. The laser beam polarization vectors are also orthogonal-oriented to one another. In the space cut with laser light, slow movement of atoms is similar to the movement of particles in a viscous liquid (e.g. in molasses). Molasses was selected as a symbol of slowness according to an English proverb: "*slow as molasses in January*". Another effect of molasses behaviour, apart from slowing down the atoms, is instantaneous concentration of atoms. In order to maintain the concentration and not to allow the slow atoms to diffuse in different directions, the coils generate a heterogeneous magnetic field around the optical molasses. Magnetic field intensity H is equal to zero in the middle of molasses space. The atoms found outside the molasses or at its peripheries are interacted by the magnetic field force directing them to the middle, i.e. to a point in which $H = 0$ and atomic energy is the least. The cooperating lasers and coils generate a magneto-optical trap. A beam of slow atoms concentrated in the trap is launched up with a speed of about 4 m/s. The slowed down atoms are interacted by the gravitational force, which slows down and stops them. So, after achieving a certain height, the atoms fall down and get into the trap—their starting point. During the movement after launching, the atoms flow through the resonant cavity twice, in an interval of about half a second, and interact with one another in the microwave electromagnetic field in the cavity. The field intensity is measured with a microwave detector.

Also the first atomic fountain was built by S. Chu. The Cs fountain clock was installed for the first time in a French Laboratory of Time and Frequency LPTF in Paris in 1995. The resonant frequency of Cs fountain atomic clock amounts to 9.192631770 GHz, i.e. it is equal to the frequency of Cs-beam clock. Atomic fountain clock was improved and put into operation in 1999 in NIST, where it was called NIST-F1. The NIST-F1 clock is the primary time and frequency standard in USA. Its uncertainty in 2005 was equal to 5×10^{-16}, then it was reduced to 3×10^{-16}. In April 2014 a second clock with a caesium fountain in NIST, the NIST-F2, reached uncertainty of 10^{-16} [15], as NIST announced [15]. Two Cs fountain clocks were installed in PTB (Germany), one clock, called the NPL-CsF1, in the National Physical Laboratory (UK) [12] and one, the NICT-CsF1, in National Institute of Information and Communications Technology (NICT) in Japan [10]. A Cs fountain atomic clock consists of components with a total cubage of a car [16].

9.3 Hydrogen Maser and Rubidium Frequency Standard

9.3.1 Hydrogen Maser Frequency Standard

The working principle of hydrogen maser frequency standard is based on stimulated emission of electromagnetic radiation with frequency v_0, corresponding to the transition of hydrogen atoms between states with quantum number $F = 1$ (level 2) of total angular momentum of atom, and with quantum number $F = 0$ (level 1).

As it was shown in Fig. 9.9, a beam of hydrogen atoms with a determined energy is selected by magnetic separator, which generated magnetic field with induction 1 T [9]. Atoms with energy E_2 ($F = 1$) from a higher level are injected to a storage

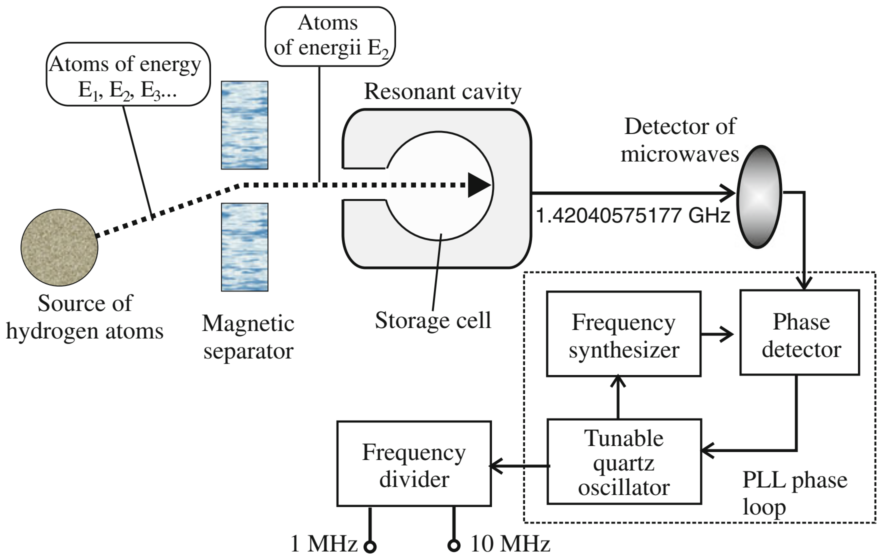


Fig. 9.9 Hydrogen maser as atomic frequency standard

cell surrounded by the resonant cavity with high quality factor Q. The electromagnetic field in the chamber is tuned to atomic resonant frequency $\nu_0 = 1.42040575177$ GHz. The resonant cavity is located inside the homogeneous electromagnetic field with induction 10^{-7} T in order to shield the space of atomic interactions from the magnetic field of the environment. Atoms with energy E_2 fall into the storage chamber and after emitting radiation with frequency ν_0 take a lower energy level E_1 ($F = 0$).

In active maser, such emission is stimulated with the microwave field generated by the atoms themselves. In passive maser the source of stimulation is an external microwave electromagnetic field. An average time of atomic interaction with the magnetic field achieved was of the order of 1 s. Hydrogen atoms are pumped out from the cavity continuously in order to keep a constant pressure, not higher than 10^{-5} Pa.

Inactive hydrogen maser the cavity volume must be big enough to generate self-sustaining oscillations. Microwave radiation with frequency ν_0 emitted by the cavity is detected with the antenna and used to close in a feedback loop of VCXO generator, which generates an external timing signal with the method of multiplying and mixing the signal (heterodyne detection). The radiation power detectable by the antenna outside the cavity is very small, of the order of 10–100 fW.

Short- and medium-term stability of hydrogen masers is currently the best of all the standard generators applied in atomic clocks. For observation time 1 s, it amounts to 2×10^{-13} for active standards and 2×10^{-12} for passive ones [8]. Long-duration stability is, however, considerably influenced by mechanical shocks and temperature changes in the microwave cavity, because the microwave signal resonant frequency depends on the cavity size, and the size—on shocks and temperature. Therefore, the resonant cavity must be protected from the mechanical and thermal influence of the environment. The structure of such shielding, multilayer and made of expensive materials, is the main reason for big size and high price of clocks with active hydrogen maser.

In passive hydrogen maser the cavity volume is smaller. The sampling microwave field must be applied to induce emission, which cannot be self-sustaining. The servomechanism dedicated to control the output frequency is similar to that used in caesium atomic clocks. Passive maser is less sensitive to the environmental changes than active maser, but it also shows a worse short-term stability.

9.3.2 Rubidium Frequency Standard

Rubidium frequency standard applies the transition of atoms of rubidium 87 isotope (^{87}Rb) between ground levels characterised by magnetic moment $F = 1$ (level 1 with energy E_1) and $F = 2$ (level 2 with energy E_2). The selection of atomic energy state and the detection of transition are achieved by optical pumping.

The working principle of rubidium standard is shown in Fig. 9.10. The light from a tube containing the ^{87}Rb atoms is filtered during the transition through the filtering cell (hyperfine filter) containing pairs of rubidium ^{85}Rb isotope that form the buffer gas. The reason for using the buffer gas is a need to slow down the ^{87}Rb atoms through many elastic collisions with buffer gas atoms. Thanks to slowing down the ^{87}Rb atoms, a longer interaction time is achieved with the microwave electromagnetic field, and inelastic collisions of atoms with the cavity walls occur.

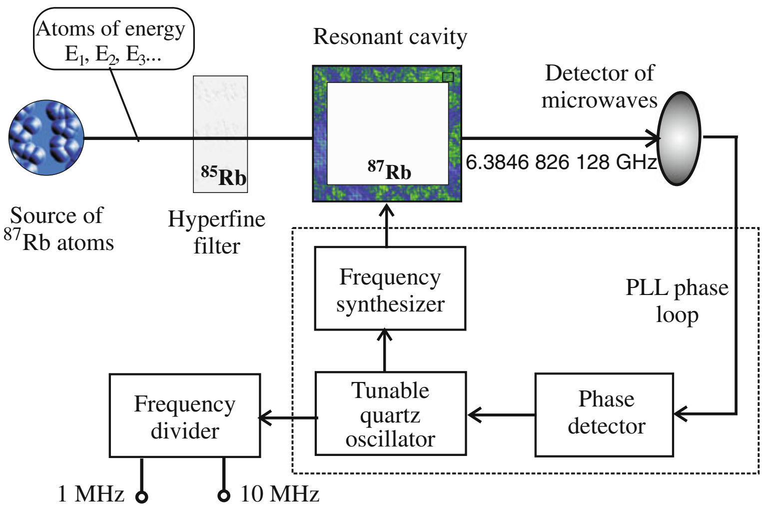


Fig. 9.10 Atomic frequency standard with rubidium gas

A cell with hyperfine filter enables the transition of only one frequency of light spectrum ν_A. Atoms with lower energy level E_1 ($F = 1$) penetrate the resonant cavity, which is the absorption cell. In the resonant cavity the number of atoms with energy E_1 decreases because the ^{87}Rb atoms absorb light radiation and, therefore, are pumped to a higher energy level E_3, and then they disintegrate, transiting onto two ground levels E_1 and E_2—Fig. 9.10. If population of level E_1 with atoms is decreased, the cell becomes transparent to radiation with frequency ν_A. Despite that, microwave sampling radiation with frequency $\nu_0 \cong 6.834\ 682\ 6109\ 043\ 24$ GHz interacts with atoms to stimulate their transitions again to level E_1. Consequently, level E_1 is again populated with a greater number of atoms and optical absorption can be launched again. In the state of resonance the microwave radiation frequency ν_P is precisely equal to ν_0. During the tuning of frequency ν_P interacting with the chamber, the signal at the photodetector output shows its maximum for the frequency of microwaves equal to ν_0. Consequently, the crystal generator VCXO is controlled according to the criterion of maximum light absorption by the electronic feedback system. The medium frequency of resonant line of Rb standard should not be considerably declined from the theoretical value ν_0 (relative change—10^{-9}), appropriate for undisturbed rubidium atoms. The deviations registered are caused by the influence of environment parameters and manifold influence of physical phenomena in the device. Because of the value of possible frequency deviations, which is by 1–2 orders of magnitude greater than in caesium or hydrogen clocks, rubidium generators are not suitable to be applied as primary frequency standards.

The generators themselves require to be calibrated by better atomic Cs generators or hydrogen masers. Rubidium generators are, in turn, suitable to be applied as secondary frequency standards. Their short-term stability is even better than that of caesium standards. After calibration the rubidium standards can reproduce frequency with inaccuracy of the order of 10^{-12} (short-term stability of $10^{-14)}$.

9.3.3 Parameters of Atomic Frequency Standards

Typical parameters of atomic frequency standards are put together in Table 9.1 (more data can be found with the manufacturers of devices). In this table the data of atomic standards are shown as: short-term stability of frequency for time interval τ given in seconds (Allan variance), linear frequency drift D (relative change per year), thermal changes and frequency accuracy during observation per 1 year.

The stability of atomic standard frequency is a function of quality factor Q of the absorption or emission process, the signal-to-noise ratio $(S/N)_{1\ \mathrm{Hz}}$ in 1 Hz band and averaging time τ [12]:

$$\sigma_y(\tau) \sim \frac{1}{Q(S/N)_{1\mathrm{Hz}}\sqrt{\tau}} \tag{9.19}$$

Table 9.1 Basic parameters of atomic frequency standards [12, 16]

Type	Rubidium standard	Cs standard	H maser standard
Resonant frequency (wavelength)	6.834682610904324 GHz (4.70 cm)	9.192631770 GHz (3.26 cm)	1.42040575177 GHz (21.1 cm)
Short-term stability (observation time τ)	5–50 × 10^{-15} for $\tau = 1$ s	1–30 × 10^{-13} for $\tau = 100$ s	2 × 10^{-13} active for $\tau = 1$ s
Linear frequency drift per 1 year	5–50 × 10^{-11}	0	1–50 × 10^{-13}
Frequency accuracy per 1 year	1–10 × 10^{-10}	Beam 5–100 × 10^{-15}	1–10 × 10^{-12}
		Fountain 1–10 × 10^{-16}	
Temperature change	1–10 × 10^{-12}/ºC	1–10 × 10^{-14}/ºC	10 × 10^{-12}/ºC
Influence of magnetic field (per 10^{-4} T)	5–20 × 10^{-12}	1–100 × 10^{-14}	1–3 × 10^{-14}

For caesium atomic clocks Q amounts to the value of 10^{10}, and for the absorption or emission of visible radiation much greater values can be achieved experimentally—of the order of 10^{17}. A limiting quality factor Q was determined theoretically for light absorption with great value $Q = 10^{23}$ [6]. The given numbers indicate possible direction of developing atomic standards—optical radiation frequency standards.

9.4 Optical Radiation Frequency Standards

9.4.1 Sources of Optical Radiation

It was found that the quality factor of selected optical radiation generators can be by 4–8 orders of magnitude (!) greater than the quality factor of caesium generators—Table 9.2. Research on such generators has been conducted, in order to build an optical radiation frequency standard with a better accuracy than in the atomic standards. The structural principle of optical radiation frequency generator is shown in Fig. 9.11.

The source of signal in an optical frequency generator is laser. Laser radiation is coherent. Laser can generate a beam with a great power density. Therefore, the laser radiation is applied as a tool to cut organic tissue (in surgery) and hard material. Laser light interacts with atoms of an element or with atomic ions, or with particles of a selected substance, ensuring selective absorption of signal with a single

Table 9.2 Parameters of microwave and optical generators of standard frequency

Clock type	Crystal	Cs beam	Cs fountain	H maser	Optical $^{199}Hg^{+}$	Optical $^{88}Sr^{+}$	Optical ^{87}Sr
ν (Hz)	$0.03–10 \times 10^{6}$	9.19×10^{9}	9.19×10^{9}	1.42×10^{9}	1065×10^{12}	448×10^{12}	429×10^{12}
Q	10^{4}	10^{8}	10^{10}	10^{9}	1.5×10^{14}	1.6×10^{11}	4×10^{17}
σ	10^{-10}	5×10^{-15}	5×10^{-16}	10^{-13}	3×10^{-15}	7×10^{-15}	1.5×10^{-14}

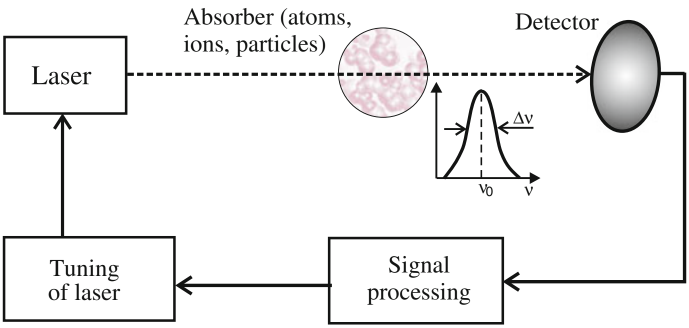


Fig. 9.11 Working principle of optical frequency standard signal generator

frequency ν_o from a narrow spectrum of laser light. The absorber can be placed in a cell, it can form a mobile beam of atoms or particles, and it can be gathered in a magneto-optical trap (ions).

Frequency ν_o of absorbed by the absorber is the demanded standard frequency. The process of selecting a labelled frequency ν_o from the laser light spectrum is called laser frequency stabilization.

Usually, emission in optical frequency generators is realised during the trapping of electrically neutral atoms or single elementary charge ions. The best and repeatable results were obtained for the following 4 frequency standards:

- mercury $^{199}Hg^{+}$ ions (ultraviolet radiation with wavelength $\lambda = 282$ nm),
- yterbium $^{171}Yb^{+}$ ions (light $\lambda = 436$ nm),
- neutral atoms of strontium ^{87}Sr (light $\lambda = 699$ nm),
- positive ions of strontium $^{88}Sr^{+}$ (light $\lambda = 674$ nm).

These 4 types of optical radiation frequency generators are recommended as *secondary standards* to reproduce the second (the fifth standard is the rubidium microwave radiation)—Recommendation CCTF 2 (2006) by the Consultative Committee for Time and Frequency (CCTF) at the General Conference on Weights and Measures [11, 12]. The parameters of those standards are set up in Table 9.3. Apart from the standards listed in the Table 9.3, also 3 other elements and isotopes are investigated as potential sources of optical radiation standard frequency. These are the standards: with indium $^{115}In^{+}$ ions (ultraviolet radiation with wavelength $\lambda = 236$ nm), with calcium $^{40}Ca^{+}$ ions (light $\lambda = 729$ nm) and with yterbium $^{171}Yb^{+}$

Table 9.3 Secondary standards of second, recommended by the consultative committee for time and frequency (CCTF) at the general conference on weights and measures (CGPM) [14]

Atom	Atomic transition	Resonant frequency/wavelength	Standard deviation
$^{199}Hg^{+}$	$5d^{10}\,6s\ ^2S_{1/2} \rightarrow 5d^9\,6s^2\ ^2D_{5/2}$	1 064 721 609 899 145 Hz/282 nm	3×10^{-15}
$^{88}Sr^{+}$	$5s\ ^2S_{1/2} \rightarrow 4d\ ^2D_{5/2}$	444 779 044 095 484 Hz/674 nm	7×10^{-15}
$^{171}Yb^{+}$	$6s\ ^2S_{1/2} \rightarrow 5d\ ^2D_{3/2}$	688 356 979 309 308 Hz/436 nm	9×10^{-15}
^{87}Sr	$5s^2\ ^1S_0 \rightarrow 5s\,5p\ ^3P_0$	429 222 004 229 877 Hz/699 nm	1.5×10^{-14}
^{87}Rb	–	6 834 682 610.904 324 Hz 47 mm	3×10^{-15}

ions (light $\lambda = 467$ nm) [6]. The latter standard uses a different atomic transition ($^2S_{1/2} \rightarrow {}^2F_{7/2}$) than the yterbium standard mentioned in Table 9.3.

Let us notice that, for example, frequency $\sim 7 \times 10^{14}$ Hz corresponds to a selected wavelength 436 nm, emitted by yterbium. The measurement of signal with so high frequency by means of electronic pulse counting is impossible, because there are no sufficiently fast electronic systems (circuits). The highest frequency of electronic system operation, equal to 70 GHz, is the frequency of photodiode of type DSC10 W/10ER, fabricated of GaAs by Discovery Semicond. To transfer the standard frequency from the light wave range 5×10^{14} Hz onto the range—first microwave (10^{10} Hz, caesium standard frequency), then to a range of 1–10 MHz (timing signal from generator VCXO)—is a very difficult task because the optical standard frequency is no less than by 8 orders of magnitude higher than the VCXO frequency.

9.4.2 Optical Frequency Comb

Two methods are applied to compare the optical standard frequencies with the caesium clock frequency. The first method is an application of the frequency multiplier chain (frequency chain) and oscillators controlled by the multipliers: lasers, masers and microwave generators (altogether, 8–20 oscillators). Non-linear diodes and non-linear crystals were used to multiply signals with highest frequencies. The frequency of signal emitted by each oscillator must be precisely matched to a concrete optical radiation frequency, which we want to compare with the Cs standard. Therefore, frequency multiplication is a complicated and expensive method (it requires several lasers in the measurement chain. The PTB laboratory in Braunschweig (Germany) performed such comparative system, in which an optical standard with ^{40}Ca atoms with frequency $\nu_{Ca} = 455\ 986\ 240\ 494\ 159$ Hz was compared with a caesium standard $\nu_{Cs} = 9\ 192\ 631\ 770$ Hz. Twenty oscillators were used in the PTB measurement chain. Two other systems of multiplier chains were set up and started in the NIST (USA) and LNE (France) laboratories.

The other method for comparing frequency standards is to apply the system of optical frequency comb. This method for measuring and reproducing standard frequency is so significant for science that the author of optical frequency comb, Theodor Hänsch, was awarded the Nobel Prize for that achievement in 2005 [7]. The co-author of optical frequency comb was T. Udem. The optical frequency comb is obtained with the use of a laser generating pulses with duration of the order of femtoseconds. Due to great significance of optical frequency comb in precise metrology, its working principle is briefly discussed below and illustrated in Fig. 9.12.

We assume that a continuous function laser generates light with frequency ν_s. The frequency spectrum of this signal is illustrated by one spectral line in the spectral-response characteristic—Fig. 9.12 (upper part). When the laser signal has 3 frequency components: ν_s, $\nu_s + \Delta\nu$, $\nu_s - \Delta\nu$, the spectral characteristic contains 3 lines—Fig. 9.12 (middle part). As a result of interference, the laser signal is emitted in the form of sinusoidal waveform with modulated amplitude, with the amplitude maximum in an interval $\Delta t = 1/\Delta\nu$ on the time scale. When there are n component frequencies of the signal emitted by the laser, and they are shifted by $\Delta\nu$, the signal spectrum contains n spectral lines separated by $\Delta\nu$ on the frequency scale. Then, the signal is emitted by the laser in the form of brief packets of sinusoidal pulses in an interval $T_p = 1/\Delta\nu$. The spectrum of such signal resembles a comb, which was used in the name of the system—optical (or laser) frequency comb.

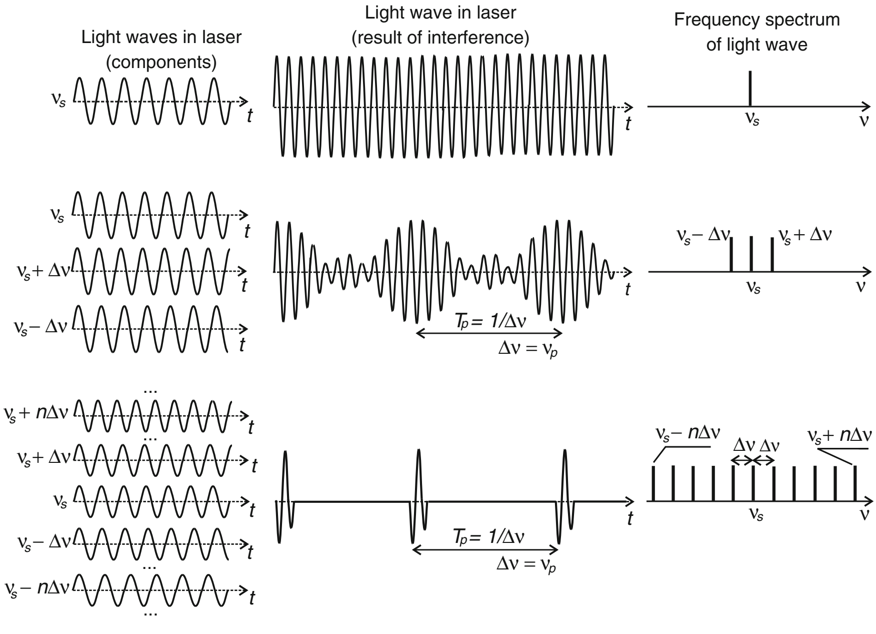


Fig. 9.12 Signal frequency spectrum in optical frequency comb

Now, we reverse the problem. Laser emits light (with wavelength λ and frequency ν_s) in the form of packets of sinusoidal pulses, repeating the packet emission with period T_p.

The duration of a packet of pulses is considerably shorter than the repetition time T_p. The frequency spectrum of such signal contains many lines (up to a million) in the following points of frequency scale: ν_s, $\nu_s \pm \Delta\nu$, $\nu_s \pm 2\Delta\nu$, …, $\nu_s \pm n\Delta\nu$, where $\Delta\nu = 1/T_p = \nu_p$ is the frequency of packet emission repetition. The repetition period T_p can be determined in a range 10^{-10}–10^{-9} s, which corresponds to the repetition frequency ν_p = 1–10 GHz. A signal with such frequency can be measured very accurately, comparing it with the caesium standard. Therefore, the femtosecond laser emits a signal with a spectrum in which the spectral lines are separated by frequency ν_p, measured very precisely and accurately. To determine the pitch on the comb frequency scale, a starting point of the scale is necessary—frequency ν_o, considerably lower than the light frequency ν_s. Frequency ν_o is the structural constant of the laser, and it is found in the GHz range; it can be measured very accurately by comparing it with the caesium standard frequency. By means of frequency comb we can determine the values of ν on the frequency scale:

$$\mathrm{\nu = \nu_o + n\nu_p} \tag{9.20}$$

where: ν_o—laser structural constant, n—number of a “tooth” of frequency comb, $\nu_p = \Delta\nu$—interval between the pitch points of the comb (“teeth”).

It is a technological problem in the optical frequency comb system to build and set working the femtosecond laser, i.e. a laser emitting light in the form of pulses with the packet length of the order of 10–40 fs. Let us notice that the light wave frequency, e.g. green light, amounts to $\nu_s = 5 \times 10^{14}$ Hz, and the wave period $T_s = = 1/\nu_s = 2$ fs. A packet of pulses emitted by the femtosecond laser comprises only 5 to about 20 light wave periods. Currently, the parameters of femtosecond lasers are still being improved.

Optical frequency comb is used to accurately determine the light radiation frequency, first of all to determine the optical radiation standard frequency—Fig. 9.13.

Light from the investigated standard and light from the frequency comb is directed onto a light detector. The beat of the investigated signal (frequency ν_{Sr}) occurs in the detector, and the signal of the m-th spectral line of the comb spectrum (Fig. 9.13). The beat frequency is stabilized in a phase locked loop PLL2 system by tuning the signal from the comb. Another phase loop PLL1 is included in the system, and its task is to stabilize frequency ν_o—the starting point of the comb scale. The optical atomic clock output signal is frequency ν_r, proportional to the investigated frequency ν_{Sr} but much lower:

$$\nu_r = \frac{\nu_{Sr}}{m + \alpha + \beta} \tag{9.21}$$

Frequency ν_r is found in the operation range of electronic systems, and it can be applied for example to measure the standard time interval (1 s) or for precise timing.

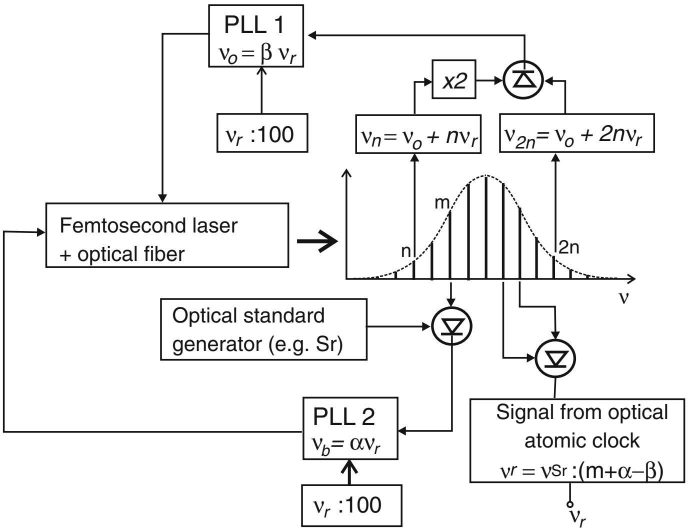


Fig. 9.13 Optical atomic clock with frequency comb for standard comparison

9.5 Time Scales

When we intend to determine an arbitrary time scale, we should select a periodic phenomenon, count its passing periods and point out indicate (display) points on the time scale. A time scale should have a starting point, determined conventionally. The specified conditions for determining a time scale: selection of a phenomenon (arbitrary), method for its measurement and determination of starting point, lead to a conclusion that there can be a number of time scales. An example of time scale is the Julian Calendar (Old Style) with the scale starting point set for the 1 day of January 4713 B.C. A demanded property of a time scale is linearity, thanks to which long time slices must be both measured and indicated on the scale. This feature is called “uniformity of time scale”, and it is achieved through ideal stability of the frequency of a phenomenon observed.

Some time scales were determined by means of astronomical measurements, e.g. UT Universal Time, Sidereal Time, Mean Solar Time or Ephemeris Time. The latter scale was the basis for defining the second in the international units of measures system in the years 1956–1967. The sidereal day (for Sidereal Time) is determined as the time of the Earth’s rotation around its axis, and it is measured as a time interval between successive “transitions” of a selected star through a local meridian. The period of sidereal day is more stable than the time of solar day, but the sidereal day is longer than a mean solar day by almost 24 s [1]. For human existence the solar time is more important than the sidereal time. The properties of the scales of both Sidereal Time and Mean Solar Time were determined by defining the Universal Time scale in three versions: UT0, UT1 and UT2. The UT scale measures the sidereal time in a given place on the Earth, converted onto mean solar time and referred to meridian 0 (Greenwich). The result of those measurements and calculations is UT0—Universal Time scale, version zero. After introducing corrections due to the position of geographical poles into UT0, a UT1 scale was determined, valid for on the whole Earth. After introducing into UT1 corrections due to the instability of the Earth's rotations in the annual cycle, we obtain the UT2 time scale.

From among known physical or astronomical phenomena, atomic energy state transitions (atomic transitions) are the effect that meets the criterion of high frequency stability in the best way. An atomic time scale was determined—TA (French: *temps atomique*); and it can be easily understood that the best atomic standards were used to count the atomic time; those standards are currently the caesium clocks. The result of international comparisons of atomic clocks was an *International Atomic Time scale TAI* (French: *Temps atomique international*). The international atomic time TAI is a statistical time scale, build on the basis of data sent from metrological centres (over 50 institutes from all over the world disposing of almost 300 atomic clocks) to BIPM, using one of two possible satellite communication tools: GPS or TWSTFT (two-way satellite time and frequency transfer). One of the centres is the Central Office of Measure (GUM) in Warsaw, represented by its Laboratory of Time and Frequency. Some research institutions, disposing of the best atomic clocks, play a leading role in generating the TAI time scale: NIST, USNO (U.S. Naval Obsevatory), PTB, NPL, NICT and Paris Observatory. The starting point of atomic time scale was conventionally set for the 1 day of January 1958, 0 h, 0 min, 0 s according to the UT2 scale [8].

Coordination of time scales and the determination of international time scale is the role of the International Time Bureau in Paris (BIH—French: *Bureau International de l`Heure*), an agenda of BIPM in Sèvres. As a result of multilateral agreements, the Coordinated Universal Time (UTC) scale was determined. The UTC scale is realised based on TAI. The UTC time is measured by atomic clocks (the same that measure TAI), but, moreover, the UTC indications must be in accordance with the UT1 astronomical scale. Because of the difference between the atomic time and the astronomical time, the UTC scale, just at its start point on 1 January 1972, was shifted in relation to the atomic time: UTC − TAI = −10 s.

Each laboratory of time contributing in creation of TAI, keeps own UTC(X) scale on basis on its atomic clocks, but periodically corrected in relation to the global UTC scale. Thus, in the U.S. there are UTC(NIST) and UTC(USNO), in Germany is UTC(PTB), in Poland UCT(PL) etc.

Coordination of the UTC scale in respect to astronomical time is made every few years by adding one second at the end of June or December. The decision of adding the leap second (additional second) to the UTC scale is recommended by the International Earth Rotation and Reference Systems Service (IERS). The last leap second was introduced into the UTC on 30 June 2012, when the last minute of the June 2012 in the UTC scale was extended to 61 s [8]. The UTC is the most important international time scale; therefore, it is used on many areas of science, technology, and navigation. The UTC scale is both created and corrected on line. We emphasize the fact that both UTC and TAI are the atomic clocks scales.

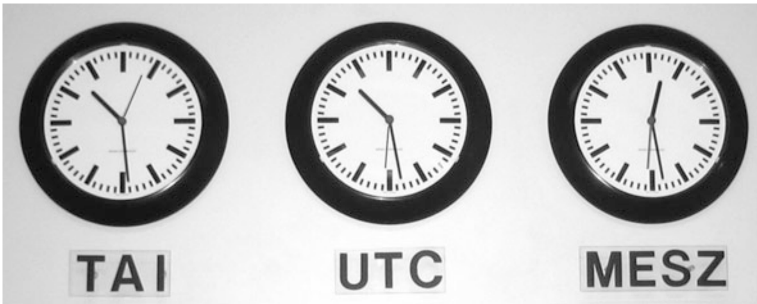


Fig. 9.14 The readings of clocks in Physikalisch-Technische Bundenanstalt in Braunschweig (Germany), counting the TAI time (*temps atomique international*), UTC (*coordinated univesal time*), and the middle-European summer time (*Mitteleuropäische Sommerzeit*—MESZ), 2007

The UTC indications differ from the TAI indications by an integral number of seconds. From July 2012 the difference is equal to −35 s (9.22)—Fig. 9.14:

$$\mathrm{UTC} - \mathrm{TAI} = -35\ \mathrm{s} \tag{9.22}$$

The International Earth Rotation and Reference Systems Service announced in November 2014 that no leap second will be introduced in 2014. The TAI time is counted uniformly, without considering the irregular Earth's rotations. The UTC scale takes into consideration the astronomical time, according to the actual Earth's position towards stars and the Sun. The organization and propagation of the TAI and UTC scales is the function of the BIH office.

The last time scale worth to mention is the GPS time scale (GPS-T). Very significant in the metrology of time are also satellite positioning systems: American Global Positioning System (GPS), Russian GLONASS, Chinese Bei-Dou and the Galileo system, designed by the European Union. The satellite positioning systems are described in Chap. 10.

Satellite positioning systems are—on the one hand—important clients of the world time service, and on the other hand—they participate formally in dissemination of time (propagating time scales), so far GPS only, or they are recommended for that task (GLONASS and Galileo)—Recommendation CCTF 5 (2006) by the Consultative Committee for Time and Frequency (CCTF) at the General Conference on Weights and Measures (CGPM) [14].

The GPS system creates and keeps its own time scale—the GPS time (GPS-T). Unification of time in the whole GPS system is crucial for its operating ability and high accuracy. Several atomic clocks are installed, mainly rubidium clocks and hydrogen masers, on a board of each satellite of GPS. Satellites send messages with information of their "on board" atomic time. They receive the correction signals according to the UTC(USNO) scale from the U.S. Naval Observatory via GPS center in Colorado Springs—Fig. 9.15. By this way the GPS time scale is created. Satellites send navigation messages containing time data and the (GPS-T − UTC)

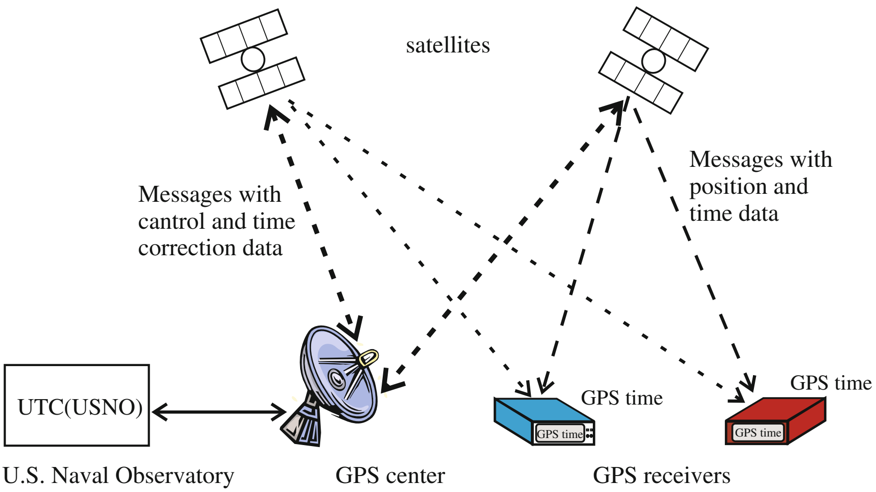


Fig. 9.15 GPS time dissemination

Table 9.4 Conversion between time scales (2014)

To time scale →	From time scale ↓		
	UTC	TAI	GPS-T
UTC	–	−35 s	−16 s
TAI	+35 s	–	+19 s
GPS-T	+16 s	−19 s	–

time correction. GPS time was set to match UTC scale in 1980, when the difference: (TAI − UTC) = 19 s, thus the difference (TAI − GPS-T) was the same, equal to 19 s. Since that year the (TAI − GPS-T) difference remains 19 s—Table 9.4.

The accuracy of realization of the GPS time scale is better than 100 ns, but at high quality receiver it can reach 20 ns [16]. An access to GPS time have all users of GPS receivers. For a measure of a position of an receiver signals from four satellites (at least) are necessary. But for read out of GPS time communication with only one satellite is sufficient, when the receiver has a known and fixed position. Time dissemination in GPS system is very useful for many institutions and companies.

9.6 National Time and Frequency Standard in Poland

Polish national time and frequency standard is installed and maintained at the Laboratory of Time and Frequency, Central Office of Measures (GUM) in Warsaw. The main tasks of the Laboratory are [13]:

Fig. 9.16 Cesium atomic clock in GUM in Warsaw

- Reproducing, maintaining and transmitting time and frequency units.
- Determining standard time in the Republic of Poland and its propagation over the country.
- Generating the UTC (PL) time scale—Polish realisation of the UTC.
- Participating in the creation of international time scales TAI and UTC.
- Calibration, research and expertise of measurement instruments.

The national frequency standard in the Laboratory of Time and Frequency, Central Office of Measures in Warsaw is a set of three Cs-beam clocks of type HP5071A (by Hewlett-Packard) (Fig. 9.16) and a hydrogen maser.

The standard generates a national standard time scale UTC (PL). The UTC (PL) scale participates in international comparisons. The atomic clocks in GUM participate in creating the Universal Time Coordinated UTC. The uncertainty of reproducing of the national time and frequency standard is equal to 2.2×10^{-14} [13]. Moreover, a hydrogen maser frequency standard is installed in the Laboratory.

To ensure the best possible stability and absolute operational continuity of the national standard, the atomic standards:

- are maintained in stable environment,
- have an emergency power generator,
- are compared continuously with standards in other national and international laboratories.

The Laboratory of Time and Frequency, GUM, propagates the standard time signals in Poland through:

- radio, broadcasting acoustic time signals in the Polish Radio programme,
- WAN network by means of two time servers with names (addresses): tempus1.gum.gov.pl and tempus2.gum.gov.pl,
- public telecommunication network by means of telephone modem; the network makes available digital time signals coded according to the European Telephone Time Code.

References

1. D.W. Allan, Statistics of atomic frequency standards. Proc. IEEE **54**, 221–230 (1966)
2. A. Bauch, Caesium atomic clocks: function, performance and applications. Meas. Sci. Technol. **14**, 1159–1173 (2003)
3. S. Bregni, *Synchronization of Digital Telecommunications Networks* (Wiley, Chichester, 2002)
4. S. Chu, *The Manipulation of Neutral Particles. Nobel Lectures* (World Scientific, Singapore, 1996–2000), pp. 122–158
5. R.H. Feynman, R.B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 3 (Addison-Wesley, Reading, 1964)
6. P. Gill et al., Trapped ion optical frequency standards. Meas. Sci. Technol. **14**, 1174–1186 (2003)
7. T.W. Hänsch, *Passion for Precision. Nobel Lectures* (World Scientific, Singapore, 2001–2005), pp. 485–509
8. Information on UTC—TAI, Bulletin C 48, International Earth Rotation and Reference Systems Service, July 2014, Observatoire de Paris
9. P. Kartaschoff, *Frequency and Time* (Elsevier, New York, 1978)
10. M. Kumagai, H. Ito, M. Kajita, M. Hosokawa, Evaluation of caesium atomic fountain NICT-CsF1. Metrologia **45**, 139–148 (2008)
11. A. Piekara, J. Stankowski, S. Smolińska, J. Galica, Ammonia maser of poznan research center (in Polish). Postępy Fizyki **15**, 565 (1964)
12. K. Szymaniec, W. Chałupczak, P.B. Whibberley, S.N. Lea, D. Henderson, Evaluation of the primary frequency standard NPL-CsF1. Metrologia **42**, 49–57 (2005)
13. http://bip.gum.gov.pl
14. http://www1.bipm.org/en/
15. http://tf.nist.gov
16. http://tycho.usno.navy.mil

Chapter 10
Standards and Measurements of Length

Abstract This chapter begins with a discussion of the evolution of the international definition of the meter, from that adopted by the Meter Convention in 1875 to the current definition relating the unit of length with the speed of light. We present three methods for the realization of the definition of the meter as recommended by the International Committee for Weights and Measures (CIPM). We discuss the design, operation and parameters of the iodine-stabilized helium-neon (He-Ne) laser with a wavelength of 633 nm, the most important device in metrology of length. We provide and discuss a table of sources of reference radiation other than the He-Ne laser recommended by CIPM. We also provide a brief overview of three global satellite systems for navigation and positioning, GPS, GLONASS and Galileo (under development), and point out the differences between them. Finally, we have a cursory look at three regional navigation systems, the Chinese BeiDou, the Indian IRNSS, and the Japanese QZSS.

10.1 Introduction

The unit of length, the meter was first defined in the French Republic in 1791 as one ten-millionth of the length of the Earth's meridian passing through Paris along the quadrant from the North Pole to the equator (see Chap. 2, Sect. 2.4). A physical prototype of the meter in the form of a platinum bar was made and deposited in the Archives of the French Republic in 1799. A definition of the meter based on the French experience was adopted by the Meter Convention in 1875. Since that time, in a 130 years' period, the meter was redefined three times. Below we present the successive definitions [1].

> One meter is the distance between the middles of the width of the main lines in the zone delimited by the directional lines on the international prototype at the temperature of 0 °C, supported at points at a distance of 0.22L from the ends, where L is the length of the prototype.

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_10

This definition was adopted by the 1st General Conference on Weights and Measures (CGPM) in 1889. A bar of a platinum-iridium alloy (Pt90-Ir10) was chosen as the physical prototype of the standard unit of length. By the next definition, adopted by the 9th CGPM in 1960:

> The meter is equal to 1 650763.73 wavelengths in vacuum of light corresponding to the transition between the levels 2p and 5d of the krypton 86 atom.

Finally, in 1983 the 17th CGPM adopted the following definition, which has remained in force to date:

> The meter is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.

As indicated by the definition of 1889, the natural standard of length (one ten-millionth of a quarter of the Earth's meridian) had to be replaced with its practical realization in the form of a platinum-iridium bar. Although for users this prototype of the meter was as arbitrary as its predecessors, its advantage was that in case of damage or loss it could be reproduced from the time-invariant dimensions of the Earth.

The definition of the meter of 1960 referred to the fixed wavelength of radiation emitted or absorbed in a quantum transition between the levels $2p_{10}$ and $5d_5$ of the atom of krypton ^{86}Kr. The fixed wavelength (λ = 605.780 2103 nm) corresponds to the constant difference between the energy levels of the krypton atom. The definition did not specify, however, how to realize the number N = 1 650 763.73 of wavelengths necessary for the realization of the meter? The fixed frequency ν_0 of the transition between the specified energy levels (and the corresponding constant wavelength λ_0) can be obtained by emission of radiation in a discharge lamp using krypton 86, by absorption of radiation in a cell filled with krypton 86, or with an oscillator with a beam of ^{86}Kr atoms. Of these techniques only the method using a discharge lamp, specially designed for this purpose, has been developed technologically and used in metrological laboratories. In 1960 a standard of the meter with a krypton discharge lamp allowed the realization of the SI unit of length with an uncertainty 10 times better than that offered by the previous standard (the platinum-iridium prototype). At that time the krypton standard was also more accurate than a laser interferometer. However, the parameters of lasers improved significantly in the following years, and the frequency of laser light was stabilized. As a result of these improvements already in the 1960s the wavelength of light emitted by certain types of lasers was determined more accurately than the reference wavelength of krypton discharge lamps.

The third definition of the meter, adopted in 1983 and still in force today, includes the assumption that the speed of light in vacuum is exactly known, equal to 299 792 458 m/s and constant in time. Even if a better determination of the speed of light were obtained, the updated value could not be used for the realization of the unit of length without formal amendment of the definition of the meter by the CGPM. The stability of the value of the speed of light is the subject of debate. It was questioned by João Magueijo, Professor of theoretical physics at the Imperial

College in London, who is of the opinion that at the beginning of the universe the speed of light was higher than at present [2].

10.2 Realization of the Definition of the Metre

10.2.1 CIPM Recommendations Concerning the Realization of the Metre

Along with the adoption of the new definition of the meter by the CGPM in 1983 the International Committee for Weights and Measures (CIPM) issued a recommendation concerning the practical realization of the unit of length by this definition. The recommendation was updated in 1992, 2003 and 2005 to take account of new scientific and technological achievements, especially the development of new lasers (Recommendation 1, 2003), in the metrology of length [3]. By this recommendation the meter should be realized by one of the following methods:

1. By means of the length l of the path travelled in vacuum by an electromagnetic plane wave in a time t; the length l is obtained from the measured time t using the relation:
$$l = ct,$$
where c = 299 792 458 m/s is the speed of light in vacuum.
2. By means of the wavelength λ in vacuum of an electromagnetic plane wave of frequency v; the wavelength λ is obtained from the measured frequency v using the relation:
$$\lambda = c/v,$$
where v is the frequency of the wave to be measured, and c is the speed of light in vacuum.
3. By means of one of the radiations from a provided list (Table 10.1); the stated wavelength in vacuum or stated frequency of these radiations can be used with the specified uncertainty on the condition that the given specifications and accepted good practice are followed.

The meter is realized accurately by the methods (2) and (3), but not necessarily by the method (1). The use of this latter technique should be restricted to lengths sufficiently short for the effects predicted by the general theory of relativity to be negligible with respect to the uncertainty of realization.

The realization of the unit of length by the method (2) recommended by the CIPM [3] involves the determination of the wavelength λ of the reference electromagnetic wave. The wavelength λ should be determined by measuring the frequency v of the reference electromagnetic wave and using the relation $\lambda = c/v$, where c is the speed of light in vacuum. Measurements of the frequency or wavelength of optical waves as well as ultraviolet (UV) and infrared (IR) radiation are costly and difficult. The encountered difficulties are discussed in detail in Sect. 9.4, devoted to optical frequency standards. It is impossible to count pulses

Table 10.1 Reference optical wavelengths recommended by the International Committee for Weights and Measures (CIPM) for the realization of the unit of length [3]

No.	Atom, ion or molecule	Wavelength (nm)	Frequency (kHz)	Relative uncertainty
1	Molecule of acetylene $^{13}C_2H_2$	1 542.383 712 38	194 369 569 384	2.6×10^{-11}
2	Atom of calcium ^{40}Ca	657.459 439 291	455 986 240 494.150	1.8×10^{-14}
3	Molecule of methane CH_4	3 392.231 397 327	88 376 181 600.18	2.3×10^{-11}
4	Atom of helium ^{1}H	243.134 624 626 04	1 233 030 706 593.55	2.0×10^{-13}
5	Ion of mercury $^{199}Hg+$	281.568 67 5919686	1 064 721 609 899.145	3×10^{-15}
6	Molecule of iodine $^{127}I_2$, transition P(13) 43-0	514.673 466 368	582 490 603 442.2	2.6×10^{-12}
7	Molecule of iodine $^{127}I_2$, transition R(56) 32-0	532.245 036 104	563 260 223 513	8.9×10^{-12}
8	Molecule of iodine $^{127}I_2$, transition R(106) 28-0	543.515 663 608	551 580 162 400	4.5×10^{-11}
9	Molecule of iodine $^{127}I_2$, transition P(62) 17-1	576.294 760 4	520 206 808 400	4×10^{-10}
10	Molecule of iodine $^{127}I_2$, transition R(47) 9-2	611.970 770 0	489 880 354 900	3×10^{-10}
11	Molecule of iodine $^{127}I_2$, transition R(127) 11-5	632.991 212 58	473 612 353 604	2.1×10^{-11}
12	Molecule of iodine $^{127}I_2$, transition P(10) 8-5	640.283 468 7	468 218 332 400	4.5×10^{-10}
13	Ion of indium $^{115}In+$	236.540 853 549 75	1 267 402 452 899.92	3.6×10^{-13}
14	Krypton ^{86}Kr (discharge lamp)	605.780 210 3	494 886 516 460	1.3×10^{-9}
15	Molecule of osmium oxide OsO_4	10 318.436 884 460	29 054 057 446.579	1.4×10^{-13}
16	Atom of rubidium ^{85}Rb	778.105 421 23	385 285 142 375	1.3×10^{-11}
17	Ion of strontium $^{88}Sr+$	674.025 590 863 136	444 779 044 095.4846	5×10^{-15}
18	Atom of strontium ^{87}Sr	698.445 709 612 694	429 228 004 229.910	2×10^{-13}
19	Ion of ytterbium ^{171}Yb, transition 6s $^2S_{1/2} \rightarrow$ 5d $^2D_{3/2}$	435.517 610 739 688	688 358 979 309.308	9×10^{-15}
20	Ion of ytterbium ^{171}Yb, transition $^2S_{1/2} \rightarrow {}^2F_{7/2}$	466.878 090 060 7	642 121 496 772.3	1.6×10^{-12}
21	Spectral lamp with ^{86}Kr, ^{198}Hg or ^{114}Cd vapor. Parameters of reference signal source are provided in Table 10.2			

with a frequency of the order of 10^{14}–10^{15} Hz (the frequency range of optical waves), since such high-frequency signals cannot be processed by the available digital electronics. The fastest electronic devices, based on gallium arsenide, can process signals with a frequency of up to 70 GHz. Frequency measurements in the range from 10^{14} to 10^{15} Hz require a measurement chain composed of non-optical atomic oscillators, or an optical frequency comb (discussed in Chap. 9). Due to the very high cost of the equipment and the complexity of the measurement process this method for the measurement of the frequency of optical signals has only been used in a few laboratories in the world, including the PTB, NIST, NPL and LNE (France).

The third method for the realization of the unit of length recommended by the CIPM is based on atomic or molecular transitions in atoms, ions or particles, used for stabilizing the frequency of electromagnetic waves. The wavelength of the electromagnetic radiation selectively absorbed or emitted by the atoms, ions or particles is determined as accurately as the frequency of the wave. The elements and compounds used in this method, along with the corresponding frequencies of their atomic or molecular transitions, are listed in Table 10.1, in the same order as published by the International Bureau of Weights and Measures (www.bipm.org).

Table 10.1 specifies the wavelength λ of the reference signal, the wave frequency ν and the achievable relative uncertainty in the determination of λ or n. The preferred wavelength range extends from ultraviolet (243 nm for ^1H) to optical and far infrared (10 μm for OsO_4) radiation.

The CIPM recommendation specifies the conditions necessary to obtain a signal with the required parameters. For example, as regards the reference radiation obtained with acetylene (item 1 in Table 10.1), the radiation source is a laser with signal stabilized as it passes through the cell. The cell contains acetylene C_2H_2 at a pressure of (3 ± 2) Pa and the radiation beam has a power density of (25 ± 20) W/cm^2. Acetylene absorbs selectively light of the specific frequency of 194 369.5693 84 GHz from the multiple frequencies in the laser light spectrum. A relative standard uncertainty of 2.6×10^{-11} can be achieved in measurements of frequency and wavelength with acetylene-stabilized signal.

It is recommended to detect the third harmonic of the signal in order to enhance the required maximum at $\nu \approx 194\,369$ GHz in the spectral characteristics of infrared radiation.

The most widely used of the recommended sources of reference radiation, the helium-neon (He-Ne) laser (item 11 in Table 10.1) stabilized by iodine I_2, emitting light with a wavelength of 633 nm, is discussed in more detail in Sect. 10.3.

As an extension of item 21 in Tables 10.1, 10.2 provides the wavelengths of light emitted in different quantum transitions of krypton, mercury or cadmium atoms in a spectral lamp containing vapor of these elements.

Note that here the relative uncertainty of realization of the wavelength is of the order of 10^{-8}, a few orders of magnitude larger than the uncertainty achieved with the radiation sources listed in Table 10.1. However, even the achievement of this worse uncertainty requires operation conditions as specified in the CIPM recommendation.

For example, for the spectral lamp with krypton vapor (items 1–4 in Table 10.2), emitting visible light, the following specification is provided: discharge lamp, with a hot cathode, comprising gaseous krypton of at least 99 % purity in an amount necessary for the presence of solid krypton at a temperature of 64 K. The lamp should contain a capillary tube of a diameter of 2–4 mm and with a wall ca. 1 mm thick. The lamp is assumed to emit radiation with a wavelength corresponding to

Table 10.2 Transitions in the spectral lamp recommended as a source of reference radiation by the International Committee for Weights and Measures (CIPM) for the realization of the unit of length [3]

No.	Atom	Quantum transition	Wavelength λ in vacuum (nm)	Expanded relative uncertainty, $U = ku_c$, $k = 3$
1	^{86}Kr	$2p_9 \rightarrow 5d_4$	654.807 20	2×10^{-8}
2	^{86}Kr	$2p_8 \rightarrow 5d_4$	642.280 06	
3	^{86}Kr	$1s_3 \rightarrow 3p_{10}$	565.112 86	
4	^{86}Kr	$1s_4 \rightarrow 3p_8$	450.361 62	
5	^{198}Hg	$6\ ^1P_1 \rightarrow 6\ ^1D_2$	579.226 83	5×10^{-8}
6	^{198}Hg	$6\ ^1P_1 \rightarrow 6\ ^3D_2$	577.119 83	
7	^{198}Hg	$6\ ^3P_2 \rightarrow 7\ ^3S_1$	546.227 05	
8	^{198}Hg	$6\ ^3P_1 \rightarrow 7\ ^3S_1$	435.956 24	
9	^{114}Cd	$5\ ^1P_1 \rightarrow 5\ ^1D_2$	644.024 80	7×10^{-8}
10	^{114}Cd	$5\ ^3P_2 \rightarrow 6\ ^3S_1$	508.723 79	
11	^{114}Cd	$5\ ^3P_1 \rightarrow 6\ ^3S_1$	480.125 21	
12	^{114}Cd	$5\ ^3P_0 \rightarrow 6\ ^3S_1$	467.945 81	

the frequency of transitions between unperturbed energy levels of ^{86}Kr atoms with an uncertainty of 10^{-8} (the reference wavelengths of the four atomic transitions in ^{86}Kr are specified in Table 10.2). The agreement to 10^{-8} between the wavelength of the emitted radiation and the wavelength calculated theoretically can be achieved by fulfilling the following conditions regarding the spectral lamp:

- The capillary tube (the light-emitting part of the lamp) is observed from the side of the anode;
- The lower part of the lamp, together with the capillary tube, is immersed in a cold bath of the temperature of the triple point of nitrogen, maintained with an accuracy of 1 K;
- The density of the electric current in the capillary is of (0.3 ± 0.1) A/cm^2.

Note that the krypton transitions listed in Table 10.2 do not include the 605.8 nm transition in ^{86}Kr. This transition figures in Table 10.1 (item 14). Its uncertainty of realization is much better than that of any of the ^{86}Kr transitions listed in Table 10.2.

The recommendations of the International Conference on Weights and Measures include four important final remarks [3]:

- The values of frequency v and of wavelength in vacuum λ of the listed radiations recommended for practical realization of the meter should exactly fulfill the relation $\lambda \times v = c$, with c = 299 792 458 m/s and the λ values rounded.
- It should be considered that a number of independent values can be obtained with different radiations, and the estimated uncertainties may not include all the sources of variation of the results.

- Some of the listed radiations can be replaced, without affecting the accuracy, by a radiation corresponding to another component of the same transition or by another radiation, when their frequency difference is known with sufficient accuracy.
- It should also be noted that meeting the technical specifications is not sufficient for the achievement of the specified uncertainties. It is also necessary to follow the best practice concerning methods of frequency stabilization, described in numerous scientific and technical publications.

The latter remark is of a more general character. To allow measurements with the best accuracy and the best resolution, quantum metrology, besides the appropriate instrumentation, requires experience and talent of the experimenter.

10.2.2 Measurements of Length by the CIPM Recommendation

In astronomy and global positioning systems (GPS, GLONASS and Galileo) length is determined by the measurement of time. Satellites used by positioning systems (discussed in Sect. 10.4) orbit the Earth at a height H of about 20 000 km above the Earth's surface. The time t necessary for an electromagnetic wave to cover this distance is $t = H/c = 2 \times 10^7$ m/3 $\times$ 10^8 m/s = 67 ms. Such a time interval can be easily measured with an uncertainty of ±1 ns, corresponding to a relative uncertainty of the order of 10^{-8}. A distance of 1 m is covered by an electromagnetic wave in vacuum or air in 3.3 ns. Time intervals of this order are too short to be measured with the required uncertainty of 10^{-10} or better by currently unavailable techniques. The other two methods for the measurement of length involve the use of an interferometer.

The length l measured with an interferometer represents a multiple of a half, quarter, etc., of the wavelength λ used: $l = N \times \lambda/2$, $N \times \lambda/4$, $N \times \lambda/8$, … , depending on the type of interferometer. The difference between the second and third methods (both are recommended by CIPM [3]) is that in the second method previous determination of the wavelength in vacuum is necessary (λ = ?), whereas in measurements by the third method the reference wavelength λ is known.

In practice length can be measured by various methods using interference of light. Interferometers for length measurements can be divided into three types:

- Interferometers with a movable mirror, which use only one (known) reference optical wavelength and do not require preliminary estimation of the length of the measured object;
- Interferometers based on the exact fraction method, using two or more reference wavelengths non-simultaneously; these interferometers do not need moving parts, but require a preliminary estimation of the length of the measured object;

- Interferometers using simultaneously two or more reference wavelengths of light; these interferometers have not moving parts, but require a preliminary estimation of the length of the measured object.

In measurements with an interferometer of the first type the length l is determined by counting the number m of fringes in the obtained interference pattern. The fringes shift as the mirror is moved along the path l between points x_1 and x_2: $l = x_1 - x_2$. The measured length l is given by:

$$l = m\lambda/2, \tag{10.1}$$

where m is the number of fringes, and λ denotes the wavelength of light in air, determined from the vacuum wavelength by taking into account the air pressure, temperature and humidity as well as the concentration of carbon dioxide.

Interferometers of the second type use two or more stable-frequency sources of light (e.g. lasers or spectral lamps) with different wavelengths. Length is determined by measuring the difference in the mutual position of interference fringes resulting from reflection by the surface of the measured object and a reference plane, using successive wavelengths. The interference patterns lead to a system of equations for the length of the measured object.

Interferometers of the third type use simultaneously two (or more) sources of light with different wavelengths λ_1 and λ_2 for illuminating the measured object. Fringes are observed on a screen as a result of interference of optical waves with the wavelengths λ_1 and λ_2 having traveled the distance l. The spacing λ between the fringes is $\lambda = \frac{\lambda_1 \lambda_2}{2(\lambda_1 - \lambda_2)}$. Thus, the measured length l is given by the formula:

$$l = m\frac{\lambda_1 \lambda_2}{2(\lambda_1 - \lambda_2)}, \tag{10.2}$$

where m is the number of fringes, and λ_1 and λ_2 denote the air wavelengths of light from the two lasers.

Interferometers of the first and second types are the most commonly used in metrology of the length. In laser interferometers the counting of fringes is performed by an image analyzer and a software. The user reads the measurement result on a display. In more advanced interferometer systems the measurement result, along with additional information, is presented on the computer screen, as shown in Fig. 10.1. Results can be also presented in a graphical form after data processing, such as statistical processing, filtering, etc.

Also diode lasers (semiconductor lasers), used as a light source, are of some practical importance in metrology. In contrast to gas lasers, diode lasers do not require a high-voltage power supply, and have the additional advantage of being much smaller and easier to use. Also the high accuracy diode lasers require stabilization of signal in absorption cells containing gas or vapor at low pressure. Interferometers (distance meters) with a semiconductor laser are applied in industry and building [4] as well as in science and high technology manufacturing

Fig. 10.1 Screenshot with the result of a length measurement ($l = 2600.742$ mm) with a resolution of 1 µm performed with an HP 5529A laser interferometer at the Central Office of Measures in Warsaw

(semiconductor industry) [5]. Numerous companies (Hewllet-Packard, Canon, Fluke, Keysight, Zygo, etc.) offer laser interferometers with a measuring range of 20–100 m with the typical accuracy of 1 mm and the best accurate of 1–50 µm (at the range up to 10 m). Laser interferometers use red-light laser diodes with wavelength of 635 nm.

The most accurate laser interferometers have a resolution better than 0.1 nm. For example, Canon declares at its website a resolution of 0.08 nm for its micro laser interferometer for the measurement of linear displacement. However an accuracy of the instruments is hundreds times worse than the resolution.

10.3 Iodine-Stabilized 633 nm He-Ne Laser

The helium-neon (He-Ne) laser stabilized by iodine $^{127}I_2$, operating at a wavelength of $\lambda \approx 633$ nm, is the most widely used of the sources of reference signal recommended by CIPM. Jokingly referred to as the work horse in the whole metrology of length, it has a wide spectrum of applications, including national standards, research spectrometers, and industrial applications.

The CIPM recommendation regarding the source of radiation with the reference wavelength $\lambda \approx 633$ nm includes the following technical specifications of the system [3]:

- Absorbing molecule $^{127}I_2$, the a16 or f component of the R(127) 11-5 transition;
- Frequency ν = 473 612 353 604 kHz,
- Wavelength λ = 632 991 212.58 fm with a relative standard uncertainty of 2.1×10^{-11}; this applies to radiation of a He-Ne laser with an internal iodine cell, stabilized using the third harmonic detection technique, in the following conditions:
- Cell-wall temperature (25 ± 5) °C;
- Cold-finger temperature (15.0 ± 0.2) °C;
- Peak-to-peak frequency modulation width (6.0 ± 0.3) MHz;
- One-way intracavity beam power (i.e., the output power divided by the transmittance of the output mirror) (10 ± 5) mW for an absolute value of the power shift coefficient ≤1.0 kHz/mW.

In the specification of operating conditions, such as the temperature, modulation width and laser power, the symbols ± refer to the tolerance, not the uncertainty.

These conditions are by themselves insufficient to ensure the achievement of the stated standard uncertainty. An appropriate technical performance of the optical and electronic control systems is required as well. The iodine cell can also operate under relaxed conditions; the larger uncertainty specified in [3], Appendix 1, applies in this case.

Figure 10.2 shows the block diagram of the iodine-stabilized He-Ne laser.

The optics in the laser system comprises a gas tube containing atoms of helium and neon emitting coherent light, an absorption cell with $^{127}I_2$ molecules, two mirrors placed on the axis of laser light, and a photodetector, which measures the intensity of radiation [4, 6]. The tube with He and Ne gases is typically 20 cm long, and the length of the iodine absorption cell is 8 cm. The photodiode used as a photodetector is an intermediate converting device between the optics and the electronics in the laser standard of length.

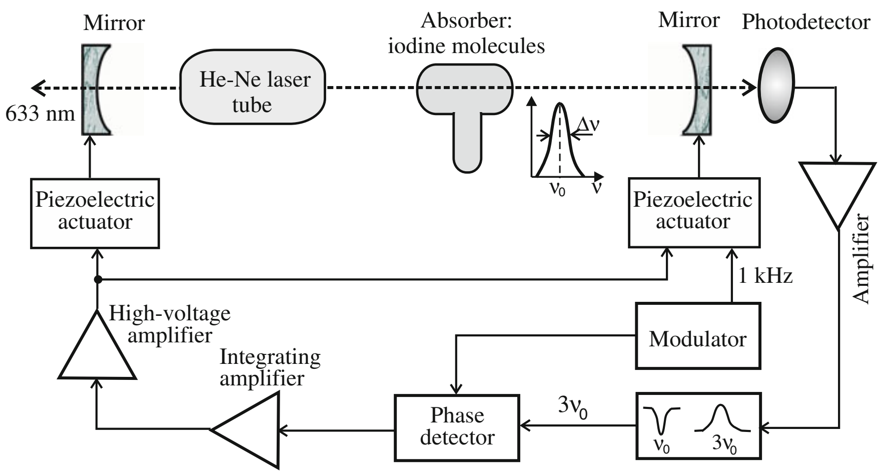


Fig. 10.2 Block diagram of the iodine-stabilized He-Ne laser emitting 633 nm light

The laser emits light in two directions, and light traveling in each direction is reflected by a mirror. The radiation spectrum is narrowed in the iodine cell due to the absorptivity of iodine. Peaks at the frequencies for which the cell saturates with iodine vapor are enhanced in the spectral power distribution characteristic. The frequency of radiation emitted by the laser can be tuned by controlling the distance between the mirrors with simultaneous monitoring of the radiation intensity. The position of the two mirrors on the axis of the light beam is controlled in a synchronous phase-locked loop using piezoelectric actuators. Since piezoelectric actuators require a control voltage of up to 100 V, a high-voltage amplifier is used for amplifying their control signal.

The third harmonic detection method, recommended by CIPM, is used for amplifying the amplitude of the signal at a selected peak [ν (3rd harmonic) $\approx 1.421 \times 10^{15}$ Hz] in the spectral characteristic of the laser light. The third harmonic is obtained by sweeping the frequency of the laser light around the resonance frequency ν_0 by moving one of the mirrors. Only for this purpose (sweeping the frequency) the mirror is shifted cyclically with a frequency of about 1 kHz (1,092 Hz in the laser at GUM in Warsaw) along the axis of the light beam. In the laser standard system the temperature of the iodine absorption cell is controlled and fixed at 15 °C, in accordance with the CIPM recommendation. The role of the actuator in this temperature control is played by a set of Peltier cells (not shown in Fig. 10.2). A polarized He-Ne lamp (not shown in Fig. 10.2) is used as a high-voltage power supply.

The radius of curvature of the mirrors in the laser standard is typically of about 1 m (also flat mirrors are used, though), and the mirror reflectance is of 98 %. The typical output power of He-Ne lasers used in interferometry and spectromicroscopy is 300 μW. The mechanical construction of the laser must meet high requirements. The components of the optics should be made of materials with a low coefficient of thermal expansion and resistant to mechanical shock.

An iodine-stabilized He-Ne laser with a wavelength of 633 nm (manufactured by Thomson, with a NEC discharge tube) is the most important instrument in the Polish national standard of the unit of length at the Central Office of Measures (GUM) in Warsaw. The laser is shown in Fig. 10.3. The laser standard at GUM participates in international comparisons. The latest comparisons, supervised by the

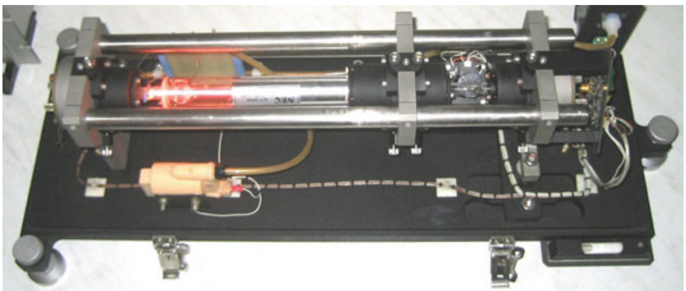


Fig. 10.3 Iodine-stabilized He-Ne laser with a wavelength of 633 nm in the Laboratory of Length at GUM in Warsaw

BIPM, took place in years 1999–2001 in Vienna, involving the laser standards from the BIPM, GUM, BEV (Austria), CMI (Czech Republic), OMH (Hungary) and NIPLPR (Romania) [7]. Comparisons were carried out by the peer process, in which the standards are compared with each other. The synthetic results of the comparison between the laser standard GUM1 and the laser standard BIPM3, after a correction by the recommended operating conditions, were the frequency difference $\Delta v = 0.8$ kHz and the standard deviation of the frequency $u = 1.8$ kHz.

The relative standard deviation calculated for the difference in the frequencies of these two lasers (GUM1 and BIPM3) is $u/v = 3.8 \times 10^{-12}$. Comparative measurements showed a good quality (at the international level) of the instrumentation and procedures used for the realization of the unit of length by the national standard at GUM in Warsaw.

10.4 Satellite Positioning Systems

10.4.1 Positioning Systems

Satellite positioning systems have become of great importance over the past 30 years [8, 9]. Positioning and navigation systems are a good example of consumer use of high-quality quantum devices, specifically atomic clocks. Developed in the United States and made available to the public in 1983, the Global Positioning System (GPS) was appreciated by the users to the point of arousing enthusiasm [9]. Soon satellite positioning system have come into use not only in maritime, air and land navigation, for which they have been designed, but also in many other areas, such as geodesy, agriculture or tourism, or for location of dogs and expensive cars.

The vital function of satellite positioning systems is a dissemination of time signal, i.e. the transfer of standard time signals to users over their operational territory. Since the GPS is owned by the United States, the American authorities have the right to exclude unauthorized users of the system or intentionally degrade the accuracy of positioning. Thus, it is not surprising that the European Union and three other countries have decided to design and build their own positioning systems, despite the huge cost of the project.

The development and implementation of a satellite system require mastery of space and rocket technologies, especially the construction of satellites and placing them into orbit. The former Soviet Union (now Russian Federation) developed a global positioning system GLONASS. A regional positioning system BeiDou was launched by China in 2012. India is developing its own regional IRNSS navigation system and Japan—an augmentation system for the GPS called QZSS. A global positioning system, Galileo, intended solely for non-military purposes, is being developed by the European Union. The construction of the Galileo positioning system is the largest EU economy project after the introduction of the euro as a common currency.

Although the principles of operation of all satellite positioning systems are similar, receivers of one system can not process navigation messages from other systems without additional software; for example, GPS receivers can not handle messages from GLONASS or BeiDou. However, receivers capable of processing messages from different positioning systems (e.g. GPS and GLONASS or BeiDou and GPS) are produced and used widely.

Global positioning or navigation systems are excellent tools especially helpful for car and truck drivers. Satellite positioning systems have limitations, though. A satellite signal cannot be received well in buildings, in streets surrounded by high buildings (so-called urban canyons), natural canyons, woods or tunnels. Although not necessary for driving in a tunnel or in a canyon, in case of accident, navigation would be very helpful for locating the vehicles involved.

10.4.2 Global Positioning System

The Global Positioning System (GPS) is an example of distributed measurement system using cutting-edge technology. This satellite system was developed in the United States to serve the American army, but since 1983 has been available to the public as well [9, 10]. The system can be used by anyone with a GPS receiver. Because of its global coverage and excellent technical parameters the GPS is widely used not only for the purposes for which is was designed, i.e., for determining the position of objects on the Earth for navigation, but also for many other purposes, the most important of which is the generation of standard time signals.

The system uses twenty-eight satellites (the required number of satellites for the entire system is twenty-four) that revolve around the Earth in six constant orbits 20 162.61 km above the equator. Each satellite is visible for about 5 h from a point on the Earth, and in this time its radio signal is available to receivers at that point. Each satellite has a rubidium or cesium atomic clocks and hydrogen maser installed on board along with a time signal transmitter and a transceiver to communicate with a ground control system. The satellites send messages in two radio channels: channel L1 with a carrier frequency of 1575.42 MHz, and channel L2 with a carrier frequency of 1227.60 MHz. Transmission in the GPS is provided by the CDMA (Code Division Multiple Access) method, applied in various wireless communication technologies. The signal in the L1 channel is intended for all GPS users [9]. The typical accuracy of determination of the position of a receiver available in the all-users sector is below 100 m, the best accuracy being 10 m. Signals and messages transmitted by the L2 channel can only be used by authorized GPS users, which are mainly army and government agencies of the United States. The capability to receive a signal in both L1 and L2 channels improves ten times the accuracy of positioning. Actually, there are also three additional, rarely used channels in the GPS: the 1381.05 MHz channel, L3, reserved for messages on nuclear detonation, the 1379.913 MHz channel, L4, being tested for an additional ionospheric correction, and the 1176.45 MHz channel, L5, being prepared for sending civilian safety-of-life (SoL) messages [9]. Since 2005 new launched GPS satellites are prepared to send messages to civilian users by the L2 channel as well.

A GPS receiver is equipped with a quartz clock, the accuracy of which is much lower than that of the atomic clock. A message sent from a satellite includes the time signal and the identification data of the satellite, which together allow an accurate determination of the position of the satellite broadcasting successive messages.

The position P of a GPS receiver is the point of intersection of spheres formed by the radio waves emitted by satellites. The point P is determined by measuring the time interval t in which the radio signals have traveled the paths from three or four satellites to the receiver. Signals may be received from many satellites (with 99 % probability from at least five satellites). Data transmission in the GPS has a rather low rate of 50 bit/s [9].

The principle of operation of the GPS is illustrated in Fig. 10.4. A receiver uses signals from four selected satellites. The criterion for selecting the satellites is the power (level) of the signal received. With a known speed of waves in vacuum (space) and air (atmosphere), the distance d between the receiver and each satellite can be calculated to provide a basis for the determination of the position P of the GPS receiver. The point P in Fig. 10.4 is the point of intersection of spheres with radii d_1, d_2, d_3. The coordinates (x, y, z) of the point P are computed from the measurement data in a coordinate system the origin of which is at the center of the Earth.

The measurement data of three satellites provide a basis for three equations, the solution of which is sufficient to determine the three coordinates x_p, y_p, z_p of the point P in the adopted coordinate system. However, the GPS uses data of four

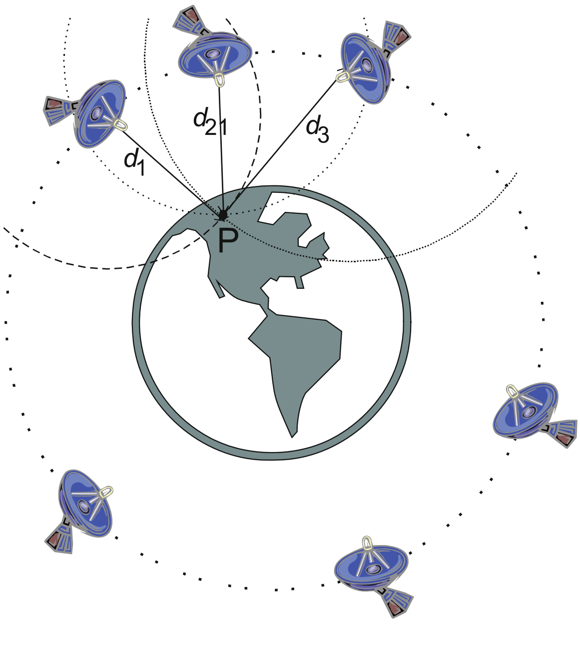


Fig. 10.4 Position P of a GPS receiver determined on the basis of the distance d between the receiver and GPS satellites

satellites, with a fourth equation obtained from the measurement data concerning the fourth satellite. As mentioned above, the accuracy of the quartz clock installed in a GPS receiver is much worse than the accuracy of the atomic clocks in the satellites. The receiver clock has an uncertainty value larger by several orders of magnitude than the uncertainty of the atomic clock. The uncertainty of the receiver clock depends on the quality of the receiver. It is assumed, however, that the clock uncertainty Δt_{CK} in a GPS receiver is fixed in relation to the readings of the atomic clocks on satellites.

The fourth equation of the position of the point P is required to calculate the GPS receiver clock uncertainty Δt_{CK}, introduced as a fourth unknown to the set of equations. The solution of the four equations (10.3)–(10.6) below yields the values of the four unknowns: the coordinates x_p, y_p, z_p of the receiver at point P with respect to the center of the Earth, and the clock uncertainty Δt_{CK}.

$$(x_1 - x_p)^2 + (y_1 - y_p)^2 + (z_1 - z_p)^2 = [\mathrm{c}(t_1 - \Delta t_{CK})]^2 \tag{10.3}$$

$$(x_2 - x_p)^2 + (y_2 - y_p)^2 + (z_2 - z_p)^2 = [\mathrm{c}(t_2 - \Delta t_{CK})]^2 \tag{10.4}$$

$$(x_3 - x_p)^2 + (y_3 - y_p)^2 + (z_3 - z_p)^2 = [\mathrm{c}(t_3 - \Delta t_{CK})]^2 \tag{10.5}$$

$$(x_4 - x_p)^2 + (y_4 - y_p)^2 + (z_4 - z_p)^2 = [\mathrm{c}(t_4 - \Delta t_{CK})]^2, \tag{10.6}$$

where x_i, y_i, z_i are the coordinates of the i-th satellite, $i = 1, 2, 3, 4$; x_P, y_P, z_P denote the coordinates of the point P (the position of the GPS receiver); t_i is the time interval in which the message is transmitted from the i-th satellite to the GPS receiver; Δt_{CK} is the GPS receiver clock uncertainty, and c is the speed of light in vacuum.

The exact coordinates x_i, y_i, z_i of the satellites, defining precisely their position with respect to the center of the Earth, are indispensable for the determination of the position P of the GPS receiver. The coordinates x_P, y_P, z_P of the receiver are subsequently converted to another coordinate system. In the GPS it is the World Geodetic System 84 (WGS-84) reference frame (other positioning systems use different coordinate systems). The coordinates in the WGS-84 include the longitude, latitude and altitude. The uncertainty Δt_{CK} of the receiver clock is taken into account in the calculation of the position of the receiver in the WGS-84 from the previously determined coordinates. The necessary correction is performed at this step of data processing. Thus, signals from four satellites are required to determine the three parameters defining the position P of the GPS receiver: the latitude, the longitude and the altitude above sea level. However, signals from more than four satellites are frequently used for positioning, which can improve the measurement accuracy. If the receiver is located at sea level (which is the case in marine navigation), one of the parameters defining its position is known: the altitude $h = 0$. Thus, two parameters, the latitude and the longitude, are necessary for positioning.

In this case communication with three satellites is sufficient to determine the position of the GPS receiver with a correction related to the clock uncertainty Δt_{CK}.

The transmission time in which the signal travels from the satellite to the receiver is calculated by subtracting the time at which the signal is sent from the time of its reception by the receiver. Radio signals propagate (are transmitted) with the speed of light in vacuum $c \cong 3 \times 10^8$ m/s. The time t_1 in which an radio wave will cover the distance d = 20 162 610 m (the altitude of the orbit above the equator) is:

$$t_1 = d/c = 20\ 162\ 610/299\ 792\ 458 = 67\ 255\ \mu\text{s}.$$

The measurement of a time interval of about 67 ms with a resolution of 1 ns and an accuracy of 10 ns does not present particular difficulties when dealing with an atomic standard of frequency, as it is the case in the GPS. In 10 ns a radio wave will cover a distance of 3.3 m. The uncertainty in determination of the horizontal position in the GPS for unauthorized users ranges from 10 to 100 m. For authorized users the accuracy is about 10 times better. The uncertainty is specified as a range (10–100 m), rather than a single value. The uncertainty of positioning is attributable to many factors which affect the measurement accuracy. The most important of these are:

- Ionosphere and troposphere, sources of error due to the delay of the radio signal passing through the ionosphere and the lower layer of the atmosphere (ionospheric and tropospheric delay). The speed of electromagnetic wave is lower in the ionosphere and troposphere than in space. The delay depends on the time of day and atmospheric conditions such as the temperature and humidity. The thickness of the ionosphere is of several 100 km, even more it is not constant. Information about current parameters of the ionosphere and the troposphere are included in messages transmitted by the L2 channel to authorized users only;
- The difference between the actual position of the satellite and the expected position assumed in the calculation (ephemeris error);
- Reflected signals reaching the antenna in an indirect way (multipath error);
- The constellation of the satellites in relation to the receiver at the time of measurement (it magnifying or reducing the uncertainty);
- The class and quality of the GPS receiver.

Figure 10.5 shows how the configuration of the satellites in relation to the receiver affects the accuracy of positioning. In Fig. 10.5a the GPS receiver is located at point Q, at a distance d_1 from the satellite S1, and d_2 from the satellite S2.

The distances d_1 and d_2 are measured with some uncertainty (measurement error), represented by the concentric rings depending on the position of the satellites. The combined uncertainties of the two satellites result in an area P1 of uncertainty in the position; this uncertainty area includes the possible position of the point Q. In Fig. 10.5b the satellite S2 is further from the receiver. Although the uncertainty in the distances d_1 and d_2 is the same as in Fig. 10.5a, the uncertainty area, labeled P2, is larger here. Also larger is the maximal uncertainty of a single measurement,

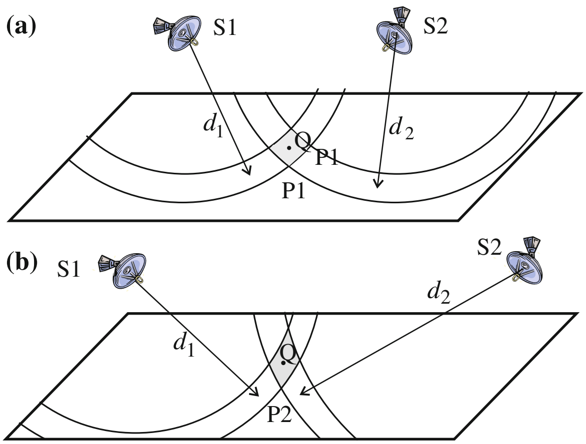


Fig. 10.5 Constellation of satellites and positioning accuracy: depending on the constellation of the satellites, the area of uncertainty is P1 or P2

represented by the distance between the point Q and the farthest point in the uncertainty area P2. Thus, the position of the receiver with respect to the satellites affects the accuracy of determination of the position of the receiver on the Earth.

GPS receivers come in various formats, integrated in smartphones or tablets (or other small computers, such as iPods, notebooks, etc.) or as GPS modules and separate receivers. Many of them are capable of communicating with two systems, e.g. GSM and GLONASS or GSM and BeiDou.

The measurement by a GPS receiver allows to read the geographical position with a resolution of 0.001 min of arc (or better), which represents 1.8 m at the latitude of the equator. The measurement accuracy is always worse than the resolution. Figure 10.6 shows screenshots of a GPS receiver (Samsung Galaxy S5 smartphone) with the measured geographical position. The screenshots present the results of successive measurements of the same position (with an uncertainty of ±0.5 cm) carried out at an interval of 2 min (differences in the positioning results: latitude $\Delta = 0.0040'$ ➩ 7.4 m; longitude $\Delta = 0.0026'$ ➩ 2.9 m. The software used affects the uncertainty of positioning as well.

10.4.3 GLONASS Positioning System

The Soviet Union began the construction of its satellite navigation system GLONASS (for Russian *Глобальная Навигационная Спутниковая Система*) in 1976. The first test satellite was launched into orbit in 1982. In 1996 twenty-four satellites operated in orbit and GLONASS covered the entire Earth. Now owned and managed by the Russian Federation, the system has twenty-four satellites in three circular orbits at a height of 19 100 km above the equator. The orbits are angularly offset by 120° [8].

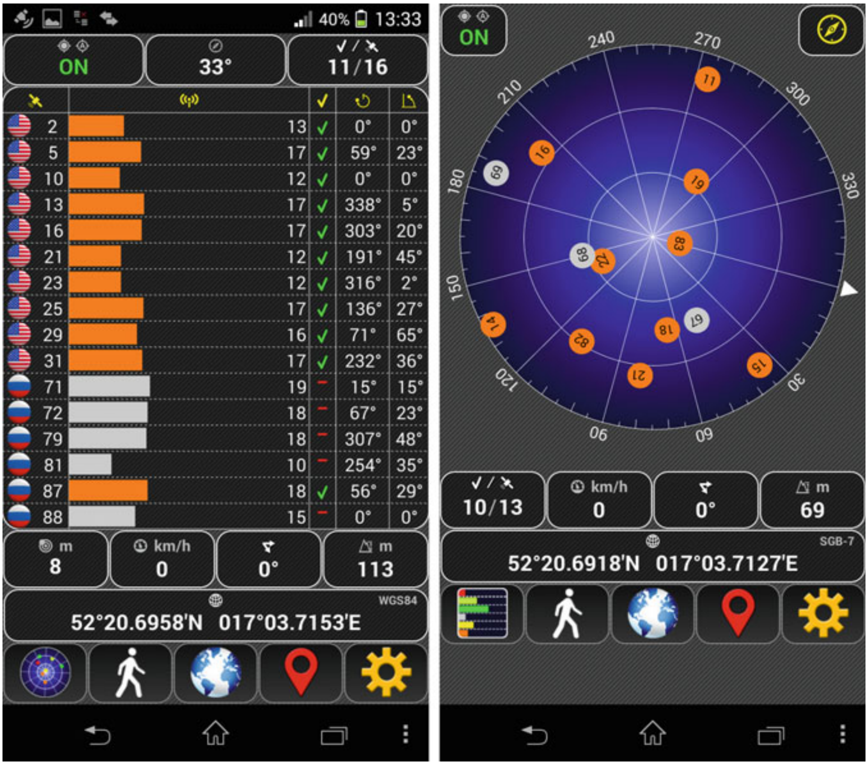


Fig. 10.6 Screenshots of a smartphone displaying the same geographical position; *left*—the WGS84 reference frame used by GPS, *right*—the SGB-7 reference frame used by GLONASS

GLONASS uses the same idea as the GPS for determining the position of a receiver on the Earth. The position is determined as the point of intersection of propagation spheres of radio waves emitted by the satellites. There are some differences between GLONASS and GPS in the principles of construction and operation, though. GLONASS satellites send messages (time signal and identification data) in two radio bands: L1 with a frequency band of 1,610 MHz, and L2 with a frequency band of 1,250 MHz [8]. Each of the twenty-four satellites has a separate carrier frequency in both the L1 and L2 bands. Signals are transmitted by the frequency division multiple access (FDMA) method, used in radio communications. GLONASS created also a third frequency band, L3, with increased positioning accuracy for communication for search and rescue purposes.

$$\begin{aligned}
&\text{Carrier frequency in the L1 band:} \quad f_{L1} = (1602 + 0.5625n)\text{MHz};\\
&\text{Carrier frequency in the L2 band:} \quad f_{L2} = (1246 + 0.4375n)\text{MHz};\\
&\text{Carrier frequency in the L3 band:} \quad f_{L3} = (1201 + 0.4375n)\text{MHz},
\end{aligned}$$

where $n = 1, 2, 3, \ldots, 24$ is the GLONASS satellite number.

The data transmission rate in GLONASS is 50 bit/s (the same as in the GPS).

GLONASS satellites are identified by the carrier frequency of the transmitted signal. This is an important difference with respect to the GPS, in which the satellite identification data are contained in messages sent by the satellites. There is only one carrier frequency in each band (L1 and L2) in the GPS.

A cesium (and/or rubidium) atomic clocks with a relative uncertainty of 5×10^{-13} in the frequency is installed in each satellite. The position of a GLONASS receiver is calculated and presented in the SGB-7 geodetic reference frame, while the GPS uses the WGS-84 reference frame. Potentially GLANOSS is more accurate than the GPS, even if only because it is some fifteen years newer than the American system and newer technologies could be used in its design. Currently the best uncertainties provided by GLONASS are:

- In horizontal position, unauthorized users: 30 m;
- In horizontal position, authorized (military, police) users: 10 m;
- In horizontal position, for search and rescue services (band L3): from 5 to 7 m;
- In vertical position: 60 m;
- In velocity: 5 km/h.

GLONASS signals can be read by a variety of commercial communication devices, including many types of smartphones and tablets. Typically, smartphones are capable of receiving both GLONASS and GPS signals, the stronger of which is used by the software for navigation or positioning. Advanced communication devices use signals from both systems simultaneously in order to improve the accuracy of positioning. Devices with dual, e.g. GPS/GLONASS, receivers are commercially available. A pioneer in the manufacturing of professional dual receivers, Ashtech has offered many types of GPS/GLONASS receivers since 1997. An interesting product of this company is the GG-RTK (GPS GLONASS Real Time Kinematic) system consisting of a central unit (CU) and satellite receivers. The CU communicates with receivers by a 915 MHz radio link. Datasheets of the GG-RTK provide information with a very good resolution of 1 cm for positioning with a period of 1 s, and 5 cm for positioning with a period of 0.2 s. The high resolution of the GG-RTK system and its good accuracy (only an order of magnitude worse than the resolution) are partly due to the simultaneous use of the two satellite systems for positioning.

In years 1996–2008 the number of active GLONASS satellites decreased to as few as ten, and one of the orbits was empty. The decrease in the number of satellites was due to their limited service life, ranging from 1 year (the oldest satellites) to 7 years. The lifetime of satellites is limited, among others, by the unreliability of the equipment on board and the depletion of fuel for the electricity generators and rocket motors. The rocket motors correct the position of the satellite in orbit. Having reached the end of their service life, GLONASS satellites ceased to operate and their replacements were not numerous enough. Currently there are twenty-four active GLONASS satellites in orbit and at least three spare ones. Newer versions are equipped with 1.5 kW solar panels.

10.4.4 Galileo Positioning System

A European satellite navigation system Galileo is at the phase of intensive development and testing. The system is being developed by the European Union and the European Space Agency (ESA), which has set up a European consortium iNavSat to develop and launch the Galileo system. The system will be based on 30 satellites, including 27 operational satellites and 3 spare ones. The satellites will be placed in three orbits at a height of 23 616 km above the Earth, inclined by 56° with respect to the equator. The first test satellite (not intended for use) was submitted to the ESA for ground-based tests in August 2005 [11]. Positioning only relying on signals emitted from Galileo satellites was first performed in March 2013. Six operational satellites were launched by 2014. The last two, launched in August 2014, did not reach their planned orbits. Presently the correction of their orbits is not sure. The 30-satellite system is expected to be completed by 2019.

Galileo satellites transmit information that includes position data, exact time, information on the reliability of these data, and data concerning possible system multifunction.

Similarly to GPS and GLONASS, Galileo has created a dedicated reference frame, the Galileo Terrestrial Reference Frame (GTRF). According to Galileo requirements the three-dimensional differences in the position compared to the International Terrestrial Reference Frame should not be larger than 3 cm.

10.4.5 Regional Positioning Systems: BeiDou, IRNSS and QZSS

The Chinese *BeiDou Satellite Navigation System* (BDS), also known as COMPASS or BeiDou-2, will be a global satellite navigation system based on 35 satellites [12]. Five of them will be geostationary (GEO) satellites orbiting the Earth at an altitude of 35 786 km, and the other thirty will be non-GEO satellites. Of the non-GEO satellites, twenty-seven are Medium Earth Orbit (MEO) satellites in orbits at an altitude of 21 528 km, and three are Inclined Geosynchronous Orbit (IGSO) satellites in orbits at an altitude of 35 786 km. At present the BeiDou is a regional navigation system, operating in the Asia-Pacific region, with 14 operational satellites in orbit. The BeiDou system has been available to customers since the end of 2012. The global satellite navigation system is planned to be completed by 2020. The BeiDou system uses three communication channels allocated in the band from 1.2 to 1.6 GHz: E2 (carrier frequency of 1561.098 MHz), E5B (1207.14 MHz), and E6 (1268.52 MHz), which overlaps with a band used by Galileo.

Like the GPS and Galileo, the BeiDou system uses the CDMA method for signal transmission. Another common feature with the GPS is that BeiDou provides two kinds of services, for unauthorized users (civilians) and for authorized users, which include government agencies, the Chinese army and the Pakistan army. The civilian service provides positioning with an accuracy of 10 m, velocity measurements accurate to 0.2 m/s, and time signal distribution with an accuracy of 10 ns. The licensed service, provided to the army and the government, has a 100 times better positioning accuracy of 0.1 m. China cooperates with the Galileo consortium by financial investments and collaboration on the satellite navigation technology.

The *Indian Regional Navigation Satellite System* (IRNSS) is an autonomous system to be fully developed and built in India [13]. The IRNSS would provide two services: the Standard Positioning Service open for civilian users, and the encrypted Restricted Service for authorized users (army). A space segment of the IRNSS will consist of seven satellites (three of which will be GEO satellites). It will be sufficient to cover the Indian subcontinent and a surrounding region as far as 1 500 km away. India plans to launch into orbit the full set of seven satellites by 2015. The IRNSS will provide a Special Positioning Service and a Precision Service. In both services communication will be carried out through the L5 channel (with a carrier frequency of 1176.45 MHz) and the S channel (2492.08 MHz). The satellites will be equipped with 1.4 kW solar electric generators. The planned positioning accuracy of the IRNSS is better than 10 m over the territory of India and better than 20 m over the entire Indian Ocean region.

The Japanese *Quasi-Zenith Satellite System* (QZSS) is a proposed augmentation system for the GPS [14]. The QZSS will use four geosynchronous satellites. At present the system uses three satellites, 120° apart, in inclined, slightly elliptical orbits. The ground traces of the satellites are asymmetrical figure-8 patterns (with a major semiaxis of 42 164 km). In such a configuration there is always one satellite almost directly overhead (in quasi zenith) in Japan. The QZSS should improve the positioning and navigation service provided by the GPS throughout Japan. The quasi-zenith location of the satellites allows to eliminate the urban canyon effect from the navigation process. Another important task of the QZSS is time signal transfer. The QZSS satellites transmit signals using channels compatible with the GPS: L1 (with a carrier frequency of 1575.42 MHz), L2 (1227.60 MHz), and L5 (1176.45 MHz). The first QZSS satellite was launched in 2010. Completion of the operational system is expected by 2017.

References

1. http://www1.bipm.org/en/
2. J. Magueijo, *Faster Than the Speed of Light: The Story of a Scientific Speculation* (Arrow Books, London, 2003)
3. Recommendation 1 of International Committee for Weights and Measures (CI 2003). Metrologia, **40**, 103–133 (2003) (Metrologia, **42**, 323–325(2005))
4. P. Hariharan, *Basics of Interferometry* (Academic Press, London, 1992)
5. S. Olyaee, S. Hamedi, in *Nano-Metrology Based on the Laser Interferometers*, Chapter 13, ed. by M. Kr. Sharma. in Advances in Measurement Systems (InTech, London, 2010)
6. J. Helmcke, Realization of the metre by frequency-stabilized lasers. Meas. Sci. Technol. **14**, 1187–1199 (2003)

7. M. Matus et al., International comparisons of He-Ne lasers stabilized with $^{127}I_2$ at $\lambda \approx 633$ nm (September 1999). Metrologia **39**, 83–89 (2002)
8. http://www.glonass-center.ru/
9. http://www.gps.gov/systems/gps/
10. A. El-Rabbany, *Introduction to GPS: The Global Positioning System* (Artech House, London, 2002)
11. http://www.gsa.europa.eu/galileo
12. http://en.beidou.gov.cn/
13. http://www.isro.org/satellites/navigationsatellites.aspx
14. http://www.qzs.jp/en/

Chapter 11
Scanning Probe Microscopes

Abstract This chapter presents scanning probe microscopy and the most important microscopes using this technique. Historically the first of these devices, the scanning tunneling microscope (STM) is used for atomic-scale imaging of the surface of samples of conducting materials. The STM is an inestimable instrument for research in solid state physics, quantum chemistry and molecular biology. Its operation is based on the conversion of the distance between its tip and the studied surface into electric current. Another question is the control of the position of the scanning tip and the recording of data for generating a map of the studied surface. The experience acquired in the construction of the STM was of much use in the development of other types of scanning microscopes. In this chapter we discuss also the atomic force microscope (AFM), electrostatic force microscope (EFM), scanning thermal microscope (SThM) and scanning near-field optical microscope (SNOM). We also present a wide range of current and prospective applications of scanning microscopes.

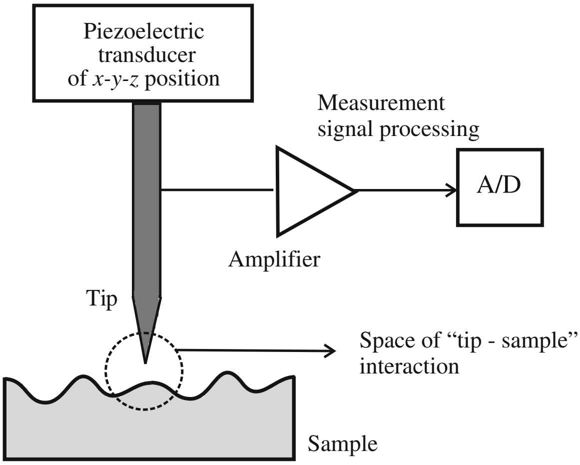
11.1 Atomic Resolution Microscopes

11.1.1 Operating Principle of a Scanning Probe Microscope

An important group of quantum instruments are scanning probe microscopes (SPM), which are of great significance for research in physics, chemistry, biology and medicine, in nanotechnology and length metrology. The SPM are used to measure objects with micro- and nanometer size, and to imaging of sample surfaces on the micrometer scale, and even on the atomic scale. In SPM we perform the spot measurement of an interaction between the investigated sample and the microscope tip—Fig. 11.1. The SPM tip is also called sampler probe or apex. Generally speaking, interactions between the sample and the tip are the forces of electromagnetic field (including thermal or optical field), electrostatic field, magnetic field (which is much weaker than the previously listed types of field) or gravitational field (very weak, irrelevant for SP microscopy).

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_11

Fig. 11.1 Operating principle of a scanning probe microscope



The microscope tip is placed between 1 Å and 100 nm high above the sample surface or it touches the surface. During the sample surface scanning, the measurement of interaction "tip-surface" is repeated in a number of thousands of points of the investigated surface, and as a result, it is possible to represent the surface image. Interactions between physical objects being so near to one another usually require to be analysed according to the principles of quantum mechanics. They are jointly called "near-field interactions" or "short-range interactions".

11.1.2 Types of Near-Field Interactions in SPM

The most important near-field interactions between the SPM tip and the sample are:

- Tunneling of electrons between the tip and the investigated sample surface; electrons are tunneling through an insulating space (vacuum, air or gas), which constitutes the potential barrier. Both the tip and the sample must be electric conductors. The tunnel effect is used in scanning tunneling microscopes STM.
- Interatomic forces between the sample and the tip; depending on the distance between the sample and the tip, they can be repulsive or attractive forces. Interatomic interactions are the basis for the operation of atomic force microscopes ATM.
- Electrostatic interactions between the electric charge on the sample surface and the charge on the tip apex. The electrostatic interaction is used in electrostatic force microscopes EFM.
- Heat transport between the tip and the sample. The heat stream flow effect is used in scanning near thermal field microscopes SThM.

- Transmission of light stream by the investigated sample or transmission of light reflection of the sample. The microscope tip takes the part of transmitter or receiver of a light wave. Transmission of light reflection is the basis for the operation of scanning near optic field microscope SNOM.

Different methods for determining the measurement signal are used in the operation of SPM. This differentiation of methods concerns especially the atomic force microscopes, which use electric or optical measurements of deflection of microscope cantilever. There are also microscopes that combine two (or more) types of near-field interactions between the tip and the investigated sample in one instrument. Such instrument is called modular scanning microscope of near-field interactions [5]. Using the modular scanning microscope, it is possible to acquire both a geometric image of the investigated surface with a resolution of single atoms and e.g. an image of temperature distribution on that surface or a map of optical properties of the sample.

11.1.3 Basic Parameters of SPM

One of the basic parameters of SPM is always geometric resolution on the coordinate axis x–y–z. In the STM and AFM the distance measurement resolution between the tip and the atoms on the sample surface on axis z is very important, i.e. the resolution of the measurement of tip height above the sample surface. In STM a resolution of $\Delta z \approx 10^{-11}$ m has been achieved. Geometric resolution on the coordinate axes of sample surface Δx, Δy informs about how distant of one another particular measurement points are on the surface of investigated sample? In these points, in scanning mode, a physical quantity appropriate for the microscope type is measured (geometric, electric, thermal, optical, magnetic quantity). Microscope resolution for x–y axis is connected with the converter parameters controlling the cantilever motion during the scanning. Sometimes such device is called microscope actuator. Piezoelectric converters are generally used as microscope actuators [x = f(V)], but the use of magnetostrictive converters is also attempted [x = f(B)].

Other important parameter of microscope is the resolution measurement of the investigated physical quantity, i.e. for example for SThM it is the temperature measurement resolution (e.g. ΔT = 10 mK) or thermal conductivity measurement, and for SNOM—measurement resolution of light stream or reflection coefficient.

The third parameter of SPM is scanning rate, given as a number of measurement points in a time unit (sampling rate) or as the speed of tip shift above the investigated surface. Scanning rate, the greater the better, is also important for measuring objects with parameters variable in the function of time. In SP microscopy a typical scanning rate v is of the order of 10 μm/s [5].

If the microscope resolution on axis x amounts to Δx, and the maximum frequency of measured signal amounts to f_{max}, then the allowable scanning rate is determined with formula (11.1)—Fig. 11.2.

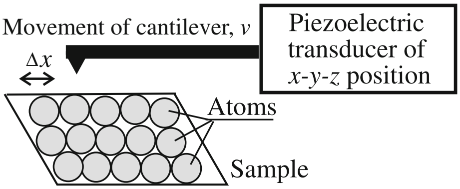


Fig. 11.2 Scanning principle

$$v = \Delta x \times f_{\max} \quad (11.1)$$

Scanning rate is an important functional parameter for a microscope used in a production process, e.g. in testing electronic microstructures.

11.2 Scanning Tunneling Microscope

In the historical development of microscopic equipment the first and still the most frequently used microscope applying short-range interactions is scanning tunneling microscope (STM). The idea of building the STM comes from a Swiss physicist *Heinrich Röhrer*. He realised the idea in 1982 together with *Gerd Binnig*; both were awarded the Nobel Prize for that achievement in 1986 [2]. The STM is used to investigate the surface of samples fabricated exclusively of a material conducting electric current (conductor, semiconductor, superconductor). The basis for the performance of STM is the tunnel effect: electric current flow through the potential barrier between the investigated sample and the microscope tip placed above the sample surface Fig. 11.3.

The sample surface is separated from the tip with an insulating space, usually air or vacuum. Current flow, i.e. transport of electrons between the surface and the tip, requires the electrons to perform a work function, E_G, and to overcome the energy gap (potential barrier). Transport of electrons is possible between the electrode 1 (tip), in which the energy levels in conductivity band are occupied and the electrode 2 (sample), in which the energy levels in this band are allowed but not filled. We denote the electronic work function of the tip material as E_{G1}, and of the sample material—as E_{G2}. In order to shift the Fermi levels E_{F1} and E_{F2} versus one another in two approaching electrodes and to excite a difference in filling the energy levels in the electrodes, it is necessary to apply electric voltage V to the electrodes. It causes electrons to move from electrode to electrode through the potential barrier, i.e. a tunnel current flow through the junction.

The potential barrier width is the function of distance b between the tip and the sample surface. Measuring signal in STM is tunnel electric current intensity I_t, which depends on the potential barrier, and indirectly—on the height b of the tip above the sample surface. The value of tunnel current is determined with formula (11.2) [2].

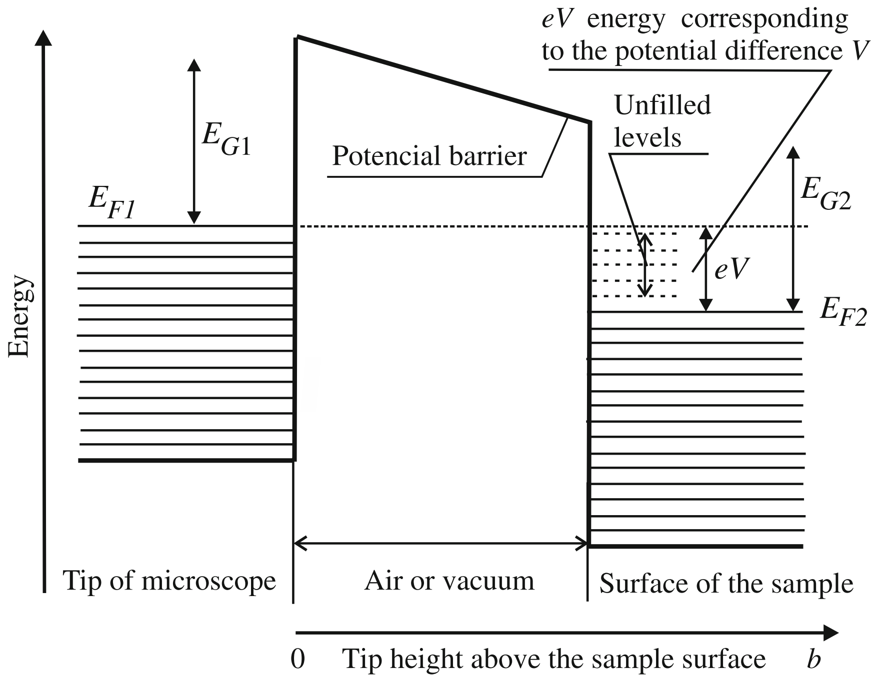


Fig. 11.3 Band energy model for tunneling electrons in STM

$$I_t = k_z V \exp\left(-C\sqrt{E_G b}\right) \tag{11.2}$$

where: I_t—tunnel current, k_z—material constant of the junction (with the dimensions of electric conductivity), m—electron mass, h—the Planck constant, V—voltage on the junction, $C = 4\pi\sqrt{2m}/h = 10.25\ (\mathrm{eV})^{-1/2}\ \mathrm{nm}^{-1}$—constant, E_G—average electronic work function (functions E_{G1} and E_{G2}), b—tip height above the sample surface.

For a voltage V of ca 1–2 V (typical values) polarizing the tunnel junction in STM, the tunnel current value I_t is of the order of 1 nA. Exponential dependence of tunnel current I_t = f(b) gives a very high sensitivity to STM. For electronic work function E_G = 4.63 eV (wolfram with surface structure 100) and initial tip height b = 1 nm, it is possible to achieve for example a tunnel current I_{t0} = 1 nA. For that tip placed at a height b = 0.9 nm above the surface, the current value is three times as high I_{t1} = 3.1 nA, and for the tip placed at the height b = 1.1 nm—current I_{t2} = 0.34 nA.

Two types of STM have been designed:

- With a variable tip height (placement) above the microscope table, and with a constant distance from each point of scanned surface; constant distance between the tip and surface makes it possible to keep a constant value of tunnel current intensity during the scanning, I_t = const;

- With a constant tip height above the microscope table, which gives a variable distance of the tip from the investigated surface and a variable value of tunnel current intensity during scanning, I_t = var.

Block diagram of STM with constant height was shown in Fig. 11.4.

The microscope tip moves over the scanned surface of the sample along the programmed path. A piezoelectric converter controls the tip move along the plane *x*–*y* according to the established programme. The piezoelectric converter has a high processing coefficient, e.g. 0.2–2 μm/100 V or 1 μm/1000 V, and is supplied with constant voltage produced by a high-voltage amplifier. Typical values of processing coefficient are defined for piezoelectric converters with control voltage of a range 0–100 V or 0–1000 V. The high-voltage amplifier itself is controlled with a signal of a digital-to-analog (D/A) converter. Between the tip and the investigated sample a voltage source *V* is switched on, which enables tunnel electric current flow I_t. The value of tunnel current I_t in each successive position of the tip over the surface depends on height *b* above the surface. Current intensity is processed into digital form in an analog-to-digital (A/D) converter. Each measuring point, i.e. coordinates *x*–*y* of the tip and its height *z* above the sample surface (calculated from the value of tunnel current intensity), is described with a digital signal. This data is a database for creating an image of scanned surface and it allows for many other operations, e.g. statistical data processing for a single sample and for a series of samples, numerical and graphic comparison of selected samples, images of samples, determination of the surface of sample section.

A very important technological issue in the building of both STM and other probe microscopes is the tip performance. Its end (apex) should be thin enough for—in ideal situation—an "atom-atom" interaction. Too thick tip gives an interaction of type "group of atoms-group of atoms"; a result is a worse resolution of both microscope and image. The STM use mainly wolfram or PtIr alloy tips, and AFM—silicon and diamond tips, and single wall carbon nanotube (SWCNT). The best apexes have tip diameter smaller than 20 nm. Examples of images of metal sample

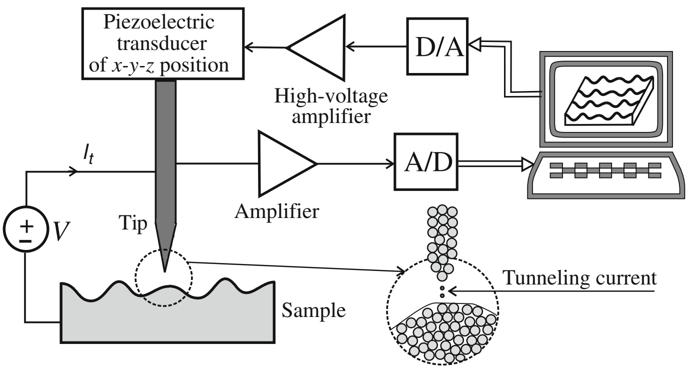


Fig. 11.4 Block diagram of STM

surfaces acquired with measurements using STM at Poznan University of Technology are shown in Fig. 11.5.

When measuring shifts and geometrical dimensions with STM, the following geometrical (linear) resolution was achieved: vertically $\Delta b = 0.01$ Å $= 10^{-12}$ m and horizontally $\Delta a = 0.1$ Å $= 10^{-11}$ m. Those very good results of STM resolution when measuring object (tip) shift are not record-breaking at all. Already 40 years ago R.D. Deslattes achieved a resolution of 10^{-6} Å $= 10^{-16}$ m, when measuring the linear shift using an interferometer with X rays [4].

STM precision makes it possible to measure the arrangement of single atoms on the sample surface and to create an image of the surface topography (using software methods). Moreover, STM makes it possible to manipulate single atoms and arbitrarily arrange them on the surface. It gives hope that particles could be created with the use of methods for manipulating atoms instead of building them by chemical methods. It should be remembered, however, that the arrangement of a few atoms one by one (e.g. of two atoms of hydrogen and one atom of oxygen) is not sufficient to create a particle (H_2O). It is necessary that the atoms be connected with atomic bonds.

11.3 Atomic Force Microscope

11.3.1 Atomic Forces

In 1986, 4 years after the implementation of STM, a team of G. Binnig, Calvin F. Quate and Christoph Gerber developed and implemented atomic force microscope (AFM) [1]. Atomic force microscope makes it possible to investigate the surfaces of samples made of every material, also not conducting electric current, and this is the difference between AFM and STM. In AFM mutual arrangement of atoms is determined by measuring an interaction force between the atoms on the microscope tip apex and the atoms on the surface of investigated sample. However, the direction of the interaction of interatomic force vector in AFM varies by 180°, depending on the range of distance b between the tip and the sample—Fig. 11.6.

For small distances b between the tip and the sample ($0 \leq b \leq 1$ nm) there is a *repulsive force* $\boldsymbol{F_o}$ between them, with a value of the order of 10^{-9} up to 10^{-7} N [5]. Mutual repulsion of two atoms, being near one to another, results from the Pauli principle, which is one of the principles of quantum mechanics. When two atoms approach, the distributions of their charges are gradually covered. The electrons of the atom in the tip cannot, however, occupy the lowest energy levels filled already by the electrons of the sample atoms (and vice versa), because an arbitrary energy level in the system (e.g. in an atom or in a molecule) can be occupied only by one electron (or other particle). It results from the Pauli principle as well. In atoms approaching to one another electrons must, then, occupy higher energy levels. Internal energy of the system is, then, higher than for distant atoms, and the

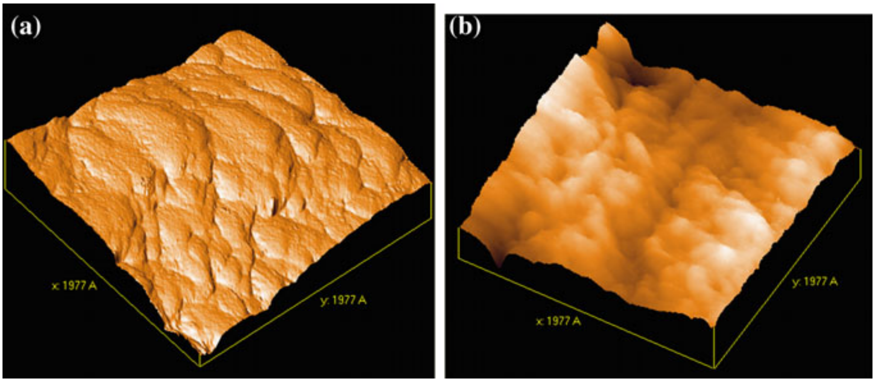


Fig. 11.5 Images of surface of metalic wires: **a** cobalt, image size: 198 × 198 × 29 nm^3, **b** nickel, image size: 399 × 399 × 46 nm^3 (measured by S. Szuba with STM of his own construction)

repulsive force is oriented in such a way that the system take the state of lower internal energy (decreasing entropy).

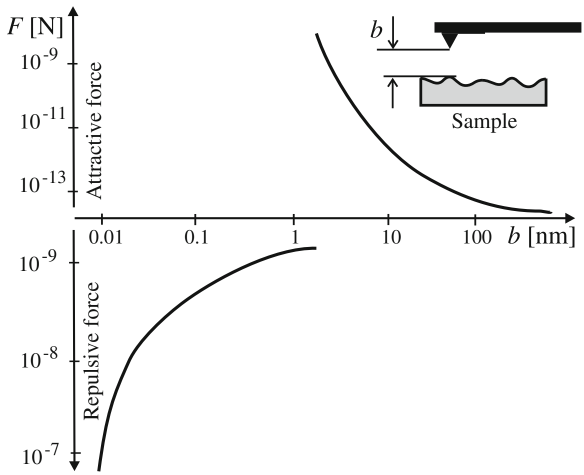
Critical distance b_{kr} = 1 nm at which interatomic forces, both attractive and repulsive, are balanced, is here only estimated, because it is dependent, among others, of the radii of atoms in the tip and atoms in the sample, and the atomic radii differ in the values for different elements. Given interval ($0 \leq b \leq 1$ nm) is the range of AFM microscope working in the static mode, i.e. Contact AFM (CAFM).

Force attracting F_p the tip to the sample surface acts for distances b greater than ca 1 nm between the tip and sample. It is a rough estimation of critical distance b, because this distance depends on the radii of interacting atoms. The forces of attractive interaction are the van der Waals forces, known in physics of solid as the forces bonding atoms in the crystal lattice99. The atoms in the microscope tip and the atoms on the sample surface mutually induce dipole moments that attract one another. The energy of attractive interaction of two atoms is proportional to R^{-6}, where R is the distance between atoms [6].

11.3.2 Performance of Atomic Force Microscope

A block diagram of the atomic force microscope (AFM) was shown in Fig. 11.7. The most important components of AFM are: cantilever, light and acute tip placed on the cantilever, system controlling the cantilever moves in three directions x, y, z (with piezoelectric position converters), system measuring the lever deflection and computer with software to control the scanning process and for imaging the sample surface. After initially positioning the tip towards the investigated sample surface an approach or a deflection of the tip towards the surface is observed, as a result of interatomic forces (interaction) F. The rising move of the tip is accompanied with lever deflection z.

Fig. 11.6 Interatomic forces in AFM in the function of distance b between the microscope tip and the sample surface



Critical element of AFM construction is a cantilever. It should be made of elastic material (spring constant of the order of 1 N/m), so that in the range of elastic deformations, according to the Hooke law, it is possible to linearly connect the interaction force F on the lever with the value of its deflection z. Most frequently the cantilevers are made of silicon or its compounds (silicon oxide SiO_2 or silicon nitride Si_3N_4). Frequent is also a monolithic construction: cantilever integrated with tip, which is obtained in the process of silicon etching. Typical cantilever dimensions are: length 30–500 μm, width ca 10–100 μm and thickness from 100 nm to 10 μm.

An important parameter of the cantilever is resonance frequency v_o of its mechanical oscillations, in a range of 5–100 kHz for many types of cantilevers with different size and mass.

11.3.3 Measurements of Microscope Cantilever Deflection

The value of cantilever deflection z is currently measured with optical or electric methods. Let us notice that for a cantilever length, e.g. 100 μm, the deflection of its end by 100 nm indicates a very small deflection angle, 3.5 min. Two methods of optical measurement are applied: using of laser interferometer and analog measurement of light ray reflected from the mirror cantilever surface. Measuring system with interferometer contains a signal source (laser diode), an optical fiber and a photoelectric detector (Fig. 11.8). A focused laser light stream, brought with an optical fiber, lightens the lever surface, is reflected from it and returns to the optical fiber (a slight deflection angle of the cantilever does not make it difficult to hit the stream to the optical fiber core) (Fig. 11.8).

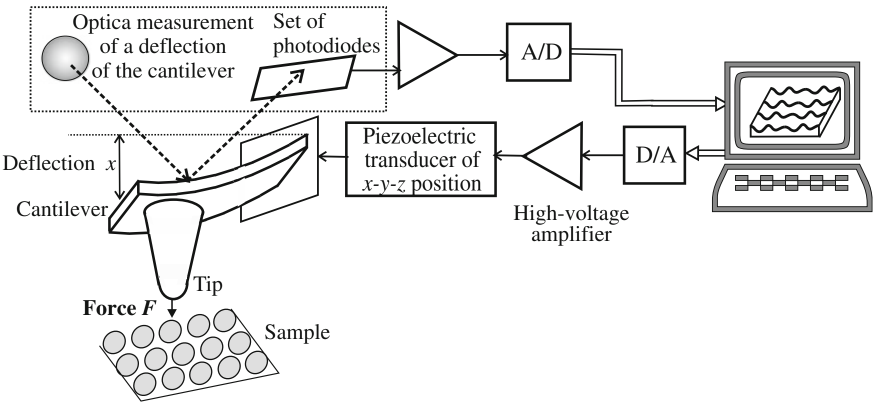


Fig. 11.7 Block diagram of scanning AFM

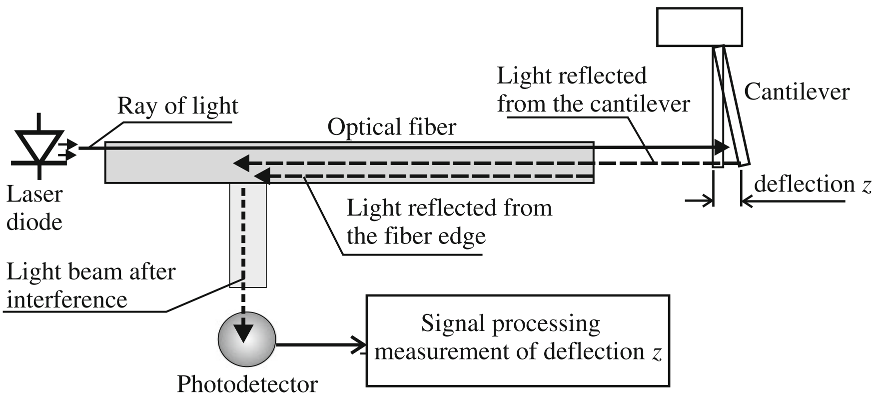


Fig. 11.8 ATM lever deflection measurement with laser interferometer

Part of the light stream is reflected earlier from the cut end of the optical fiber and runs to the light source. Two light streams: reflected from the cantilever and reflected from the end of the optical fiber end interfere with one another. Because of the interference those two streams give as a result a light stream pulsating with a frequency proportional to difference $(l_1 - l_2)$ of the way covered by the streams, and calculated modulo (λ), where λ is the laser light wave length, i.e. $l_R = (l_1 - l_2)$ mod (λ). For instance, for wave $\lambda = 650$ nm (red light) and difference $(l_1 - l_2) = 2000$ nm the calculated difference modulo (650 nm) is equal to: $l_R = (l_1 - l_2)$ mod (650 nm) = 50 nm. The cantilever deflection is a double section of the way covered by the stream reflected from the cantilever. The pulsating light stream is into electric pulses in the photoelectric detector (photodiode). After adjusting the system, the light pulsing period is the measure of both the difference of way for light streams and the microscope cantilever deflection. The measurement deflection resolution achieved by means of laser interferometer is equal to 10^{-11} m.

Optical, analog measurement of microscope cantilever deflection is also frequently applied. In such a system, the laser light falls on the mirror cantilever surface and reflected away from it. After being scattered on the mirror, a circular light spot illuminates the photoelectric detector made up of 4 photodiodes. For a non-deflected cantilever all photodiodes are illuminated with the same intensity and the electric current of each photodiode has a similar value—Fig. 11.9.

When the cantilever is deflected up or down, the light spot moves on the photodetector, and therefore the photodiode illumination and their currents are differentiated. The photodiode current measurement makes it possible not only to measure the cantilever deflection on axis z, but also to twist the cantilever (at shearing stress). Electrical measurement of lever deflection is carried out indirectly. Piezoresistance tensometers, which are sprayed on the cantilever or etched in its structure, are used to measure the lever mechanical stress. Knowing the lever stress and its elastic constant, we can calculate the cantilever deflection.

Resistive tensometers perform usually in the system of full Wheatstone bridge, which means a system with active resistive tensometer in each branch of the bridge —Fig. 11.9. The system of full bridge gives a compensation of temperature influence on the output bridge voltage, and such compensation is necessary at every electric tensometric measurement. The resistance value of the tensometer is equal to several hundred Ω up to a few kΩ. Typical bridge sensitivity, when measuring the cantilever deflection, is equal to 1 up to 10 μV na1 nm, and when measuring the force acting on the cantilever—from 0.1 to 0.5 μV to 1 nN [5].

No detailed proof is needed to show how important it is to limit the mechanical oscillations in the microscope contraction (both AFM and different types of SPM), in order to conduct measurement with geometric resolution of the order of 1 nm. The construction oscillations amplitude is attenuated by coefficient $(v_k/v_o)^2$, where v_k is the frequency of mechanical construction oscillations. Typical values of

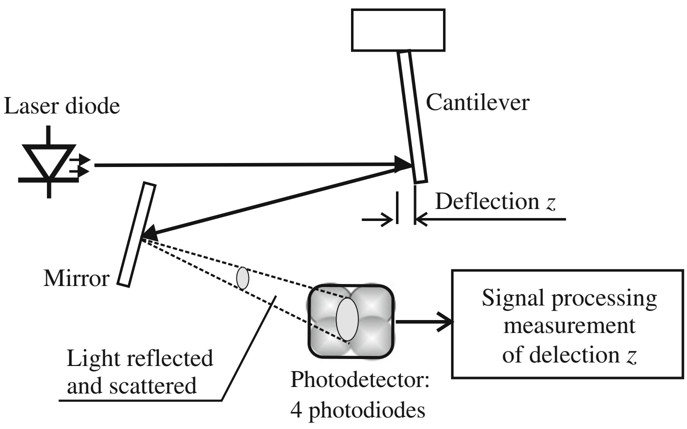


Fig. 11.9 Cantilever deflection measurement by means of AFM illumination intensity

frequency and oscillations amplitude for a building amount to, respectively: v_k = 20 Hz and 1 μm. Therefore, for a cantilever with resonance frequency v_o = 20 kHz, a reduction of such oscillations, transferred from the building to the cantilever, is achieved, to the amplitude cantilevel 1 pm.

11.3.4 AFM with Measurement of Cantilever Resonance Oscillation

To measure atomic forces in AFM the measurement of cantilever resonance oscillations (in fact of oscillation amplitude or frequency) is also used, instead of measuring the cantilever deflection. The AFM cantilever being subject to stress, as a result of atomic forces, has a smaller frequency resonance of oscillations than the free (non-loaded) cantilever—Fig. 11.10. This type of microscopy is called "resonance microscopy of short-range interactions".

Intrinsic oscillations of microscope cantilever (noise) are too weak to be the measuring signal. The cantilever oscillations must be artificially forced in the AFM system, so that their amplitude is sufficient to measure. The cantilever oscillations are induced with electric or thermal signal (by means of piezoelectric converter). The frequency of forced oscillations f_p is different form the resonance frequency of the beam f_{r1}. For a particular cantilever both the resonance characteristic of the lever $A \sim f$ and the frequency change in cantilever resonance oscillations in the function of atomic forces: $f_{r2} \sim F$ (F—force) are known—Fig. 11.11.

Atomic forces can be measured by means of measuring the oscillation amplitude (amplitude detection) or by measuring the offset resonance frequency (frequency detection). For amplitude detection, it is necessary to force the cantilever oscillation with frequency f_p. The measuring signal is the amplitude of forced oscillations f_p of loaded cantilever, which after loading the lever varies from value A_1 to A_2. A more

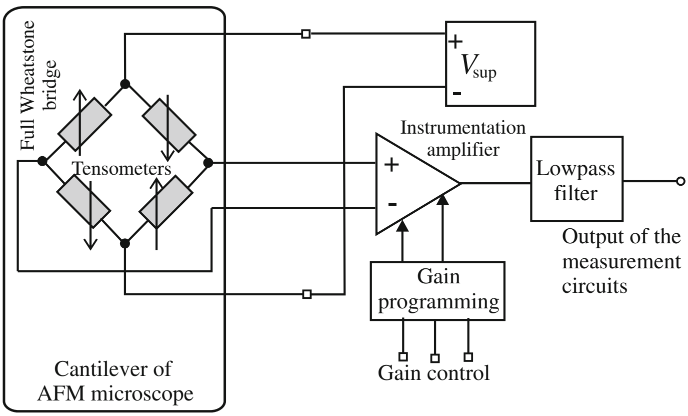


Fig. 11.10 Cantilever deflection measurement for piezoresistive tensometers

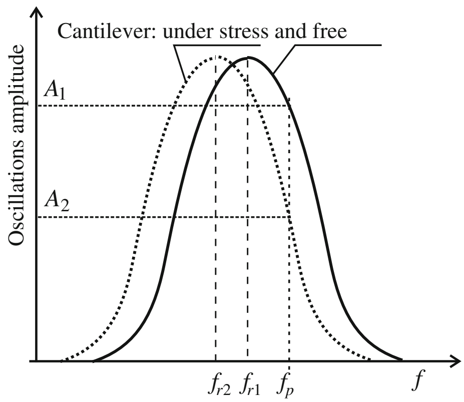


Fig. 11.11 Oscillations amplitude of AFM lever in the function of frequency for free cantilever and loaded cantilever, respectively

complex is the system of AFM resonance with frequency detection. In the system a self-excited oscillation generator is set up, and the microscope cantilever is a device in the circuit of generator feedback. The generator is induced to oscillate with cantilever resonance frequency f_r. By measuring the frequency of generator oscillations, it is possible to determine the atomic forces acting to its cantilever. AFM resonance is applied in the range of atomic forces both attractive and repulsive. The AFM are used for imaging the sample surfaces.

Apart from measuring the surface topography, the resonant AFM are also applied in biology to measure particle mass. The investigated particle placed on the microscope cantilever increases its mass and decreases the frequency f of resonance oscillations. The frequency f of cantilever oscillations is determined with formula (11.3).

$$f = \frac{\lambda^2}{2\pi}\left(\frac{k}{3m}\right)^{1/2} \tag{11.3}$$

where: λ—coefficient dependent on the number of modes of cantilever oscillations, m—effective cantilever mass (together with the investigated particle), k—cantilever elastic constant.

11.4 Electrostatic Force Microscope

Electrostatic force microscope (EFM) is a useful measuring instrument applied in research, production technology and in equipment operation. It is applied in areas so distant from one another as manufacturing electronic components, and high voltage technique (electric power engineering). The EFM tip is placed at the height b ($b \geq 5$ nm) above the sample surface. Electric charge Q_1 on the tip apex interacts with the charge on the sample surface Q_2. The force of electrostatic interaction of apex charges is determined with the Coulomb law (11.4).

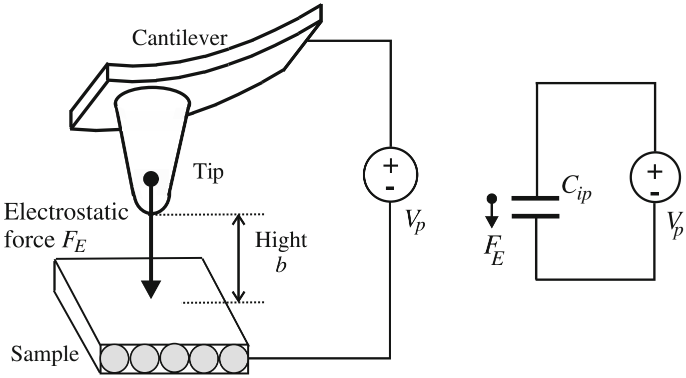


Fig. 11.12 Block diagram of an electrostatic force microscope (EFM)

$$F_E = \frac{1}{4\pi\varepsilon} \times \frac{Q_1 Q_2}{b^2} \tag{11.4}$$

where ε—permittivity of medium (e.g. air), Q_1 and Q_2—electric charge, b—distance between charges.

Between the microscope tip and the sample there is an electric capacity C. Polarization of this capacity with electric voltage V_p means conducting electric charge: $Q = C_{ip}V_p$ to both “plates” of “tip-sample” capacitor. After polarization of capacitor a “tip-sample” voltage, opposite polarized charges $+Q$ and $-Q$ attract one another. Electric strength of the air is of the order of 100 kV/mm. For EFM working in the air, the approach of the tip to the sample at the height of 5 nm enables maximum capacitor polarization with voltage 0.5 V.

Controlling the EFM cantilever and measuring the cantilever deflection is carried out similarly as for the ATM. An EFM system was shown in the diagram in Fig. 11.12.

11.5 Scanning Thermal Microscope

Scanning thermal microscope (SThM) is applied to measure temperature distribution on the sample surface and to measure thermal properties of samples by means of atomic probes with micrometer- or nanometre-size [3]. Such measurements are of great significance to the development of technology of semiconductor integrated circuits. One of major limitations for further miniaturization of integrated circuits and for an increase in the packing of electronic devices is the problem of heat exchange. The smallest size of single electronic devices and conductive paths contained in integrated circuits are currently equal to ca 14 nm (2014), but 15-year technological forecasting predicts a decrease in these dimensions beneath 5 nm (see Table 7.1). More and more complex integrated circuits are investigated with microscope tests, as a whole or in part of their structure, and the fields of scanned samples are still smaller.

Heat flow in objects of nanometre size (or between such objects) can require taking into account the quantum effects, including the quantization of thermal conductivity (see Chap. 7). Quantization of thermal conductivity in one-dimensional system 1-DEG, e.g. in nanowire, was theoretically predicted by P. Středa [10] with the use of the Landauer theory of conductivity. The effect of thermal conductivity quantization was confirmed experimentally by Schwab [8], who also measured the quantum of thermal conductivity G_{T0}. The value of thermal conductivity quantum is a linear function of temperature: $G_{T0} = (\pi^2 k_B^2/3h)T = 9.5 \times 0^{-13}T$; for $T = 300$ K $G_{T0} = 2.8 \times 10^{-10}$ [W/K].

The measuring signal in SThM is the heat stream flowing between the sample and microscope tip. Two versions of SThM are known: passive SThM (static) and active SThM.

In *passive scanning thermal microscope* the temperature sensor is integrated with the tip, and the microscope itself is used to measure the temperature distribution on the investigated surface—Fig. 11.13.

During the measurement the SThM tip touches the investigated sample surface. The temperature sensors used in the SThM are thermocouples or sensing resistors. Because the sensitivity of thermocouple does not depend on the diameter of wires that constitute the device, the thermocouple wires can be arbitrarily thin and the size of temperature sensor integrated with the SThM tip—very small. Thermocouple can be also produced through spraying thin-film metal electrodes or etching the electrodes.

In [7] SThM sensor was described, in the form of a thermocouple Au-Ni, placed on a cantilever made of nitride Si_3N_4. The sensitivity of this thermoelement is equal to 14 μV/K. The achieved resolution of the temperature measurement with thermocouple in SThM was equal to 5 mK. In [9] Shi gives an example of record-breaking geometric resolution in the measurement of temperature distribution for

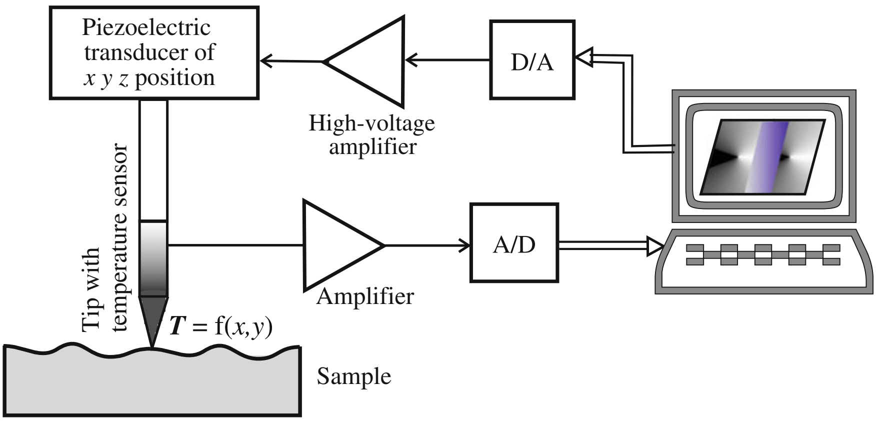


Fig. 11.13 Block diagram of passive SThM

carbon nanotube. With SThM with thermocouple a temperature distribution along the carbon nanotube 4 µm long and 70 nm in diameter was determined, and the temperature difference 20 K was measured at the ends of the nanotube.

The resistance sensors of SThM are platinum wires with diameter not greater than 10 µm, sprayed onto the SThM cantilever with a beam of electrons. Another version of resistance sensors are the sensors made of so-called Wollaston wire, i.e. platinum wires in silver covering. The Wollaston wire has good elastic properties (constant ca 5 N/m), and the SThM sensor made of such wire is part of microscope cantilever—Fig. 11.14. T. Gotszalk gives the size of resistance tip-sensor, produced at the Wroclaw University of Technology out of platinum wire with diameter of 70 nm [5]. The tip has the form of V letter, with added apex (apex made of Pt wire, 70 nm long, 2 µm wide and 3 µm high. The tip-sensor is placed on aluminium beam.

In *active SThM* (Fig. 11.15) the tip is at the same time the temperature sensor and a radiator supplying energy for heat exchange between the tip and probe. Active

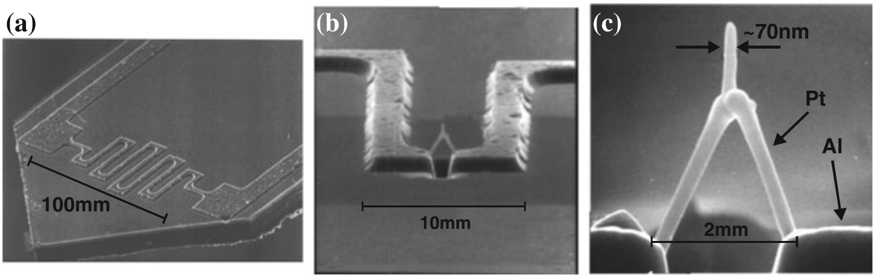


Fig. 11.14 SThM lever with Pt tip placed at the end: **a** lever end, **b** enlarged image of lever with visible temperature tip-sensor, **c** platinum tip (picture from [5] published upon of T. Gotszalk's consent)

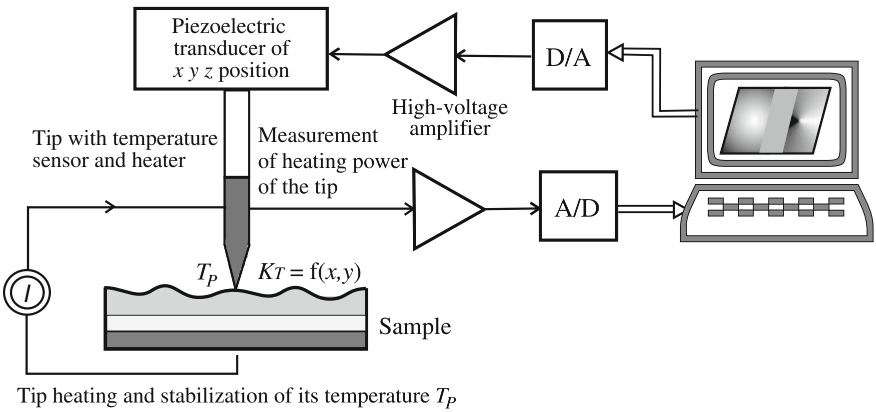


Fig. 11.15 Block diagram of SThM

SThM is used for imaging the thermal conductance K_T of the sample. During the scanning a given temperature T_P of the microscope tip is maintained, higher than the sample temperature. For tips made of Wollaston wires a given temperature can amount even to 700 K. The heat flow from tip-radiator to the sample requires a differentiated heating power in particular scanning points, if the sample thermal conductivity in those points takes different values. The measurement of heating power in points with coordinates *x–y* enables to build an image of thermal conductivity of the investigated surface.

Testing thermal conductance K_T of a sample with SThM can concern either the sample surface only, i.e. its upper layer, or also the conductivity of some selected layers of the sample. In the latter case, the sample is heated by an electric current with different frequency. Differentiated frequency results in a variable depth of penetration of heating electromagnetic wave, delivered by the tip-radiator, into the sample. Linear (geometric) resolution and thermal resolution of SThM depend on:

- Size of the end of microscope tip (the thinner is the tip apex, the better resolution),
- Thermal properties of a whole probe, i.e. lever with tip,
- Thermal contact of the tip with the investigated sample,
- Measuring frequency of electric temperature measuring system (with resistance sensor or thermocouple),
- Proper noise of measuring signal amplifiers in the system.

The achieved resolution values of SThM amount to 10 nm (linear resolution) and 5 mK (thermal resolution).

11.6 Scanning Near-Field Optical Microscope

Scanning near-field optical microscope (SNOM) is dedicated to investigate optical properties of samples with linear resolution of the order of several dozen nanometers. The properties under investigation are light reflection coefficient or transmission. Those measurements are of great significance in electronic technology (in manufacturing optoelectronic devices), in chemical technology and biology (Fig. 11.16).

The SNOM tip is a source of light illuminating the investigated sample or it is a receiver of light transmitted or reflected away from the sample. The tip is placed 100 nm high above the sample surface. Note that the distance between the tip and the sample is a few times shorter than the length of wave of the activating signal—light wave (380–780 nm), which enables a better geometric (linear) resolution of SNOM. Depending on the role of the tip in the system (light transmitter or receiver), the microscope works either in transmission mode or in reflection mode.

In *transmission operating mode* of SNOM the light source illuminating the sample is placed on the tip—Fig. 11.17.

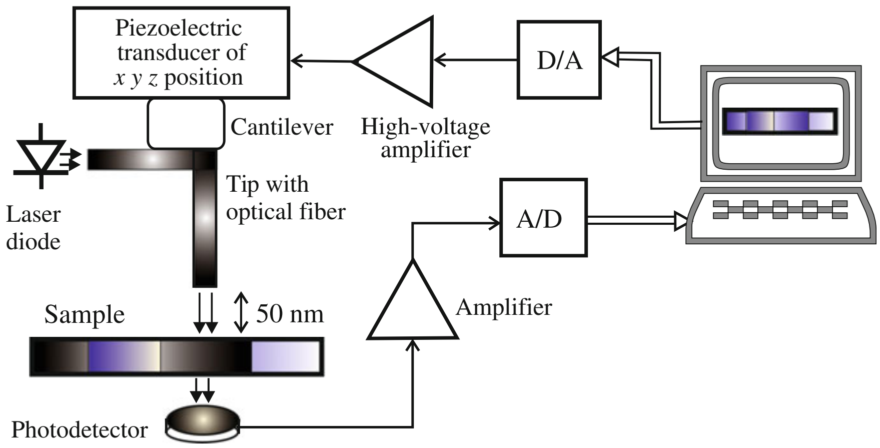


Fig. 11.16 Block diagram of scanning near field optical microscope SNOM with transmission working mode

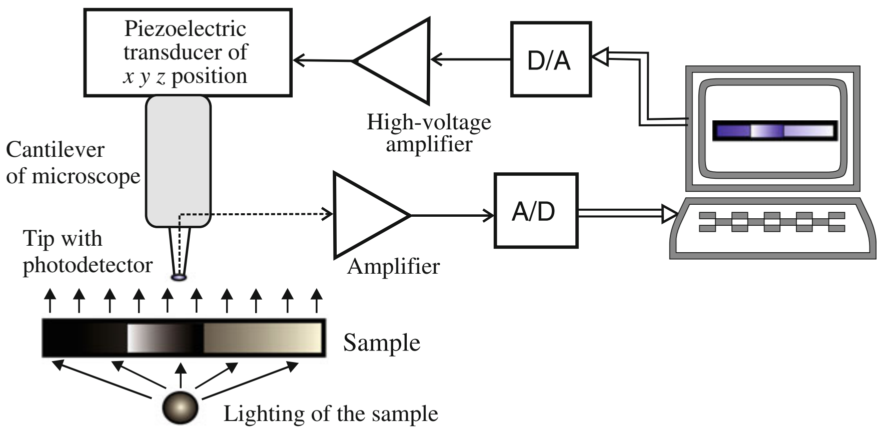


Fig. 11.17 Block diagram of scanning near field optical microscope SNOM with in the transmission operating mode

The light source for the microscope cantilever can be a thin optical fiber connected with the lever, with a diameter of the order of several dozen nanometers, which takes the role of tip. The optical fiber itself gains light from a laser diode, or from a set of diodes producing light with differentiated colour. An advantage of the fiber integrated with the tip is a possibility of precise orientation of the ray illuminating the sample.

A laser diode of a SNOM microscope can be placed at the end of the cantilever, performed in the form integrated with the cantilever.

In another version of the *transmission operating mode* of SNOM (Fig. 11.17) the investigated sample is illuminated with a light stream of uniform intensity. After light penetration through the sample its intensity is measured on the whole surface investigated, with the use of a photodetector (usually a photodiode), installed at the end of the microscope tip. Such investigation in the transparent mode of SNOM is useful in chemistry, in molecular biology, in medicine and material nanotechnology.

11.7 Opportunities of Scanning Probe Microscopy Development

Now several types of scanning probe microscopes are known [NATO ASI], which use different short-range interactions. Apart from the types of SPM discussed above, the following probe microscopes have been developed, among others:

- Magnetic force microscope (MFM)
- Scanning capacitance microscope (SCM)
- Scanning confocal fluorescence microscope (SCFM)
- Scanning electrochemical microscope (SECM)
- Scanning spread impedance microscope (SSRM)
- Scanning surface potential microscope (SSPM) or High-resolution Scanning
- Kelvin probe force microscopy in corrosion science
- Scanning impedance microscope (SIM), Kalinin 2002
- Nanoscale impedance microscope (NIM), Shao 2003
- Piezo-response force microscope (PFM) dedicated to observe local changes in ferroelectric domains—scanning gate microscope (SGM) dedicated to observe, among others, quantum rings in low temperatures and strong magnetic fields.

References

1. G. Binnig, C.F. Quate, Ch. Gerber, Atomic force microscope. Phys. Rev. Lett. **56**, 930–933 (1986)
2. G. Binnig, H. Röhrer, *Scanning Tunneling Microscope—from Birth to Adolescence*. Nobel Lectures, vol. 1981–1990. (World Scientific, Singapore, 1993), pp 389–408
3. D. Cahill, K. Goodson, A. Majumdar, Thermometry and thermal transport in micro/nanoscale solid-state devices and structures. J. Heat Trans. **124**, 223–241 (2002)
4. R.D. Deslattes, Optical and x-ray interferometry of a silicon lattice spacing. Appl. Phys. Lett. **15**, 386–388 (1969)
5. T.P. Gotszalk, *Near-Field Microscopy of Micro- and Nanostructures (in Polish)* (Publishing House of Wroclaw University of Technology, Wrocław, 2005)
6. C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996)
7. K. Luo, Z. Shi, J. Lai, A. Majumdar, Nanofabrication of sensors on cantilever probe tips for scanning multiprobe microscopy. Appl. Phys. Lett. **68**, 325–327 (1996)

8. K. Schwab, E.A. Henriksen, J.M. Worlock, M.L. Roukes, Measurement of the quantum of thermal conductance. Nature **404**, 974–977 (2000)
9. Z. Shi, S. Plyasunov, A. Bachtold, P. McEuen, A. Majumdar, Scanning thermal microscopy of carbon nanotubes using batch fabricated probes. Appl. Phys. Lett. **77**, 4295–4297 (2000)
10. P. Středa, Quantised thermopower of a channel in the ballistic regime. J. Phys.: Condens Matter **1**, 1025–1028 (1989)

Chapter 12
New Standards of Mass

Abstract In this chapter we discuss the efforts to redefine the unit of mass, the kilogram, and to develop a standard adequate to the new definition. The kilogram is likely to be redefined in relation to Planck's constant. The work in progress discussed in this chapter aims to determine Planck's constant by means of a watt balance and the Avogadro constant by counting the number of atoms in a mass standard of monocrystalline silicon ^{28}Si (XRCD method). For the new definition to be adopted the uncertainty of the results of measurements performed by three different teams should be better than 5×10^{-8} and the results obtained by different methods should agree. Apart from the two methods mentioned above, we discuss the development of a mass standard based on the accumulation of heavy metal ions of gold (^{197}Au) and bismuth (^{209}Bi).

12.1 Introduction

The accuracy of measurement of each physical quantity in the International System of Units (SI) is limited by the uncertainty with which the SI base units are realized by the standards. This makes understandable the striving for maximum accuracy and simplicity of the standards of base units. The simplicity of the standards is required for their dissemination throughout the world in a large number of calibration laboratories. Complicated and expensive standards will not come into common use.

One of the seven basic units of the SI system, the kilogram, is defined and realized by the international prototype of the kilogram (IPK) kept at the BIPM in Sèvres. The stability of the mass standard and the uncertainty of realization of the unit of mass are of importance not only for measurements of mass, but also for the realization of three other SI base units: the mole, the ampere and the candela (see Sect. 2.2). The kilogram is directly referred to in the definition of the mole. The ampere is defined in relation to force, which is expressed in Newton, N = m × kg × s^{-2}. The definition of the candela refers to power, measured in watts, W = m^2 × kg × s^{-3}. Besides its effect on the mole,

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9_12

the ampere and the candela, the quality of realization of the kilogram is of importance for the definition and realization of a large number of derived units of the SI system, including the units of all the mechanical quantities.

For 25 years scientists have been working on a new definition of the kilogram that would relate it to a fundamental physical constant, similarly to the already developed new definitions of some other units, such as those of time, length or voltage. There is a need for a standard of the kilogram (or, in general, a standard of the unit of mass) based on a fundamental physical constant and more accurate than the current standard, providing an accuracy of 5 parts per 10^8 (long-term stability). Much effort has been made to improve the status of the kilogram in the SI system and render it equal to that of the other SI base units. The new standard should ensure accurate realization of the unit of mass at any place and any time in a laboratory equipped with an appropriate apparatus operated by experienced staff.

Three physical constants have been considered for the redefinition of the kilogram: the electron mass m_e, the Planck constant h and the Avogadro constant N_A. The idea to define a standard mass in relation to parameters of molecules or atoms dates back to James Clerk Maxwell, which said: *If the Earth changes, it will be the same Earth, if a molecule changes, it will be a molecule of another substance*. On this basis Maxwell concluded that standards should not be found in the macroscopic world, but on the microscopic level, where characteristics remain unchanged. A standard mass could be the mass of a specific number of molecules (or atoms) of a chosen substance.

The idea to define the kilogram in relation to the electron mass was the first to be abandoned, because of the great technical difficulties encountered in the development of the appropriate standard. Arguments for redefining the kilogram were presented in 2005 in *Metrologia*, the official journal of the BIPM, along with a discussion of many arguments in favor of relating the unit of mass with either the Planck constant or the Avogadro constant in the new definition [8]. The article pointed out that further improvement of the uncertainty of realization of the unit of mass is impossible with the current arbitrary standard, the international prototype of the kilogram. Two SI base units defined in relation to physical constants, the second and the meter, are realized with an uncertainty better by a few orders of magnitude than the uncertainty of realization of the kilogram. Moreover, the units of the electrical quantities: voltage and resistance, in spite of being only derived units in the SI system are realized with better repeatability and lower standard deviation than the kilogram. The units of both voltage and electrical resistance are defined in relation to quantum phenomena and fundamental physical constants. Thus, the unit of mass, the kilogram, should be similarly defined in relation to a fundamental physical constant.

The mass of a particle can be determined directly or by measuring the Avogadro constant N_A (for example, with a monocrystalline silicon sample), or by a method based on accumulation of ions from an ion beam (for example, a beam of gold ions). In the approach based on the Avogadro constant the number of atoms and their total mass are determined in a silicon block, the dimensions of which is well known from measurements. The ion accumulation technique consists in determining the mass of accumulated ions, the number of which is determined on the basis of measurements of the electric current carried by the ion beam, and the time of accumulation.

As mentioned above in Chap. 3, the redefinition of the kilogram is still under discussion. It remains to be decided whether the unit of mass will be defined in relation to the Planck constant or the Avogadro constant. The latest discussions at the 24th CGPM in 2011 incline towards the Planck constant [11]. However, considering the previous proposal, with the Avogadro constant as the first candidate, and the related experience, the planned presentation of a new definition relating the kilogram with the Planck constant at the 26th CGPM in 2018 seems uncertain. Two versions of the new definition of the kilogram are proposed [9]. One version refers to mass-energy equivalence:

> The kilogram is the mass of a body whose equivalent energy is equal to that of a number of photons whose frequencies sum to exactly [(299 792 458)^2/66 260 693] × 10^{41} Hz.

This definition relates the kilogram to the Planck constant h, the value of which results from the Einstein formula:

$$E = mc^2 \tag{12.1}$$

and the Planck formula:

$$E = hv = hc/\lambda \tag{12.2}$$

for photons with a frequency v and a wavelength λ (c denotes the speed of light in vacuum).

In the other version the considered new definition of the kilogram refers to the de Broglie-Compton frequency (also known as the Compton frequency) of a body with a mass of one kilogram:

> The kilogram is the mass of a body whose de Broglie-Compton frequency is equal to exactly [(299 792 458)^2/(6.6260693 × 10^{-34})] Hz.

Referred to by this definition, the de Broglie-Compton frequency of a body with a mass of one kilogram is defined by the formula:

$$v_m = \frac{c}{\lambda_{C.m}} = \frac{mc^2}{h}, \tag{12.3}$$

where $\lambda_{C.m} = h/(mc)$ is the Compton wavelength of the body, analogical to the Compton wavelength λ_C of the electron: $\lambda_C = h/(m_e c)$.

Each of the considered definitions of the kilogram has its advantages. The first one is based on well-known fundamental relations of physics and can be understood by a wide audience. The advantage of the other definition is that it refers to the de Broglie-Compton frequency and the *h/m* ratio, which is often measured in molecular physics. Two examples of devices sensitive to the *h/m* ratio are:

- The atomic interferometer, which measures the ratio $h/m(^{A}X)$ for molecules of elements, where ^{A}X is the chemical symbol of the element with its atom number specified, e.g. ^{28}Si;
- The watt balance, which measures the ratio h/m_s, where m_s denotes the mass of a macroscopic standard, typically in the range from 100 g to 1 kg.

Studies aimed at the development of a new standard of the kilogram are carried out in many directions. Comparable quality of realization of the kilogram has been obtained with a standard based on the Planck constant h and a standard referring to the Avogadro constant N_A. Further in this chapter we shall discuss laboratory setups for the realization of the kilogram by new standards.

12.2 Mass Standards Based on the Planck Constant

12.2.1 Watt Balance Standard

In the currently recommended direction of development of mass standards the unit of mass is related to the Planck constant h through electric energy, according to the proposed redefinition of the kilogram. Three solutions for the standard in accordance with the proposed new definition are being developed. In the considered standards:

- The mass of a superconductor is levitated by a magnetic field generated by a current flowing in a coil;
- The gravitational force on the mass is balanced by the attractive electrostatic force of an electrometer plate;
- A watt balance with a moving coil measures the force acting on the coil with current in a magnetic field. The required coefficient of mechanical force exerted on the coil by the magnetic flux is determined directly by the measurement of the voltage generated when the coil is in motion. It is also necessary to measure the velocity of the coil; an interferometer and a stopwatch are used for this purpose. The realization of the unit of mass by the watt balance standard is based on a virtual comparison of the mechanical power with the electrical power. The standard holds promise to realize the unit of mass with an uncertainty of the order of one part per 10^8.

Figure 12.1 shows the block diagram of the standard of the unit of mass with mechanical and electrical energy balance installed at the National Physical Laboratory in the UK. In the NPL system a very large permanent magnet generates a magnetic field with a 'horizontal' induction vector $\boldsymbol{B}$.

The magnetic flux acts on a horizontal circular coil suspended from one arm of the balance. The current I in the coil generates a 'vertical' force, which is balanced by the gravitational force $g \times M$ on the mass M. After the weighing experiment the coil is set in vertical motion (along the z axis) by rotating the arm of the balance,

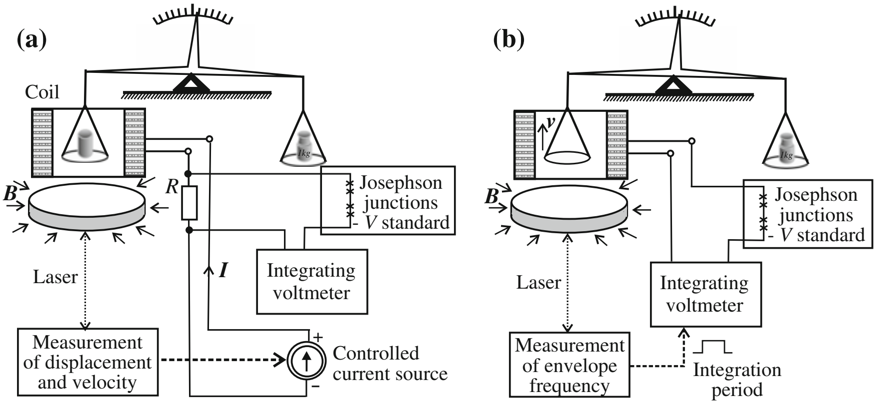


Fig. 12.1 Standard of the unit of mass based on mechanical and electrical energy balance: **a** weighing experiment, **b** moving experiment

and passes through the initial position with a velocity v, as a result of which a voltage V is generated in the coil. Since the energy E of the interaction between the current I and the magnetic flux Φ is $I\Phi$, the measured vertical component of the interaction force is:

$$\frac{\partial E}{\partial z} = I\frac{\partial \Phi}{\partial z} = M \times g, \tag{12.4}$$

where g is the acceleration of gravity experienced by the mass M.

The voltage V and the magnetic flux Φ fulfill the relation:

$$V = \frac{\partial \Phi}{\partial t} = \frac{\partial \Phi}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial \Phi}{\partial z}v. \tag{12.5}$$

The elimination of the rate of change $\partial\Phi/\partial z$ of the magnetic flux between the weighing and moving experiments leads to the basic equation of the standard:

$$M \times g \times v = I \times V. \tag{12.6}$$

Variation of the gradient of the magnetic flux and the components of the linear and angular velocities must be taken into account or reduced by careful positioning (alignment in the axis) of the coil and the other parts of the equipment with respect to the vertical axis so that the effect of these additional factors is negligible. In practice this means that the individual angular deviation must be below 10^{-4} rad.

The electric current I in the coil is determined from Ohm's law, $I = V/R$, with the voltage measured with a quantum standard based on the Josephson effect, and the resistance with a standard using the quantum Hall effect. Constant potential difference ΔV of up to a few volts can be measured with an uncertainty of 1 part per 10^9.

The voltage V realized by the Josephson standard is a function of the frequency f of irradiation of the Josephson junctions, the physical constants e and h, and a natural number n_1 (see Chap. 3):

$$V = n_1 f\,(h/2e). \tag{12.7}$$

Resistance can be also measured with an uncertainty of 1 part per 10^9 as a function of the quantum Hall resistance R_H (see Chap. 6):

$$R_H = (h/e^2)/n_2. \tag{12.8}$$

Denoting the plateau number in the characteristic of the quantum Hall effect sample, the integer n_2 ranges from 2 to 4. If $n_2 = 2$, then $R_H \approx 12\,906\ \Omega$; this value can be scaled by a factor k in order to measure any resistance in the range from 1 Ω to 100 kΩ with a Harvey cryogenic current comparator. With a Josephson effect standard and a quantum Hall effect standard the current I can be determined in the range from a few milliamperes to a few amperes, by the formula:

$$I = \frac{V}{R} = \frac{n_1 n_2 f e}{2k}. \tag{12.9}$$

The relation (12.6) for the standard of the unit of mass can be converted to the (12.10)–(12.12) below:

$$Mgv = \frac{V}{R}V \tag{12.10}$$

$$Mgv = \left(\frac{n_1 h}{2e}f\right)^2 \frac{n_2 e^2}{h} \tag{12.11}$$

$$M = k_W \times h, \tag{12.12}$$

where h is the Planck constant and the coefficient k_W includes g, v and f, quantities determined by the measurement of length and time and defined by physical constants and the integers n_1 and n_2.

The relation (12.12) indicates that in measurements using the standard with mechanical and electrical energy balance the measured value of mass M is mainly determined by the Planck constant, and to a lesser degree by the parameters of the equipment.

The formula (12.6) compares the *virtual* electrical and mechanical power. No *actual* energy dissipation, such as that due to friction in the bearings of the balance or electrical heating of the coil, enters this equation. This provides grounds for adopting the (12.6) with an uncertainty required in the metrology of standards.

The measurement technique used in the NPL [6] differs slightly from the general concept presented at the beginning of this section. The balance is biased with a mass of 500 g. The current I in the coil generates a force that balances the gravitational force on this mass, $g \times 500$ g. Then, a reference mass (e.g. a mass of 1 kg) is placed on the same scale as the coil, as a result of which the current in the coil changes direction to maintain the equilibrium. The watt balance is kept in equilibrium by an adjusted current I, the intensity of which is controlled by a computer. Thus, the gravitational force $g \times M$ corresponds to the total change in current in the coil. This substitution method eliminates many systematic errors of electrical, magnetic and mechanical origin. The effect of the hysteresis of the balance pointer (due to its inelastic deformation) is minimized by reducing the deflection of the balance in the weighing experiment. Also, before each observation the arm of the balance is set in oscillation with decaying amplitude to reduce and randomize the offset related to friction.

The mass samples used at the NPL are made of gold-plated copper to reduce the interaction by the dispersed magnetic flux from the magnet. The samples are stored in vacuum, with the exception of the short time interval necessary for their transfer to and from vacuum containers. The samples are compared with the BIPM prototype of the kilogram by the agency of its NPL copy.

The coil is moved in the vertical direction by rotating the arm of the balance to induce a voltage V in the coil. This causes also slight horizontal movements of the coil and, as a consequence, its swinging as well as a slight extra voltage in the coil. Contributing to the measurement error, this extra voltage is minimized by careful alignment of the coil in the axis. The effect of the horizontal swinging is reduced by averaging the voltage generated as a result of this motion over one period of oscillation or a longer time. Substrate vibrations present another problem, destabilizing the motion of the coil (causing velocity noise) and generating as much as ca. 0.2 % of correlated voltage. These effects are reduced to a minor level also by averaging both the velocity of the coil and the voltage it generates.

The density of the magnetic flux from the permanent magnet varies with temperature with a coefficient of 400 parts per 10^6 K^{-1}, and with the magnetic field in the environment. Even though the parameters of the environment are stabilized in order to reduce the fluctuations in the magnetic flux density, the magnetic flux can still change slightly in the time necessary for the weighing and moving experiments. The effect of changing magnetic flux can be taken into account on the assumption that the changes in the flux are continuous and slow. The results of the weighing and moving experiments are used for the determination of the constant reference mass, which is the final result of the experiment. The expected uncertainty of realization of the unit of mass by the system used at the NPL is below 1 part per 10^8 in a single series lasting the whole night. Similar performance is expected at the NIST and METAS (Bundesamt für Metrologie, Switzerland). The understanding of the sources of the type B uncertainty in this standard of the unit of mass and the correction of their effects will certainly take some time.

12.2.2 Levitation Standard and Electrostatic Standard

Another mass standard under development uses *magnetic levitation*. In this system a superconductor is suspended at a precisely measured height by an electrodynamic force which counteracts the effects of the gravitational force on the mass of the superconductor. Resulting from the Meissner effect, discussed in Chap. 2, the levitation of a superconductor in a magnetic field requires cooling the superconductor to a temperature below its critical temperature T_c. The mass standard system measures the electric energy generating a magnetic field necessary for the magnetic levitation. Each undesirable energy loss, e.g. due to deformation of the levitated object, must be taken into account in the measurement uncertainty balance. In a system built at the National Research Laboratory of Metrology (NRLM) in Tsukuba, Japan [4, 5], a high-purity 99.9 % niobium disc with a mass M = 170 g was used as a reference mass. The disc was cooled with liquid helium to 4.2 K in order to obtain the superconducting phase. Also other important elements of the system, such as coils with current for the generation of the magnetic field and an optical interferometer for the measurement of the height at which the niobium disc was suspended, were placed in helium bath. The coils were made of aluminum, which is a nonmagnetic metal. As a result of magnetic levitation the disc was suspended at a height of 15 mm, depending on the current in the coils and the magnetic induction they generated. The uncertainty of realization of the reference mass by this NRLM standard was of the order of 10^{-6} [5].

The *electrostatic mass standard* with a moving-plate air/vacuum capacitor has been studied for many years at the University of Zagreb [2]. A number of versions of the standard were built. In the first version the gravitational force on a mass of 2 g was counterbalanced by the electrostatic force produced by electric charge accumulated on the plates of a cylindrical air capacitor at a voltage of 10 kV. The distance between the capacitor plates was 7 mm, and the capacitance was changed by moving coaxially the movable plate. The generation of an electrostatic force adequate to the controlled capacitance and to different voltage across the capacitor was assumed in the measurement procedure. In the last version the system was fit for comparing a mass of 2 × 500 g. A multiple-plate air capacitor with adjustable capacitance was used in this version. The polarization voltage was increased to 100 kV. The electrostatic force acting on the plates of the capacitor was counterbalanced by the gravitational force for two values of capacitance of the air capacitor. This method, however, involved difficulties related to the low value of the electrostatic force and the necessity to use expensive components and demanding high-voltage technologies. For example, rectification and filtering of alternating voltage to obtain a direct voltage of up to 100 kV with the allowable pulsation amplitude of a few millivolts were problematic in this technique. The target uncertainty of realization of the mass by this method was estimated at 10^{-8} [2].

12.3 Silicon Sphere Standard

12.3.1 Reference Mass and the Avogadro Constant

The realization of the definition of the kilogram in terms of the Avogadro constant involves a standard with a known number of identical atoms or molecules, determined ('counted') on the basis of its size. One of the methods that have been developed for many years in a few laboratories in this direction of research consists in counting atoms in a silicon sphere of monocrystalline ^{28}Si. The number of atoms is determined from the size of the sphere, which is measured with an optical interferometer. In another version of this strategy, also still in the phase of development, ionized gold atoms, compressed in vacuum to form a beam, are accumulated by a collector, which is weighted subsequently. The number of ions accumulated by the collector can be determined from the electric current carried by ions in the beam and the time of accumulation (see Sect. 12.4).

Studies towards the definition of the kilogram as the mass of a specific number of silicon atoms in an almost perfect crystal have been conducted for many years [5]. The spherical shape chosen for the silicon standard has the advantage to allow precise measurements of the diameter of the sphere and control of its size, i.e. deviations from the assumed perfectly spherical form. In contrast to mass standards in the form of a cylinder or a cube, a spherical standard does not have edges, the possible damage of which would result in a loss of mass.

With the currently available measurement techniques it is not possible to count directly all the atoms in a crystal. Thus, indirect means must be resorted to. Two measurements are necessary to determine the number of atoms:

- The size of an 8-atom unit cell of the silicon crystal is measured with an X-ray interferometer. First of all, the lattice constant d, i.e. the spacing between atoms in the crystal lattice, is determined in precision measurements. The lattice constant of monocrystalline silicon is 0.235 nm, approximately, and the X-ray wavelength ranges from 10 pm to 10 nm.
- The diameter and volume of the silicon sphere are measured with an optical interferometer. The spherical shape of the silicon standard allows to measure precisely its size (diameter of the sphere) and check its shape for irregularities representing deviations from the assumed form (the perfect sphere, in this case).

Silicon used for the construction of the standard must have a monocrystalline structure, which ensures a uniform and regular distribution of Si atoms in the crystal lattice. Natural silicon is almost always polycrystalline, i.e. composed of variously oriented crystallites. Used in the semiconductor industry, monocrystalline silicon is obtained by crystal growth. The major method of growing monocrystals was developed in 1916 by the Polish technologist Jan Czochralski.

The results of the two measurements mentioned above can be used for calculating the number N_X of atoms of the given element in the determined volume of the sample (object). This, in turn, allows the calculation of the reference mass M by multiplying the mass m_S of a single atom by the total number N_X of atoms in the sample:

$$M = N_X m_S. \tag{12.13}$$

Perfect crystals are necessary for precise determination of the number N_X. This was the reason behind choosing silicon, the monocrystalline form of which is superior in terms of both structure and purity. The sought number N_X can be obtained by calculating the total number of silicon atoms in one mole. For a silicon sample (object) the Avogadro constant N_A is given by the formula:

$$N_A = \frac{M_{mol}}{m_{Si}} = \frac{n M_{mol} v}{V_0 m}, \tag{12.14}$$

where v is the volume of the sample, m denotes its mass, M_{mol} is the molar mass of silicon, and V_0 the volume of a silicon unit cell containing n atoms. The molar mass M_{mol} equals the molecular mass in number, but is expressed in grams.

The (12.14) implies that the Avogadro constant N_A defines the relationship between the macroscopic quantities v and m and the atomic quantities V_0 and M_{mol}. Being the number of atoms in one mole, the Avogadro constant represents the ratio of the molar volume to the volume of a single atom. The determination of the value of the Avogadro constant involves the measurement of the following quantities [2]:

- The volume occupied by a single silicon atom. The lattice structure and lattice constant of silicon crystal must be known for the determination of the atomic volume. Although monocrystalline silicon is an almost perfect crystal, the measurements should also take account of the impurities and defects in the silicon sample.
- The macroscopic density of the crystal that is the material of the sample.
- The isotopic composition of the silicon sample. Silicon is a mixture of three stable isotopes: ^{28}Si, ^{29}Si and ^{30}Si. Their relative masses have been determined with an accuracy of 1 part per 10^{10} by mass spectroscopy, but the exact proportion of each isotope in the sample depends on the geological conditions of the deposit and is difficult to measure.

The atomic volume and the mass density of silicon are measured in many laboratories, including the NIST, PTB, Istituto di Metrologia (Italy), NRLM (Japan) and the National Measurement Laboratory (Australia). The isotopic composition of a silicon crystal is measured by refined mass spectrometry or thermal neutron capture spectroscopy. The combination of these two methods reduce to 2 parts per 10^7 the contribution of the uncertainty related to the presence of different isotopes of silicon in the sample. Additional uncertainties are due to the presence of impurities in the sample and possible defects in the crystal lattice. Defined and measured in this way, the mass of a silicon sample can be compared to a platinum-iridium copy of the prototype of the kilogram with an accuracy of 1 part per 10^8, with the substitution weighing method employed. The weighing experiment is carried out in vacuum to reduce the effect of air buoyancy by Archimedes' principle on the measurement results. Differences in volume of the samples are of no importance under such conditions. The best results of density measurements have been obtained at the Institute for Reference Materials and Measurements of European Commission's science service in Geel, Belgium.

12.3.2 Measurement of Volume of Silicon Unit Cell

The volume of a single silicon atom is determined on the basis of measurements of the lattice constant *d*, i.e. the atomic spacing in the crystal lattice. The lattice constant *d* is used for the determination of the volume of a single unit cell by geometrical calculation. The sample should be checked for homogeneity of the crystal structure and chemical purity of the material. The first measurements of the lattice constant, performed 50 years ago, were based on X-ray diffraction in the sample. The diffraction angle was measured to provide the basis for the determination of the lattice constant from the Bragg law:

$$\lambda = 2d \sin \theta, \tag{12.15}$$

where λ is the wavelength of the X-rays, *d* the lattice constant, and θ the angle between the direction of the X-ray beam and the diffraction plane.

This method can be used on the condition that the lattice constant and the wavelength are comparable. The X-ray wavelength ranges from 10 pm to 10 nm. Measurements by the diffraction method are not precise.

Today the lattice constant is measured by X-ray interferometry, which is a much more precise measurement technique. The principle of measurement of the lattice constant of a silicon crystal is shown in Fig. 12.2 [1].

Three silicon plates, labeled C1, C2 and C3, cut from the same crystal are stacked so that their crystal planes are parallel. The construction is carefully designed so that the plate C1 can be moved along the *y* axis, in the plane of the other two plates. X-rays are incident on the plate C3 at an angle adjusted so that the diffracted X-rays pass through the plate along the direction perpendicular to its plane. The X-rays pass through all three plates to reach a detector. Due to interference, a slight shift of the plate C1 along the *y* axis will change the intensity of the detected radiation with maximums (and minimums) separated along the *y* axis by a distance equal to the lattice constant *d*. These maximums, detected by the X-ray detector, correspond to fringes observed in an optical interferometer.

The movement of the plate C1 is measured with an optical interferometer using laser light with a precisely known wavelength. Measurements of the shift along the *y* axis and the radiation maximum are repeated thousands of times in a series, and the results averaged to yield the lattice constant *d*. The measurement resolution is of the order of 10^{-15} m. The unit cell of a silicon crystal is assumed to contain eight

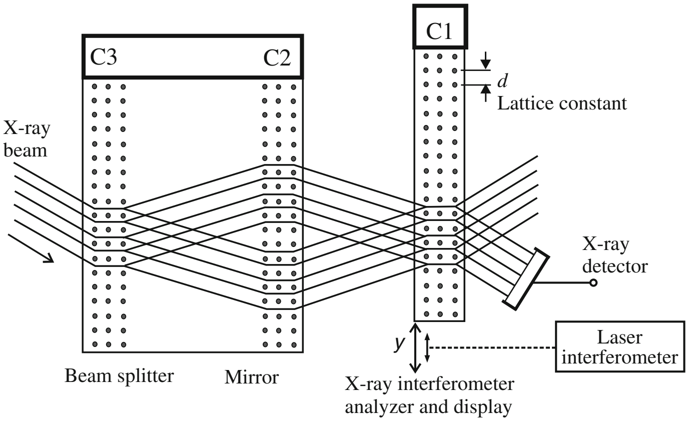


Fig. 12.2 Measurement of the lattice constant of a silicon crystal with an X-ray interferometer [1]

atoms and have the form of a perfect cube with a side *d*. The volume V_0 of the silicon unit cell:

$$V_0 = d^3, \tag{12.16}$$

where *d* is the lattice constant of the silicon crystal.

Unit cells measured in this way are not identical in size in separately grown crystals or even in plates cut from different parts of the same crystal block. Measurements of the parameters of the silicon crystal lattice have reached the accuracy of ±4 parts per 10^8. The results of measurements carried out in different laboratories agree within 1 part per 10^7. The average volume of a unit cell of the silicon crystal has been found to only slightly depend on the isotopic composition [6].

12.3.3 Measurement of Volume of Silicon Sphere

The measurement of the volume of the silicon sphere is one of three issues that are of equal importance in the fabrication of a silicon sphere standard (the other two issues are the measurements of the volume and mass of the unit cell). Of major importance is also the shape of the produced sphere, which should be as close as possible to the perfect sphere. The measurements are paramount, though. The results of measurements of the diameter of the sphere are used for perfecting its shape of for final determination of its volume. Scientists in a number of research centers, including the PTB in Germany, the NML in Australia and the Mendeleev University of Chemical Technology in Russia, work on the production of silicon spheres with a mass of 1 kg and a diameter of ca. 93 mm. Such a silicon sphere produced at the PTB is shown in Fig. 12.3.

Fig. 12.3 Silicon sphere with a mass of 1 kg and a diameter of 93 mm at the PTB, Germany (in the back—a 1 kg titanium sphere)

Figure 12.4 presents the block diagram of the PTB system [1] for the measurement of the diameter of the sphere with a spherical Fizeau interferometer. At the NML silicon spheres are produced by the conventional method that consists in grinding and polishing the surface of the sphere to perfection which is only limited by the presence of discrete particles (atoms). However, major difficulties are encountered in the attempt to reduce the uncertainty to the level of 1 part per 10^8. The diameters of the silicon sphere are measured in many directions with laser interferometers using lasers with a known wavelength. At the PTB the diameter of the silicon sphere is measured in 16 000 directions. The volume of the sphere is calculated from these diameters.

The accuracy of the volume calculated on the basis of these measurements is limited by the presence of films of silicon oxide formed on the surface, rather than by the imperfections related to the surface roughness of the sphere. The effect of the silicon oxide films must be taken into account in the calculation. These surface films affect also the results of the interferometric measurements of the diameter by changing the phase of the front of the reflected optical wave. Because of these two effects the uncertainty of definition of the kilogram is limited to a few parts per 10^7, approximately [6].

A silicon sphere closest to the perfect sphere is described in [7]. The deviations of the diameter of this sphere from the perfect shape were of 30 nm (for a diameter of 93 mm), and the surface roughness was 0.5 nm. The uncertainty of definition of the unit of mass can be reduced below 10^{-7}, since the thickness to the surface film can be determined by measuring the phase shift of optical waves by ellipsometry, which provides high sensitivity. As the average density of silicon oxide films is close to the density of silicon, the error related to the presence of these films is negligible in the determination of the volume.

A total uncertainty of ca. 4 parts per 10^7 can be achieved with all the specified uncertainties taken into account and the results of a large number of measurements

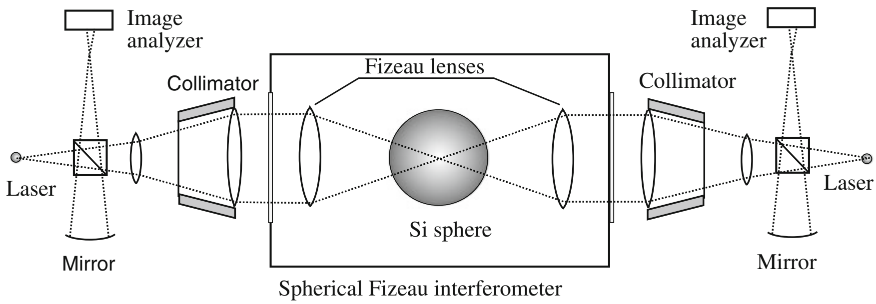


Fig. 12.4 Measurement of the diameter of a silicon sphere with a Fizeau interferometer

considered. Since all the uncertainty components are close to the best results in their measurement ranges, the 40-fold reduction of the uncertainty to the level of 1 part per 10^8, required for the redefinition of the kilogram, seems very difficult. A stable silicon sphere could replace the platinum-iridium prototype of the kilogram as a transition standard. Unfortunately, silicon standards have the same disadvantages (though perhaps to a lesser degree) as the present platinum-iridium international prototype of the kilogram: they can be damaged or destroyed, and their mass can change as a result of surface pollution, subsurface absorption of impurities, etc.

Current studies within the project aimed at the development of standards of the unit of mass defined in relation to the Avogadro constant tend to focus on direct determination of the Avogadro constant N_A rather than on the method of monitoring the stability of mass. Each laboratory can only undertake one or two tasks within the project.

12.4 Ions Accumulation Standard

An ion is characterized by an electric charge q and a mass m_a. An ion beam is a stream of electric charge carriers with an electric current I. A beam of emitted ions is used for the realization of the unit of mass by ion accumulation and counting. Ions accumulating on a collector increase its mass. The total charge Q transferred by the ion beam over a time T can be calculated if the current I and the accumulation time T are known. The electric charge allows to determine the number of ions accumulated on the collector and their total mass, representing a reference mass. Gold (^{197}Au) and bismuth (^{209}Bi) ions are used in studies aimed at the development of such a standard. Gold and bismuth were chosen because of their large atomic numbers and masses. Another advantage of gold in the mass standard application is that it only has one isotope, ^{197}Au. As determined in the most precise measurements to date, the atomic mass of gold is 196.966551, and the unified atomic mass unit m_u (representing one twelfth of the mass of the carbon atom) is $1.6605387313 \times 10^{-23}$ kg [10].

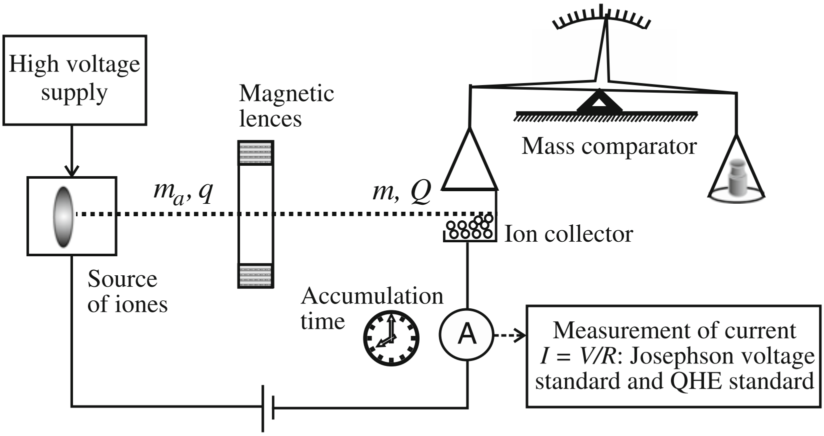


Fig. 12.5 Ion accumulation mass standard

The electric current I carried by the ion beam is measured with a Josephson effect voltage standard and a quantum Hall resistance standard. The current is determined from the relation (12.9) (see Sect. 12.2):

$$I = \frac{V}{R} = \frac{n_1 n_2 f e}{2k}.$$

If each ion carries a single charge e, then the number of transferred ions is given by:

$$N = \left(\int I \,\mathrm{d}t\right) \Big/ e, \tag{12.17}$$

where the integration is performed over the time T of accumulation of electric charges, or ions with a mass m_a each, on the ion collector. The total accumulated mass M is given by the formula:

$$M = m_a \frac{n_1 n_2}{2k} \int f \,\mathrm{d}t. \tag{12.18}$$

Figure 12.5 presents the principle of determination of mass by the discussed method. Some experimental difficulties are encountered in the employment of this technique. To achieve an accuracy of 1 part per 10^9, all the electrically counted ions must be accumulated without loss and all of them must have the same mass. This means that all other atomic particles, such as ions of other elements, must be set apart by means of a good mass spectrometer using the high-vacuum technology. Gold was chosen as the material for the standard of mass because it only has one stable isotope, its atoms are massive and its evaporation pressure is negligible at room temperature. If an element has more than one isotope, the proportion of each isotope in the ion beam should be specified, since isotopes always differ in the atomic mass. Metrologists hope to obtain electric current of ca. 10 mA from a gold ion beam, which will accumulate a mass of 10 g on the collector in 6 days. Such a mass could be compared with a reference mass with an accuracy within ±0.1 μg (which represents $\pm 3 \times 10^{14}$ gold atoms) by weighing the collector in vacuum.

References

1. P. Becker, M. Gläser, Avogadro constant and ion accumulation: steps towards a redefinition of the SI unit of mass. Meas. Sci. Technol. **14**, 1249–1258 (2003)
2. V. Bego, J. Butorac, D. Ilić, Realisation of the kilogram by measuring at 100 kV with the voltage balance ETF, IEEE Trans. Instrum. Meas. 1999, 48212–48215
3. U. Bonse, M. Hart, An X-ray interferometer. Appl. Phys. Lett. **6**, 155–156 (1965)
4. Y. Fujii, Y. Miki, F. Shiota, T. Morokuma, Mechanism for levitated superconductor experiment. IEEE Trans. Instrum. Meas. **50**, 580–582 (2001)
5. Y. Fujii, Possible contribution of levitating mass to force measurement. Metrologia **38**, 83–84 (2001)
6. B.P. Kibble, I.A. Robinson, Replacing the kilogram. Meas. Sci. Technol. **14**, 1243–1248 (2003)
7. A.J. Leistner, W.J. Giardini, Fabrication and sphericity measurements of single-crystal silicon sphere, Metrologia, **31**, 231–234 (1995)
8. M.I. Mills, P.J. Mohr, T.J. Quinn, B.N. Taylor, E.R. Williams, Redefinition of the kilogram: a decision whose time has come. Metrologia **42**, 71–80 (2005)
9. M.I. Mills, P.J. Mohr, T.J. Quinn, B.N. Taylor, E.R. Williams, Redefinition of the kilogram, ampere, kelvin and mole: a proposed approach to implementing CIPM recommendation 1. Metrologia **43**, 227–246 (2006)
10. P.J. Mohr, B.N. Taylor, CODATA recommended values of the fundamental physical constants: 2002. Rev. Mod. Phys. **77**, 1–106 (2005)
11. M. Stock, The watt balance: determination of the Planck constant and redefinition of the kilogram. Phil. Trans. Royal Soc. A **369**, 3936–3953 (2011)

Index

W. Nawrocki, *Introduction to Quantum Metrology*,
DOI 10.1007/978-3-319-15669-9