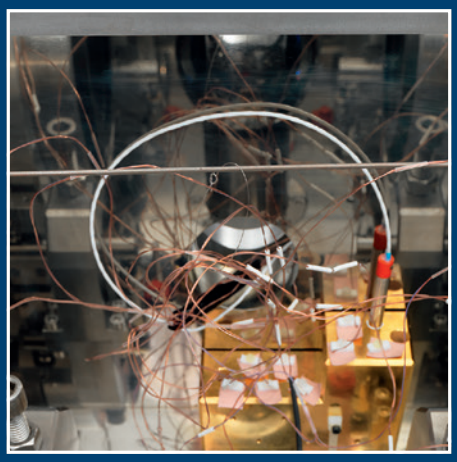
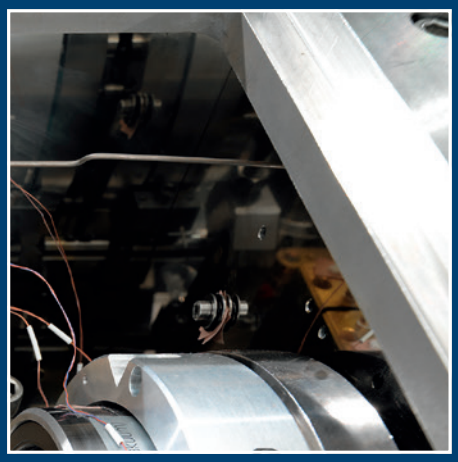
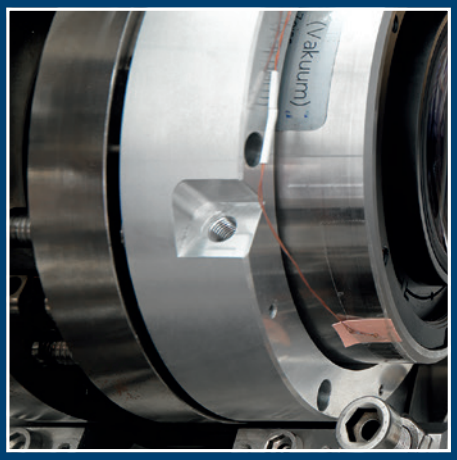
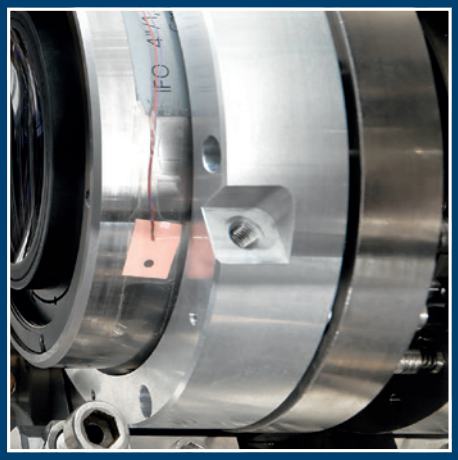
PTB mitteilungen S 2016

Special Issue
Experiments for the New SI

2nd Edition 07/2018

Physikalisch-Technische Bundesanstalt ■ Braunschweig and Berlin

National Metrology Institute

PTB mitteilungen

Experiments for the New SI

Cover picture:
On the way towards a new kilogram definition. In the sphere interferometer,
the diameter of the silicon sphere lying in the center is measured extremely accurately.
Source: PTB

PTB mitteilungen

Special Journal for the Economy and Science
Official Information Bulletin of the
Physikalisch-Technische Bundesanstalt
Braunschweig and Berlin

Volume 126 (2016), No. 2

Contents

Experiments for the new SI

Foreword

Jens Simon*

Today, we can already say with certainty that the autumn of 2018 will earn an entry in the history books of science. And perhaps it will be more than just the history of science that takes note of this event, but the history of civilization itself.

For it is in the autumn of 2018 that an undertaking will be signed and sealed that is the sum total of years and even decades of work at national metrology institutes in the very highest art of measurement: a fundamental revision of the International System of Units (*Système international d'unités*, or SI for short).

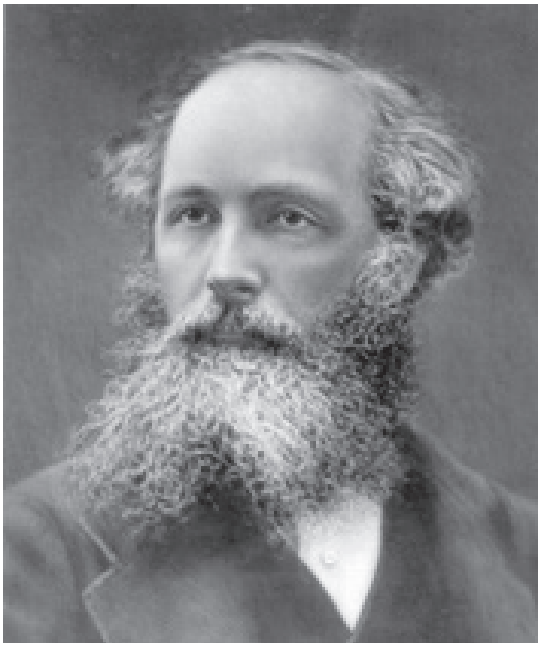
The new definition of the SI (base) units is so fundamental in nature that it is not an exaggeration to call it a paradigm change. Instead of a small, selected number of base units (with all of their obsolete, arbitrary and idealized notions), it is a set of fundamental constants that will determine the world's measurements from that point on. In contrast to all material measures, such constants represent those "objects" that are truly immutable.

In the system that comprises the units currently in use, the values of the fundamental constants are specified – leading to the remarkable situation that the values of the fundamental **constants** are in a permanent state of flux, since our measurement capabilities are reflected in these values. Beginning in the autumn of 2018, this relationship will become inverted: from the specified values of the fundamental constants, the units will be derived as an inference. Provided that the fundamental constants are truly constant, our system of units will then have the most solid and reliable foundation conceivable. The units will be universal, in the literal sense of the word: in principle, it will be possible to apply them throughout the entire universe. To put it simply, even a Martian could then understand what a kilogram is (a feat that, currently, is not possible, unless we send the Martian the prototype of the kilogram – the "metrologically holy" piece of metal from the safe of the International Bureau of Weights and Measures in Sèvres, France).

The notion of defining the units in such a universal way is not a new one, but was born in the closing decades of the 19th century. As early as 1870, James Clerk Maxwell was concentrating more on atomic quantities to provide a definition of the units:

"If, then, we wish to obtain standards of length, time, and mass which shall be absolutely permanent, we must seek them NOT in the dimensions, or the motion, or the mass of our planet, but in the wave-length, the period of vibration, and the absolute mass of these imperishable and unalterable and perfectly similar molecules."

Address to the Mathematical and Physical Sections of the British Association



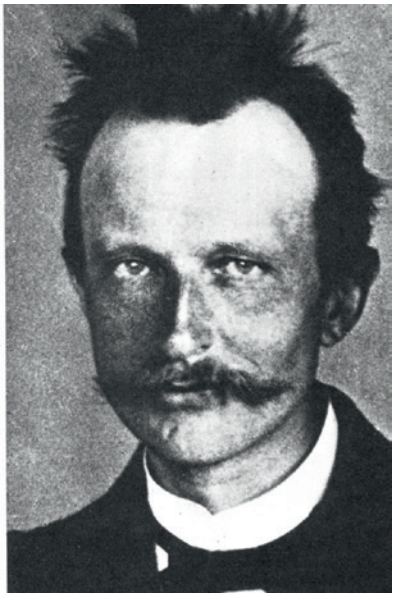
James Clerk Maxwell, 1870

* Dr. Dr. Jens Simon, PTB Press and Information Office; e-mail: jens.simon@ptb.de

After Maxwell, it was above all the great Max Planck who made use of "Constanten" (constants) when he formulated his law of radiation:

"On the other hand, it should not be without interest to say that the use of the two [...] constants a and b offers the possibility of establishing units for length, mass, time and temperature which necessarily maintain their significance for all times and for all cultures (this also includes extraterrestrial and non-human cultures) – independent of special bodies and substances – and which can, therefore, be called "natural mass units."

Ann. Physik 1, 69 (1900)



Max Planck, 1901

In certain scientific circles, particularly the field of theoretical physics, Planck's natural units actually did gain prominence, albeit more as a theoretical pursuit and not as practical units.

By contrast, the revision of the International System of Units that is currently under negotiation does indeed envisage being suited for daily use; it will also live up to this aspiration, not least because – on the basis of the new definitions – the realization and dissemination of the units can (in principle) be continuously improved. No technological barriers of any kind will be inherent in the new system of units.

In this respect, the new system of units that is planned represents a milestone in the history of science – as well as in the history of technology in the foreseeable future after the new definition takes effect. Yet, due to its universal validity, the new system represents considerably more than this: it is a milestone in the history of civilization itself. From the Middle Ages until well into the 18th and 19th centuries, the units were decreed by the sovereign of a country, and used regionally for the most part. With the French Revolution at the close of the 18th century came the abandonment of feet, ells and miles, lines, fathoms and rods in favor of a measure wrested from the planet Earth – the meter was born, and with it came the kilogram. The Metre Convention and its Member States put these units into use throughout the world. Today, our life on this planet is characterized by a single, uniform measuring system (with a few exceptions). In 2018, the step will be taken that reaches beyond our little planet: The units are about to cast aside their anthropomorphic mantle, and I am certain that they will not freeze, but instead enjoy their new freedom.

Planck Units

If fundamental constants are multiplied and divided by each other in such a way that a dimensional observation yields length, time and mass, then Planck units are thereby defined:

Planck mass	$m_P = \sqrt{\frac{\hbar c}{G}}$	$= 2.176 \cdot 10^{-8}$ kg
Planck length	$l_P = \sqrt{\frac{\hbar G}{c^3}}$	$= 1.616 \cdot 10^{-35}$ m
Planck time	$t_P = \frac{l_P}{c}$	$= 5.391 \cdot 10^{-44}$ s
Planck temperature	$T_P = \frac{m_P \cdot c^2}{k}$	$= 1.417 \cdot 10^{32}$ K

h = Planck's constant

c = speed of light

G = gravitational constant ("Big G")

k = Boltzmann constant

Conversely, if the natural constants are expressed in these Planck units, they assume the numerical value of 1.

A Paradigm Change in the International System of Units (SI)

Rainer Scharf*, Thomas Middelmann**

The significance of measurement

Measurement is one of the foundations of our modern-day civilization. It is a key prerequisite of trade, technology and science – fields that demand ever more precise and reliable measurement procedures. During measurement, the existing state of a quantity is compared to a reference state that is made available by means of a measurement standard such as the prototype of the kilogram. In order to ensure that different measurements can be compared to one another, a binding agreement on a suitable reference quantity is necessary.

These reference quantities are the units in which measurements take place. The "binding agreement" illustrates the legal nature of the units, as they can fulfill their purpose only if they are used in the same way at all times.

The result of the measurement of a quantity Q is the product of a numerical value {Q} and a unit of measurement [Q]. By virtue of geometric and physical laws, many correlations exist between different physical quantities; for this reason, the way in which we select their units of measurement cannot be totally arbitrary. In fact, many units can be derived from other units. For example, the liter, the unit of volume, cannot be independent of the meter, the unit of length – and the newton, the unit of force ($1\ \mathrm{N} = 1\ \mathrm{kg\ m\ s^{-2}}$), cannot be defined independently of the kilogram, the meter and the second.

In the International System of Units (French: *Système international d'unités,* SI), the units that are used to measure all known physical quantities can be derived from the seven base units: the meter, the second, the kilogram, the ampere, the kelvin, the mole and the candela. These units, which the SI is based on, do not form a basis in the mathematical sense of an orthonormal system. Instead, the fundamental units (the meter, the second, the kilogram and the ampere) are used side by side with the practically oriented units (the kelvin, the mole and the candela); in principle, this latter group could also be expressed by means of the fundamental units. The reasons for selecting these base units are primarily practical in nature, as the SI serves to facilitate communication between different groups such as retailers, consumers, engineers and researchers. Thus, the selection of a particular base unit may seem useful to one group and strange to another, but the SI endeavors to reconcile these divergent interests.

At present, the base units are defined in various ways. The reference quantities used include artifacts (the kilogram), idealized measurement regulations (the ampere), material properties (the kelvin and the mole), defined factors (the candela) and physical constants (the second and the meter). If one of these quantities changed (as, for example, appears to be happening with the international prototype of the kilogram), the correlations given by the physical constants would seemingly have to change as well – leading to preposterous consequences.

For this reason, it is more meaningful to make use of invariable correlations between different (physical) constants in order to define the reference quantities. A fundamental revision of the SI to achieve this very goal is due to be completed in the near future. Now that the second, the meter and the candela have already been defined by means of fixed (physical) constants, the intent of the new SI is to also define the kilogram, the ampere, the kelvin and the mole by linking them to one another by means of such constants.

Units through the ages

The history of measurement goes back a long way. As early as the dawn of the Classical era, weights, lengths and time were measured, although the measurement units used were obvious-seeming ones. Thus, weights were measured in grain kernels and lengths in inches or ells, while time was measured according to the diurnal passage of the sun. In most cases, different units were used in different territories. For example, in the different small territorial states which comprised Germany at that time, dozens of ells (cubits)

* Dr. Rainer Scharf, science journalist, e-mail: r.scharf@rz-online.de

** Dr. Thomas Middelmann, Semiconductor Physics and Magnetism Department, e-mail: thomas.middelmann@ptb.de

Figure 1:
Symbol of the Metre Convention and the International Bureau of Weights and Measures (BIPM).

of different lengths existed, ranging from 40.38 cm in Erfurt to 79.90 cm in Munich. Merchants who were familiar with these ells and who were skilled at converting them could use this situation to their advantage. Until the 18th century, little about this matter changed. Yet, in the course of industrialization, with the emergence of manufacturing enterprises and the expansion of trade, the multitude of different measures of length resulted increasingly in trade barriers.

In 1789, the French Revolution brought a solution to these problems, as demands were raised that uniform measures be introduced in addition to the decimal system. This led to the birth of the meter and the kilogram; although both were based on artifacts, they nonetheless laid claim to being objectively and universally valid on Earth. Hence, the prototype of the meter embodies one ten-millionth of the distance from the North Pole to the Equator, as measured along the meridian through Paris, while the prototype of the kilogram corresponds to the mass of one liter or cubic decimeter of water at a temperature of 4 °C. [1]

In 1799, both of these "revolutionary" artifacts – the prototypes of the meter and the kilogram – became the "measures of all things" in France; ultimately, in 1875, they formed the foundation of an international agreement within the scope of the Metre Convention. Under this agreement, 17 countries joined together with the goal of establishing and ensuring the further development of uniform measures. In addition to France, the 17 founding Member States of the Metre Convention included the German, Russian and Ottoman Empires and the United States. In the years that followed, other countries were added, such as Great Britain in 1884 and Japan in 1885. [2]

As of 17 August 2016, the Metre Convention has 58 Member States, and an additional 41 countries and international organizations are associated members. Yet the use of metric units has not yet gained acceptance in all of the Member States – such as in the United States, where length, mass and temperature continue to be measured in miles, ounces and degrees Fahrenheit.

The Metre Convention led to the establishment of the following three organs:

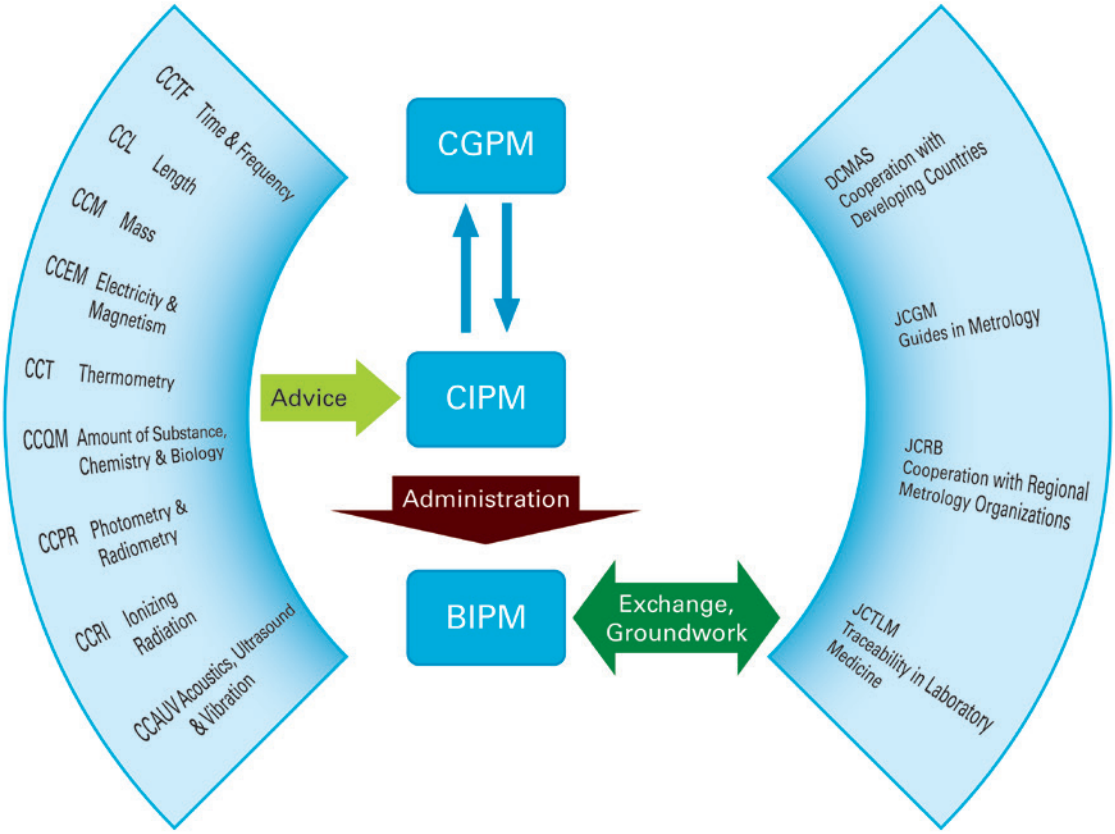
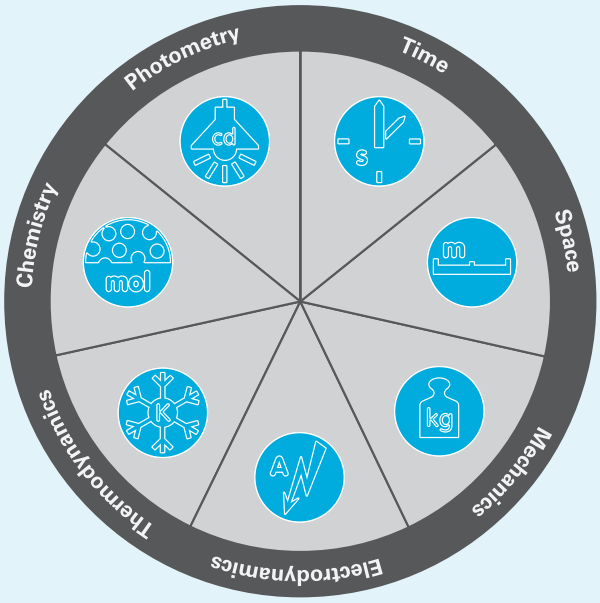


Figure 2:
Structural diagram of the institutions/organs established by the Metre Convention.

The idea of the old and the new SI

The International System of Units (SI) was established between 1948 and 1960 and dubbed the *Système international d'unités*, or SI for short. Developed and established by the organs of the Metre Convention, the SI is based on the metric system. Its basic idea is to form all SI units solely by means of multiplication or division from a few base units (at first six, currently seven). The SI provides that only decimal factors are to be used; these factors are identified by means of corresponding prefixes such as "k" for "kilo", meaning 1,000. The seven base units are: the second (s), the meter (m), the kilogram (kg), the ampere (A), the kelvin (K), the candela (cd) and the mole (mol). Each base unit contributes an additional "dimension" – i.e., an additional physical and metrological field – to the system of units.

In the new SI, the seven base units are defined by determining seven "defining constants" that contain these units. Instead of using artifacts, as is still the case today in the definition of the kilogram, these constants provide an "anchor" for the SI and form its new basis. The dimensional relationships of the SI units among one another will be preserved, and it will be possible to form all of the units by multiplying or dividing the "base constants" by a prefactor. The prefactor in each case is derived from the numerical values determined for the base constants concerned. Needless to say, all of the units can still be formed from the base units as well; however, it is the underlying base constants which are the actual points of reference.



Explanations, Figure 2

CGPM: General Conference on Weights and Measures (Conférence générale des poids et mesures)
CIPM: International Committee for Weights and Measures (Comité international des poids et mesures)
BIPM: International Bureau of Weights and Measures (Bureau international des poids et mesures)
CCs: Consultative Committees of the CIPM
DCMAS: Network on Metrology, Accreditation and Standardization for Developing Countries
JCGM: Joint Committee for Guides in Metrology
JCRB: Joint Committee of the Regional Metrology Organizations and the BIPM
JCTLM: Joint Committee for Traceability in Laboratory Medicine

1. The International Bureau of Weights and Measures (French: *Bureau international des poids et mesures*, BIPM) in Sèvres near Paris is the international center for units of measurement.
2. The General Conference on Weights and Measures (French: *Conférence générale des poids et mesures*, CGPM), held every four to six years at the BIPM and attended by delegates of all signatory states, is the highest committee of the Metre Convention.
3. The International Committee for Weights and Measures (French: *Comité international des poids et mesures*, CIPM) is an administrative committee which is subjected to the BIPM and meets every year at the BIPM.

Measuring by means of physical constants

The motto of the Metre Convention is "*A tous les temps, à tous les peuples*" – "for all men and all times". Initially, the metric system failed to live up to this principle, as it drew on the size and rotational period of the Earth – factors subject to constant change – in order to define the fundamental units of the meter, the kilogram and the second. Yet, as early as 1870, the Scottish physicist James Clerk Maxwell (1831–1879) [3] proposed using the invariable properties of atoms to define the measurement units for length, time and mass. Among these properties are the wavelength and frequency of a given emission line of an atom, as well as its mass.

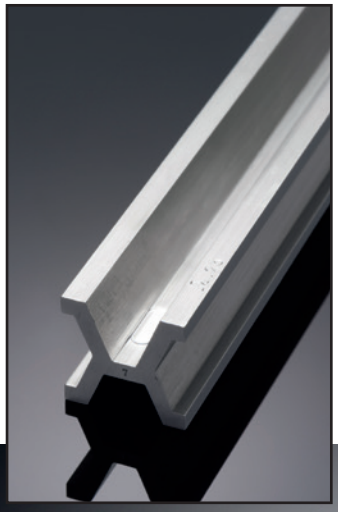


Figure 4:
One of the national copies of the prototype of the meter; prototype #23 is kept in a safe at PTB as a museum piece.

This idea was taken up in 1960, when the prototype of the meter – a bar made out of platinum and iridium that had a special cross-section – was abandoned in favor of the definition of the meter as 1 650 763.73 wavelengths in vacuum of the radiation corresponding to a specific transition of atoms of the krypton-86 nuclide. In 1967, the second was defined in a similar way as 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.

Because the product of the frequency and the wavelength of a monochromatic electromagnetic wave in vacuum is equal to the speed of light c, the above definitions of the meter and the second are not independent of one another but linked via the physical constant c. The speed of light c was specified by measuring the period of time needed by light to traverse a reference distance. However, the limited precision with which the reference distance was measured determined how accurately c was specified. For this reason, the procedure was reversed in 1983: The speed of light was specified as 299 792 458 m/s (the best measured value), and the length of a distance became determined by measuring the period of time that light needs to traverse the distance.

Can units of measurement also be traced directly to physical constants? This question was pursued by the Irish physicist George Johnstone Stoney (1826–1911) [4], who was a contemporary of Maxwell's. In 1874, Stoney devised a system of units that he based on three physical constants: The speed of light c, the gravitational constant G and the elementary charge e. From these constants, he obtained the unit of length $G^{1/2}e/c^2 \approx 10^{-37}$ m, the unit of time $G^{1/2}e/c^3 \approx 10^{-46}$ s and the unit of mass $e/G^{1/2} \approx 10^{-7}$ g. As is evident, Stoney's units of length and time are much too short to be used in practice.

Max Planck [5] had a similar idea after discovering (in 1900) the radiation law which would later bear his name and which contained two new

fundamental constants: Planck's constant h and the Boltzmann constant k_B. Planck founded his system of units on the four constants c, G, h and k_B. From these constants, he derived (among other things) the unit of length $(Gh/c^3)^{1/2} \approx 10^{-35}$ m, the unit of time $(Gh/c^5)^{1/2} \approx 10^{-43}$ s and the unit of mass $(hc/G)^{1/2} \approx 10^{-5}$ g. Planck's units of length and time, like Stoney's, are also much too small for practical use.

Yet both Planck's and Stoney's "natural" units have two significant advantages: First, they are invariable, to the extent that the physical constants are truly constant. Second, they are universally valid, meaning that even an extraterrestrial civilization with sufficient knowledge of physics could define them in the same way.

However, the fact that even standards that have a very good reproducibility and are of great practical importance can be directly traced to fundamental physical constants can be seen in the quantum Hall effect and the Josephson effect. Both are electronic quantum effects, meaning that they both involve the elementary charge e and Planck's constant h.

By means of the quantum Hall effect, for whose discovery the German physicist Klaus von Klitzing was awarded the 1985 Nobel Prize in physics, electrical resistance can be measured with a very high degree of precision. This effect is observed in thin, current-carrying semiconductor layers that are exposed to a strong magnetic field. In this process, a voltage occurs which is perpendicular to the current flow; the voltage's relationship to the current is characterized as the Hall resistance R_H. This resistance can only take on specific, quantized values: $R_H = R_K/m$, with an integral m and the von Klitzing constant $R_K = h/e^2$, which was defined in 1990 as the then-best measurement value $R_{K-90} = 25\,812.807\ \Omega$. Electrical resistance can be measured by means of comparison to this standard. In this way, the unit of the ohm (Ω) can be directly traced to physical constants.

In 1973, British physicist Brian Josephson received the Nobel Prize for his prediction of the Josephson effect; with the aid of this effect, electric voltages of a precisely defined size can be generated. The Josephson junctions used for this purpose are two superconductor components that are separated from one another by means of a thin normal-conductor layer. If a microwave of the frequency f is irradiated onto such a junction, a direct current flows in this junction, as a result of which an electrical voltage U occurs between the two ends of the junction. Here, the following applies: $U = nf/K_J$, with an integral n and the Josephson constant $K_J = 2e/h$. This constant was defined in 1990 as the then-best measurement value of $K_{J-90} = 483\,597.9$ GHz V^{-1}. In this way, the unit of the volt (V) can be traced to the frequency of cesium by means of a frequency measurement.

Because the von Klitzing constant and the Josephson constant were precisely specified as the values (measured in 1990) of R_{K-90} and K_{J-90}, respectively, the physical constants h and e were then also determined. Yet, in the SI that is valid today, these physical constants are merely quantities whose best measurement values have changed since 1990. For this reason, it is necessary to distinguish between the constants specified as R_{K-90} and K_{J-90} on the one hand, which are not part of the current SI, and the quantities R_K and K_J on the other hand, which are dependent on h and e.

Strengths and weaknesses of the old SI

The old SI (which is still in effect) is in keeping with the times as concerns the definitions of the second and the meter; for this reason, no revision is planned for these base units for the time being. The second is thus defined as the duration of 9 192 631 770 oscillation periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of a cesium-133 atom at rest, while the meter is defined as the length of the distance traveled by light in vacuum during a time interval of 1/299 792 458 of a second. In addition to the meter and the second, the candela (the unit of luminous intensity) will also keep the definition currently used as its basis in the new SI. Apart from being reformulated, these three base units will ultimately be adopted in an unchanged form.

However, for the remaining units – the kilogram, the mole, the kelvin and the ampere – there is

Figure 5:
Held in safekeeping – PTB holds several prototypes of the kilogram, such as the two specimens issued to the Federal Republic of Germany (#52, top level in safe) and the German Democratic Republic (#55, middle left) in the early 1950s, as well as an additional prototype (#70, middle right) acquired by the Federal Republic of Germany in 1987. Kilogram number 22 (lower level in safe) is from 1889, was damaged during wartime and is no longer an official prototype.

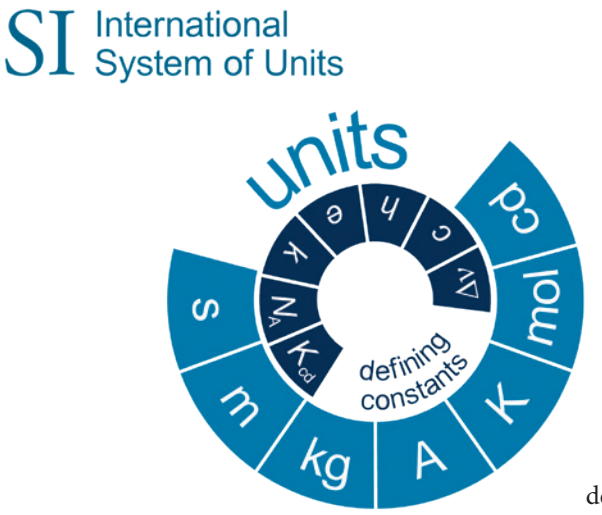


Figure 6:
In the new system of units, the values of selected constants specify all of the units – both the current base units and the derived units.

demand for a new definition, as the SI definitions that have been used to date demonstrate considerable shortcomings. For example, the kilogram, the unit of mass, is equal to the mass of the international prototype of the kilogram – a cylinder made of a platinum/iridium alloy. Since 1889, this prototype has been kept in a safe at the International Bureau of Weights and Measures (BIPM) in Sèvres near Paris. Yet, compared to most of the official copies and mass standards, the prototype of the kilogram is becoming lighter and lighter. Measurements have revealed that the prototype had lost about 50 micrograms between 1950 and 1990 – presumably due to aging processes or mass loss during cleaning of the cylinder.

The unit of electric current, the ampere, is defined by means of an idealized and unrealistic measurement regulation. This definition states that the ampere is that constant current which, if maintained in two straight, parallel conductors of infinite length, of negligible circular cross-section, and placed one meter apart in vacuum, will produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per meter of length. Apart from the fact that such a measurement regulation can only be realized approximately, one significant disadvantage of this definition is that it links the ampere to the kilogram by means of force.

The unit of temperature, the kelvin, is defined as 1/273.16 of the absolute temperature of the triple point of water, at which water vapor, liquid water and ice are in equilibrium. Yet this triple point is influenced by impurities and by the isotopic composition of the water. As temperature is not an additive quantity, additional definitions are necessary in order to expand the temperature scale beyond the triple point of water. The new SI remedies these flaws of the old SI, which are primarily associated with the problematic definitions of the kilogram, ampere and kelvin.

The seven constants of the new SI

In accordance with the resolutions of the General Conference on Weights and Measures, a fundamental change is to be made to the SI in 2018. To this end, the numerical values of seven unit-related constants – the "defining constants" – are to be specified exactly. In this way, the seven base units (s, m, kg, A, K, cd, mol) will no longer be directly but indirectly defined. This is vividly illustrated using the example of the speed of light c and the meter. After the second was defined by means of the frequency of the hyperfine structure transition of cesium-133, the speed of light was specified exactly: c = 299 792 458 m/s. Since then, one meter is the length of the distance traveled by light in 1/299 792 458 s. Thereby, the best measured value of the speed of light was used so that agreement with the definition of the meter used to date was guaranteed.

The new SI will be defined by means of a binding specification of the following seven constants (CODATA 2017 adjustment):

- The frequency $\Delta\nu(^{133}\mathrm{Cs})_{\mathrm{hfs}}$ of the hyperfine structure transition of the ground state of the cesium atom is exactly 9 192 631 770 Hertz (Hz).
- The light velocity in vacuum c is exactly 299 792 458 m s^{-1}.
- Planck's constant h is exactly $6.626\,070\,15 \cdot 10^{-34}$ joule seconds (J s).
- The elementary charge e is exactly $1.602\,176\,074 \cdot 10^{-19}$ coulombs (C).
- The Boltzmann constant k_B is exactly $1.380\,649 \cdot 10^{-23}$ joules per kelvin (J K^{-1}).
- The Avogadro constant N_A is exactly $6.022\,140\,76 \cdot 10^{23}$ reciprocal moles (mol^{-1}).
- The photometric spectral luminous efficacy K_{cd} of monochromatic radiation of the frequency $540 \cdot 10^{12}$ Hz is exactly 683 lumens per watt (lm W^{-1}).

The numerical values indicated above may still change if such changes are necessary due to improved experimental results being available by the time the specifications take place. The goal of this approach is to ensure that, if possible, there are no "artificial leaps" in the measurement values during the transition from the old SI to the new SI. Exactly specifying the seven defining constants by using the units Hz = s^{-1}, m, J, C, K, mol, lm and W = J s^{-1} means that these latter units are also specified and, with them, all further SI units such as the kilogram (kg = J m^{-2} s^2) or the ampere (A = C s^{-1}).

The quantities chosen as defining constants are quantities that can be measured with great precision in the old SI, and whose relative uncertainties are as low as possible – ideally around 10^{-8}. For this reason, the gravitational constant is not one of these quantities, as it is known only with a relative uncertainty of 10^{-4}. This represents a fundamental difference between the new SI and Stoney's and Planck's systems of units.

As it is frequencies which can be measured with the greatest precision, the definition of the second via the cesium frequency was kept. This frequency is not a fundamental constant but an atomic parameter that is influenced by external disturbances such as electrical and magnetic fields. Nevertheless, since these disturbances can be controlled well, the cesium frequency can be reproduced with great precision; as a result, the systematic uncertainty of present-day cesium clocks is around 10^{-16}. Optical atomic clocks even have stabilities of 10^{-18}. Yet, of the optical frequency standards that have been developed, none has proven to be clearly superior to the others thus far; consequently, no redefinition of the second is planned before 2018.

The next three defining constants are actually fundamental physical constants that cannot be traced to other quantities: The speed of light c, Planck's constant h and the elementary charge e. As early as 1983, after the speed of light had been measured with a relative uncertainty of 10^{-9}, it was specified as the same value which the new SI is also based on. As the von Klitzing constant $R_K = h/e^2$ and the Josephson constant $K_J = 2e/h$ are known with relative uncertainties of 10^{-10} and 10^{-8}, respectively, the uncertainties of h and e are correspondingly small; for this reason, both were also made into defining constants.

The remaining three constants are actually conversion factors. The Boltzmann constant k_B makes conversion possible between the unit of temperature, the kelvin, and the unit of energy, the joule; historically, the numerical value of k_B was obtained from the former specifications for the temperature scale. The Avogadro constant N_A specifies the number of particles to be contained in the amount of substance of 1 mol as an amount of substance. The luminous efficacy K_{cd} makes it possible to convert the power, indicated in watts, of a green monochromatic radiation source with a wavelength of 555 nm into a light flux measured in lumens (lm), which is a measure of the light emitted in all spatial directions. It is linked to the base unit of the candela (cd) due to the fact that a radiation source which generates a light flux of 4π lm has a luminous intensity per unit of solid angle (steradian) of 1 cd = 1 lm/steradian.

Clearly, these three base units, which are defined by means of conversion factors, are an acknowledgement on the part of the new SI of the practical demands of industry, technology and research. The candela, for example, serves as a unit of reference for the lighting industry. Temperature could be expressed just as well by means of the mean particle energy in joules, the amount of substance by means of the number of particles and luminous intensity by means of radiant power per unit of solid angle. However, this would mean abandoning the conventional units of measurement.

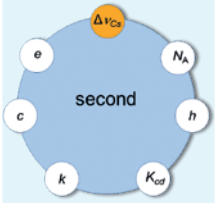
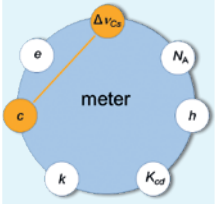
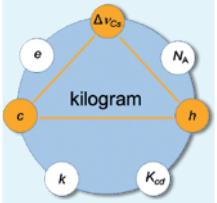
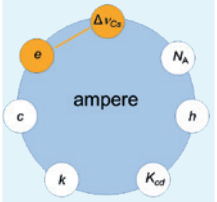
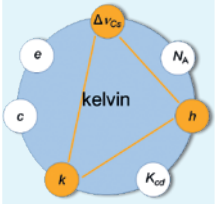
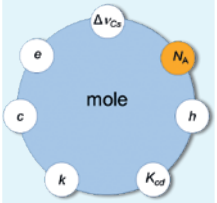
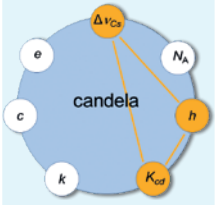
The seven base units of the new SI

By specifying the seven defining constants, the seven base units are defined. The table below provides an overview of how the base units are defined and how they can be calculated from the constants [6, 7]. The diagrams show which constants each of the base units depend on.

After the seven base units have been defined by specifying the constants, all other "derived" units can also be shown by means of reference to these constants. Furthermore, some units do not even have to take the "indirect route" via a base unit; instead, they can be directly related to one or more of the specified constants.

This is true, for example, of the volt, the unit of electric potential. It can be related by means of the Josephson effect to the cesium frequency $\Delta\nu_{Cs}$ and the Josephson constant $K_J = 2\,e/h$; the Josephson constant, in turn, is a combination of the elementary charge e and Planck's constant h. Similarly, the ohm, the unit of electrical resistance, can be related by means of the quantum Hall effect to the von Klitzing constant $R_K = h/e^2$, which is also a combination of e and h. In the same way, the unit of energy, the joule, can be directly related to h and $\Delta\nu_{Cs}$ without having to take an "indirect route" via the definition of the unit of mass, the kilogram.

Only the base units of the second and the mole are directly related to a single defining constant. All other base units are defined by means of two (m, A), three (kg, K) or even four (cd) constants. Thus, the difference between base units and derived units which was present in the old SI no longer exists in the new SI. For example, the coulomb (C = A s), which was previously a derived unit, will now be defined directly by means of a fundamental constant, $C = 1/(1.602\,176\,565 \cdot 10^{-19})\,e$, thereby specifying it more directly than the base unit of the ampere, $A = e/(1,\,602\,176\,565 \cdot 10^{-19}) \cdot \Delta\nu_{Cs}/9\,192\,631\,770 = 6{,}789\,68711... \cdot 10^8\,\Delta\nu_{Cs}\,e$.

Unit	Definition	Conversion	Diagram
second (s)	The second is the SI unit of time. It is defined by taking the fixed numerical value of the cesium frequency $\Delta\nu_{Cs}$, the unperturbed ground-state hyperfine transition frequency of the cesium-133 atom, to be 9 192 631 770 when expressed in the unit Hz, which is equal to s^{-1}.	$1\ s = 9\ 192\ 631\ 770/\Delta\nu_{Cs}$	
meter (m)	The meter is the SI unit of length. It is defined by taking the fixed numerical value of the speed of light in vacuum c to be 299 792 458 when expressed in the unit m s^{-1}, where the second is defined in terms of the cesium frequency $\Delta\nu_{Cs}$.	$1\ m = (c / 299\ 792\ 458)\ s$ $= 30.663\ 318\ldots\ c/\Delta\nu_{Cs}$	
kilogram (kg)	The kilogram is the SI unit of mass. It is defined by taking the fixed numerical value of the Planck constant h to be $6.626\ 070\ 040 \cdot 10^{-34}$ when expressed in the unit J s, which is equal to kg m^2 s^{-1}, where the meter and the second are defined in terms of c and $\Delta\nu_{Cs}$.	$1\ kg = (h / 6.626\ 070\ 15 \cdot 10^{-34}) m^{-2}\ s$ $= 1.475\ 521\ldots \cdot 10^{40}\ h\ \Delta\nu_{Cs}/c^2$	
ampere (A)	The ampere is the SI unit of electric current. It is defined by taking the fixed numerical value of the elementary charge e to be $1.602\ 176\ 620\ 8 \cdot 10^{-19}$ when expressed in the unit C, which is equal to A s, where the second is defined in terms of $\Delta\nu_{Cs}$.	$1\ A = e / (1.602\ 176\ 634 \cdot 10^{-19})\ s{-}1$ $= 6.789\ 686\ldots \cdot 10^{8}\ \Delta\nu_{Cs}\ e$	
kelvin (K)	The kelvin is the SI unit of thermodynamic temperature. It is defined by taking the fixed numerical value of the Boltzmann constant k to be $1.380\ 648\ 52 \cdot 10^{-23}$ when expressed in the unit J K^{-1}, which is equal to kg m^2 s^{-2} K^{-1}, where the kilogram, the meter and the second are defined in terms of h, c and $\Delta\nu_{Cs}$.	$1\ K = (1.380\ 648\ 52 \cdot 10^{-23}/k)$ kg m^2 s^{-2} $= 2.266\ 665\ \Delta\nu_{Cs}\ h/k$	
mole (mol)	The mole is the SI unit of amount of substance of a specified elementary entity, which may be an atom, molecule, ion, electron, any other particle or a specified group of such particles. It is defined by taking the fixed numerical value of the Avogadro constant N_A to be $6.022\ 140\ 857 \cdot 10^{23}$ when expressed in the unit mol^{-1}.	$1\ mol = 6.022\ 140\ 857 \cdot 10^{23}/N_A$	
candela (cd)	The candela is the SI unit of luminous intensity in a given direction. It is defined by taking the fixed numerical value of the luminous efficacy of monochromatic radiation of frequency $540 \cdot 10^{12}$ Hz, K_{cd}, to be 683 when expressed in the unit lm W^{-1}, which is equal to cd sr W^{-1}, or cd sr kg^{-1} m^{-2} s^3, where the kilogram, the meter and the second are defined in terms of h, c and $\Delta\nu_{Cs}$.	For light with a frequency of $540 \cdot 10^{12}$ Hz, the following applies: $1\ cd = (K_{cd}/683)$ kg m^2 s^{-3} sr^{-1} $= 2.614\,830\ldots \cdot 10^{10}\ (\Delta\nu_{Cs})^2\ h\ K_{cd}$	

Realization of the units kg, mol, A and K

Aided by the defining constants, the base units of the new SI are "anchored" and thus specified in a binding way. However, the question remains open as to how these units are to be "realized" – meaning how they are to be "disseminated" – in order to perform calibrations and in order to actually measure the corresponding physical quantities in these units. For the base units of the second and the meter, and for the derived units of the ohm and the volt, extremely precise measurement procedures exist which employ the cesium atomic clock, light propagation and the quantum Hall effect/Josephson effect; these procedures are used to trace the units listed above directly to the (fundamental) constants $\Delta\nu_{Cs}$, c, h and e. However, for the kilogram, the mole, the ampere and the kelvin, it was neccessary to develop measurement procedures of comparable precision.

There are two fundamentally different methods for the realization of the kilogram, the "silicon sphere" and the "watt balance", thereby allowing a reciprocal control to take place. The silicon sphere produced in Braunschweig at PTB is the contemporary form of the kilogram prototype and is an almost perfectly sphere-shaped silicon single crystal with a diameter of approximately 9.4 cm and a mass of around 1 kg. The crystal, made of highly enriched ^{28}Si, is virtually defectless and free of impurity atoms. The element of silicon was chosen primarily because the semiconductor industry has decades of experience with the manufacture of virtually defectless silicon crystals. After the lattice parameter of the single crystal (and thus the distances between the atoms) had been measured with great precision by means of X-ray diffraction, and after the sphere volume had been obtained with sufficient accuracy by means of measurements in the sphere interferometer, it was possible to indicate the number of atoms contained in the sphere with great precision as well.

This allows the macroscopic mass of the silicon sphere (which can be measured by comparing it to the prototype of the kilogram) to be traced to atomic masses and physical constants such as Planck's constant h. As a result, it has become possible, on the one hand, to determine the Avogadro constant N_A (i.e. the number of atoms contained in one mole), thus allowing the mole to be realized. On the other hand, using the silicon sphere, it has become possible to determine Planck's constant with a relative measurement uncertainty of $2 \cdot 10^{-8}$.

Since the new SI bindingly specifies the numerical value of h in the unit of the joule second, it will now be possible to produce a silicon sphere with a defined mass at any location where the technical prerequisites are fulfilled. In fact, this new mass standard is to be disseminated in another way, however: by having PTB deliver the copies of the silicon sphere that it produces to its international partners.

The "watt balance" being constructed at the National Institute of Standards and Technology (NIST) in the United States, among other places, creates an additional relation between the kilogram and Planck's constant. To this end, the gravitational force of a mass is compensated by the magnetic force which acts on a current-carrying coil in a magnetic field. The magnetic force is calculated from the electric current and the magnetic field strength. The electric current, in turn, is determined from the electrical resistance of the coil and the applied voltage, which can be measured very precisely with the aid of the quantum Hall effect and the Josephson effect.

The magnetic field strength of the watt balance can be obtained by moving the coil at a constant speed through the magnetic field. The voltage induced in the coil in this process, which is a measure of the magnetic field, is measured again using the Josephson effect. Finally, it is necessary to measure the gravitational acceleration. Since the voltages measured and the resistance are proportional to the Josephson and von Klitzing constants, which contain the elementary charge e and Planck's constant h, a relation between h and the mass in the watt balance is created. If a copy of the kilogram prototype is used as a mass, Planck's constant h can be specified directly. By contrast, if the value of h is set, the kilogram can be realized with the aid of the watt balance.

The measurements of Planck's constant with the aid of the silicon sphere and by means of the watt balance have now both reached a relative uncertainty of around $2 \cdot 10^{-8}$. Within the scope of this uncertainty, both of the values measured for h are in agreement. Nevertheless, they deviate significantly from the value $h_{90} = 4/(R_{K-90}\, K_{J-90})$, which results from the values (specified in 1990) of the von Klitzing constant R_{K-90} and the Josephson constant K_{J-90}. Thus, both of the constants R_{K-90} and K_{J-90} are outside of the new SI. As it is very complicated to construct and run experiments on the watt balance, the silicon sphere is likely to be given preference during the practical realization of the kilogram. Although producing the silicon sphere involves a great deal of effort, using it as a mass standard is not problematic. By now, the silicon sphere is ready for use as a new mass standard, and can replace the prototype of the kilogram in the autumn of 2018 [7–10].

In the old SI, the ampere, the unit of electric current, is defined by means of an idealized measurement regulation, which can only be realized approximately and is therefore unsatisfactory. The new SI, by contrast, has two procedures which are compatible with one another in order to realize the ampere with greater precision. The first procedure

is based on Ohm's law, $U = R \cdot I$, which creates a relation between the voltage U applied to a resistance R and the flowing current I. In this way, the current I can be determined by measuring the resistance R using the quantum Hall effect and the voltage U using the Josephson effect.

The second procedure is more direct and based on the fact that, in the new SI, the ampere will be traced to the defining constants of the elementary charge e and the cesium frequency $\Delta\nu_{Cs}$. Here, the ampere will be realized with an electronic circuit that measures the electric current by counting the electrons that pass the circuit in a certain time interval. A circuit of this kind consists of several sequentially arranged, nanostructured "single-electron pumps" that transport the electrons one at a time according to a set pulse at GHz frequency. As a result of this procedure, currents in the nanoampere range can be reached. Between these electron pumps, special single-electron transistors are located which check whether exactly one electron is actually being transported in each pulse. Four series-connected electron pumps and three single-electron detectors allow unavoidable errors to be identified reliably; as a result, an electron current with a relative uncertainty of a few 10^{-8} has been obtained.

Together, both procedures allow the consistency of the definitions of the volt, the ohm and the ampere to be checked. Whereas, in the first procedure, the volt and the ohm are realized by means of the Josephson constant and the von Klitzing constant, respectively, the second procedure realizes the ampere by counting individual charges $Q = e$. This results in Ohm's law on a quantum basis: $K_J\, R_K\, Q = (2e/h)(h/e^2)e = 2$. This "quantum metrology triangle" has been experimentally confirmed with an uncertainty of $9 \cdot 10^{-7}$.

To realize the kelvin, the unit of temperature, the Boltzmann constant is determined by means of two fundamentally different measurement methods in which a relative uncertainty of around 10^{-6} should be achieved. In acoustic gas thermometry, the temperature-dependent speed of sound is measured in a gas; this speed of sound is proportional to $(k_B\, T)^{1/2}$. The relative uncertainty for k_B which is obtained as a result is $1 \cdot 10^{-6}$.

As an alternative, dielectric constant gas thermometry (DCGT) is being investigated at PTB by determining at a constant temperature the pressure-dependent density of helium with reference to its dielectric constant. To this end, measurements are being taken of how the gas changes the capacity of a special capacitor. From the pressure and density of the gas, it is possible to obtain T with the aid of the thermal equation of state k_B [11]. In this way, k_B was determined at the triple point of water with a relative uncertainty which is currently $4 \cdot 10^{-6}$. PTB's researchers are confident that they will achieve the necessary uncertainty of around $2 \cdot 10^{-6}$ prior to the redefinition of the units.

Table 1: Relative uncertainties of selected constants in the current SI and in the new SI, based on the data published in CODATA 2014 [12].

Uncertainties of the constants/10^{-8}					
Constant	Previous	New	Constant	Previous	New
m(K)	0	1.2	R	57	0
T_{TPW}	0	57	F	0.62	0
$M(^{12}C)$	0	0.045	σ	230	0
μ_0	0	0.023	K_J	0.62	0
ε_0	0	0.023	R_K	0.023	0
Z_0	0	0.023	$N_A h$	0.045	0
$\Delta\nu(^{133}Cs)_{hfs}$	0	0	m_e	1.2	0.033
c	0	0	m_u	1.2	0.045
K_{cd}	0	0	$m(^{12}C)$	1.2	0.045
h	1.2	0	α	0.023	0.023
e	0.61	0	J ↔ m^{-1}	1.2	0
k_B	57	0	J ↔ Hz	1.2	0
N_A	1.2	0	J ↔ K	57	0

Outlook

In the currently valid SI, a few constants already have specified values (such as the speed of light, the magnetic constant and the electric constant). However, only when the new SI takes effect will a set of physical constants (or, more precisely, their specified values) create a coherent framework for all units. To prepare for the new SI, the defining (physical) constants will be measured as precisely as possible. These measurements will require a relative measurement uncertainty of a few 10^{-8}; this uncertainty can be achieved by exploiting quantum effects for the measurements. By always using several different measurement procedures, and by comparing them to one another (e.g. "silicon sphere" and "watt balance"), the risk of undetected systematic errors will be reduced.

In 2018, the new SI will permanently specify the defining constants h, e, k_B and N_A as the best measurement values available at that time, with the result that they will no longer have any measurement uncertainty. By contrast, other numerical values which have been exactly specified to date will become quantities which are affected by a measurement uncertainty in the new SI. For example, the mass of the prototype of the kilogram will become a quantity, as will the molar mass of carbon $M(^{12}C)$ and the temperature of the triple

point of water. In the same way, the magnetic constant μ_0 and (due to $c = (\mu_0\varepsilon_0)^{-1/2}$) the electric constant ε_0 will then become quantities. For this reason, the fine-structure constant $\alpha = e^2/(2\,\varepsilon_0\,c\,h)$ will also become a quantity, one with a relative uncertainty of $3.2 \cdot 10^{-10}$.

In the future as well, it will become possible to measure physical quantities with ever greater precision thanks to advances in metrology and new measurement procedures. However, as the defining constants are specified in a binding way, more precise measurements will allow the base units to be realized with greater precision – without it being necessary to make changes to the definitions on which the new SI is based. In the case of the unit of time, a different reference transition frequency may be selected in the future by introducing an optical atomic clock instead of the cesium clock. It will then be possible to realize the second with greater precision, thereby allowing time intervals to be measured more accurately.

The new SI is based on constants whose validity is universal and timeless – yet it is still open to future improvements. In so doing, it fulfills the promise that the Metre Convention made so many years ago: to exist "for all men and all times".

Literature

[1] Bureau international des poids et mesures: "Le Système international d'unités. The International System of Units." 8th edition (2006). (9th edition in preparation. Draft dated 11 December 2015: http://www.bipm.org/en/measurement-units/new-si/#communication).

[2] *J. Hoppe-Blank*; Vom metrischen System zum Internationalen Einheitensystem: 100 Jahre Meterkonvention, PTB-Bericht PTB-ATWD-5 (1975)

[3] *J. C. Maxwell*; The scientific papers. Part 2 (W.D. Niven, ed.), University Press, Cambridge (1890) p. 225.

[4] *J. D. Barrow*; Das 1×1 des Universums. Neue Erkenntnisse über die Naturkonstanten. Campus Verlag, Frankfurt 2004.

[5] *M. Planck*; Ann. Physik **1** (1900) 69.

[6] *R. Wynands and E. O. Göbel*; Die Zukunft von Kilogramm und Co., Spektrum der Wissenschaft (March 2010) p. 34.

[7] *J. Stenger and J. H. Ullrich*; Für alle Zeiten ... und Culturen. Physik Journal **13** (2014) no. 11, p. 27.

[8] PTB Recipe for the New Kilogram, PTB Press Release (03-26-2015)*.

[9] A New Basis for all Measures; PTB Press Release (03-26-2015)*.

[10] PTB's Mass Standards for the New Kilogram Are Now Ready for Use, PTB Press Release (03-26-2015)*.

[11] High-Precision Measurement of the Boltzmann Constant, PTB-News 1.2016, http://www.ptb.de/cms/presseaktuelles/zeitschriften-magazine/ptb-news.html (retrieved: 14.07.2016).

[12] *Peter J. Mohr, David B. Newell, Barry N. Taylor*; CODATA Recommended Values of the Fundamental Physical Constants: 2014, Preprint at http:// http://arxiv.org/abs/1507.07956 (retrieved: 14.07.2016).

* PTB's Press Releases online:
http://www.ptb.de/cms/en.html >
Press & What's New > For journalists >
Press releases > Archives of press releases

How Does an Atomic Clock Tick? – Realization of the Second from 1955 to the Present Day

Andreas Bauch*, Stefan Weyers, Ekkehard Peik*****

Introduction

The SI base unit the "second" has a special position among the units: since 1967, it has been defined with reference to an atomic constant – here the hyperfine structure transition frequency in the cesium atom (^{133}Cs). Even today, the unit of time is the SI unit which has been realized by far with the highest accuracy – and this is why other base units are defined or realized with reference to it. The definition of the base unit the "meter", for example, refers to the second: the meter is defined as the distance traveled by light in vacuum in 1/299 792 458 seconds. The realization of the volt – the unit of electric voltage – exploits the Josephson effect, which links the volt with a frequency via the ratio of two fundamental constants, $h/(2e)$ (h: Planck's constant, e: elementary charge). The interactions between the units are the subject of the article *Counting Electrons to Measure Current* in this publication, written by Scherer and Siegner.

The decision made in 1967 was obviously very farsighted. The original text of the definition, "*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 atoms of the caesium 133 atom*", was formulated so generally that it has never been questioned to this day, despite the numerous developments atomic clocks have experienced since 1955. In this period, the relative uncertainty with which the SI unit can be realized was reduced from approx. 10^{-10} down to nearly 10^{-16}.

In this article, we will try to answer five questions:

- What were the prerequisites for the first cesium atomic clocks to be built and the subsequent definition of the unit of time in 1967?
- What were the prerequisites for the "traditional" cesium atomic clocks to become cesium fountain clocks – the most accurate clocks of our times?
- What are the limits to these clocks' accuracy?
- How do so-called "optical clocks" surmount these limitations?
- Do we need a new definition of the second?

A whole section is dedicated to each of these questions, whereas more specific topics are dealt with in info boxes.

The path to the "traditional" cesium atomic clock

Three indispensable achievements heralded the cesium atomic clock: the generation of beams of free atoms in vacuum, the understanding of directional quantization (i.e. of the orientation of magnetic moments of atoms in space and the possibility of manipulating them), and the generation of signals in the gigahertz frequency range (i.e. microwave radiation) by multiplying the frequency of HF signals in the kilohertz range. The latter had been available with high frequency accuracy since quartz clocks were developed in the 1930s [1]. We will provide a simple overview of the individual steps which are described in detail in [2, 3].

The first two achievements are inseparably connected with the name of Otto Stern who was a professor first at the University of Frankfurt and from 1923 on at the University of Hamburg [4]. In his first atomic beam experiments, he was able to determine the mean velocity of atoms and their velocity distribution. Basically every physicist has heard of the "Stern-Gerlach Experiment" either in an introductory lecture on atomic physics or in a physics text book. This experiment, which was carried out jointly by Stern and Gerlach, furnished proof of the fact that the orientation of atomic rotational impulses with regard to an external magnetic field can only take discrete values (directional quantization). Since the angular momenta are connected with magnetic momentums, the forces exerted on an atom in an inhomogeneous magnetic field also take on discrete values. A silver atomic beam for instance will, after passing through an inhomogeneous magnetic field, divide itself into two partial beams of atoms with different orientations (symbol "+" or "-" of the magnetic momentum). In one of their experiments, two members of Stern's staff

* Dr. Andreas Bauch, Working Group "Dissemination of Time", e-mail: andreas.bauch@ptb.de

** Dr. Stefan Weyers, Working Group "Unit of Time", e-mail: stefan.weyers@ptb.de

***Dr. Ekkehard Peik, Department "Time and Frequency", e-mail: ekkehard.peik@ptb.de

from Hamburg (Otto Frisch and Emilio Segrè) blocked one of the partial beams behind magnet "A" (polarizer). With the aid of a second magnet "B" (analyzer), the atoms remaining in the beam were directed towards a detector due to the orientation of their angular momentum. Then, transitions between the directional quantization states in the intermediate region "C" were induced by means of a static magnetic field with a fast change in direction, so that the number of atoms diverted towards the detector by the analyzer decreased. In 1938, Isaac I. Rabi, another of Stern's staff members who had meanwhile joined Columbia University in New York, induced the transition between directional quantization states with the energies E_1 und E_2 in the static C field by irradiating the atoms with HF frequency at the frequency $f_0 = (E_2 - E_1)/h$. The sketch of Rabi's apparatus was taken from one of his early publications (see Fig. 1). Its working principle is explained in more detail at the bottom of Fig. 1: without transitions in the "C" region, the atoms follow the paths represented as continuous lines and reach detector D. If transitions into states are induced with the opposite angular momentum orientation, the atoms "miss" the detector (paths represented by dotted lines). To date, the term "C field" has been used for the region of the weak static field in which the hyperfine structure transition (between states with opposite orientations of their angular momentum) is induced in cesium atoms by interaction with microwave radiation, even though there are no more magnets A and B in a cesium fountain clock, as will be explained later.

As early as 1940, Rabi suggested his method of *molecular beam magnetic resonance* be used not only to investigate atomic properties, but that the transition frequency between two selected states be used, so to speak "in reversal", as a reference for a frequency standard. He identified the hyperfine structure states in the ^{133}Cs atom as being particularly well-suited for this purpose [5]. Here, the transition between the states with the magnetic quantum number $m_F = 0$ is particularly well-suited and is therefore referred to as the "clock transition". Rabi's proposal, which was awarded the Nobel Prize in 1944, made it to the headlines of the New York Times as "*radio frequencies in hearts of atoms would be used in most accurate of timepieces*" [6]. Back then, however, no one had yet succeeded in synthesizing and controlling the frequency $f_0 \approx 9.2$ GHz which is necessary for this purpose. This development was only achieved in the decade after World War II, during which tremendous progress had been made in the field of radar technology [3].

This was also the period in which another considerable development was achieved, namely Norman Ramsey's method of *separated oscillatory fields* [6, 7] by means of which the advantages of resonance spectroscopy in an atomic beam first came into their own. Similar to the sketch shown in Figure 1, the change in the state population is also recorded in a cesium atomic clock (for more details, please see info box 1) after the atomic resonance frequency has been irradiated. If the excitation frequency varies around the frequency of the clock transition, a resonance curve is recorded whose spectral width, due to the absence of spontaneous transitions, solely depends on the duration of the interaction with the HF field. To this end, also the technical prerequisites must be fulfilled; these consist in the static field "C" being sufficiently homogeneous over the entire region

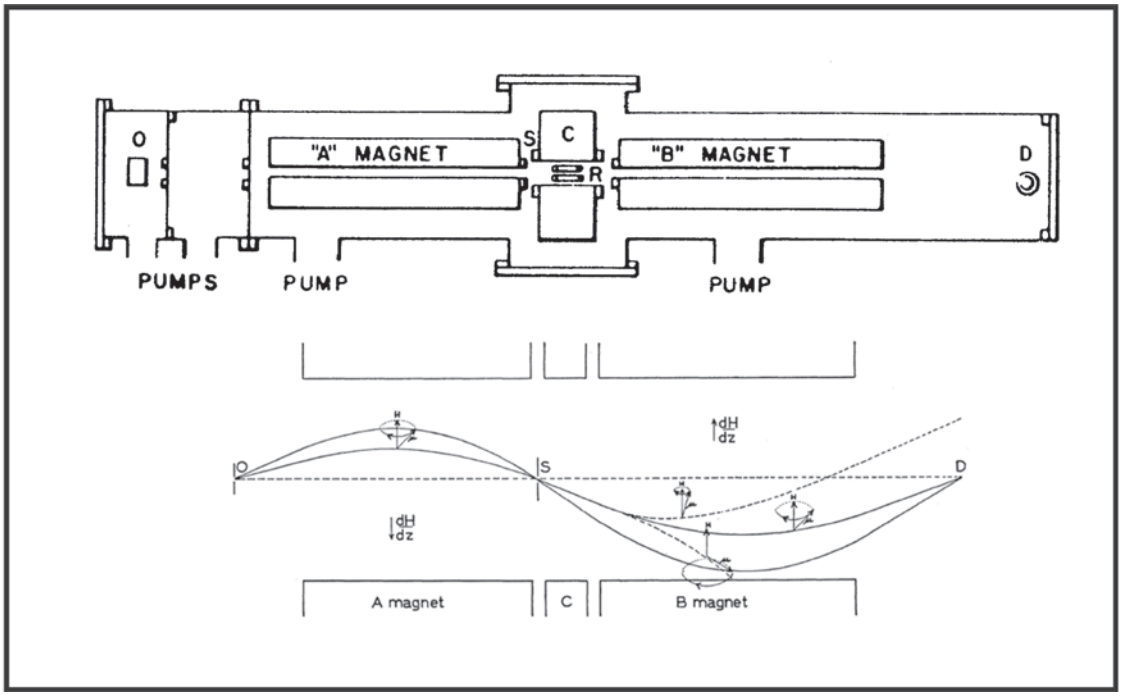


Figure 1:
Sketch of the Rabi apparatus from the 1930s (from [2]) and illustration of the atomic trajectories (from [4]); explanations in the article.

i 1 Principle of the atomic clock

Atomic clocks exploit the property of atoms to emit or absorb electromagnetic waves with a characteristic oscillation frequency. At the transition between two energy states of an atom, E_1 and E_2, a photon is emitted or absorbed with the frequency f_0, where $f_0 = (E_2 - E_1)/h$; h being Planck's constant. Atomic clocks are based on atomic transitions between two energy levels which have a long natural lifetime and are hardly influenced by electric and magnetic fields. For technical and practical reasons, the transitions used in the first decades after 1950 first had resonance frequencies in the microwave range. Suitable atoms are thus alkaline metals with their hyperfine splitting of the ground state, but also ions with an alkaline-like electron configuration.

The transition is induced by electromagnetic irradiation acting on the atom from the outside. For this purpose, the atoms are prepared in *one* state, and after the irradiation, the population of the *other* state is registered. The change in state shows resonance behavior with a spectral "linewidth" W of approx. $W \approx 1/T$, where T is the time during which the atoms interact with the irradiation. Figure K1.1 illustrates the functional principle of PTB's primary atomic beam clocks. Based on a quartz oscillator, a signal of frequency f_p with $f_p \approx f_0$ is generated periodically around the center frequency by means of a frequency generator and is coupled into the resonance apparatus.

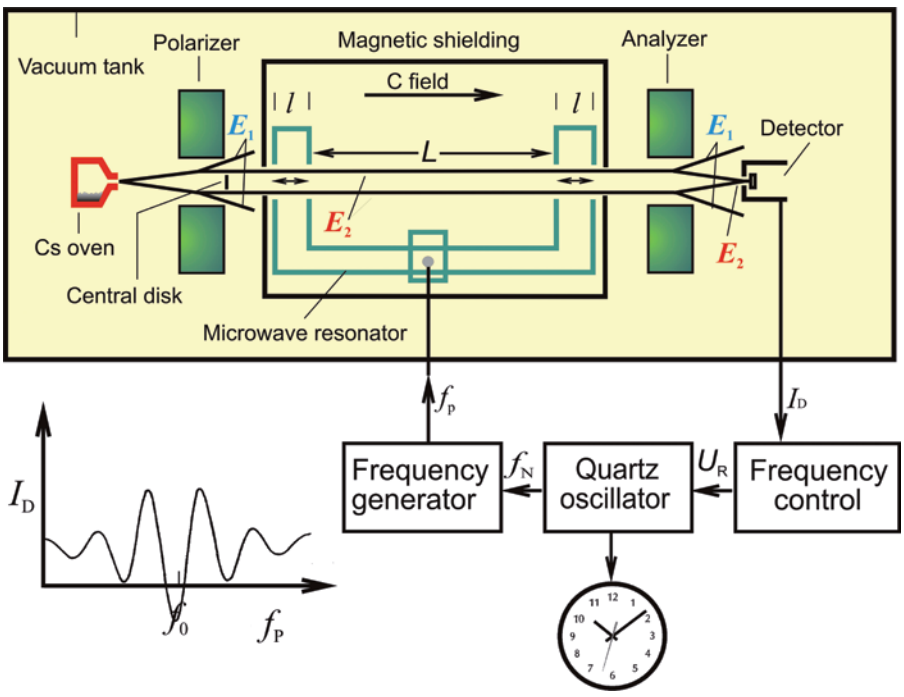


Figure K 1.1: Principle of PTB's cesium atomic beam clocks CS1 and CS2. Bottom left: Scheme of the detector signal I_D as a function of the irradiated frequency f_P corresponding to Fig. 4 in the article.

The signal I_D contains the information looked for, namely whether the frequency f_p is in agreement with the transition frequency of the atoms f_0. The detection of the signal I_D synchronous with the modulation provides a signal U_R for the adjustment of the quartz oscillator. Its natural frequency fluctuations and its frequency drift (aging) are corrected to the extent defined by the control time constants that have been set, and the inherent stability of the atomic resonance determines the quality of the output signal. Starting from the quartz oscillator, the standard frequency f_N and, via a splitter, 1-pulse-per-second signals are generated.

The most important specifications of a clock (its frequency instability and systematic uncertainty) are generally stated as relative quantities. The symbol y designates the relative frequency difference between the clock and a reference standard. A change in the time difference between the clock and the reference amounting to 0.864 ns per day corresponds to a value of $y = 10^{-14}$. The time-dependent change in the output frequency of the clock (i.e. its *frequency instability*) is usually described by means of the Allan standard deviation $\sigma_y(\tau)$, which is calculated according to the relation

$$\sigma_y(\tau) = \left(\sum_{i=1}^{n-1} \left(y_{i+1} - y_i \right)^2 / 2(n-1) \right)^{1/2}$$

if n relative frequency differences y_i occur over the averaging period τ. Figure K1.2 shows the relation between $\sigma_y(\tau)$ of the atomic frequency standards mentioned in this paper and the averaging period τ (the so-called "σ-τ diagram").

In the text, the term *systematic uncertainty* is also used: it describes the estimated agreement between the duration of the second that is realized by the clock and the duration that is laid down in the definition (see text); it is, strictly speaking, only applicable to cesium atomic clocks. For the best commercially available cesium atomic clocks, the manufacturer states an uncertainty of $5 \cdot 10^{-13}$.

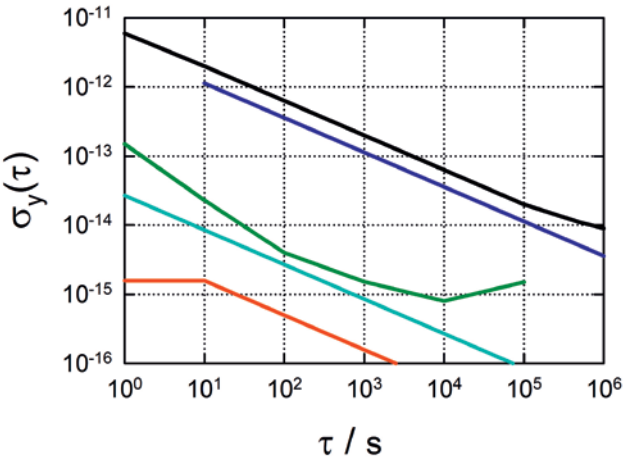


Figure K1.2:
Relative frequency instability $\sigma_y(\tau)$ as a function of the averaging period τ for various atomic frequency standards mentioned in the article (schematically): Commercially available cesium atomic clock (black), primary atomic beam clock PTB-CS2 (blue), hydrogen maser (green), fountain clock PTB-CSF2 (cyan blue) and PTB's optical Yb clock (red).

Info 1: Principle of the atomic clock

and in the phase of the excitation field being sufficiently constant.

At a wavelength of approx. 3 cm, the latter can only be attained over a very small range. Ramsey therefore had the idea of guiding the HF field in a two-arm microwave resonator. In the first arm of the resonator, the atoms (velocity v) are then transferred into a coherent superposition state over the length of approx. 2 cm and, after the free drift section with the length L, they are irradiated again with the HF field in the second resonator arm. The resonance of the transition probability observed then has the spectral width $W \approx v/(2L)$. A fixed phase relation between the two irradiation fields is necessary; in addition, the mean value of the C field in the drift section must agree with the magnetic fields in the two irradiation sections. Ramsey's method allows the realization of interaction sections of approx. 1 m in length. Due to the velocity of the atoms in a thermal beam (≈ 200 m/s), linewidths on the order of ≈ 100 Hz are obtained. In the next section, it will be shown how considerably smaller linewidths $W < 1$ Hz can be attained without significantly extending the drift section, by simply reducing the velocity v.

The first cesium atomic clock (see info box 1) was developed at the same time at the American *National Bureau of Standards* (NBS) and at the British *National Physical Laboratory* (NPL). You can read about the circumstances in which these clocks were developed in [2] as well as in Louis Essen's memories [8] which contain numerous

entertaining anecdotes. The NPL was ahead of NBS in presenting the first functioning atomic clock in 1955 [9]. Essen writes – not without British understatement: "*The atomic clock was made possible through the brilliant theoretical and experimental work of a number of scientists, several of whom received Nobel prizes, but the clock itself is very simple, as can be seen from the sketch below*" (p. 66 in [8]). From 1955 to 1958, the duration of the unit of time valid at that time – that of the ephemeris second – was determined in cooperation with the United States Naval Observatory, Washington, to be 9 192 631 770 periods of the Cs transition frequency [10]. The value stated as the uncertainty of this numerical value was 20, although no one actually knew the duration of the ephemeris second with a relative uncertainty of $2 \cdot 10^{-9}$. Nevertheless, this measurement result provided the basis for the definition of the unit of time in the International System of Units (SI) which was decided in 1967 by the 13th General Conference of Weights and Measures (CGPM) and which is still valid today. The frequency of the clock transition, f_0, was laid down as being 9 192 631 770 Hz.

The significance of the atomic clock for science, technology and, last but not least, for military applications soon became obvious. As early as 1958, the first commercial cesium atomic clock, "Atomichron", was presented [3]. In the early 1960s, PTB purchased such an *Atomichron* and used it to monitor the rates of the quartz clocks in Braunschweig and at the DCF77 emitter in Mainflingen [11].

In the mid-1960s, PTB started to make efforts to develop a cesium atomic clock of its own and these efforts were – eventually – successful. The CS1 clock (see Fig. 2) implemented new ideas of Holloway and Lacey [12]. The following points were regarded as essential advantages [13]:

- reduced frequency instability due to a two-dimensional focusing of the atoms with magnetic lenses (instead of the dipole magnets used until then);
- axial-symmetric geometry of the atomic beam with small radial extension, and
- reduced inhomogeneity of the C field by using a long cylindrical coil and cylindrical shielding – instead of a magnetic field transverse to the beam direction.

CS1 was used for the first time in 1969 [14] and has been ticking ever since. There are only two primary clocks with a thermal atomic beam operating worldwide today: PTB's CS1 and CS2.

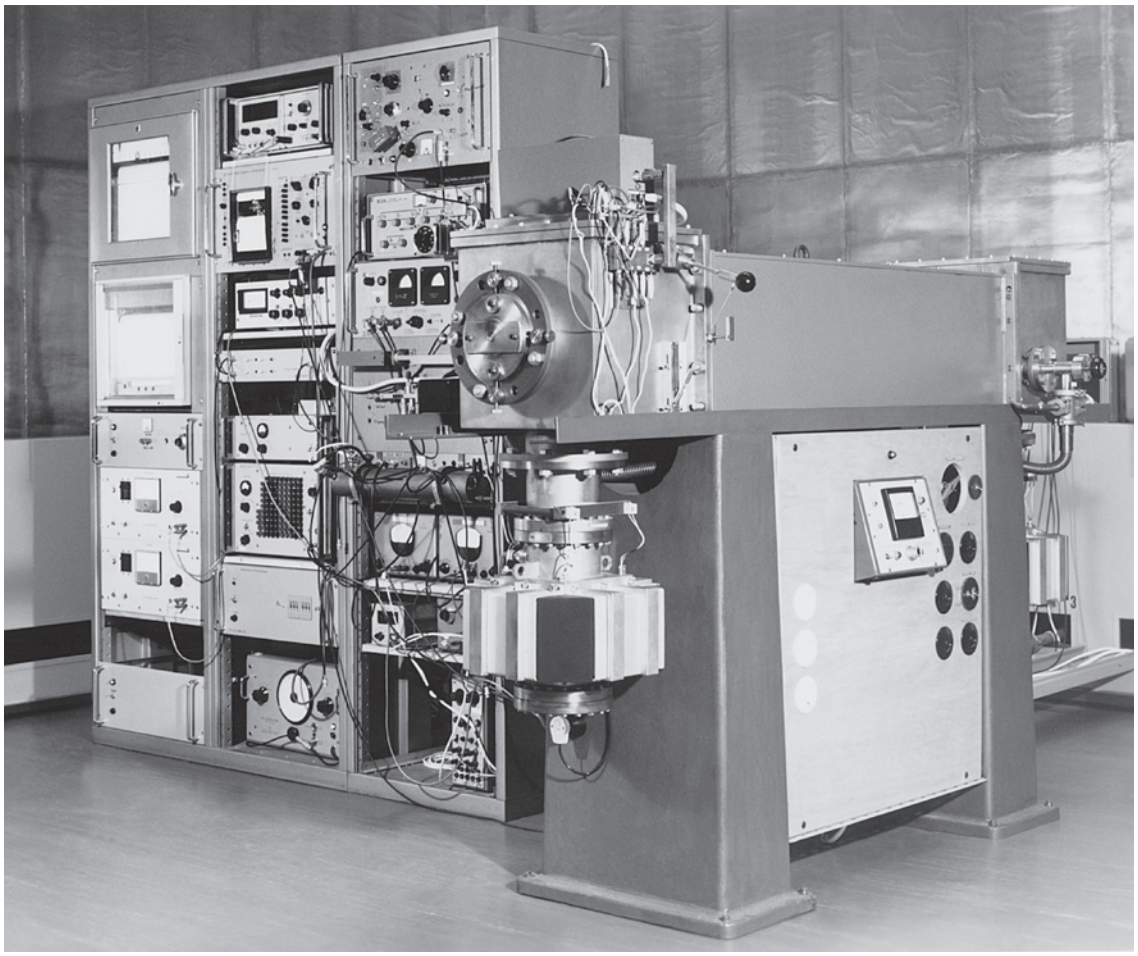


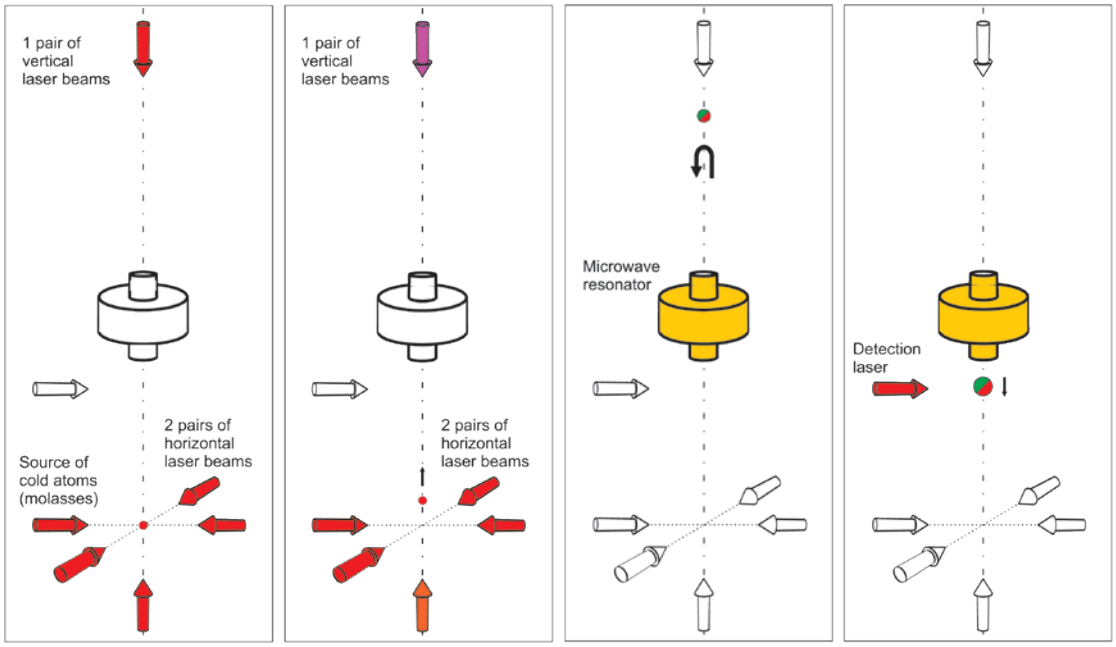
Figure 2:
Primary atomic beam clock CS1 in PTB's atomic clock hall (1969).

CS2 was completed in 1985; its constructional principle is very similar to that of CS1. The estimated uncertainty for CS2 amounts to approx. $1.2 \cdot 10^{-14}$ [15]. In the past two years (until September 2015), the mean relative deviation of CS2's frequency from the SI unit, as approximately disseminated by means of TAI (see info box 3), has amounted to $3.3 \cdot 10^{-15}$ with a standard deviation of the monthly mean values of $3.1 \cdot 10^{-15}$. The uncertainty estimation from the late 1980s is thus very probably correct.

On the way to the fountain clock

While atomic clocks based on the design which was introduced by Rabi and Ramsey were being developed, Jerome Zacharias from the *Massachusetts Institute of Technology* was working on surmounting the limitations of this design as early as in the 1950s. From secondary sources [3, 7], we know of ambitious – but unfortunately failed – experiments with an upward-facing thermal atomic beam. Under the influence of gravity, the atoms in such a beam first become slower before being accelerated downward after passing the reversal point. Zacharias wanted to be able to detect the few atoms in the thermal beam that were particularly slow after they had interacted with the microwave field while going up and down. In an apparatus of several meters in height, a linewidth of less than 1 Hz would have been attained (i.e. considerably less than the 100 Hz achievable until then), and the center of the line would have been much easier to determine with 0.1 Hz accuracy, which – relatively – corresponds to 10^{-11} of the resonance frequency of 9.2 GHz. However, the researchers did not succeed in detecting atoms which had passed the microwave resonator a second time on their ballistic flight path. Due to collisions of atoms around the oven nozzle, the tiny proportion of extremely slow atoms in the thermal velocity distribution which was exploitable for the experiment was further reduced [16]. Zacharias called his project "Fallotron" or "Fountain".

The latter denomination was adopted by a group at the *École Normale Supérieure*, Paris, when its members presented their "Zacharias Fountain" with laser-cooled cesium atoms in 1990 [17]. Due to the laser cooling (which is dealt with separately in info box 2), it is possible to generate a cloud of cold atoms in a cell with cesium vapor where the thermal motion of the atoms is, to a large extent, "frozen". The temperature of the cold atoms is around 1 µK, and the distribution of the relative velocities occurring is approx. 10 000 times more narrow than the thermal velocity profile of a gas made of the same type of atoms at room temperature. This allows a fountain with a pre-defined flight height of the atoms to be realized with a cold atom cloud; hereby, the spatial expansion of the cloud with time remains small due to the narrow velocity distribution. Contrary to a continuous



Loading a cloud of cold cesium atoms; the atoms are prepared in a hyperfine structure (HFS) state (1).

Starting the cloud in state (1) by detuning the frequencies of the vertical laser beams.

Flight of the atoms after their first interaction with the microwave field in the resonator; the population of the second HFS state (2) is shown by the color green.

Detection of the population of the atomic states (1) and (2) in the cloud; exactly in resonance, all atoms would be in the state (2) and the cloud would be totally "green".

Figure 3:
Temporal cycle of the functional steps of a fountain clock.

thermal beam, here cesium atoms must be periodically trapped, cooled and then accelerated upward. To be able to excite the atomic resonance transition with as few disturbances as possible, free atoms must be observed and the trapping and cooling light field must be switched off.

The sequence of the individual function steps is illustrated in Fig. 3. The atoms are first prepared in one of the two hyperfine structure states (red). They are then accelerated upward to the speed of a few meters per second within 1 ms, so that they fly about one meter high before falling back down

i 2 Laser cooling

To give a simplified explanation of laser cooling, let us consider an atom moving longitudinally to two counteracting laser beams of the same frequency ν_L and intensity. The frequency of the laser light is tuned slightly lower (i.e. tending to red) than the optical excitation frequency ν_R at which the atom at rest absorbs a photon resonantly (i.e.: most probably). Since it is in motion, the atom considers the frequency of the oncoming light as enhanced due to the Doppler effect, i.e. shifted towards the resonance frequency. Simultaneously, the frequency of the light propagating in the same direction as the atom, on the contrary, seems off-tuned further away from the resonance frequency (towards red). This is why it is most probable that a counter-propagating photon will be absorbed. During the absorption, the momentum of the photon from the laser beam is transmitted onto the atom, whereas the photon is re-emitted in a random spatial direction. Thus, the momentum transfer is, on average, directed against the original direction of motion when many photons are scattered. Due to the aforementioned red shift (ν_L-$\nu_R < 0$), the energy of the emitted photons is higher by h (ν_R-ν_L) than that of the absorbed ones; h being Planck's constant. The energy balance shows that the atoms lose kinetic energy during the scattering. Since after 30 ns, the cesium atoms, to give an example, go back from the optically excited state to their ground state, the absorption/emission process repeats itself very fast so that the atom can be brought from room temperature to nearly absolute 0 within fractions of a second. To understand the details and the full theory of laser cooling better, it is useful to read the Nobel prize award contributions of the laureates from 1997 or articles in more recent textbooks.

In order to form a source of cold atoms for a fountain clock, one should proceed as follows: three orthogonal pairs of laser beams with opposite radiation directions and matching frequency, polarization and intensity are made to overlap in a vacuum chamber which contains the atomic gas at room temperature. Inside the overlapping volume, so-called "optical molasses" is formed which consists of laser-cooled atoms with velocities in the range of a few centimeters per second. The atoms move in the molasses like in thick syrup since the force acting upon the atoms is similar to friction: the force is proportional to the momentary velocity and its direction is opposite to this velocity. Optical molasses, however, does not represent an actual trap for atoms since there is no force acting upon the atoms which is directed toward the intersection point of the laser beams. To reach higher numbers of atoms in a shorter time, it is also possible to populate the molasses not via background gas, but via a beam of pre-cooled atoms which is made available by means of an additional mechanism. Only with the low temperatures from 1 µK to 2 µK attained in molasses is it possible to realize a cesium fountain clock as described in the article.

under the influence of gravity. Acceleration to the initial speed v is done by detuning the frequency of the two vertical laser beams that act in the opposite direction to each other: the frequency of the upward-facing laser beam is detuned by $\delta\nu_L$ and that of the downward-facing one by $-\delta\nu L$. The following thus applies: $v = \lambda\, \delta\nu_L$, where λ is the wavelength of the laser radiation used.

Taking a closer look at this procedure, the sequence of laser frequency detunings and changes in laser intensity, which is necessary to attain perfect cooling and acceleration of the atoms, is rather complicated [18]. On their ballistic trajectory, the atoms cross a microwave field with a frequency close to the clock transition ($f_p \approx f_0$) twice – once when flying upward and once when falling back down. During that time, the laser beams are blocked. In the detection zone, the population of the two hyperfine structure levels is recorded as a function of the microwave frequency fp. In Fig. 3, the transition probability is only 1/2 since the microwave frequency was increased by half the linewidth ($f_p = f_0 + W/2$). In the next cycle, the frequency is decreased by half the linewidth ($f_p = f_0 - W/2$). The difference in the transition probabilities of successive recordings provides the position of f_p with respect to f_0. The time between the two microwave irradiations (i.e. the effective interaction time) amounts to approx. 0.6 s in practically all fountain clocks that have been realized to date, and the linewidth of the transition is then 0.8 Hz. Figure 4 illustrates the gain obtained with PTB's fountain clock CSF1, compared to a clock with a thermal atomic beam. The corresponding values in PTB's atom-beam clocks are namely 8.5 ms and 60 Hz, respectively.

Challenges and limitations in 2016

At the national metrology institutes worldwide, more than a dozen cesium fountain clocks are currently in operation or under development. SYRTE (*Systèmes de Référence Temps-Espace*) led the way at the *Observatoire de Paris* where the fountain clock FO1 was working as early as 1994 [19] and where three fountain clocks are now operated. PTB's fountain clocks are shown in Fig. 5. With the fountain clock CSF1 [20], the first ever resonance line was recorded in October 1998 and since mid-1999, frequency measurements have been performed on a regular basis. Based on the findings from the development and the first years of operation of CSF1, a new, enhanced version was developed at the turn of the millennium: the fountain clock CSF2 [21, 22]. CSF2 has now been available for frequency measurements for six years and with five other fountain clocks, it is part of the leading group of the (currently) most precise clocks worldwide with relative uncertainties on the order of a few parts per 10^{-16} [23]. Moreover, CSF1 and CSF2 act as a reference for one another, for monitoring purposes on the one hand, and to investigate systematic effects on the other.

Which of the predicted advantages have been verified to date and what problems occurred? Let us first look at the frequency instability. For a frequency measurement with a statistic uncertainty of $1 \cdot 10^{-14}$, the primary beam clock CS2 needs 36 hours whereas it takes the fountain clock CSF2 seven seconds only. If a quartz oscillator is set to the atomic resonance line, then its frequency instability is proportional to $(Q\, S/R)^{-1}$ where Q is the line quality factor (the resonance frequency divided by the linewidth) and S/R is the signal-to-noise ratio. Since – while S/R remains the same – the frequency instability decreases with decreasing linewidth, CSF2 shows a clear advantage thanks to the approx. 65 times higher line quality factor compared to CS2. However, since in a fountain clock, the duration between the two irradiations (which is responsible for the linewidth) in the Ramsey resonator only increases by the root of the

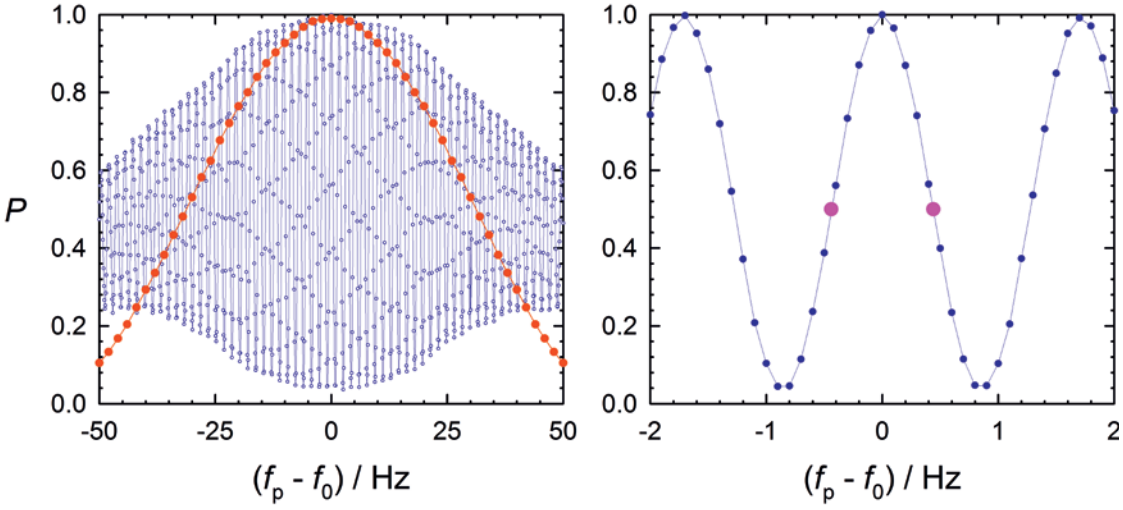


Figure 4: Transition probabilities P between the states 1 and 2 of the clock transition in ^{133}Cs as a function of the microwave frequency f_p (Ramsey resonance curves), recorded with PTB's fountain clock CSF1 (blue) and beam clock CS2 (red).

launch height, some sort of practical limitation is reached by fountain clocks with regard to the linewidth in the gravitational field of the Earth: no one is willing to build a fountain clock of 10 meters! A fountain clock of such a height would, with atoms at the same temperature, have the disadvantage that fewer atoms could be detected because a larger proportion of atoms would not reach the detection zone due to the larger expansion of the atomic cloud on the longer path.

This leads us to the second decisive factor for the frequency instability, namely the signal-to-noise ratio S/R. After the second irradiation with microwaves, two signals are obtained in the detection zone by exciting the atoms with laser radiation and recording the fluorescence they then emit; from these two signals, the excitation probability of the microwave transition from state 1 into state 2 can be derived. The two detection signals, N_1 and N_2, are each proportional to the number of atoms detected in the states 1 and 2 and to the number of photons detected per atom. A fountain is usually operated in such a way that N_1 and N_2 are equal (see Fig. 3 and the magenta dots in Fig. 4, right), since the transition probability corresponding to the signal S, $P = S = N_2/(N_1+ N_2) \approx 1/2$, is then most sensitively dependent on frequency changes in the alternating field used to irradiate the atoms. Due to the statistic variations in the distribution of the atoms onto the two different states (quantum projection noise), the noise R of the signal S is $R = (N_1+ N_2)^{-1/2}(P(1\text{-}P))^{1/2}$. Thus, for normal operation as a fountain with $P = 1/2$, the simple relation $S/R = (N_1+ N_2)^{1/2}$ is yielded, i.e. a signal-to-noise ratio which is proportional to the root of the total number of atoms detected.

To attain this – to some extent fundamental – lowest noise in a fountain clock presupposes that all other technical contributions to noise which normally exist in the detection signal must be thoroughly reduced. One of the problems is that the fountain operates in a pulsed mode: the time interval between the two flights of the atoms through the microwave resonator is followed by a period for atom detection and then by a phase in which the atoms are loaded and started anew. During this dead time, the control of the quartz oscillator is ineffective since there is no information on the time-dependent behavior of the oscillator phase. Part of the phase fluctuations of the quartz oscillator is transmitted in the form of frequency variations also at averaging times that are high compared to the cycle time of the fountain (typically 1 s to 1.5 s). John Dick from the ion trap group of the *Jet Propulsion Laboratory*, Pasadena, described this behavior for the first time; it is therefore called the "Dick Effect" [24]. Due to this effect, the best frequency instability that can be attained with a fountain clock whose 9 GHz signal is derived from an oscillator with one of the lowest-noise oscillating quartzes available is slightly less than $10^{-13}\ (\tau/1\text{s})^{-1/2}$, with τ being the averaging time.

In contrast, to obtain a frequency instability in the lower range of $10^{-14}\ (\tau/1\text{s})^{-1/2}$, the 9 GHz signal must either be derived from a cryogenic oscillator or from a microwave oscillator which is stabilized onto a narrow-band laser by means of a frequency comb [25]. It is interesting to note that in the latter case, technologies which have been developed in connection with optical clocks (see last but one section – narrow-band lasers, frequency comb) contribute to the performance of fountain clocks. In turn, this enhanced performance in frequency measurement of potential optical clock transitions with reference to the SI second as realized with fountain clocks brings about considerable advantages.

The curve plotted for the fountain clock CSF2 in Fig. K1.2 $\sigma_y(\tau) = 2.7 \cdot 10^{-14}\ (\tau/1\text{s})^{-1/2}$ results, in principle, exclusively from the quantum projection noise of approx. $2.8 \cdot 10^6$ detected atoms per fountain cycle. To reach this frequency instability, the loading of the atom source of CSF2 had to be effected from a beam of pre-cooled atoms in order to send such a high number of atoms through the

Figure 5:
PTB's cesium fountain clocks CSF1 (in the background) and CSF2 (in the foreground) in 2005.

fountain that even the *S/R* is higher by a factor of approx. 2 than in CS2. This, in combination with the linewidth which is reduced by a factor of 65, yields a frequency instability improved by a factor of 130. The main obstacle to further reducing it by using an even greater number of atoms in order to attain an even lower quantum projection noise, is the increasing frequency-shifting effect of collisions among the cold atoms, which we will discuss in the following together with the achievable systematic uncertainty of fountain clocks.

What are the advantages of the fountain clock with respect to accuracy? Table 1 lists the main frequency shifts and the resulting uncertainty contributions of the primary clock CS2 [15] and of the fountain clock CSF2 [21, 22]. The advantage of the fountain clock takes effect in particular where the shifts are proportional to Q or where they are directly scaled with the velocity v or v^2 [26] (marked with (*) in Table 1).

Some of the uncertainty contributions of CSF2 (total uncertainty: $2 \cdot 10^{-16}$) are smaller by more than one order of magnitude than those of CS2 (total uncertainty: $120 \cdot 10^{-16}$). For example, CS2 needs a static magnetic field of approx. 8 µT so that the frequency separation of neighboring transitions that are dependent on magnetic fields is high, compared to their linewidth. It then becomes possible to excite the transition between the Zeeman sublevels with the magnetic quantum number $m_F = 0$ (the clock transition) as the transition between two isolated atomic levels selectively. In the fountain clock, a field strength of approx. 0.15 µT suffices due to the reduced linewidth. The shifting of the resonance frequency of the clock transition amounts to 2.9 Hz in the first case and only to 0.0009 Hz in the second case; it is then correspondingly easier to determine this shift and to correct it down to field zero. Moreover, in a conventional atomic clock, the magnetic state selection leads to an asymmetric population of neighboring Zeeman sublevels with $m_F = \pm 1$ with respect to the $m_F = 0$ state population. In this way, undesirable transitions in the hyperfine structure (HFS) multiplet in the vicinity of the clock transition are excited to a different degree. In general, this leads to a frequency shift of the clock transition [12]. In the case of a fountain clock, however, the population of neighboring sublevels is suppressed to a large extent after the laser cooling step [18], and it is symmetric. The corresponding uncertainty contribution is therefore negligible.

Since CSF2 does not require the cesium reserve to be heated (as is the case for CS2) in order to generate a sufficiently intensive atomic beam, the temperature distribution all in all is much more homogeneous here. This leads to an uncertainty of the frequency shift due to thermal radiation (Stark effect) that is lower by one order of magnitude. Similarly, the uncertainty caused by the quadratic Doppler effect is practically negligible [18], since the mean atomic velocity between the two resonator passages by the atoms reaches 1.5 m/s only in CSF2, whereas it amounts to approx. 100 m/s in CS2.

To understand the varying influence of the phase difference of the resonator in CS2 and in CSF2, you have to keep in mind that the successive excitation of the atoms in the *single* microwave resonator in a fountain clock is equivalent to the successive interaction of the atoms in the *two* spatially separate irradiation sections of the resonator of an atomic beam clock. In the *single* resonator of the fountain clock, the irradiation fields located in the same spot are, by nature, always in phase, whereas in the two-arm resonator of an atomic beam clock, power is lost to the walls in addition to mechanical asymmetries, which leads to a phase difference. Determining these experimentally is difficult, which limits the achievable accuracy (e.g. of CS2). In the case of CSF2, similar to other fountain clocks, a challenge remains: namely, to control the influence of spatial phase gradients that are transversal to the moving direction of the atoms [22, 27]. Such phase gradients lead to frequency shifts, also in fountain clocks. Averaged across the different expansion and the possibly slightly different horizontal position of the atomic ensemble each time the cesium atoms rise and fall, phase differences of the acting microwave field generally occur. These also result from power that is lost to the resonator walls or from asymmetries occurring when the field is coupled into the resonator [27]. Evaluating such effects requires demanding computations and experiments involving the tilting of the fountain's axis against the vector $\boldsymbol{g}$ of the gravitational acceleration [22].

In fountain clocks, particular attention has always been paid to the effect of collisions among the cold cesium atoms. At the prevailing temperatures (≈1 µK), the shifts of the energy levels involved in the clock transition due to impacts are relatively high and have therefore been investigated thoroughly – both theoretically and experimentally (e.g. [28, 29]). These investigations have led to various measures aiming to get the resulting frequency shift and its uncertainty under control. In principle, this requires the atomic density to be varied during the operation of the fountain clock in order to determine the undisturbed transition frequency from the transition frequencies measured at diverse densities by linear extrapolation to vanishing density. To vary the density, it is possible, for example, to change the loading time of the atomic cloud. A simple method to reduce systematic errors caused by shifts due to collisions as much as possible is to use a low number of atoms in a relatively wide atomic cloud. The disadvantage

in this is, however, the resulting low S/R .

In the case of detected atoms amounting to several 10^6 (as are needed for frequency instabilities in the lower 10^{-14} $(\tau/1s)^{-1/2}$ range), however, relative collisional shifts of 10^{-14} occur, so that a very precise procedure must be used to vary the atomic density in order to keep extrapolation errors at the level of 10^{-16} or lower. The so-called "*Rapid Adiabatic Passage*" *method* [18] enables the density of the atoms used for the clock transition in the entire atomic cloud to be varied by a factor of exactly 2 by means of an additional microwave resonator located below the Ramsey resonator. The resulting frequency difference (measured between operation at full density and operation at half density) can then be used for the correction of the collisional shift. With the number of detected atoms increasing further (above the range of several 10^6), this method would, however, also lead to inacceptable systematic uncertainties of several 10^{-16}. In this respect, we have to deal with limitations which force us to make a compromise between the achievable frequency instability and the systematic uncertainty.

To conclude, one can keep in mind that since the times of the pioneers (Rabi, Ramsey, Zacharias and Essen), the relative uncertainty of cesium clocks with a thermal atomic beam has been reduced from approx. $1 \cdot 10^{-10}$ to approx. $1 \cdot 10^{-14}$. With laser-cooled atoms in fountain clocks, uncertainties in the range of $1 \cdot 10^{-16}$ have been attained, but it seems difficult to reach values well below this threshold.

Even more precise: optical clocks

Parallel to the development of the cesium fountain clocks, a new type of atomic clock has been intensively investigated for more than 20 years – the optical clock [30]. Optical clocks also rely on an atomic transition as a reference, however, not in the microwave range, but in the optical spectral range. The oscillator used for these clocks is – correspondingly – a frequency-stabilized laser. By increasing the clock frequency by five orders of magnitude (from approx. 10^{10} Hz in a cesium clock to approx. 10^{15} Hz in an optical clock), considerable stability and accuracy gains can be obtained. Thus, the short-term stability can be improved due to the considerably higher number of oscillating cycles per time unit. Enhanced accuracy is obtained due to the fact that some of the disturbing external influences on the atom (e.g. collisions with other atoms or the influence of magnetic fields) cause a characteristic shift magnitude ΔE of the energy levels which, as a relative frequency

Table 1:
Main uncertainty contributions of PTB's primary clock CS2 [15], fountain clock CSF2 [21, 22] and optical Yb clock (see below).

Cause of a frequency shift	**Relative frequency shift δF and relative uncertainty $\delta\delta F$ in multiples of 10^{-15}**					
	Cs clock CS2		Fountain clock CSF2		Optical clock (Yb^+ E3)	
	δF(CS2)	$\delta\delta F$(CS2)	δF(CSF2)	$\delta\delta F$(CSF2)	δF(Yb^+ E3)	$\delta\delta F$(Yb^+ E3)
Magnetic field* (quadratic Zeeman effect)	317 500	1	99.85	0.01	–0.0404	0.0006
Stark effect thermal radiation)	–17	0.5	–16.51	0.06	–0.0705	0.0018
Quadratic Doppler effect* (relative time dilation)	–50	0.5	–0.01	<0.01	–0.0037	0.0021
Excitation of undesired transitions in the HFS multiplet of ^{133}Cs	0	4	0	0.0013		
Resonator phase difference*	255	10	0.03	0.15	–0,0037	
Atomic collisions	0	<0.5	–10	0.05	0	0.0005

i 3 Time system

Under the coordination of the International Bureau of Weights and Measures (BIPM – *Bureau International des Poids et Mesures*), 72 time institutes are participating in the realization of the international atomic time scale at the time of writing (September 2015). By averaging more than 400 atomic clocks, approx. 300 commercial cesium atomic clocks and 100 hydrogen masers of all these institutes, the BIPM obtains a so-called "free atomic time scale". Its basic scale unit is compared with the primary cesium fountain clocks (as well as CS1 and CS2), provided their measured values are available. In the past 24 months (until September 2015), data from 11 different fountain clocks (from China, Germany, France, the UK, India, Italy, Russia and the USA) were supplied. Such comparisons reveal by how much the basic scale unit of the free atomic time scale has to be changed to agree with the SI second as realized by the primary clocks. The resulting correction is applied by calculation, and one then obtains the international atomic time TAI (*Temps Atomique International*) The TAI second can currently be aligned with the SI second with 10^{-15} s accuracy on average.

This degree of agreement would have been practically unimaginable only 10 years ago; it has only become possible thanks to the quality and the availability of fountain clocks. Eventually, UTC (*Coordinated Universal Time*) is obtained by adding leap seconds to TAI.

Each of the 72 institutes realizes a replica of UTC called UTC(k). Depending on how sophisticated the realization and the quality of the clocks locally available are, this replica is more or less in agreement with UTC. In 2010, PTB was the first time institute to derive its UTC(k) realization from one of its cesium fountain clocks. Since then, the deviation between UTC and UTC(PTB) has never exceeded 10 ns. Other institutes have now followed suit. SYRTE (*Observatoire de Paris*) has thus been using fountain clocks since the end of 2012 to control UTC(OP). Figure K3 shows UTC(PTB) and UTC(OP) which are some of the most stable time scales currently realized with respect to UTC. One can observe how the deviations in UTC – UTC(k) have become considerably smaller due to the use of fountain clocks, first in the blue curve and then in the orange one.

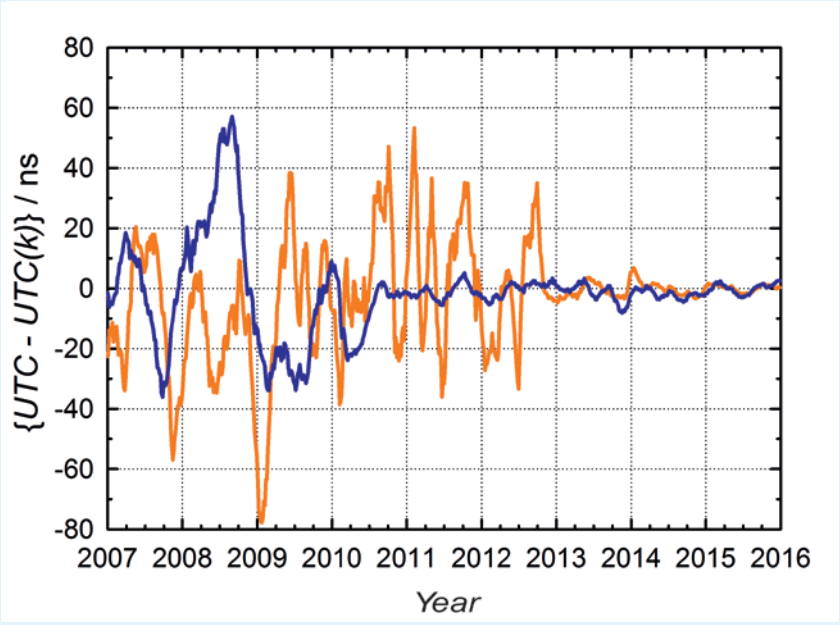


Figure K3:
Evolution of two of the most stable time scales realized to date: UTC(PTB) (blue) of PTB and UTC(OP) (orange) of SYRTE (Observatoire de Paris) with reference to UTC over the period from 2007 until the end of 2015.

shift $\Delta f/f=\Delta E/(hf)$, has less influence at the higher frequency of the optical clock than for a clock operating in the microwave range. The higher clock frequency, however, raised three challenges which were more difficult to handle:

1) designing a stable interrogation oscillator;
2) counting the frequency, and,
3) last but not least: localizing the atoms.

(1) The linewidth and the frequency stability of the interrogation laser are of great significance for the short-term stability of optical clocks. In most setups, diode lasers with unstabilized linewidths of a few MHz are used. To reduce the linewidth, the laser is stabilized onto an external optical resonator. This resonator consists of two high-reflectivity mirrors placed on a stable spacer which, under vacuum, are excellently shielded from temperature variations and mechanical vibrations. With reflectivities of 99.9997 % of the best available mirrors and a spacer of approx. 10 cm in length, resonances with a width of only a few kHz are obtained. Since these resonances can be detected very fast with very little noise, the frequency fluctuations of the lasers can be reduced down to linewidths of less than 1 Hz by means of an electronic

i 4 Applications

The accuracy of cesium atomic clocks has been improved by a factor of 10 every decade since the first clock of its kind was commissioned in 1955. It is fascinating to notice how improved clocks are used in numerous areas of science and are still needed to investigate topics such as the validity of the theory of relativity, the constancy of fundamental constants or the properties of millisecond pulsars. This requires a relatively small number of clocks in selected places, contrary to applications of daily life such as the operation of telecommunications or energy supply networks. Such applications necessitate networks of oscillators that are reliably synchronized with each other. These networks are often clocked to UTC (or another UTC(k)) using the signals of satellite navigation systems. Such systems (GNSS, *Global Navigation Satellite Systems*) help answer the question “Where am I?” in everyday life, and interest in them will keep on growing with the integration of navigation functions into cell phones and cars. The use of GNSS, however, goes far beyond such functions and can be used to transmit high-precision time information.

The way that GNSS work is based on measuring the signal’s propagation times between several satellites equipped with atomic clocks and the ground receiver. Atomic clocks are required both in the satellites and on the ground: the quality of the signals sent by the satellites directly depends on the stability – more specifically: on the predictability – of the clocks aboard the satellites. The prediction refers to the system time generated in the GNSS which is realized by means of stationary clocks on the ground and is, in turn, kept in excellent agreement with UTC by comparison with the best terrestrial time scales. To this end, the GPS system time is steered towards UTC(USNO) of the U.S. Naval Observatory. The time of the Russian GNSS (GLONASS) is approximated to the time scale UTC(SU) realized at the Russian metrology institute VNIIFTRI which operates two cesium fountain clocks. Last but not least, the system time of the European Galileo system is presently approximated to the average of five UTC(k) time scales, among these UTC(PTB) and UTC(OP), whose excellent properties are described in info box 3.

feedback circuit. The achievable relative stability is then approx. 10^{-16} for times from 0.1 s to 100 s. Over a longer period of time, glass spacers exhibit a frequency drift due to material aging. Resonators made of monocrystalline silicone with cryogenic cooling are thus currently being investigated as a further improvement.

(2) The problem with counting the optical frequency was solved in the year 2000 by the invention of the optical frequency comb generator. It consists in using a laser which emits an extremely uniform series of very short (fs range) optical pulses. The spectrum of this laser resembles a comb with a high number (>10^6) of optical frequencies arranged as a uniform array. It can be used to divide the frequency of a reference laser (similar to the oscillator in an optical clock) in one step down into the microwave range and thus to make it electronically countable. This procedure has been thoroughly tested with various systems and has turned out to be so reliable that numerous (more than 10^{18}) optical oscillations can be tracked without errors, i.e. without losing track of a single cycle. For this invention and its first groundbreaking applications, Theodor Hänsch and John Hall were awarded the physics Nobel Prize in 2005 [31].

(3) To avoid frequency shifts due to the conventional, linear Doppler effect, the movement of the atom in the exciting radiation field must be limited to a range whose dimensions are smaller than the wavelength. In cesium fountain clocks, this wavelength is approx. 3 cm, and the aforementioned criterion can be met even with free-falling atoms. With an optical clock with a wavelength on the order of a few 100 nm, however, this can only be done if the atoms are caught in a trap. Localizing the atoms in an optical clock is done by combining a trapping potential (an atom trap) and laser cooling by means of which the atom is deprived of its kinetic energy. Especially charged atoms (ions) can easily be trapped with electric fields (Fig. 6). Their charge acts as a strap by means of which the atom can be fixed without significantly disturbing its inner structure – and thus the resonance frequency which is critical for the clock. The first optical clock proposals came from Hans Dehmelt (physics Nobel laureate in 1989, together with Norman Ramsey and Wolfgang Paul, the inventor of the "Paul trap" for the storage of ions) in the 1970s and were based on a single ion which was stored and cooled in an ion trap [32]. Neutral atoms lack such straps so that forces can only be exerted by attacking the charge distribution inside the atom, which inevitably has an influence on the electronic structure. It was only in 2001, following a proposal by Hidetoshi Katori, that it became clear that it was possible to build a very precise optical clock also with neutral atoms, namely by setting up a trap which shifts the two energy levels of the transition which determines the clock's reference frequency exactly equally [33]. The trap is formed by the electric field in the interference pattern of several laser beams whose "magic" wavelength is selected in such a way that the afore-

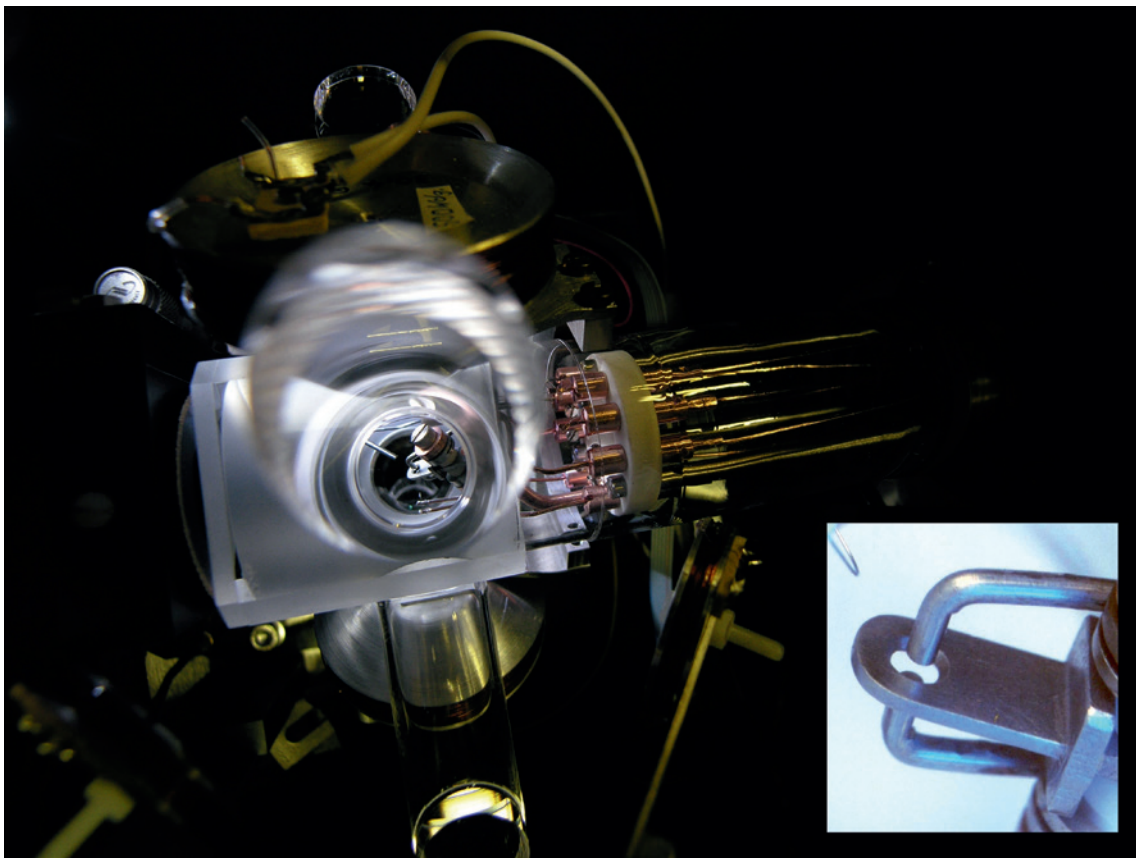


Figure 6:
The ion trap of PTB's optical Yb clock. In the glass vacuum chamber, one can see the electrodes of the Paul trap where an Yb^+ ion is stored. Bottom right: Detailed view of the trap. The ring electrode has an inner diameter of 1.4 mm.

mentioned condition is fulfilled. This arrangement is called an "optical lattice", since many attractive potentials which trap the atoms periodically form in space (Fig. 7). This is where a considerable advantage of a clock with neutral atoms comes to fruition: even at low densities in ultrahigh vacuum, ensembles of several thousands of atoms can easily be cooled and stored. Since identical ions repel each other electrostatically, the ion clock uses a single ion, which leads to the signal achievable to control the frequency of the clock being correspondingly weaker. In return, the ion trap is considerably deeper than the optical lattice for atoms, so that the storage is more stable: a number of experiments have shown that the same ion can be observed for several months in a row.

The storage and laser cooling methods are applicable to a great variety of atoms and ions in different charge states. When selecting an atom for an optical clock, the properties of the reference transition therefore play an important role. In Dehmelt's work [32], the Tl^+ (thallium) ion is mentioned, and this choice is based on the advantageous properties of its electronic shell: in the ground state and also in the first excited state, the two valence electrons are coupled to form a pair with vanishing angular momentum. The shifts of energy levels caused by external electric and magnetic fields are thus significantly smaller than in states with a single valence electron. The transition between the two levels violates the selection rules for electric dipole radiation and therefore has the narrow natural linewidth which is required for an atomic clock. Due to diverse practical disadvantages, no experiments have been carried out with a Tl^+ clock yet; most of the systems investigated to date, however, have the same structure with two valence electrons: in the atoms Sr, Hg, Yb and the ions In^+ and Al^+. The optical clock developed at PTB represents an exception to this rule since it uses Yb^+ [34], a rare earth ion which has an exceptionally narrow line between two levels with orbital angular momenta 0 and 3 that is also hardly influenced by external fields. Figure 8 shows an excitation spectrum of this line. The linewidth is determined by the selected laser pulse duration and is here at 2.4 Hz slightly larger than in a fountain clock. Since the interrogated frequency is, however, higher by a factor of 70 000 (blue light at a wavelength of 467 nm), this signal can be used for a clock with considerably better stability.

Based on these fundamental concepts, spectacular progress has been made over the past 15 years when it comes to the development of optical clocks, from the spectroscopic resolution of linewidths in the range of a few Hz at optical transitions, to measurements of optical transition frequencies with respect to cesium fountain clocks in the uncertainty range of 10^{-16}, to systematic uncertainties in the range of 10^{-18}. On short time scales, optical clocks are now more stable by a factor of 100 and also more accurate by a factor 100 than cesium fountain clocks when it comes to the reproducibility of their reference frequency. Dehmelt concretely mentioned the objective of attaining an accuracy of 10^{-18} as early as 1981 [32]. To reach six orders of magnitude beyond the then best realization of atomic clocks – and this with totally novel methods such as laser cooling and the observation of individual quantum jumps – this vision seemed quite ahead of its time then, however, it became the main motivation of the scientists working in this field. Until mid-2016, four groups of scientists carried out different experiments (two with the ions Al^+ and Yb^+ and two with Sr atoms) and succeeded in reaching this accuracy range. Table 1 shows the most significant uncertainty contributions of PTB's Yb^+ optical clock compared with those of the cesium clocks. The total uncertainty of this clock is currently $3.2 \cdot 10^{-18}$.

The second – on its way to a new definition

Since the frequency comparisons and the measurement or frequency ratios of optical clocks are now possible with lower uncertainties than the current realization of the SI second allows, the debate about a possible new definition of the second is gaining in importance. Looking back at the fundamental principle of the definition of the SI second from 1967, it can be considered as a successful model: an appropriate atomic transition frequency is defined under ideal conditions, and a fixed number of oscillation periods is laid down as the unit of time. It therefore seems logical to pursue this model at an appropriate point in time by selecting a suitable optical transition. The new numerical value of the oscillation periods per time unit would be determined from measurements of the frequency ratio of the optical clock to the cesium clock. In preparation of this approach, a working group of the consulting committee for time and frequency has therefore been collecting and evaluating the results of frequency measurements of suitable transitions since 2001 and has now derived recommended frequencies for so-called "secondary realizations" of the second. Eight different transitions are currently recommended: the hyperfine structure frequency of ^{87}Rb which can be realized in a rubidium fountain, the systems ^{87}Sr and ^{171}Yb investigated in optical lattices, and five transitions of the ions $^{27}Al^+$, $^{88}Sr^+$, $^{171}Yb^+$ (with two transitions) and $^{199}Hg+$. In the coming years, an increasing number of optical frequency ratios will be measured to be used for consistency tests in the uncertainty range of 10^{-18}. Due to the variety of promising candidates and the fast pace of developments in these new technolo-

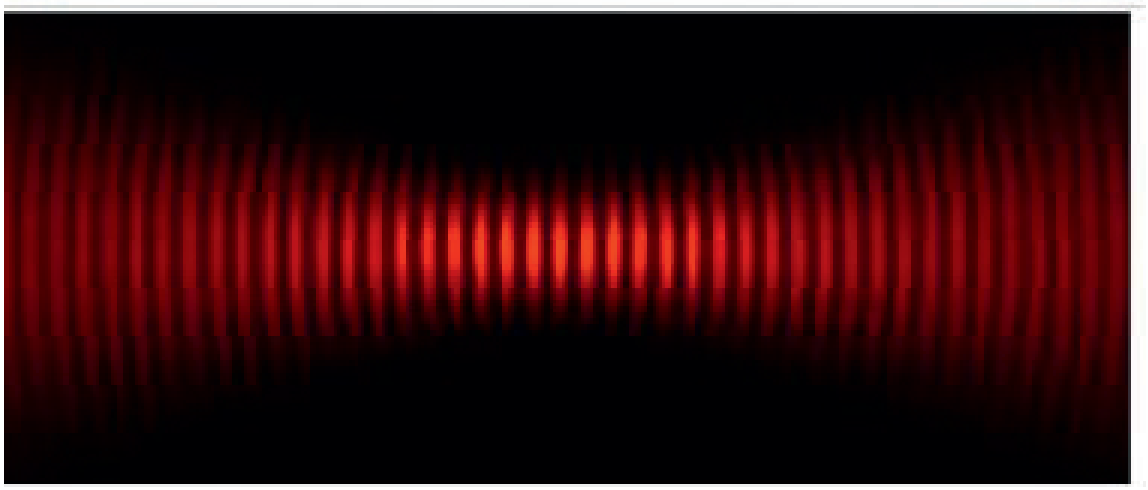


Figure 7:
The most simple atomic trap for an optical lattice clock is formed by a laser beam reflected back into itself. Here the intensity distribution of such a standing wave is shown schematically: in the direction of the beam, the intensity maxima which trap the atoms are 400 nm in width, the lateral extension is 150 µm.

gies and methods, it is not possible yet to determine a clear favorite as the successor of cesium. We can assume that different optical clocks will be used, depending on the application – as is the case today with cesium clocks, small commercially available rubidium clocks and hydrogen masers. Since the vast majority of practical applications of atomic clocks and atomic frequency standards are in the microwave range today and the currently valid requirements in this range are easily met, it is very improbable that cesium clocks might lose their importance in the near future, in the context of the new SI system.

The fact that the ranking list of the "best" clocks changes every year shows how dynamic this field of research really is. The most interesting applications of the new clocks are currently to be found in science, especially in the search for "new physics" at the limits of fundamental theories such as the theory of relativity and quantum physics. Due to the different structures (such as mass and nuclear charge) of the atomic reference systems used, new effects (such as changes in the coupling constant of the fundamental interactions or an additional coupling to dark matter), which have, to date, only be assumed, would express themselves in the form of frequency shifts between the different atomic clocks. The high measurement accuracy of frequencies could thus open up a window on new insights into the fundamentals of physics. This type of fundamental research does not require the second to be redefined. On the other hand, if one of the candidates for the planned redefinition were to be nominated precociously, this would somewhat limit the dynamics and diversity of this field.

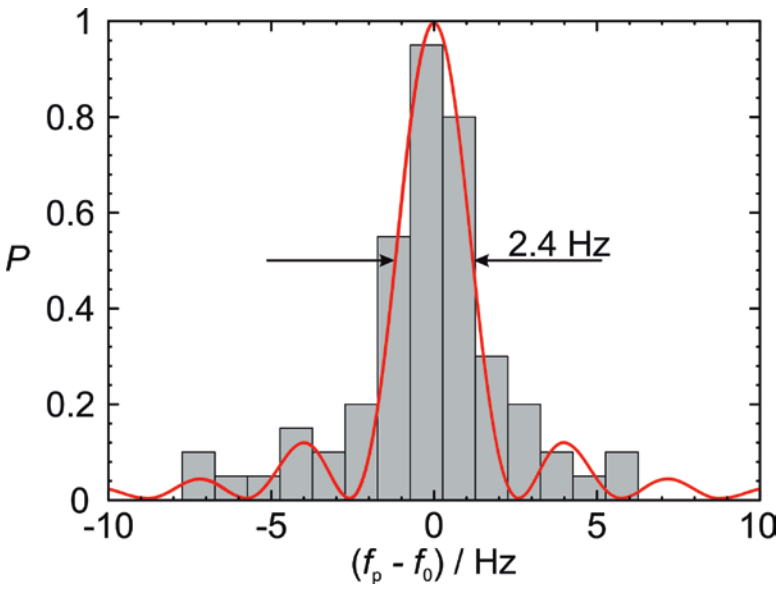


Figure 8:
Transition probability *P* in PTB's optical Yb clock as a function of the laser frequency. With laser pulses of a duration of 0.335 s, a linewidth of 2.4 Hz is attained at an optical frequency of 642 THz. At each frequency, the number of excitations having occurred was counted in 20 tests (gray bars). The red curve shows the spectrum which is theoretically expected in the absence of any oscillator noise.

Literature

[1] *A. Scheibe;* Genaue Zeitmessung, Ergeb. Ex. Naturw. 15 (1936) pp. 262–309 (with numerous original quotations)

[2] *R. E. Beehler;* A historical review of atomic frequency standards, NBS Monograph 140 Chapter 4, Boulder (1974)

[3] *P. Forman;* Atomichron: the atomic clock from concept to commercial product, Proc. IEEE 73 (1985) pp. 1181–1204

[4] *H. Schmidt-Böcking, K. Reich;* Otto Stern, Physiker, Querdenker, Nobelpreisträger, Biographienreihe der Goethe-Universität Frankfurt [series of biographies of Goethe University Frankfurt], Societäts-Verlag, 2011

[5] Ref. 19 in [3]

[6] *N. F. Ramsey;* A molecular beam resonance method with separated oscillatory fields, Phys. Rev. 73 (1950) 695

[7] *N. F. Ramsey;* Molecular Beams, London, New York, Oxford University Press (1956)

[8] National Physical Laboratory, Tedddington , UK, and Ray Essen, "The Memoirs of Louis Essen, father of atomic time" (2015), Chapter V

[9] *L. Essen und J. Parry;* Atomic standard of time and frequency, Nature 176 (1955) 280

[10] *W. Markowitz, R. G. Hall, L. Essen, J. V. L. Parry;* Frequency of cesium in terms of ephemeris time, Phys. Rev. Lett. 1 (1958) 105

[11] *A. Bauch, P. Hetzel, D. Piester;* Zeit- und Frequenzverbreitung mit DCF77: 1959–2009 und darüber hinaus [English edition: 50 Years of Time and Frequency Dissemination with DCF77], PTB-Mitt. 119 (2009) pp. 217–240

[12] *J. H. Holloway, R. F. Lacey;* Factors which limit the accuracy of cesium atomic beam frequency standards, Proc. Intern. Conf. Chronometry (CIC 64) (1964) pp. 317–331

[13] *G. Becker;* Stand der Atomuhrentechnik, Jahrbuch der Deutschen Gesellschaft für Chronometrie 18 (1967) pp. 35–40

[14] *G. Becker, B. Fischer, G. Kramer, E. K. Müller;* Neuentwicklung einer Casiumstrahlapparatur als primäres Zeit- und Frequenznormal an der PTB, PTB-Mitteilungen 69 (1969) pp. 77–80

[15] *A. Bauch;* The PTB primary clocks CS1 and CS2, Metrologia 42 (2005) p. 43

[16] *A. De Marchi;* The Optically Pumped Caesium Fountain: 10^{-15} Frequency Accuracy?, Metrologia 18 (1982) 103

[17] *A. Clairon et al.;* Ramsey resonance in a Zacharias fountain, Europhys. Lett. 16 (1991) 165

[18] *R. Wynands and S. Weyers;* Atomic fountain clocks, Metrologia 42 (2005) S64

[19] *A. Clairon et al.;* A Cesium Fountain Frequency Standard: Preliminary Results, IEEE Trans. Instrum. Meas. 44 (1995) 128

[20] *S. Weyers, U. Hübner, R. Schröder, Chr. Tamm, A. Bauch;* Uncertainty evaluation of the atomic caesium fountain CSF1 of PTB, Metrologia 38 (2001) 343

[21] *V. Gerginov et al.;* Uncertainty evaluation of the caesium fountain clock PTB-CSF2, Metrologia 47 (2010) 65

[22] *S. Weyers, V. Gerginov, N. Nemitz, R. Li and K. Gibble;* Distributed cavity phase frequency shifts of the caesium fountain PTB-CSF2, Metrologia 49 (2012) 82

[23] For current information on the primary clocks and TAI, see: http://www.bipm.org/jsp/en/TimeFtp.jsp (last accessed on 28 June 2016)

[24] see, e.g., IEEE Trans. UFFC 45, July 1998, special issue on the "Dick Effect"

[25] *Chr. Tamm et al.;* Cs-based optical frequency measurement using cross-linked optical and microwave oscillators, Phys. Rev. A 89 (2014) 023820

[26] *J. Vanier und C. Audoin;* The Quantum Physics of Atomic Frequency Standards, Adam Hilger, Bristol and Philadelphia, 1989; J. Vanier und C. Tomescu; The Quantum Physics of Atomic Frequency Standards – Recent Developments, CRC Press, Boca Raton, London, and New York, 2016

[27] *R. Li and K. Gibble;* Evaluating and minimizing distributed cavity phase errors in atomic clocks, Metrologia 47 (2010) 534

[28] *S. Ghezali, Ph. Laurent, S. N. Lea and A. Clairon;* An experimental study of the spin-exchange frequency shift in a laser-cooled cesium fountain frequency standard, Europhys. Lett. 36 (1996) 25

[29] *P. J. Leo, P. S. Julienne, F. H. Mies, and C. J. Williams;* Collisional Frequency Shifts in 133Cs Fountain Clocks, Phys. Rev. Lett. 86 (2001) 3743

[30] *A. D. Ludlow, M. M. Boyd, Jun Ye, E. Peik, P. O. Schmidt;* Optical Atomic Clocks, Rev. Mod. Phys. 87 (2015) 637

[31] *J. L. Hall;* Nobel Lecture: Defining and measuring optical frequencies, Rev. Mod. Phys. 78 (2006) 1279; T. W. Hänsch; Nobel Lecture: Passion for precision, Rev. Mod. Phys. 78 (2006) 1297

[32] *H. Dehmelt;* Coherent spectroscopy on single atomic system at rest in free space, J. Phys. (Paris) 42 (1981) C8–299

[33] *H. Katori;* Spectroscopy of Strontium atoms in the Lamb-Dicke confinement, in Proc. of the 6th Symposium on Frequency Standards and Metrology, World Scientific, Singapore (2002)

[34] *N. Huntemann, M. V. Okhapkin, B. Lipphardt, Chr. Tamm, and E. Peik;* High-accuracy optical clock based on the octupole transition in 171Yb+, Phys. Rev. Lett. 108 (2012) 090801

Interferometry – How Do I Coax a Length Out of Light?

René Schödel*

1. History and introduction

The search for a universal measure of length goes back many centuries. In the Middle Ages, a large number of different measures of length became established, some of which are still commonly used today. This diversity was an obstacle to trade, which was becoming increasingly supraregional. One of the first ideas for a universal measure of length was the seconds pendulum proposed by Jean-Félix Picard in 1668. This proposal was based on the physical relation between the length of a pendulum and its oscillation period. However, due to regional differences in the gravitational field of the Earth, the oscillation period of such a pendulum exhibited considerable differences. This definition of length was thus not universal enough. After the French Revolution, the concept of a metric system of units started to gain acceptance. In 1799, the original meter was created as a material measure of length. The length of this material measure was to correspond to one ten-millionth of the distance between the equator and the North Pole (Earth's meridian quadrant). This statement proved wrong only later when the Earth turned out not to be exactly spherical and thus not suitable for defining the meter; consequently, this original meter was actually too short by 0.2 mm. In 1889, the International Bureau of Weights and Measures (BIPM) introduced the International Prototype of the Meter for the unit *meter*, which was a rod made of a platinum/iridium alloy with a cross-shaped section. The length of the meter was defined as the distance between two center lines, each belonging to a group of lines, at a temperature of 0 °C.

It was only in 1960 that the definition of the meter stopped relying on a material measure (and its copies) that needed to be maintained. At the 11th Conférence Générale des Poids et Mesures (CGPM), the meter was defined as the length equal to 1 650 763.73 wavelengths in vacuum of the light radiated at the transition between the levels 5d5 and 2p10 of ^{86}Kr. This definition was based primarily on a krypton-86 spectral lamp with a wavelength of approx. 606 nm developed at PTB by Johann Georg Ernst Engelhard.

By fixing a numerical value for the speed of light in vacuum c_0 in 1983, the currently internationally valid definition of the meter was laid down:

"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."

This definition presupposes a measuring method that allows length to be realized along a spatial dimension or along material measures for length. These measures have not diminished in significance: they are still used to disseminate the unit of measurement, the meter, to industry and society. The accuracy of the realization of lengths by means of suitable length measurement methods is in perpetual development as a result of new requirements placed on their precision. The following chapters of this paper contain a description of the fundamentals of these procedures.

2. Direct measurement of a time difference used to realize the SI base unit *meter*

The exact wording of the definition of the meter is as follows: "The meter is defined as the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second". This definition assumes a direct time measurement in connection with a length measurement (1 m corresponds to 3.335 640 952 ns). In fact, the distance between the Earth and the Moon is measured by means of the time needed by a laser pulse from the Earth to the Moon and back (approx. 2.6 seconds). To this end, a total of five retroreflectors were placed on the Moon within the scope of American and Soviet Moon missions. In this way, it was possible to demonstrate that the average distance between the Moon and the Earth increases by approx. 3.8 cm each year.

* Dr. René Schödel, Department "Interferometry on Material Measures", e-mail: rene.schoedel@ptb.de

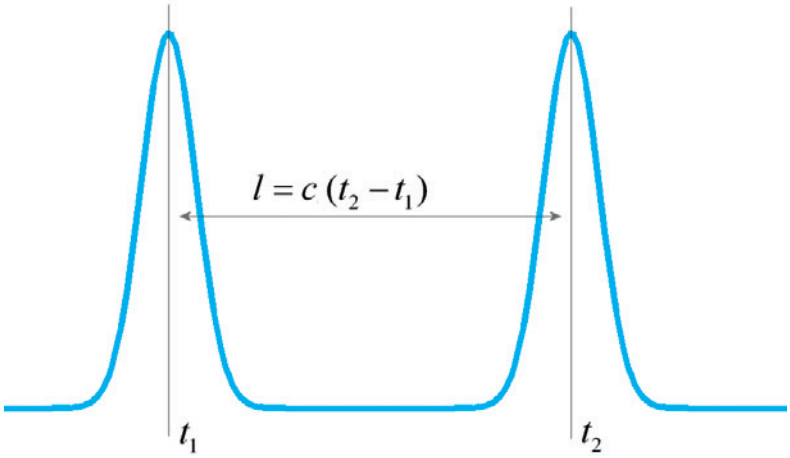


Fig. 1:
Relation between a length and the difference in the travelling time of light

Distance measurements based on the speed of light are used not only in space, but also in applications such as geodesy (tachymeters, global satellite navigation systems) and in commercial distance meters. Depending on the requirements for precision, different modulation and pattern-detection technologies are used. The influence of the air on the speed of light must always be taken into account, since the refractive index of the air n scales down the speed of light ($c = c_0/n$). If this is neglected, the length measured under atmospheric conditions is too long by approx. 0.3 mm per meter of measured length. For precise measurements, it is therefore absolutely indispensable to acquire the parameters of the air (pressure, temperature, humidity) along the measured path in order to be able to correct the influence of the refractive index of the air. Since the refractive index of the air depends on the wavelength (dispersion), it is also important when considering velocity-based measurements to use the signal velocity that can actually be detected (group velocity) into account. This signal velocity is associated with the so-called "group refractive index" and can be computed from the "normal refractive index".

Even in the case of shorter distances, it is relatively easy to demonstrate that an increase in the path travelled by the light leads to a temporal delay – for example, of a light impulse. For precise length measurements in the range of a few meters and less, this method soon reaches its accuracy limits. For example, to measure 1 m with an accuracy of 1 mm, the time difference would have to be measured with approximately 3.3 ps accuracy.

3. Interferometry for the realization of the SI base unit *meter*

Interferometric measurement procedures lend themselves well for length measurements in the range of a few meters and less, but also to realizing a length in general as precisely as possible, since such procedures allow length measurements to be performed with sub-nanometer accuracy. In the following, the basic idea of interferometric length measurement and its relation to the definition of the meter quoted above will be explained.

Interferometry is a measurement method that consists in making waves overlap. The electric field of a single light wave spreading in z direction can be described mathematically as $E(z,t) = A\cos[\varphi] = A\cos[\omega t - kz + \delta]$, with the characteristics A : amplitude, φ : phase, ω : angular frequency, k : wave number, δ : initial phase. The relation between the quantities ω and k and the parameters of *wavelength* λ and *frequency* ν is given via $k = 2\pi/\lambda$ and $\omega = 2\pi\nu$, respectively. The speed of light c corresponds to the spreading velocity of the wavefronts, the so-called "phase velocity". During the period $T = 1/\nu$, the wavefront moves by a single wavelength in the direction of spreading, which is yielded via the consideration φ = const. ($\omega t = kz$). From this, the important correlation $c = \lambda\nu$ is yielded. Since the frequency of light waves lies in the range from 300 THz to approx. 600 THz, their time periodicity cannot be acquired directly with a detector. The only measurable parameter is the mean intensity I of a light wave, which is essentially[1] proportional to the square of a time-dependent mean value of the electric field. Mathematically speaking, the following correlation is yielded between the intensity and the amplitude:

$$I = \left\langle E^2 \right\rangle_t = \lim_{t\to\infty} \frac{\int_0^t \left(E(t,z)\right)^2 dt}{t} = \frac{A^2}{2} \qquad (1)$$

Due to the extremely high frequency, a single light wave does not disclose its wave properties, apart from the property of "color". Only by the overlapping (interference) of several waves can parameters such as the wavelength be accessed. To illustrate this, the intensity of two overlapping light waves will be considered in the following:

$$\left.\begin{aligned} E_1 &= A_1\cos[\varphi_1] \\ E_2 &= A_2\cos[\varphi_2] \end{aligned}\right\} \rightarrow I = \left\langle \left(E_1 + E_2\right)^2 \right\rangle = \frac{A_1^2}{2} + \frac{A_2^2}{2} + A_1A_2\cos[\varphi_1 - \varphi_2] \qquad (2)$$

$$= I_1 + I_2 + 2\sqrt{I_1 I_2}\cos[\varphi_1 - \varphi_2]$$

$$= I_0\left(1 + \gamma\cos[\varphi_1 - \varphi_2]\right),$$

where γ describes the so-called interference contrast

$$\gamma = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

and $I_0 = I_1 + I_2$ was selected. The individual phases φ_1 and φ_2 each contain the parameters of *frequency* and *wavelength* ($\varphi_{1/2} = \omega_{1/2}\, t - k_{1/2}\, z + \delta_{1/2}$).
If two light waves overlap that have the same frequency (and thus the same wavelength) and that are spreading along the same axis (in z direction), the resulting wave will have the same properties with regard to these parameters. This is very easy to demonstrate[2] with two waves of the same amplitude, where

$$E_1 + E_2 = A\cos\left[\omega t - kz + \delta_1\right] + A\cos\left[\omega t - kz + \delta_2\right] = 2A\cos\left[\omega t - kz + (\delta_1 + \delta_2)/2\right] \cdot \cos\left[(\delta_1 - \delta_2)/2\right]$$

is yielded. The bottom wave is thus not a "standing wave"; it spreads with a certain phase shift in the same direction as the single waves. The intensity of this wave is essentially determined by the shift of the two waves in relation to each other,

in general $I = I_0\left(1 + \gamma\cos\left[\delta_1 - \delta_2\right]\right)$ or

$I = I_0\left(1 + \cos\left[\delta_1 - \delta_2\right]\right)$ for

$A_1 = A_2 = A \Rightarrow \gamma = 1$.

Fig. 2 illustrates the spatial overlapping of two identical waves for three different phase shifts $\Delta\varphi = \delta_1 - \delta_2$. To the right of each of the gray-shade representations, the intensities of the single waves and those of the resulting interference are plotted. This example includes the case of mutual full extinction (destructive interference) for phase differences that correspond to the odd-numbered multiples of π and the case of constructive interference for even-numbered multiples of π.

Interestingly, it becomes visible here that, depending on the shift of the waves in relation to each other, the intensity of the overlapping in the case of constructive interference ($\Delta\varphi = 0, 2\,\pi, ...$) is twice as high as the sum of the intensities of the single waves. However, this apparent contradiction can be easily explained by considering how the single waves that are able to interfere with each other are generated. Fig. 3 shows an arrangement called an "interferometer", in which a light wave of intensity I hits a beam splitter. For the sake of

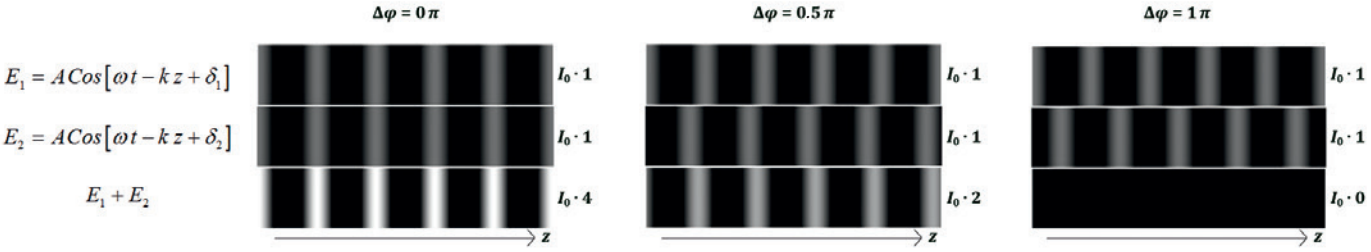


Fig. 2:
Spatial overlapping of two waves at a given point in time.

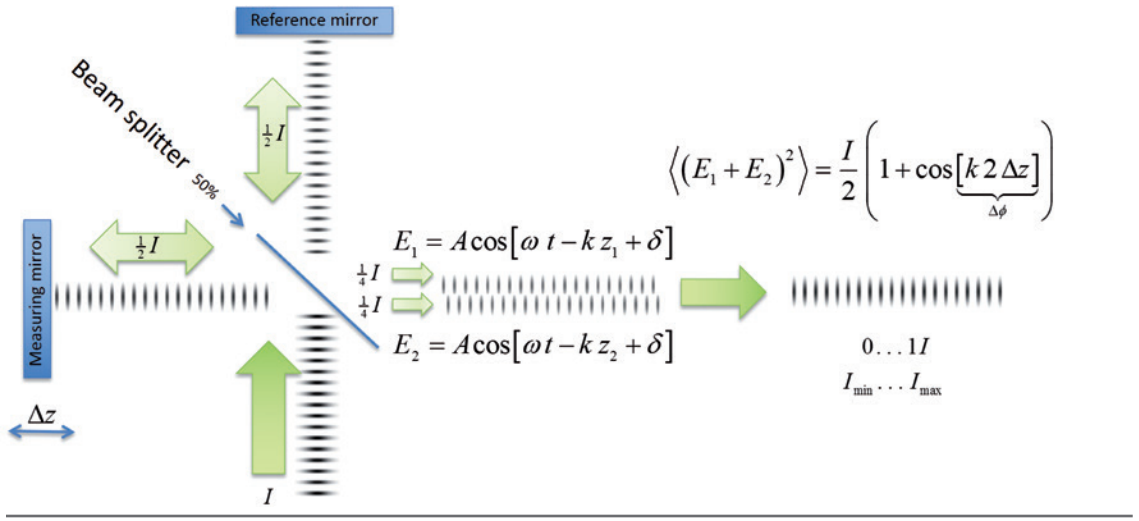


Fig. 3:
Generation (by means of an interferometer) of two light waves that are able to interfere, considered at a given point in time.

[1] Strictly speaking, the intensity of an electromagnetic wave, i.e. its power density, is defined as the time-dependent mean value of the so-called "Poynting vector" $\vec{S} = \frac{c}{4\pi}\vec{E} \times \vec{H}$. The density of the electric field $\vec{E}$ is usually proportional to that of the magnetic field $\vec{H}$. Here, for reasons of convenience, all proportionality factors are set as being = 1.

[2] $\cos x + \cos y = 2\cos\frac{x+y}{2}\cos\frac{x-y}{2}$

convenience, the beam splitter assumed here has a ratio of 50 % transmission to 50 % reflection, and the mirrors hit by the light waves are identical and "adjusted" to be exactly vertical to the direction of the beam. The beam splitter is set in such a way that the light waves reflecting on the two mirrors overlap each other exactly when exiting the interferometer. By passing twice through the beam splitter, the intensity of each of the single beams is reduced to 25 %. Differences in the optical path travelled by the two waves lead to a phase shift $\Delta\varphi = k\,2\,\Delta z$; this, in turn, determines the intensity of the light exiting the interferometer. However, the maximum intensity is never higher than that of the wave entering the interferometer, as shown in Fig. 3[3].

To explain why it is possible to realize the SI definition of the meter quoted above by using an interferometer, we will show an interferometer again – this time, however, without the idealized limitations contained in Fig. 3. Fig. 4, left, illustrates the general case of a so-called "two-beam interferometer", i.e., for any splitting ratios between the intensity of the reference beam and the so-called "measuring beam". Whereas the reference mirror is considered as being invariable (stable), the position of the measuring mirror can be changed. Here, since the phase difference in Equation (2) is caused by the length difference $2\,\Delta z$ ($\Delta\varphi = k\,2\,\Delta z = 2\pi\,\Delta z / \tfrac{1}{2}\,\lambda$), the intensity, which is measured by means of a detector, changes periodically (see Fig. 4, right). The intensity of the maximum and of the minimum values depends on the ratio $\alpha = I_1/I_2$, which determines the interference contrast $\gamma = 2\sqrt{a}\,/\,(a+1)$. What is noteworthy is the fact that the interference itself, even at an extremely unfavorable intensity ratio of 1/1000, still exhibits visible variations that can be metrologically detected. In the easiest case, it suffices to count the number of periods (called "interference orders") while the mirror is being shifted ($\Delta\varphi$ / 2π = 1, 2, ...).

In each case, a shift or a length can be expressed as a multiple of the half wavelength multiplied by the number of interference orders; thus, it can also be expressed as a product of the speed of light and a time difference Δt:

$$\Delta z = \frac{\lambda}{2}\frac{\Delta\varphi}{2\pi} = \underbrace{\tfrac{1}{2}\,c}_{\text{Speed of light}} \cdot \overbrace{\Delta t}^{\text{Difference in the travelling time of light}}, \qquad (3)$$

where Δt describes the time difference for the occurrence of the wavefronts (or of light impulses if conducted into the same interferometer) on the detector:

$$\overbrace{\Delta t}^{\text{Result of measuring by means of interferometry}} = \frac{1}{\underbrace{\nu}_{\text{Frequency of the light}}}\,\overbrace{\frac{\Delta\varphi}{2\pi}} \qquad (4)$$

Equation (3) clearly represents the relation between the travelling time of light and the associated length – exactly as formulated in the definition of the meter according to the SI. Equation (4) states what needs to be known or measured in order to determine the time difference:

1.) The frequency of the light ν.

2.) The phase difference of the interference resulting from a change in an interval inside the interferometer or from different lengths.

Length-measuring interferometers are mostly used in air. Similar to the direct measurement of time delays (see Section 2 of this article), the medium of air leads to a reduction of the speed of light, i.e., in Equation (3), λ must be replaced by λ/n and c by c_0/n. The reduction of the speed of light is compensated for by increasing the phase difference or the time

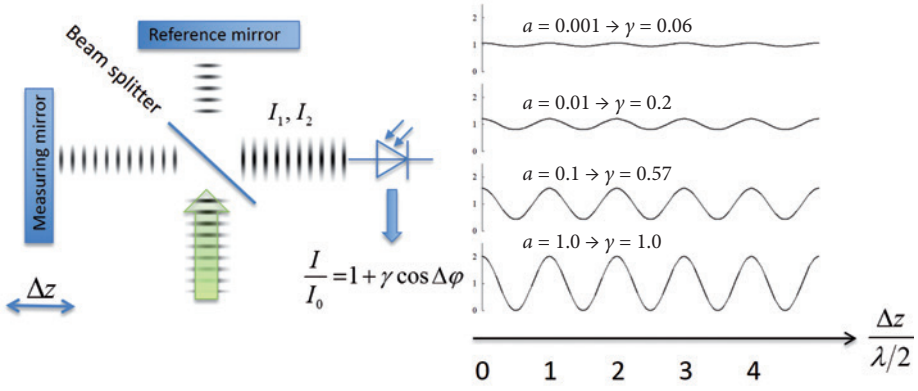


Fig. 4:
Generation (by means of an interferometer) of two light waves that are able to interfere (left). What is represented on the right-hand side is the intensity as a function of the measuring mirror shift for various intensity ratios ($a = I_1/I_2$).

[3] The question as to where the energy remains in the event of full extinction at the interferometer exit reveals that the beams reflected back to the entrance of the interferometer are not plotted in Fig. 3.

difference to be measured, so that a length can be realized irrespective of whether the measurement is performed in vacuum or in air.

For the sake of completeness, we will write out the equation (which is fundamentally important for the realization of length by means of interferometry) in full:

$$\Delta z = \frac{\lambda_0}{2n}\frac{\Delta\varphi}{2\pi} = \frac{\lambda_0}{2n}(i+q), \qquad (5)$$

where λ_0 is the size of the so-called "vacuum wavelength", i is the integer value of the interference orders, and q is the so-called "interference fraction".

3.1. Optical frequency standards

It is absolutely necessary to know the frequency ν of the light in order to realize length by means of interferometry (see Equation (4)). However, the value of the vacuum wavelength derived from this ($\lambda_0 = c_0/\nu$) is purely notional since it does not describe any geometrically relevant length, but the distance between the wavefronts under idealized conditions and in vacuum.

In order to realize the frequency of a suitable light source, this frequency must be compared with the primary time standards (e.g. the frequency of the cesium atomic clocks of PTB). Today, this is done by means of so-called "frequency comb generators". These allow the realization of frequencies of optical frequency standards with relative uncertainties of 10^{-15} and less [1]. To connect the frequency of a secondary frequency standard (such as a laser for use in interferometric length measurements) to that of a suitable comb line of the synchronized frequency comb, the different light sources are overlapped on a detector, i.e., made to interfere. Fig. 5 illustrates the result of the interference of two waves – whose frequencies differ slightly – in a fixed place, as a function of time (as in Equation (2)). The intensity is shown in the lower part of the figure. The green curve shows the mean intensity within a period. This corresponds to the so-called "beat signal" that can be detected due to its frequency, which is considerably lower than the frequencies of the individual light sources ($\nu_{beat} = |\nu_1 - \nu_2|$). To be able to represent this relation clearly, unrealistic assumptions were made in Fig. 5 with regard to the single waves ($\lambda_1 = 630$ nm, $\lambda_2 = 670$ nm, $\nu_{½} = c/\lambda_{½}$). The frequencies in real frequency comparisons lie much closer to each other, which results in observable beat frequencies in the Hz, kHz or MHz ranges. The frequency of a light source to be calibrated (ν_{EUT}) is obtained from the frequency of a standard ($\nu_{standard}$) as $\nu_{EUT} = \nu_{standard} \pm \nu_{beat}$.

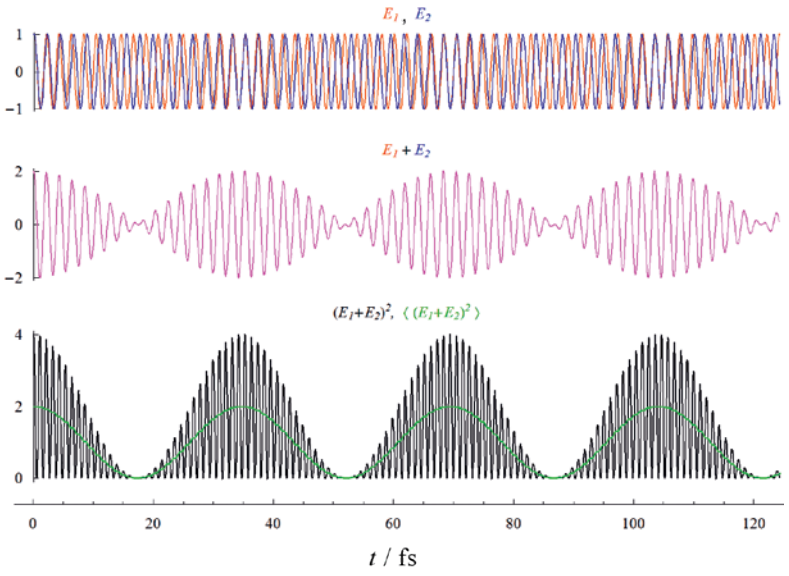


Fig. 5: Overlapping of two light waves whose frequencies differ, considered at a given point in space. To illustrate both the fundamental mode (the high-frequency fraction) and the beating waves (the low-frequency fraction) of a light wave, the following parameters were used:
$\lambda_1 = 630$ nm,
$\lambda_2 = 670$ nm,
$\nu_{½} = c_0/\lambda_{½}$.

The ambiguity of this relation is due to the fact that the beat frequency only indicates the amount of the frequency difference ($\nu_{beat} = |\nu_{standard} - \nu_{EUT}|$). To clarify this relation, it is therefore necessary to have several standard frequencies at one's disposal.

Practicable light sources for use in interferential length measurement are the sources recommended by the *Comité International des Poids et Mesures* (CIPM) whose frequency is established by laser stabilization on molecular hyperfine structure transitions [2]. Here, relative frequency uncertainties of typically < 10^{-11} are achieved.

Today, optical frequency standards are being further developed mainly so that the future realization of time can be carried out with so-called "optical clocks" that are more precise than the cesium atomic clocks currently used. At present, it is already possible to attain relative frequency uncertainties of < 10^{-16} [4], depending on the averaging period considered. In contrast to the situation that predominated until approx. 30 years ago, the realization of length by means of interferometry can no longer benefit from this, since the uncertainty contribution from the frequency of such sophisticated light sources makes a negligibly small contribution to the overall uncertainty of length measurement.

3.2. Distance-scanning interferometers

The basic principle of distance-scanning interferometry was already mentioned in the previous section (see Fig. 4 and the corresponding explanations): namely, the observation of the periodically changing detector signal of the interference intensity during a constant change in the distance of a measuring mirror, with the reference mirror being considered stable. Each period corresponds to one interference order (i.e., to one change in

distance) by half the wavelength of the light used (Equation (5)). This means that, when using red laser light of wavelength 633 nm, approx. 3 million interference orders would have to be counted for one meter. During this time, however, shocks and air turbulences must be ruled out. In principle, with the laser light sources available today (coherence length in the kilometer range), it is possible to measure very large distances. In order for them to be insensitive to tilting, the mirrors are designed as retroreflectors.

However, since the interference is primarily determined by the cosine function (see Equation (2)), a periodic signal measured by the detector cannot indicate the direction in which the measuring mirror has been moved. In extreme cases, moving the mirror to and fro can even simulate uniform motion along an axis. In the following, various possibilities of how to establish a clear connection between the detector intensity and the position of the measuring mirror are described.

3.2.1. The quadrature procedure

In the so-called "quadrature procedure" (see Fig. 5), a laser beam whose polarization axis is tilted by 45° hits a polarizing beam splitter. The beam splitter splits the incident light into the two polarization directions that are vertical to each other, i.e., into the polarized component that is parallel to the incident plane (p-beam), which passes entirely through the beam splitter, and into the vertically polarized component (s-beam), which is fully reflected. These two beams are reflected by retroreflectors and then joined together again at the polarizing beam splitter. At the exit of the interferometer, two beams that are polarized vertically to each other are then available[4]. Polarizers are located behind another beam splitter, in front of a detector; their axis is tilted by 45°, so that they project both polarization components (s and p) onto the polarization axis of the corresponding polarizer. Only in this way is it possible for the beams that are first polarized vertically to each other to lead to an observable interference signal. The beam reflected at the beam splitter traverses a λ/4 delay plate in front of the polarizer and therefore generates a signal at the upper detector that is shifted by π/2 compared to the signal of the other detector. This procedure thus provides two interference signals that are shifted by π/2 towards each other and are called sine and cosine signals, respectively. When processed electronically (i.e., freed from the constant fraction and suitably scaled), these signals $\tilde{I}_{\sin}$ and $\tilde{I}_{\cos}$ can be represented as shown in Fig. 6 on the right, e.g. by means of an oscilloscope in the xy mode. A shift of the measuring mirror then leads to a circular or elliptical figure being displayed. Depending on the direction of displacement, the corresponding vector rotates in one direction or the other. Here, a full revolution corresponds to one interference order ($\Delta\varphi = 2\pi$) and thus to a shifting of the measuring mirror by $\Delta z = \lambda/2$, which can be clearly detected with this procedure. The quadrature procedure is ideal for calibrating distance meters and line scales (e.g. by means of PTB's 50 m baseline) [5]. This baseline is equipped with a cart moving on rails that are rooted in a concrete base which is firmly connected to the floor of the basement.

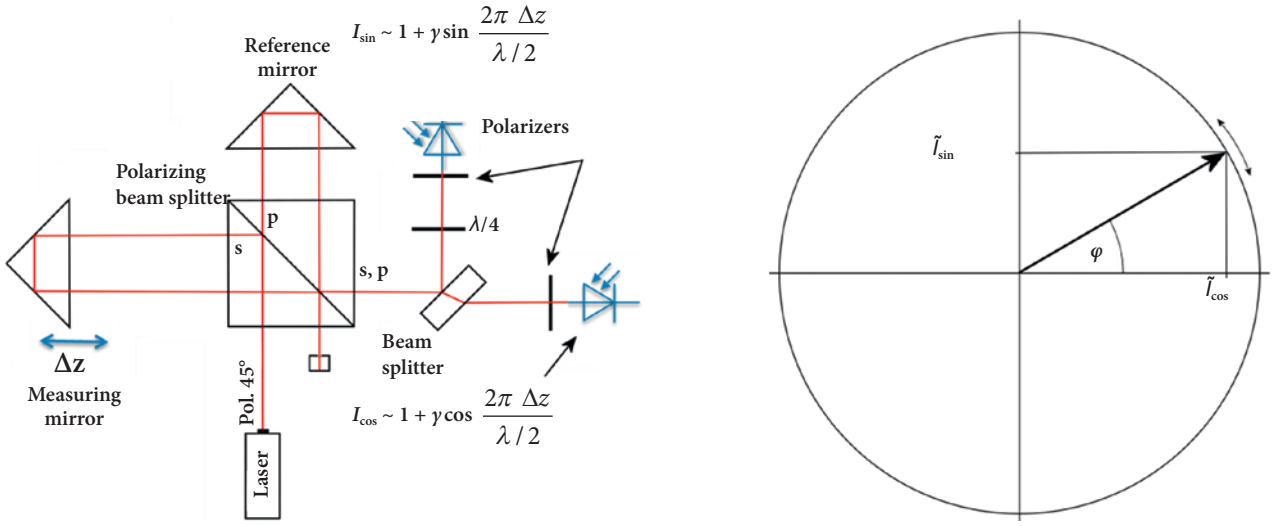


Fig. 6:
Extension of an interferometer by polarization components to generate two interference signals whose phases are shifted by π/2. On the right, the vector yielded from the signals $\tilde{I}_{\sin}$ and $\tilde{I}_{\cos}$ is plotted.

[4] Making light waves overlap that are polarized vertically to each other does not lead to an observable interference.

3.2.2. The heterodyne procedure

Certain lasers emit two light waves that are polarized vertically to each other and have different frequencies (i.e., two different laser modes). Furthermore, there are possibilities for shifting the frequency of a light wave by a defined value, e.g. by means of acousto-optic modulators (AOMs). The availability of light waves whose frequencies differ from each other by a known value is exploited by so-called "heterodyne interferometers"[5]. Fig. 7 shows the typical design of a heterodyne interferometer. Here, the light source used is a laser with two different frequencies. The light wave whose electric field is polarized vertically to the image plane is called "s"; the wave which is polarized parallel to the plane is designated as "p". First, both light waves are split by means of a beam splitter. The two reflected waves then hit a polarizer whose axis is tilted by 45° with regard to the two polarization directions. At the detector located behind the polarizer, this leads to a beat interference (see Section 3.1 of this article) of the now equally polarized waves of different frequencies (ν_1 and ν_2), the so-called "reference signal". The two light waves transmitted at the beam splitter later hit a polarizing beam splitter where the s-polarized wave is reflected and the p-polarized wave is transmitted. After being reflected by the retroreflectors, the two waves are joined together again in the polarizing beam splitter and pass a polarizer whose axis is again tilted by 45° with regard to the two polarization directions. As a result, a second beat interference, the so-called "measuring signal", occurs at the detector located behind the polarizer.

The phase of the reference signal, Φ_{Ref}, is yielded from the phase difference of the individual phases belonging to the different frequencies, $\varphi_{1/2}^{\mathrm{ref}} = k_{1/2} z^{\mathrm{ref\text{–}detector}} - \omega_{1/2} t + \delta_{1/2}$, where $z^{\mathrm{ref\text{–}detector}}$ describes the joint light path of the individual waves up to the reference detector. In contrast, the phase of the measuring signal, Φ_{meas}, is yielded from the overlapping of single waves which have covered different distances. If Δz is defined as the length difference between the measured length and the distance from the reference mirror, $z^{\mathrm{ref\text{–}mirror}}$, then the measuring signal is obtained from the difference between

$\varphi_2^{\mathrm{meas}} = k_2 \left(z^{\mathrm{ref\text{-}mirror}} + \Delta z \right) - \omega_2 t + \delta_2$ and

$\varphi_1^{\mathrm{meas}} = k_1 \, z^{\mathrm{ref\text{-}mirr}} - \omega_1 t + \delta_1$. If $z^{\mathrm{ref\text{-}Detector}}$ and $z^{\mathrm{ref\text{-}mirror}}$ are considered constant, and if they are considered equal WLOG, then the phases of both detectors and their difference can be expressed as follows:

$$\begin{aligned} \Phi_{\mathrm{ref}} &= \left(k_2 - k_1\right) z^{\mathrm{ref}} - \left(\omega_2 - \omega_1\right) t + \delta_2 - \delta_1 \\ \Phi_{\mathrm{meas}} &= \left(k_2 - k_1\right) z^{\mathrm{ref}} + k_2 \Delta z - \left(\omega_2 - \omega_1\right) t + \delta_2 - \delta_1 \\ \Phi_{\mathrm{meas}} - \Phi_{\mathrm{ref}} &= k_2 \Delta z \end{aligned} \tag{6}$$

This means that the shift Δz of the measuring mirror can be determined from the comparison of the phase position between the beat signals at the reference detector and at the measuring detector. Both signals can be freed from the DC component by capacitive coupling, which makes the procedure insensitive to variations of the light intensity, including influences due to the ambient light. Interpolating an interference signal – as is usually done in homodyne interferometers – becomes superfluous, and reversing the direction is no longer an issue, since the phase position of the beat signals always increases. In this way, the phase difference between the measuring signal and the reference signal can be determined unambiguously. One of the most accurate heterodyne interferometers to date was set up at PTB a few years ago [6].

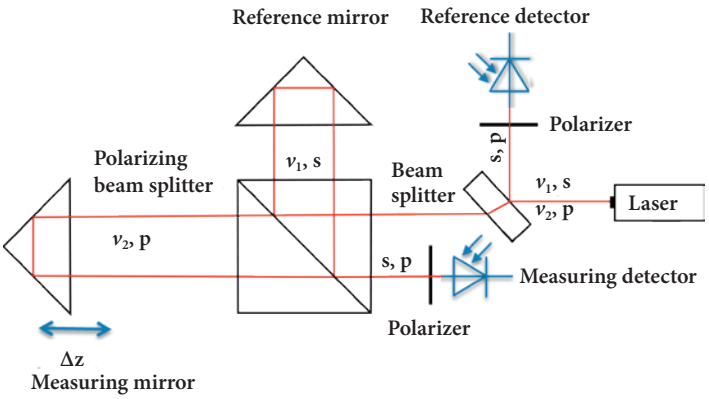


Fig. 7:
Schematic representation of a heterodyne interferometer operated with laser light from two waves (s and p) that are polarized vertically to each other and have different frequencies (ν_1 and ν_2).

[5] As a counterpart to this, interferometers are used that use a light source with a single frequency only (also called "homodyne interferometers").

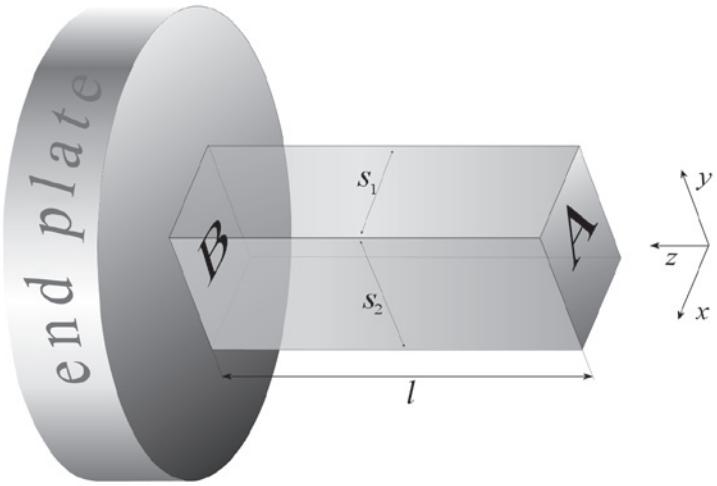


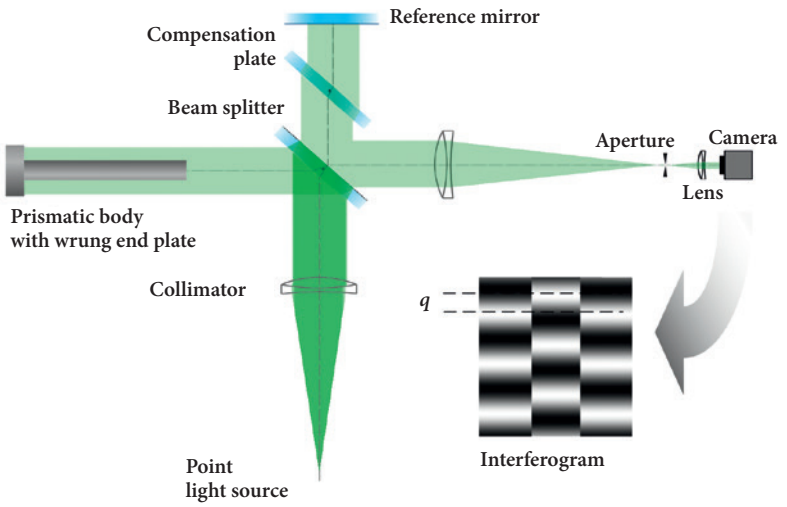
Fig. 8:
Prismatic body with terminal points A and B that are parallel to each other, wrung with an end plate

3.3 Imaging interferometers used to measure the length of prismatic bodies

To date, length has been considered to be a distance along an axis. By shifting an interferometer mirror along the spreading direction of light, this length can be determined by counting interference orders – as described above – directly by means of interferometry, and therefore traceably to the SI definition. The following section describes a different measurement task: determining the length of bodies, in particular those that are suitable as material measures for the dissemination of length: 1) bodies with end faces that are parallel to each other (prismatic bodies), mostly in the form of so-called "gauge blocks"; and 2) spherical bodies, i.e., spheres. The length and the diameter of these material measures can be determined interferometrically; in most cases, the length is represented as a step height. Fig. 8 shows a prismatic body whose reflecting end faces A and B are parallel to each other, and whose face B coincides with the plane of an end plate. Here, the body and the platen cling together due to adhesion or cohesion. In fact, this state can be achieved by so-called "wringing" (which is also called "optical contacting"). Thus, in principle, the step height is identical to the length of the body.

In the assembly shown in Fig. 8, it would be possible to use an interferometer instead of a measuring mirror (as shown in Fig. 4). Depending on whether the light beam were reflected by the front surface or by the end plate of the body, the interference signal obtained at the detector would be different. An attempt could be made to compare the interference of several spatially separate beams, which are reflected both by the front surface and by the end plate of the body, by means of different detectors in order to eventually obtain a length. However, utilizing a large bundle of beams, as described below, clearly seems more appropriate. In both cases, the integer interference orders between the front surface and the platen, which are usually very numerous, cannot be accessed directly; that is, it is first only possible to determine the difference in the non-integer residual of the interference orders, the interference fraction q.

Fig. 9:
Twyman-Green interferometer for the realization of the length of prismatic bodies. The interferogram shown (bottom right) symbolizes the interference fraction that can be read out from the fringe mismatch when the reference mirror is slightly tilted.



3.3.1. Imaging Twyman-Green interferometers

Figure 9 shows an imaging Twyman-Green interferometer for the realization of the length of prismatic bodies. A point light source set in the focus of a collimator generates a large bundle of parallel beams that cover the prismatic body together with its end platen. The light, after being reflected and returning to the plane of the beam splitter, follows a "quasi-joint" path to a camera that records a so-called "interferogram". In this way, the interference intensities for each partial beam of the beam bundle are resolved laterally. Since the reference beam is reflected by the optically denser medium (outer surface of the beam splitter) once more than the measuring beam, the sign of the interference term is reversed – compared to Equation (2) – as follows:

$$I = I_0\left\{1-\gamma\cos\left[\frac{2\pi}{\lambda/2}\left(z_1(x,y)-z_2(x,y)\right)\right]\right\}, \quad (7)$$

where $z_{½}$ (x, y) represents the distribution of the geometric paths (measuring: 1, reference: 2) vertically to the optical axis. If the reflecting surfaces are perfectly flat, then the path difference $z_1(x, y) - z_2(x, y)$ describes a plane whose tilt depends on the orientation of the surfaces to each other. The cosine function is then responsible for the typical fringe pattern with maxima for path differences of $m\lambda/2$ and minima for $(m+ ½)\lambda/2$.

3.3.1.1. Phase-stepping interferometry

In phase-stepping interferometry, the length of the interferometer reference arm is shifted stepwise by the equidistant value α. From the intensities resulting for each of these positions, the phase φ of the interference can be determined. Many different algorithms can be used to this end. In the case of the Tang algorithm [7], for instance, five positions are considered. The figure on the right is intended to illustrate the relation between the intensities $I_{1\,\ldots}\,I_5$ and the interference phase φ:

The theoretical relation

$$I_k = I_0\left\{1-\gamma\cos\left[\varphi+\left(k-3\right)\alpha\right]\right\}$$

corresponds – according to the Tang algorithm – to the following interference phase:

$$\tan\varphi = \frac{\sqrt{\left[\left(I_2-I_3\right)+\left(I_1-I_4\right)\right]\left[3\left(I_2-I_3\right)-\left(I_1-I_4\right)\right]}}{I_2+I_3-I_1-I_4}$$

This is exploited experimentally in order to determine the interference phase from a series of measured intensities. In imaging interferometry, a camera records interferograms for each position. The interference phase can be calculated for each pixel coordinate by means of the Tang algorithm.

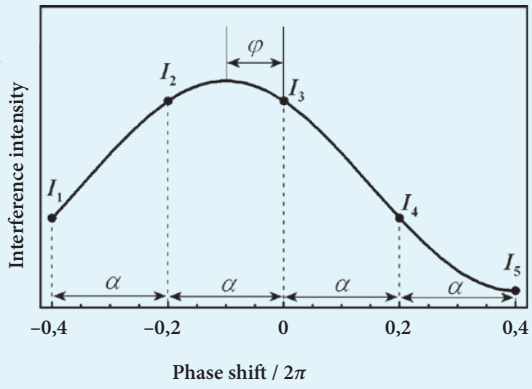


Fig. 10

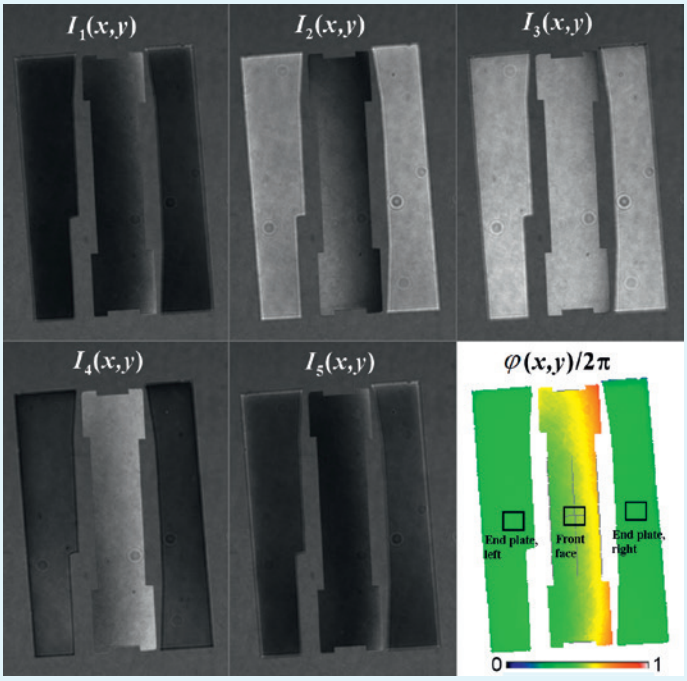


Fig. 11

The interference fraction is determined based on the resulting phase topography $\varphi(x, y)$. The interference fraction corresponds to the phase difference between the front surface and the end plate and is calculated by averaging within well-defined ranges (see the rectangles in the phase topography):

$$q = \frac{1}{2\pi}\left[\frac{1}{2}\left(\varphi^{\text{end plate, left}} + \varphi^{\text{end plate, right}}\right) - \varphi^{\text{front surface, center}}\right]$$

For length measurements with sub-nm accuracy, the position of these ranges must be assigned to the center of the front surface with sub-pixel accuracy, since even the smallest parallelism deviations can otherwise lead to errors.

3.3.1.2. How sharp can (or must) the body be imaged?

For a sharp image of an object (e.g. the edges of a prismatic body), the deflected light must be able to reach the image plane on the sensor (CCD). The setup shown in Fig. 9 includes an imaging system which is dimensioned in such a way that the front face of the sample is imaged "sharply". The sharpness of this image depends on the size of the aperture in the focus of the output collimator. This aperture is designed to suppress disturbing secondary reflections. The smaller the aperture, the less light deflected at the edges can reach the camera and contribute to sharpness. This generally known fundamental principle of optical images is illustrated in Fig. 12. For the sake of convenience, an optical system with only one lens (with an effective focal length f) was assumed. It appears that diffraction orders outside the angular sections represented in green at the aperture are blocked out.

The optical imaging of the front face of the body by means of a lens causes the light diffracted at the edges to contribute to the sharpness of the image. The effective angular section of the diffracted light also influences the measured topography of the interference phase. The larger the aperture, the less the phase topography is distorted around the edges. When very small apertures are chosen, but also when no real optical imaging is used (i.e. letting the beam expand to just match the size of a CCD), "diffraction rings" become clearly visible along the edges of the body which manifest themselves both as a blur and as phase errors.

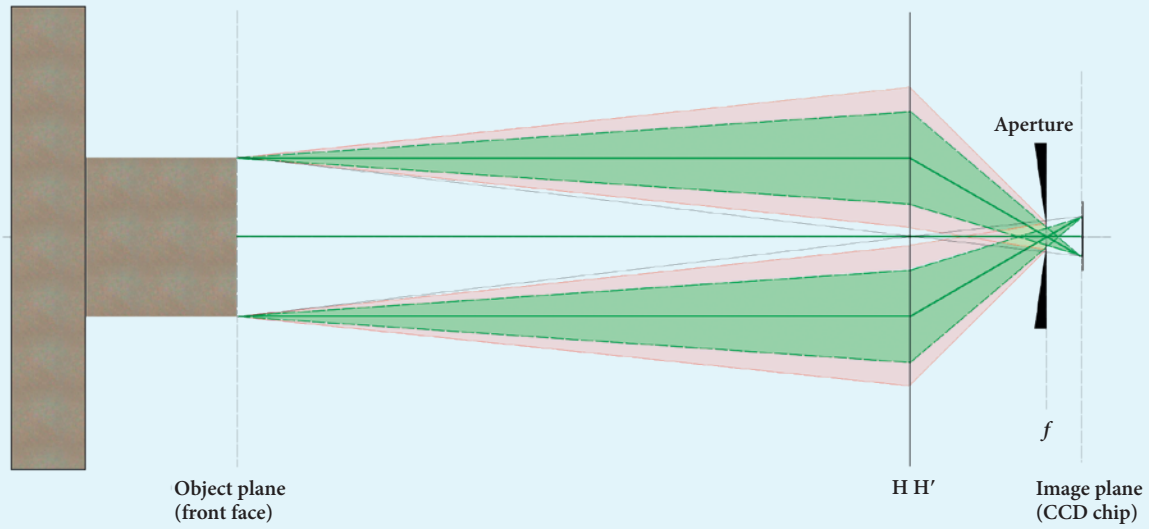


Fig. 12

The prismatic body in the measuring arm of the interferometer represented in Fig. 9 causes the appearance of two mismatched fringe systems. The value of this mismatch corresponds to the interference fraction q and represents the direct result of the interferometric measurement for the calculation of the length, $l = (i+q)\lambda/2$ (Equation (5)).

The integer interference order i can, however, be determined indirectly using additional information. If only one wavelength is available as light source, then there must be an exactly known value of the nominal length l_{nom}. By inserting $q = 0$ and inversing Equation (5), it is possible to determine an estimate of the integer orders from the rounded value of $l_{nom}/\frac{1}{2}\lambda$. Here, however, the risk of "miscounting" is relatively high, since even the smallest deviations from the target temperature (usually 20 °C), to which l_{nom} refers, lead to changes in length, which cannot be corrected with sufficient accuracy if the thermal expansion coefficient of the sample is not known. It is therefore actually indispensable to use at least two separate wavelengths when determining the integer interference orders. The advantage is that the lengths obtained are independent of each other and can be compared, allowing the whole length measurement to be checked and possible errors in the determination of q – such as an incorrectly set wavelength λ of the light source – to be detected.

If N different wavelengths $\{\lambda_1, ..., \lambda_N\}$ are available, the requirement can be defined that the same length $l = l_k$ be obtained for each wavelength. Under this condition, the integer interference orders $\{i_1, ..., i_N\}$ can be determined as follows: i_k is varied around the integer values δ_k, and the resulting lengths $l_k = (i_k + \delta_k + q_k)\,\lambda_k/2$ are checked in terms of whether they coincide.

Here, it is useful to consider the mean deviation of the lengths l_k from their mean value $\bar{l} = \frac{1}{N}\sum_{k=1}^{N} l_k$:

$$\Delta = \frac{1}{N}\sum_{k=1}^{N}\left|\bar{l} - l_k\right| \qquad (8)$$

If a mean deviation Δ is assigned to each mean length $\bar{l}$, a dataset of the form $\{l, \Delta\}$ is obtained for a number of variations $\{\delta_1, ..., \delta_N\}$; this dataset, when represented as a scatterplot, yields a so-called "coincidence pattern". Fig. 13 shows a typical example for $N = 2$ and data points with $\Delta < 20$ nm.

In Fig. 13, the two notional wavelengths $\lambda_1 = 532.3$ nm and $\lambda_2 = 548.6$ nm, a nominal length $l_{nom} = 10$ mm and a length l that deviates from l_{nom} by the value $\delta l = 2.5$ µm were assumed. The corresponding interference fractions q_1 and q_2 were exactly calculated. These input quantities are reflected in the dataset $\{l, \Delta\}$ as follows:

- For $\bar{l} - l_{nom} = \delta l$, a minimum is yielded with $\Delta = 0$ (marked with red dashes).
- Further minima exist at the interval of half the synthetic wavelength λ_{synt}, i.e., with $\bar{l} - l_{nom} = \delta l + m \cdot \lambda_{synt}/2$. Here, m describes integer numbers and $\lambda_{synt} = \lambda_1\,\lambda_2/|\,\lambda_2 - \lambda_1|$.
- The next minima with $\Delta = 0$ are distant from the minimum at $\bar{l} - l_{nom} = \delta l$ by certain multiples of half the synthetic wavelength. In the example shown in Fig. 13, these multiples are located at $m = \pm 3$. However, it is not possible to make a general statement, since the number m strongly depends on the selected wavelengths λ_1 and λ_2 used.
- The lengths that are closest to the minimum at $\bar{l} - l_{nom} = \delta l$ are located one interference order away from this length, i.e., at $\bar{l} - l_{nom} = \delta l = \pm\,(\lambda_1 + \lambda_2)/2$. The corresponding value for Δ is $|\lambda_1 - \lambda_2|\,/4$!

The last item mentioned makes it clear that this method – based on finding coincidences – requires sufficient spacing of the wavelengths. If, for example, the wavelengths λ_1 and λ_2 are only 2 nm apart from each other, then $\{\delta_1, \delta_2\} = \{\delta_1^{opt} + 1, \delta_2^{opt} + 1\}$ yields a neighboring coincidence $|l_1 - l_2|$ of only 1 nm (i.e., $\Delta = 0.5$ nm). In real interferential length measurements, this coincidence can hardly be differentiated from the "actual" coincidence at the optimal variation numbers $\{\delta_1, \delta_2\} = \{\delta_1^{opt}, \delta_2^{opt}\}$ ($\Delta = 0$ nm).

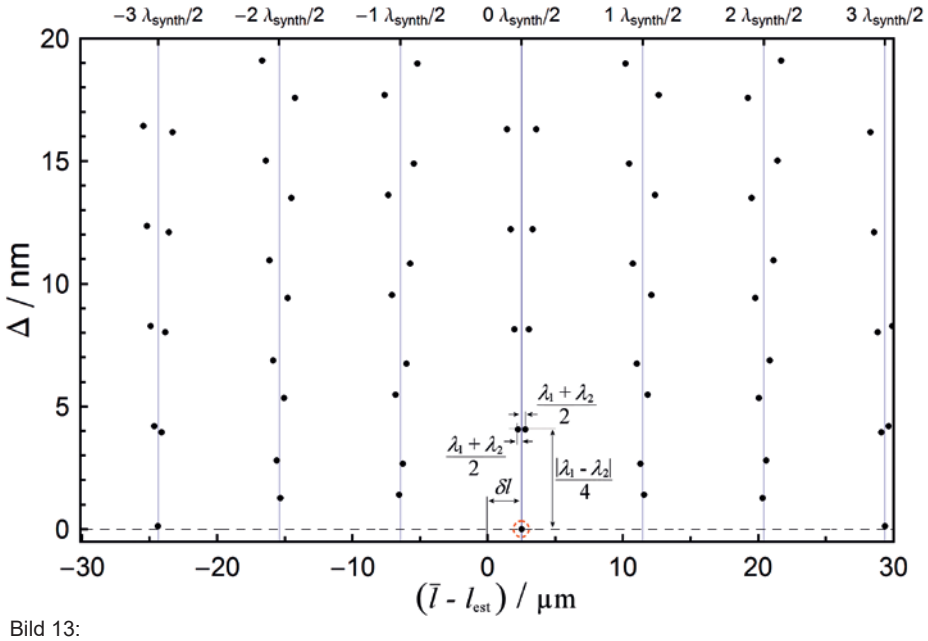


Bild 13:
Typical coincidence pattern when using two wavelengths.

The unambiguous range that can actually be achieved and that is covered by the above-mentioned variation method thus depends primarily on the achievable uncertainty of Δ – and thus on the uncertainty of the determination of the interference fractions [8]. The achievable unambiguous range can, in principle, be enlarged by using more than two wavelengths, especially for measurements in vacuum – as is the case at PTB's Ultra Precision Interferometer [9].

3.3.1.3. Cosine error and aperture correction

The so-called "cosine error" occurs in interferential length measurements due to non-vertical light incidence on the surfaces of a body. This results in a length that is not measured for a long enough time by the factor cos α. For small angles α, the relation

$$\tilde{l} = l \times \cos\alpha \overset{\alpha<<1}{\cong} l \times \left(1 - \tfrac{1}{2}\alpha^2\right)$$

is yielded between the measured length $\tilde{l}$ and the actual length l of a body. The error thus depends on the length and is approximately proportional to the negative square of the angle α. In order to minimize the cosine error, the interferometer must be adjusted in such a way that the light is as vertical to the optical surfaces as possible.

An autocollimation procedure developed at PTB seems best suited for this purpose [10]. Its basic idea consists in observing the light returning to the entrance of the interferometer, especially the proportion of light fed back into the optical fiber from which the light is guided into the interferometer. This fraction of light is measured as a function of the fiber position. If the fiber is located inside the focal plane of the collimator, the intensity of this retroreflected light thus represents the area where two discs overlap whose diameter corresponds to that of the fiber. The intensity distribution is measured as a function of the lateral fiber position (x, y). Then, the position of the maximum is approached.

The fact that the light source, i.e., the optical fiber, has a finite size leads to the so-called "aperture correction B". This correction is yielded as a cosine

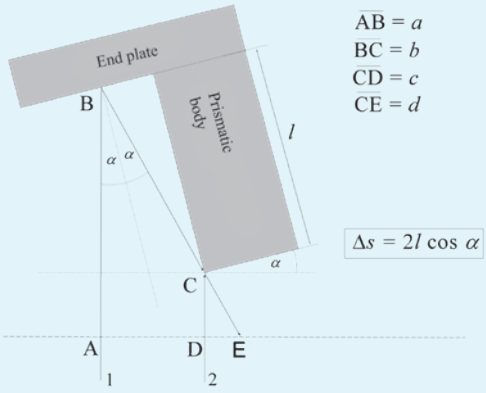


Fig. 14

error averaged over the entire fiber cross section. For a circular, homogeneous light source (e.g. multimode fiber), $B = d^2/(16f^2)$ is yielded by integration, where d is the diameter of the light source and f is the focal length of the collimator. To correct a measured length, the focal length must be multiplied by $1 + B$. For $d = 0.25$ mm and $f = 600$ mm, this means, for example, an aperture correction of approx. 10 nm per meter (10^{-8}).

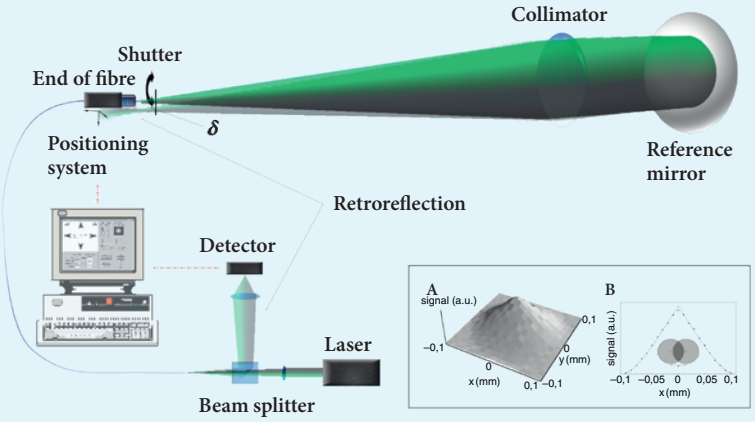


Fig. 15

3.3.2. Multiple-beam interferometers

Multiple-beam interference occurs wherever light is reflected several times and eventually overlaps. A well-known example of this is the colored stripes generated by white light falling onto an oily patch on water. This phenomenon is used in its most simple form in surface testing. For this purpose, an almost perfectly smooth glass plate is laid onto the surface of the test object. Here as well, colored patterns become visible where the surface exhibits a tilt; the number and the direction of these patterns depend on the tilt of the surfaces to each other. The straightness of these patterns is a measure of the flatness of the specimen.

A multiple-beam interferometer is also called a "Fizeau interferometer" when the distance between the two plane surfaces on which the reflection takes place is large compared to the wavelength of the light used. Fig. 16 shows a particular case of a Fizeau interferometer in which the interference of the light is observed in reflection. The light passes through an optical plate with a semi-transparent surface. Part of the light reflected by the mirror is, in turn, reflected by the semi-transparent surface and goes back to the mirror. The higher the reflectivity R of the semi-transparent surface, the more often this sequence repeats itself.

Contrary to two-beam interference, the intensity of the multiple-beam interference observed at the output of the interferometer is not cosine-shaped, but complies with the Airy formula. This characteristic of the interference is plotted in the insert in Fig. 16 for different reflectivities. With increasing reflectivity, the structure becomes increasingly sharp. As in the case of the two-beam interferometer, the periodicity of this structure is given by half the wavelength of the light used. Fizeau interferometers are frequently used for flatness measurement. In principle, an imaging Fizeau interferometer could be used similarly to a corresponding Twyman-Green interferometer to measure the length of prismatic bodies (e.g. using a gauge block instead of the mirror and determining the fringe mismatch as shown in Fig. 9). In interferometric length measurement, a Fizeau interferometer offers particular advantages in a two-sided measuring setup, as represented in Fig. 17.

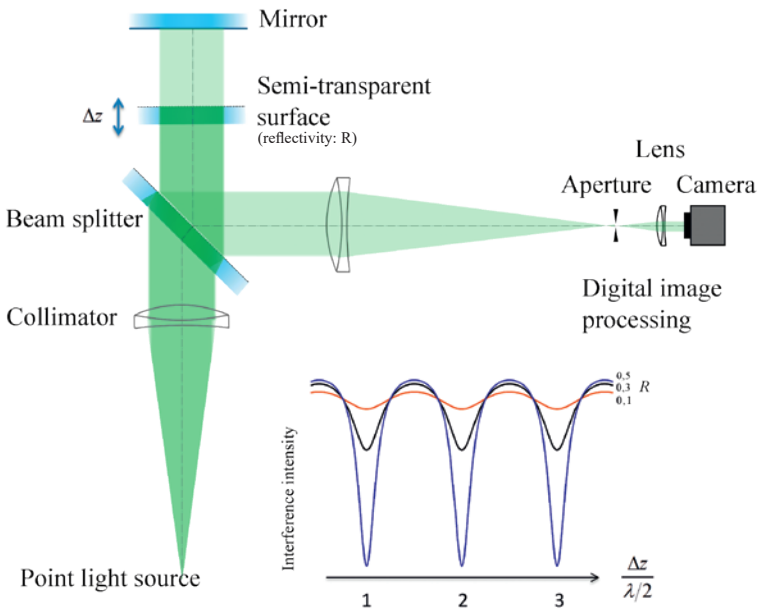


Fig. 16: Example of a Fizeau interferometer, observing the reflection. The characteristics of the interference intensity as a function of the variation of the light's path Δz are influenced by the reflectivity of the semi-transparent plate (the higher R is, the more dark fringes appear that become increasingly narrow).

Here, the distances Δz_1, Δz_2 and ΔZ shown in the figure are determined by a sequential measurement from both sides. This would allow the length of a prismatic material measure to be determined from $l = \Delta Z - \Delta z_1 - \Delta z_2$. Two-sided Fizeau interferometry is used at PTB in the sphere interferometers, which are unique worldwide. In a sphere interferometer, spherical reference surfaces are used instead of a flat semi-transparent surface in order to measure the diameter of spheres with sub-nm accuracy [11].

A Fabry-Pérot interferometer is actually a Fizeau interferometer that is operated with two semi-transparent mirrors while observing the transmitted light; however, most of the time, the reflectivity of semi-transparent mirrors is higher. As an example, a parallel glass plate with

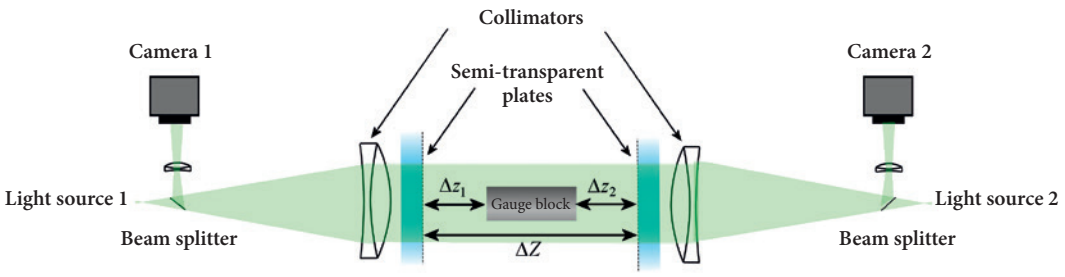


Fig. 17: Bidirectional Fizeau interferometer to measure the length of a gauge block.

reflective coatings is such an interferometer. The resonator thus obtained can be used as a material measure for the optical path length (the product of the refractive index of the glass plate and its thickness) and is also known as a "Fabry-Pérot etalon". Since Fabry-Pérot interferometers filter a narrow-band spectrum out of a broad-band radiation, they are often used as optical filters. The so-called "finesse" is used to characterize the resonator. It is defined as the ratio of the so-called free spectral range $\Delta\lambda$ to the full width at half maximum $\delta\lambda$ of an individual maximum of the interference intensity: $\mathcal{F} = \Delta\lambda/\delta\lambda = \pi\sqrt{R}/(1-R)$. The higher the finesse (i.e., at high reflectivities R), the narrower the band of the filtered light. An extremely stable Fabry-Pérot resonator with particularly high finesse has recently been developed and set up at PTB. This resonator, which is made of monocrystalline silicon, allows the light of commercial laser systems to be stabilized at the inconceivably low frequency of 0.04 Hz, i.e., better than 10^{-16} [12].

3.3.2.1. The importance of gauge blocks in the "traceability chain"

Since the 19th century, gauge blocks have become established as material measures. To date, numerous length-measuring instruments can be calibrated using gauge blocks. With a set of steel gauge blocks consisting of 103 pieces, more than 20 000 measures between 1 mm and 201 mm can be realized with a gradation of 0.005 mm by combining them.

The length of gauge blocks is traceable to the SI unit "meter" (as described above) by means of optical interferometry. The most precise gauge blocks serve as reference standards in mechanical differential measurements, which are a service provided by accredited calibration laboratories for length measurement.

Fig. 18:
Steel gauge block set (consisting of 103 pieces).

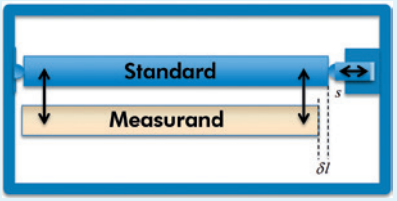


Fig. 19:
Principle of mechanical differential measurements of gauge blocks.

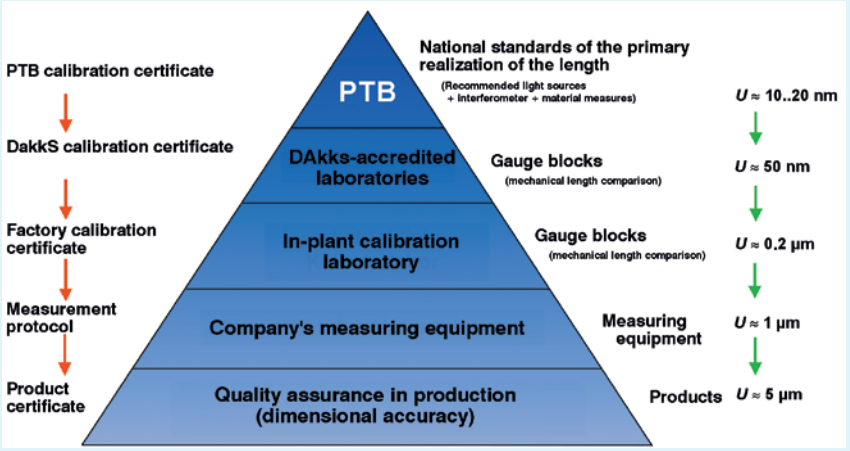


Fig. 20:
Chain of traceability to the SI unit *meter*, illustrated with the example of gauge blocks, DAkkS: German accreditation body, NMI: national metrology institute (in Germany: PTB).

3.4. Determining the refractive index of air

Realizing length by means of interferometry (e.g. for primary calibrations of gauge blocks) is nearly always done in air, mostly due to the fact that material measures are also used under atmospheric conditions when they serve as standards for length. The presence of air considerably reduces the wavelength of the light used (λ) compared to the wavelength in vacuum (λ_0), which is significant for the interferometrically measured length:

$$\lambda = \frac{\lambda_0}{n(\lambda_0, p, t, f, x)} \quad . \tag{9}$$

The refractive index of air, n, depends both on the wavelength itself and on the ambient conditions (the air pressure p, the air temperature t, the air humidity f, and the CO_2 concentration x). Since n is close to 1, the relative influence of the refractive index of air on the length is scaled by the so-called "refractivity", $n-1$ (approximation: $1/n\,|_{n\approx 1} \simeq 1-(n-1)$).

Determining the refractive index of air accurately is an important limitation to the precision of a length measurement; it can be done in two different ways:

A) high-precision measurement of the above-mentioned air parameters and determination of the refractive index of air by means of an empirical formula, $n(\lambda, p, t, f, x)$ [13]

B) interferometrically, by means of an air refractometer. The basic principle of the most accurate interferometric determination of the refractive index is shown in Fig. 21.

An evacuated cell, closed from both sides by means of large-area windows, is located inside the measuring arm of an imaging Twyman-Green interferometer (as in Fig. 9).
The collimated bundle of light beams traverses the evacuated inner part of the cell, and light beams traverse the windows outside the cell along the same geometric path. The length of the vacuum cell, expressed as the product of the interference orders $i + q$ (i.e., integer orders + interference fraction) by (half) the wavelength, can thus be expressed in two different ways:

$$l_{\mathrm{cell}} = (i_{\mathrm{vac}} + q_{\mathrm{vac}})\frac{\lambda_0}{2} = (i_{\mathrm{air}} + q_{\mathrm{air}})\frac{\lambda_0}{2n} \quad , \tag{10}$$

where the suffix "vac" represents the light path in vacuum and the suffix "air" represents the light path in air. Equation (10) yields the refractive index of air:

$$n = 1 + \frac{1}{l_{\mathrm{cell}}}\left(\overbrace{i_{\mathrm{air}} - i_{\mathrm{vac}}}^{\tilde{i}} + \overbrace{q_{\mathrm{air}} - q_{\mathrm{vac}}}^{\tilde{q}}\right) \Big/ \lambda_0/2 \tag{11}$$

where l_{cell} is the length of the vacuum cell which must be known with an accuracy of a few micrometers. The integer numbers $\tilde{i}$ can be estimated from $\tilde{i}^{\mathrm{est}} = (n^{\mathrm{est}} - 1) l_{\mathrm{cell}} \big/ \frac{1}{2}\lambda_0$ whereby n^{est} is the estimated value of the refractive index according to method A) above. $\tilde{q}$ is the interference fraction which can be determined from the interference phase topography (see Fig. 21, right):

$$\tilde{q} = \tfrac{1}{2\pi}\left[\tfrac{1}{2}\left(\phi^1_{\mathrm{air}} + \phi^2_{\mathrm{air}}\right) - \phi_{\mathrm{vac}}\right] ,$$

where ϕ^1_{air}, ϕ^2_{air} and ϕ_{vac} are the mean phase values within the circular areas of the light path in vacuum "vac" and in air "air".

If several wavelengths are used for the interferometric length measurement, then the refractive index of air must (of course) be determined for each of these wavelengths. Similar to length measurement itself, coincidence procedures are also applied in the case of the interferometric determination of the refractive index (for more details, see [8]).

When measuring large lengths, a third procedure called "refractive index compensation" is used:

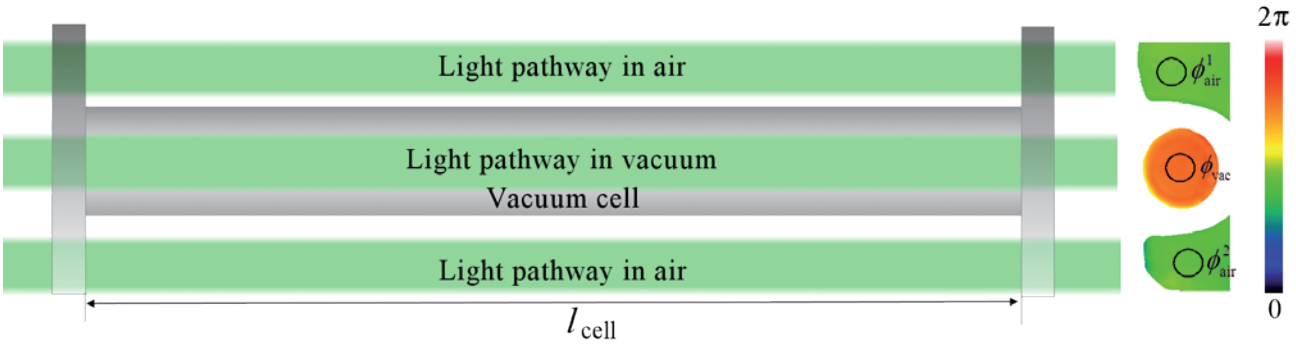


Fig. 21:
Schematic representation of the interferometric determination of the refractive index of air using a vacuum chamber.

C) In the refractive index compensation procedure, two different wavelengths are used (e.g. 532 nm and 1064 nm) such as those that can be supplied simultaneously by a Nd:YAG laser system.

The basic idea of refractive index compensation consists in considering the difference between the two measurements, providing that the wavelength dependence of the refractive index of air (dispersion) is known (according to the empirical formula, see A). In dry air, the refractivity scales with the so-called dispersion term: $n-1 \sim K(\lambda)$; therefore, the following correlation is given for two wavelengths:

$$\frac{n_1-1}{n_2-n_1}=\frac{K(\lambda_1)}{K(\lambda_2)-K(\lambda_1)}=A(\lambda_1,\lambda_2), \qquad (12)$$

where parameter A depends solely on the two wavelengths used, and not on the air parameters. If one considers two interferometric length measurements along the same path l that were performed with two different wavelengths, then l can be expressed either as a multiple of half the wavelength in air and of the measured interference phase ($\varphi = 2\pi(i+q)$), i.e., as $l = \varphi_1 ½\lambda_1/n_1$, or as $l = \varphi_2 ½\lambda_2/n_2$. However, in accordance with Equation (12), the same length can also be expressed as follows:

$$l = \varphi_1 ½\lambda_1 - A(\lambda_1, \lambda_2)(½\lambda_2\varphi_2 - ½\lambda_1\varphi_1). \qquad (13)$$

Accordingly, Equation (13) allows the length of a path to be determined by means of two wavelengths, without knowing the refractive index of the air. The disadvantage of this method is, however, that errors in the phase measurement by scaling with the large factor A lead to increased measurement uncertainty of the length. Moreover, Equation (13) is strictly valid only in dry air. A description of the extension required in humid air is given in [14].

3.5. Final remarks

The realization of a given length according to the definition of the meter in the International System of Units requires a measurement principle that establishes a relation between the travelling time of light in vacuum and the length to be measured. Two procedures are available for this purpose: a) measuring a travelling time difference directly (as described in Section 2), and b) using interferometry with light (see Section 3), which is essentially also a measurement of travelling time differences. Some of the most important fundamental principles of interferometric procedures for the realization of a given length are described in this article. All of the principles are based on the wave properties of light. Interferometric length measurement is a principle that can be realized relatively easily in practice. However, the achievable accuracy limits are reached very quickly. For the interferometric realization of lengths that is relevant in practice, a relative measurement uncertainty on the order of 10^{-7}, 10^{-8} or even 10^{-9} (1 nm in 1 m) is required. Given the availability of modern laser light sources whose frequencies usually exhibit a measurement uncertainty of better than 10^{-10}, it becomes clear that the frequency of the light used today is the smallest of challenges when it comes to reducing the measurement uncertainty of length measurements. In order to make the realization of a given length more precise, other limitations come to the fore, namely those that have a direct influence on the result of the interferometric measurement. Such factors include the accuracy of the interference phase measurement, which is mentioned in this article, as well as the influence of the refractive index of the air, but also limits set by the quality and the adjustment of optical components that cannot simply be manufactured with any desired level of precision. Evaluations are often based on assumptions which, in view of new requirements, can no longer be adhered to. This makes the struggle for improvements a difficult one that demands a great deal of resilience to frustration from the persons involved, and also openness with regard to accepting their own errors. The forefathers of interferometric procedures surely never imagined that the measurement uncertainties achievable today would ever become possible.

Literature

[1] *N. Huntemann, M. Okhapkin, B. Lipphardt, S. Weyers, C. Tamm and E. Peik*; 2012, High-accuracy optical clock based on the octupole transition in 171Yb+, Phys. Rev. Lett. **108** (2012) 090801.

[2] Recommended values of standard frequencies: http://www.bipm.org/en/publications/mises-en-pratique/standard-frequencies.html (last accessed on 19 September 2015).

[3] *P. Cordiale, G. Galzerano and H. Schnatz*; International comparison of two iodine-stabilized frequency-doubled Nd:YAG lasers at 532 nm, Metrologia **37 (**2000) 177–182.

[4] *C.W. Chou, D.B. Hume, T. Rosenband, D.J. Wineland*; Optical Clocks and Relativity, Science **329** (2010) 1630–1633.

[5] *M. Wedde, K. Meiners-Hagen, A. Abou-Zeid*; Interferometrische Kalibrierung von Strichmaßen, Laserinterferometern und Entfernungsmessgeräten: Die Geodätische Basis der PTB, PTB-Mitteilungen **2/2010** (2010) 120–123, http://www.ptb.de/cms/?id=2744 (last accessed on 16 December 2015).

[6] *C. Weichert, P. Köchert, R. Köning, J. Flügge, B. Andreas, U. Kuetgens and A. Yacoot*; A heterodyne interferometer with periodic nonlinearities smaller than ±10 pm, Meas. Sci. Technol. **23** (2012) 094005.

[7] *S. Tang*; Self-calibrating five frame algorithm for phase shifting interferometry Proc. SPIE **2860** (1996) 91–7, http://dx.doi.org/10.1117/12.276293.

[8] *R. Schödel*; Utilisation of coincidence criteria in absolute length measurements by optical interferometry in a vacuum and in air, Meas. Sci. Technol. **26** (2015) 084007, http://dx.doi.org/10.1088/0957-0233/26/8/084007.

[9] *R. Schödel, A. Walkov, M. Zenker, G. Bartl, R. Meeß, D. Hagedorn, D. Gaiser, G. Thummes and S. Heltzel*; A new Ultra Precision Interferometer for absolute length measurements down to cryogenic temperatures; Meas. Sci. Technol. **23** (2012) 094004 (19pp), http://dx.doi.org/10.1088/0957-0233/23/9/094004.

[10] *R. Schödel, G. Bönsch*; Highest accuracy interferometer alignment by retroreflection scanning, Appl. Opt. **43** (2004) 5738–5743.

[11] *R. A. Nicolaus, G. Bönsch*; A novel interferometer for dimensional measurement of a silicon sphere, IEEE Trans. Instrum. Meas. **46** (1997) 563–565.

[12] *T. Kessler, C. Hagemann, C. Grebing, T. Legero, U. Sterr, F. Riehle, M.J. Martin, L. Chen, J. Ye*; A sub-40-mHz linewidth laser based on a silicon single-crystal optical cavity, Nature Photonics **6** (2012) 687–692.

[13] *G. Bönsch and E. Potulski*; Measurement of the refractive index of air and comparison with modified Edlen's formulae, Metrologia **35** (1998) 133–9.

[14] *K. Meiners-Hagen and A. Abou-Zeid*; Refractive index determination in length measurement by two-colour interferometry, Meas. Sci. Technol. **19** (2008) 084004, http://dx.doi.org/10.1088/0957-0233/19/8/084004.

Counting Electrons to Measure Current

Hansjörg Scherer*, Uwe Siegner**

1. Introduction

In every aspect of our daily lives, we are surrounded by electricity – every household receives an electricity bill at regular intervals. However, the "product" we are charged for on this bill is not the electric *current*, but the electric *energy* supplied for use in lighting, heating, cooling and mechanical work. The physical quantity of "current" is defined as the amount of electric charge that flows through the cross-section of an electric conductor, divided by the length of an interval of time. In order to measure the electric current, and to define its corresponding physical quantity (the ampere), different effects of the electric current can be used, as history shows us. For example, if current is conducted through a metallic salt solution, the metal ions are discharged and the metal is separated at the cathode. Until the mid-20th century, the unit of electric current was defined on the basis of an electrolytic process of this kind. This so-called "international ampere" was also defined according to the 1898 laws of the then German Empire; the definition reads as follows:

"The ampere is the unit of electric current. It is realized by means of the unchanging electric current that, upon passing through an aqueous solution of silver nitrate, deposits 0.001118 grams of silver in one second."

In 1948, at the 9th *General Conference of the Metre Convention* (the international agreement on the development and usage of a metric system of units), a new definition for the ampere as one of the base units of the International System of Units (SI) was adopted. This definition is based on the fact that electric current generates a magnetic field, and that another current-carrying conductor meets with a force (the *Lorentz force*) within this magnetic field: Two current-carrying conductors attract or repel each other – depending on the directions of the currents. This definition of the ampere (illustrated in Fig. 1), which was laid down at that time and is still valid in the current SI, reads as follows:

"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 meter apart in vacuum, would produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per meter of length."

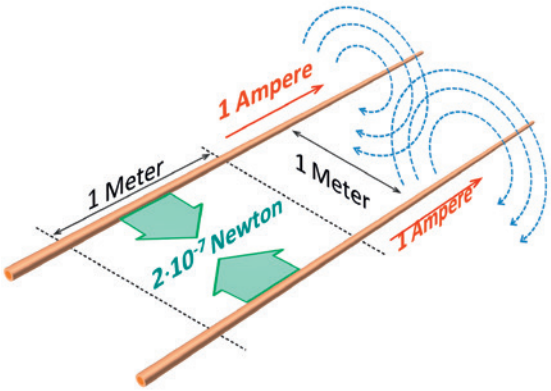


Fig. 1:
Illustration of the SI ampere definition. The forces marked in green, caused by the magnetic fields of the conductors (field lines indicated by blue arrows), each have the absolute value of $2 \cdot 10^{-7}$ N per meter of conductor length if a current of 1 ampere is flowing.

This definition was realized in practical implementation by means of electromechanical apparatus such as the *current balance*. This device was developed at the beginning of the 19th century by *André-Marie Ampère*, after whom the unit of electric current (the ampere) was named. Using this device, the Lorentz force between two conductive coils (approximating the requirement of conductors of *infinite length*) is balanced out or counterbalanced by means of a mechanical weighing device via the gravitational force of a mass in the gravitational field of the Earth, as shown schematically in Fig. 2.

Until as late as the second half of the 20th century, different variations on the current balance were used for *direct* realization (i.e., realized in accordance with the definition) of the ampere. Here, it was possible to achieve relative uncertainties of a few parts per million. The precision was limited due to the uncertainty that arises when determining the distance between the current

* Dr. Hansjörg Scherer, Working Group "SET, Current and Charge", e-mail: hansjoerg.scherer@ptb.de

** Dr. Uwe Siegner, Division "Electricity", e-mail: uwe.siegner@ptb.de

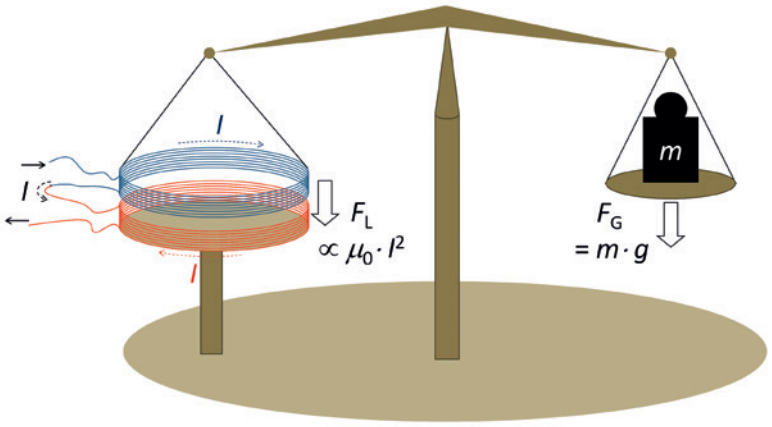


Fig. 2:
Principle of the current balance. The Lorentz force F_L, caused by the coil currents I, acts between a non-moving coil (red) and a moving coil (blue). This force is compensated for by the gravitational force F_G exerted on the other side of the balance beam by the mass m.

paths as, in accordance with the definition above, this distance contributes to the uncertainty. Better results were achieved with *indirect* realizations, by means of which the ampere was realized "via detours", in a certain sense. In these indirect realizations, Ohm's law ($I = U/R$) is exploited and the ampere is realized via the electrical voltage U and the electric resistance R. One example of the way in which these quantities were realized is by means of a *current balance* and a *calculable cross capacitor* (also known as a *Thompson-Lampard capacitor*). As with the current balance, these experimental set-ups are based on electromechanical devices, yet achieve lower uncertainties of a few parts per 10 million.

What the two definitions above have in common is the fact that they link the *electric* unit of the ampere to a *mechanical* quantity (mass). The earlier "international ampere" was based directly on a determination of mass by weighing electrolytically deposited silver. In the SI definition of the ampere that is still valid today, mass enters via force. As regards the accuracies that can be obtained, both definitions of the ampere depend on the accuracy of the realization of the kilogram – a fact that is generally seen as problematic in light of the assumed instability of the international prototype of the kilogram (explained in further detail in the article titled *Counting Atoms for Mass and Amount of Substance* in this publication). Furthermore, neither of the two definitions takes into consideration the definition of the physical quantity of electric current, or the flux of electric charge per time. This is the most direct method for determining the unit of the ampere and will be applied in the new SI. The following section explains why this new definition is favorable and how the development of the electrical units led to this new definition. Section 3 then explains the "future ampere" and its possible realization.

2. Electrical quantum effects and "modern" electrical units

A new chapter in electrical metrology began with the discovery of two electrical quantum effects in the second half of the 20th century:

The first discovery was *Brian D. Josephson's* prediction in 1962 of an effect (later named after him) between weakly coupled superconductors. When microwaves are irradiated, this quantum-mechanical tunnel effect leads to the formation of constant voltage levels in the current-voltage characteristic of a tunnel junction. Shortly after Josephson's prediction, these voltage levels were experimentally observed; they enable voltage values to be generated with fundamental precision.

The second discovery was made by *Klaus von Klitzing* in 1980; during investigation of the Hall effect in two-dimensional (i.e., extremely thin) conductive layers, von Klitzing discovered B levels of constant resistance in high magnetic fields, thereby also discovering a new effect that can be used to realize quantized (i.e., discrete) resistance values. This effect was later named after him, and is also known as the "quantum Hall effect". Both effects are illustrated in greater detail in Fig. 3. A detailed explanation of the physical fundamentals of these effects can be found in reference [1].

Not long after being discovered, both the Josephson effect and the quantum Hall effect were put to use at PTB and at other national metrology institutes due to their excellent suitability for generating well-defined values for the electrical quantities of voltage and resistance. In the case of the Josephson effect, the voltage values are

$$U_n = n \cdot K_J^{-1} \cdot f_J \qquad \text{with } n = (1, 2, ...),$$

where f_J is the frequency of the microwave radiation and K_J is described as the *Josephson constant*. According to the underlying theory, this constant is given by $K_J = 2e/h$, i.e., by the elementary charge e and Planck's constant h, two so-called *fundamental constants*. The Josephson constant is equal to the reciprocal value of the magnetic flux quantum, and its value is around $5 \cdot 10^{14}$ Hz/V. For the quantum Hall effect, the quantized resistance values are

$$R_i = 1/i \cdot R_K \qquad \text{with } i = (1, 2, ...),$$

where R_K is the *von Klitzing constant*. According to theory, the von Klitzing constant is also connected to e and h via $R_K = h/e^2$.

It quickly became apparent that the reproducibility of the values U_n and R_i (generated with quantum effects) was considerably better than

the results obtainable with other, more conventional methods of electrical metrology. A high degree of reproducibility means that the values of the electrical quantities generated in different experiments scarcely deviate from one another – not even when (for example) different material systems are used. This discovery provided a new foundation for an internationally uniform realization of the units of the volt and the ohm. Not least for this reason, the discoverers of these effects were each awarded the Nobel Prize in Physics some years later. In particular, the reproducibility of the quantized voltage and resistance values was better than the accuracy with which the unit of the ampere could be realized – a matter that, incidentally, has not changed to date. For this reason, the International Committee on Weights and Measures decided in 1988 to establish exact values for the Josephson constant and for the von Klitzing constant, and to recommend that these values be used for the maintenance and dissemination of the electric units. The exact values established at that time, which have been in use since 1990, are $K_{J\text{-}90} = 483\,597.9 \cdot 10^9$ Hz/V for the Josephson constant, and $R_{K\text{-}90} = 25\,812.807\ \Omega$ for the von Klitzing constant.

In the decades thereafter, this development proved itself to be extremely favorable for increasing the accuracy and comparability of electrical measures in metrology, yet also entailed a certain dilemma: Basing the units of the volt and the ohm on the two quantum effects with the established values of $K_{J\text{-}90}$ and $R_{K\text{-}90}$ also meant leaving the SI. From that point on, and to this day, the electrical units derived in this way have existed in what is technically a "parallel" system of units; this system is also referred to as a *conventional* electrical system in order to distinguish it from the SI. For the units that have been traced to $K_{J\text{-}90}$ and $R_{K\text{-}90}$ around the world since the 1990s, this means that they are not referred to in terms of their *realization* (i.e., the practical implementation of the definition of the units into reality that, according to definition, must take place in the SI) but only in terms of their *reproduction*. This means the realization of a unit based on a physical effect that can be reproduced extremely well – in this case, on the Josephson effect for the volt and on the quantum Hall effect for the ohm.

Since that time, electric current has also been traced to the electrical quantum effects for resistance and voltage with an incomparably high degree of accuracy; here, in turn, advantage is taken of the fact that the current is linked to both of these quantities via Ohm's Law. For this reason, however, the ampere is merely *reproduced*, in the sense explained above. The planned redefinition of the SI reveals an elegant way out of this unsatisfactory situation.

3. The future definition of the ampere

In the future SI, the numerical value of the elementary charge e will be defined in the unit of "coulomb equals ampere times second" (1 C = 1 As).

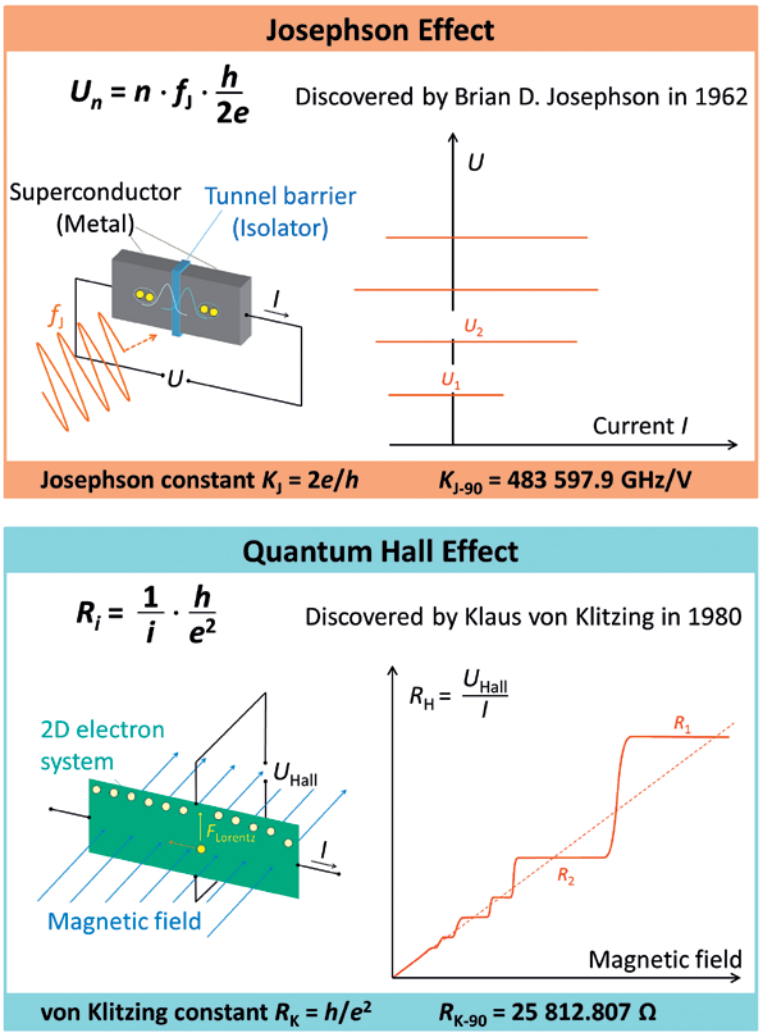


Fig. 3:
Profiles of the Josephson and quantum Hall effects. The effect above, named after Brian D. Josephson, leads to steps of constant voltage in the current-voltage curves of superconducting tunnel junctions that are irradiated with microwaves of the frequency f_J. In present-day metrology, series arrays made up of thousands of microstructured tunnel junctions are used to realize DC voltages of up to 10 volts, thereby reproducing the electrical unit of the volt. The effect below, discovered by Klaus von Klitzing, occurs in two-dimensional semiconductor structures in which the movement of electrons is limited to one level. In the case of high magnetic fields of several tesla, the effect leads to plateaus in the course of the Hall resistance as a function of the magnetic field. In metrology, the quantum Hall effect is used to reproduce the unit of the ohm. Both effects necessitate the use of low-temperature technology: the electrical circuits typically have to be operated at very low temperatures around the boiling point of liquid helium – i.e., 4 kelvin (–269 °C) – or lower.

The corresponding definition reads as follows:

"The International System of Units, the SI, shall be the system of units in which the elementary charge e is exactly 1.602 176 620 8 · 10^{-19} coulombs."

Thus, the elementary charge as one of the seven defining constants of the new SI will be the reference point for the ampere. As the ampere was chosen as the electrical base unit of the SI, an additional, *explicit* definition was formulated for this purpose. This definition reads as follows:

"The ampere, represented by the symbol A, is the SI unit of electric current. It is defined by means of the numerical value of the elementary charge e that is established as 1.6021766208 · 10^{-19}, expressed in the unit C, which is equal to As, where the second is defined in relation to $\Delta\nu_{Cs}$"

Here, $\Delta\nu_{Cs}$ is the frequency that corresponds to the transition between the two hyperfine levels of the undisturbed ground state of atoms of the cesium 133 atom. Thus, this corresponds to the definition of the second that is based on optical clocks. As discussed in the introduction, the new definition of the unit of current is thus based directly on the physical definition of electric current as the flux of electric charge. Current can thus be determined by counting electrons (charge quanta with the charge -*e*) that flow through a conductor per unit of time. In order to realize this "quantum ampere", an electric circuit is therefore needed that can transport single electrons in a controlled way, as shown schematically in Fig. 4. If this transporting is cyclic and clocked at the frequency *f*, and if *n* electrons are conveyed in every cycle, the current can be expressed as $I = n \cdot e \cdot f$.

The possible realization of the future ampere by means of single-electron circuits – as these special circuits are called – is a prime example of *quantum metrology*. This metrological discipline pursues the realization of physical units by counting quanta, i.e., precisely defined amounts of quantities that are themselves traced to fundamental constants. These constants, according to present knowledge, are unchangeable in terms of time and space. As a result, it is possible to create a universal system of units, as Max Planck, the founder of quantum physics, had the foresight to express in 1899 [2]:

"... establishing units ... which necessarily maintain their validity for all times and for all cultures (this also includes extraterrestrial and non-human cultures) – independent of special bodies and substances – and which can, therefore, be called 'natural measurement units'."

Although the generation of quantized currents by means of single-electron transport (SET) is as natural as it is elegant, it is not surprising that this approach was not implemented earlier – after all, to do so, single electrons have to be deliberately manipulated in conductor structures. This is possible by means of special electronic circuits, as explained in the following section. The technological methods required to manufacture circuits of this kind were not developed until the late 1980s.

4. Single-electron circuits

In very small electrical circuits, effects occur that are caused by the repellant coulomb interaction between electrons: The repulsion between particles with the same kind of electrical charge increases if they are brought closer together. If electrons are "locked in" very closely together in circuits (see Fig. 5), this is also expressed in their electronic properties: The electrons can then only take on discrete states of energy that are separate from one another. This is the basis of the so-called *Coulomb blockade* effect that is exploited in single-electron circuits in order to control the flux of single electrons. To do so, extremely small structures of typically 1 µm or smaller are required, in addition to extremely low temperatures. In order to manufacture circuits that are this small, the same modern nanotechnology methods are used that are also used for purposes such as the production of highly integrated electronic circuits.

Thus, in order to be able to use the effect described above specifically for single-electron circuits, the electrons have to be "trapped" – or, more precisely, *localized* in such a way that they can be controlled – in sections of a conductor that are very small (so-called "charge islands", or "islands" for short). This is accomplished by means of potential barriers that can be generated in a direction perpendicular to the current flow using nanotechnology. Here, two categories of single-electron circuits are distinguished that require different technologies for their manufacture and are based on different physical principles.

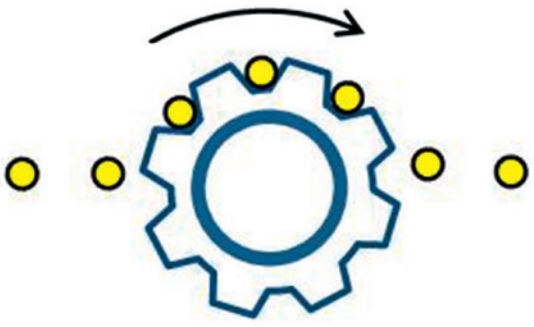


Fig. 4:
Illustration of the principle of a single-electron current source, depicted as a gear for the clocked transport of single electrons (charge quanta, yellow). In each recess between two neighboring "teeth", one electron is transported.

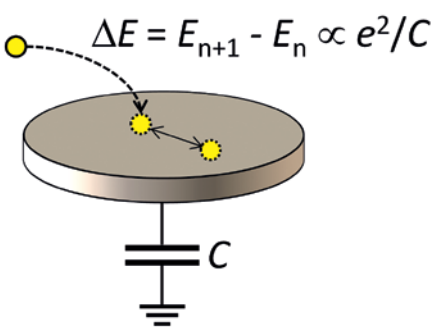


Fig. 5:
Explanation of the Coulomb interaction between electrons in a conductor, depicted here as a disc. This "charge island" has a total capacitance C to the surroundings. In order to add an electron ($n \rightarrow n+1$), energy has to be applied that matches the electrostatic energy variation $\Delta E \sim e^2/C$. If the charge island is made smaller, its capacitance C decreases, and the "added energy" ΔE increases. Thus, for smaller dimensions of the charge island, more energy has to be applied in order to charge it with additional electrons. If the temperature of the system is so low that it is not possible to apply this added energy by means of thermal excitation, this is expressed in the so-called "Coulomb blockade" effect. To this end, temperatures under 0.1 kelvin (corresponding to 0.1 degree above absolute zero) are usually required for structures in the order of magnitude of 1 µm.

The first category of single-electron circuits is based on so-called "tunnel junctions" – very thin insulation layers that realize potential barriers in a metallic conductor [3]. The "height" of these barriers is given by material parameters and is thus unchangeable. However, if the insulation layers are only a few nanometers thick, electrons can pass through them due to the quantum-mechanical *tunnel effect* (see Fig. 6 above). The flow of electrons across a metallic island between two tunnel junctions can be controlled by means of an electrical voltage applied to a gate. By means of the voltage, the potential of the island can be electrostatically displaced, with the result that the *Coulomb blockade* is raised. In accordance with the laws of quantum physics, the barriers are "tunneled through" with a certain probability per unit of time – or, to put it crudely, the electron must be given a certain amount of time to allow the tunneling process to take place. For this reason, the charge transfer via a tunnel junction is naturally subject to the laws of statistics.

The second category of single-electron circuits is based on the use of *controllable* potential barriers in semiconductor materials [4]. These potential barriers are generated electrostatically by means of two negatively charged gate electrodes that cross a thin conducting bridge. Here, the heights of the barriers can be varied by changing the gate voltages (Fig. 6 below). The island that develops into a "trough" in the potential landscape between the two barriers is also known as the "quantum dot". Via the gate voltages, it is possible to control the occupancy of the quantum dot with electrons. The methods used to manufacture circuits of this kind are similar to the methods used to manufacture modern field effect transistors.

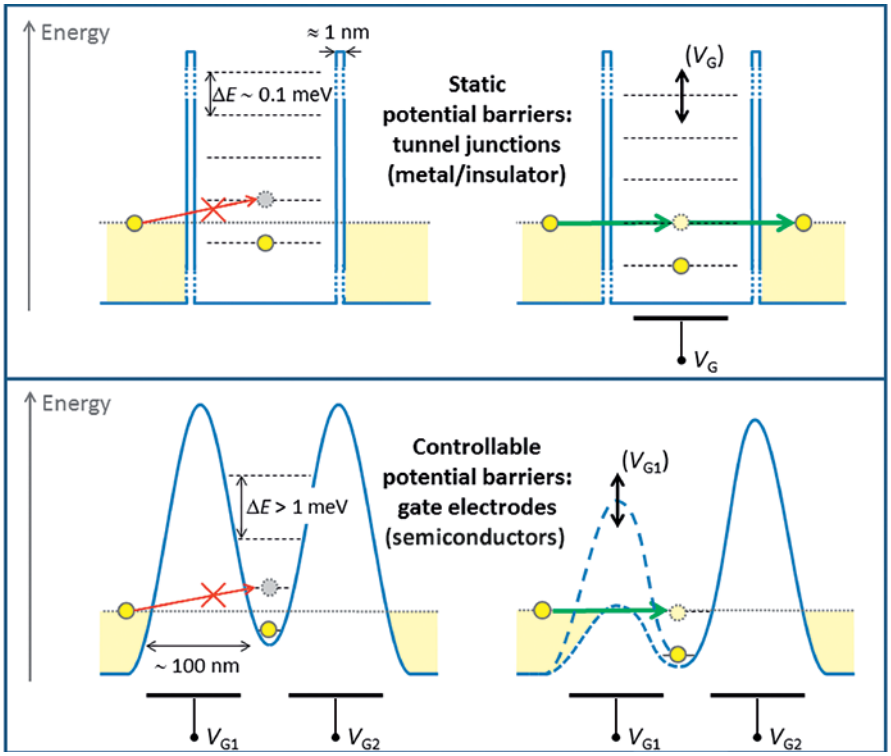
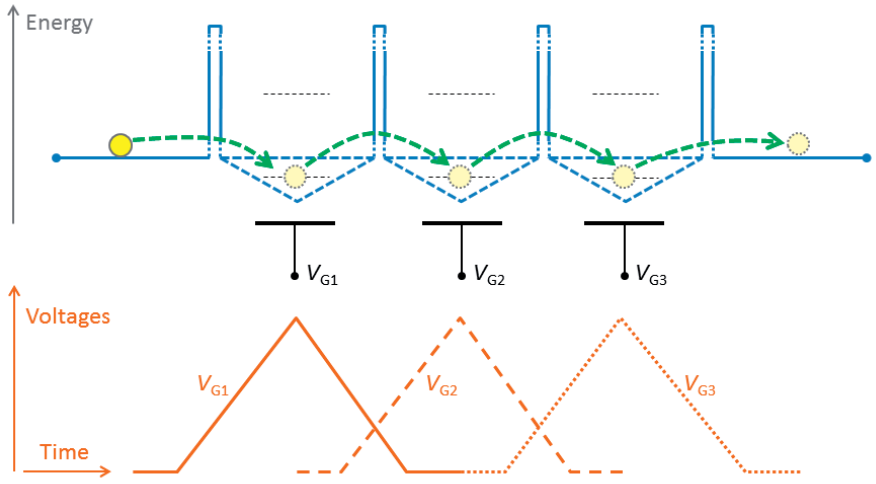


Fig. 6:
Schematic representation of different single-electron circuits. The type shown above is based on a small "charge island" formed by the insulation barriers of two tunnel junctions. Due to the Coulomb interaction in the small island, "island electrons" can only take on discrete states of energy (depicted by the dashed lines). At very low temperatures (usually well below 1 kelvin), this means that the flow of electrons across the island is blocked, as shown at the top left. This "Coulomb blockade" can be lifted if the potential of the island is set by means of the gate voltage V_G in such a way that electrons can "slip" through (or, more precisely, quantum-mechanically tunnel through) the tunnel junctions without a change in energy, as shown at the top right. The circuit type shown below is based on two controllable potential barriers. These barriers are caused by electrostatic potentials that occur when negative voltages V_{G1} and V_{G1} are applied to the two gate electrodes of the circuit. Here as well, no flow of additional electrons onto the island is possible at first (as shown at the bottom left). By lowering the barrier (i.e., by raising the voltage V_{G1}, as shown at the bottom right), transport of electrons onto the island becomes possible. In this way, the electron occupancy of the island can be varied in a controlled way. Here as well, if the island is very small, discrete energy levels will develop: Only certain energy values in the potential trough can each be occupied by one electron, referred to as a "quantum dot". Thus, the Coulomb blockade effect occurs here as well.

These two types of single-electron circuit are different not only in terms of the technology used to manufacture them, but also in a very basic sense in terms of their modes of operation and their characteristics. This will now be explained with the aid of a particular circuit type – the so-called *single-electron pump* or *SET pump*. SET pumps make it possible to transport single electrons in a controlled manner, thereby facilitating quantized current generation in accordance with $I = n \cdot e \cdot f$, as shown in Fig. 4. These pumps therefore have a particularly important role concerning the future realization of the "quantum ampere".



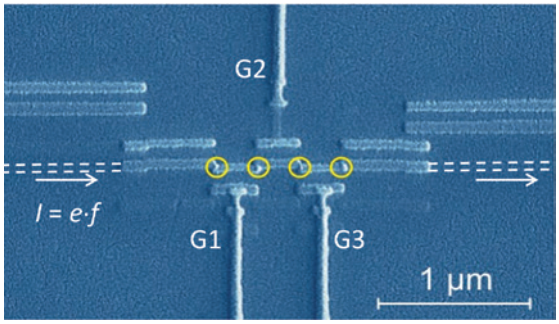
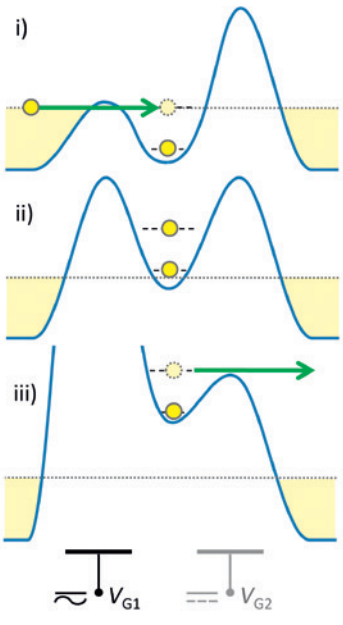


Fig. 7:
Top: Operation of an SET pump with four tunnel junctions (three "charge islands" between two tunnel junctions) in a schematic representation. The transport cycle of an electron through the circuit is shown. Bottom: Electron microscope image of an SET pump of this type with four tunnel junctions connected in series (circled in yellow) and three gate electrodes (G1-G3), which trigger the potentials of the islands.

SET pumps with static or controllable potential barriers

The construction and operation of an SET pump with tunnel junctions – i.e., with static potential barriers – are shown in Fig. 7. This SET pump is made up of a series array of at least three tunnel junctions; here, each of the islands between each set of two neighboring tunnel junctions is equipped with one gate electrode. In this way, the potentials of the islands can be electrostatically controlled. If all gate voltages are zero, no electrons can flow through this circuit due to the Coulomb blockade. If a stream of voltage pulses is then sent across the gate electrodes, the Coulomb blockade of the islands that are arranged in series will be removed one after the other, and an electron will follow the electrical "polarization wave" of gate voltages from island to island through the circuit. Here, the Coulomb blockade prevents one island from being occupied by two (or more) electrons. If this transport cycle is repeated at a frequency f, this SET pump will supply a current $I = e \cdot f$. As mentioned above, however, the transport through the tunnel barriers is subject to the laws of statistics. As a result, errors due to "missed" tunneling events occur much more often at frequencies above roughly 100 MHz. In practice, this means that the maximum current attainable is limited to around 10 pA to 20 pA (1 pA = 10^{-12} A).

The operation of an SET pump with controllable potential barriers is shown in Fig. 8. Here, by means of an AC voltage V_{G1} applied to the gate electrode, the height of the left barrier is periodically modulated in such a way that, in an alternating manner, single electrons coming from the left side of the conductor are trapped in the "dynamic quantum dot" between the barriers and then released to the other side. In this transport mechanism, no "slow" tunneling processes due to high potential barriers occur (which would restrict the repetition frequency f). For this reason, an SET pump with controllable potential barriers can provide markedly higher currents than the tunnel-junction pump described above – frequencies into the gigahertz range are possible that correspond (according to $I = e \cdot f$) to maximum currents of over 160 pA. Another advantage of this pump type is that it is only necessary to operate *one* gate electrode with an AC voltage. This simplifies the layout of the circuit and facilitates the pump operation.



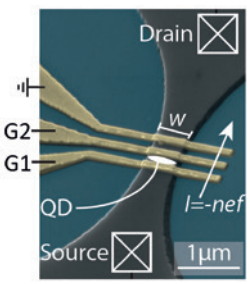


Fig. 8:
Right: Operation of an SET pump with controllable potential barriers. The image shows a transport cycle through the "dynamic quantum dot"; in this cycle, an electron coming from the left is first trapped (i) and isolated in the quantum dot (ii) before then being released to the right (iii). Here, only the height of the left barrier is modulated. Bottom left: Electron microscope image of an SET pump with a "quantum dot" (QD) between gate electrodes G1 and G2.

Challenges and the current state of research

Two issues number among the most important challenges in the metrological use of SET pumps:

1. Strength of the current generated: SET pumps available to date provide only very **small currents** that are typically less than 1 nA (10^{-9} A).
2. Accuracy of the current generated: Single-electron transport in SET pumps typically results in (statistical) **errors**.

Work is currently being done in research and development to find a solution to the problems associated with these issues, with results that are already very promising:

Concerning issue 1: As mentioned above, the currents that can be obtained from tunnel-junction SET pumps are limited to around ten picoamperes. Pumps that are based on "dynamic quantum dots" can provide currents that are up to ten times greater by means of pump frequencies in the gigahertz range. This has been demonstrated with different semiconductor SET pumps. Here, it was confirmed that, according to $I = e \cdot f$, quantized currents with an intensity of approximately 100 picoamperes can be generated at a level of accuracy that is sufficient for metrological applications as the future current standards. Nevertheless, endeavors are being pursued to investigate and develop new concepts for SET pumps that open up the nanoampere range [5]. Appropriate research activities require methods for *high-precision measurement* of currents that are this small. For the planned use of SET pumps as the future current standards, and for their practical use in metrology based on this, extremely *precise methods of current amplification* are required in addition. Thanks to innovations achieved at PTB in instrumental metrology, considerable progress has been made recently concerning both of the above points: Researchers at PTB recently developed and verified a novel amplifier for small currents [6, 7]. This novel instrument, whose performance over time is extremely stable, makes it possible to amplify currents (by a factor of 1000 or more) with very high precision. In addition, input currents in a range from sub-fA ($< 10^{-15}$ A) to 5 µA ($5 \cdot 10^{-6}$ A) can be transformed into voltage signals with extremely high accuracy by means of a network of resistors. This means that small currents can now be measured with unmatched accuracy, traced to the quantum Hall resistance and to the Josephson voltage standard – relative uncertainties of 10^{-7} for 100 pA have already been demonstrated.

Concerning issue 2: The intended use of SET pumps as future current standards requires that relative uncertainties of 10^{-7} or better be achieved when current is generated. As a reminder, the indirect, "conventional" realization of the SI ampere mentioned in the introduction that takes place by means of electromechanical apparatus already achieved a relative uncertainty of a few parts in 10^7. In principle, however, errors occur during single-electron transport in SET pumps; these errors can cause deviations in the supplied current from the quantized value $e \cdot f$, thereby limiting the accuracy of the generated current. In the case of SET pumps that are based on metallic tunnel junctions, errors may be caused by the static nature of the tunneling processes or by thermal or electromagnetic excitation of electrons (among other things). Pumps of this type are susceptible to these latter errors as, typically, the Coulomb blockade in them is relatively weak: The corresponding energy distances ΔE on the charge islands are in the range below around 0.1 meV (see Fig. 6 above). In the case of semiconductor pumps based on "dynamic quantum dots", errors may occur during the "loading-in phase" of electrons into the quantum dot (Phase i in Fig. 8) – for example, if an electron falls back to the side it came from before it can be isolated stably in the potential trough (Phase ii in Fig. 8). All of the errors mentioned above typically occur statistically and have to be taken into consideration quantitatively when current is generated in order to make a statement on the accuracy attained. This, in turn, requires that individual errors in the SET circuits be "counted". To do so, ultrasensitive charge detectors are used that can measure the charge at a resolution of smaller than e and give evidence of individual electrons [8]. These so-called "SET detectors" can also be realized by means of special single-electron circuits (so-called "SET electrometers" or "SET transistors"); however, in this article, such circuits will not be examined in detail.

Recently, the state of the art of error detection used in SET pump circuits was advanced considerably at PTB by developing new "counting procedures" [9, 10]. What makes these procedures special is the fact that they do not count every electron transported by the SET pumps. At high pump frequencies, this would not be possible anyway due to the limited range of the detectors; instead, the method is based on counting only the errors that occur much less frequently than the transport of the electrons. Here, an arrangement of several SET pumps in sequence is used together with SET detectors, as shown schematically in Fig. 9. In this way, pump errors can practically be detected *in situ* while current is being generated, and can be taken into account in order to correct the current supplied. For this purpose, a correlation analysis is performed of the detector signals that are read out at the same time; by means of this analysis, certain error types can be identified unambiguously. This

makes it possible to identify which of the pumps has made an error, and whether the error concerns a "missing" or an "additional" electron.

Experiments performed at PTB have already successfully demonstrated that, by means of this procedure, the accuracy of the current generated using an SET current source can be increased considerably [10]. A prototype of the "self-referenced" (i.e., *self-inspecting*) single-electron current source used for this purpose is shown in Fig. 10.

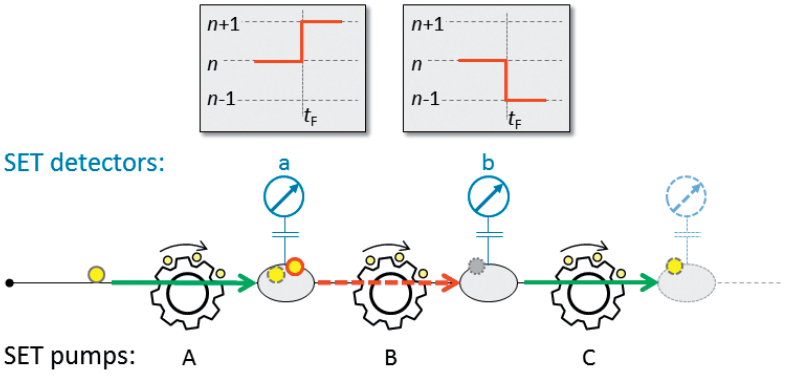


Fig. 9:
Detection/counting of SET pump errors in a series array of SET pumps by means of SET detectors that detect the charge states of the islands between two pumps. While the pumps are in continuous pumping operation, an error caused by pump B is caused at a point in time t_F – here, an electron is "left behind" on the island between pumps 'A' and 'B'. This is seen in the signatures of the two SET detectors ('a' and 'b'), which monitor the charge states of each of the islands: Following the error, the signal from detector 'a' (topmost left) shows one excess electron (red outline) on the first island ($n \rightarrow n$+1). At the same time, detector 'b' registers that there is now one electron missing on the next island ($n \rightarrow n$-1), as the correctly functioning pump 'C' has removed one electron.

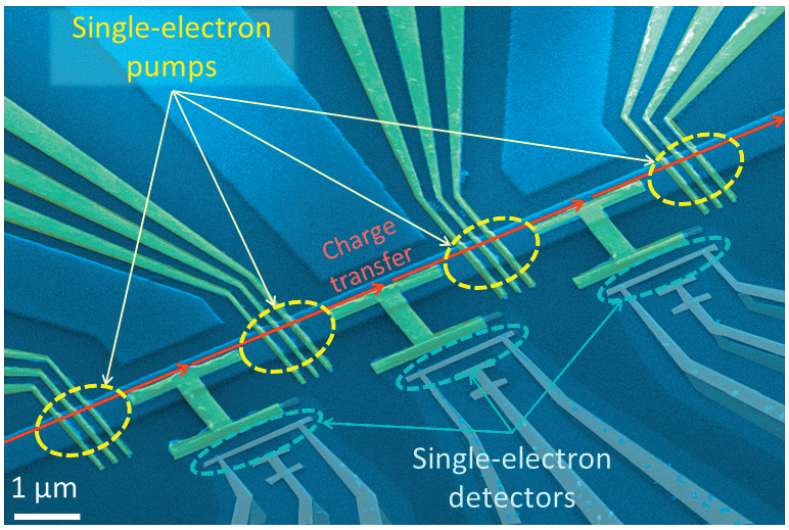


Fig. 10:
Prototype of a "self-referenced" single-electron current source (operation scheme in accordance with the illustration in Fig. 9) that uses four series-connected SET pumps (semiconductor structures with controllable potential barriers) and SET detectors that monitor the charge states of the islands between the SET pumps.

Concerning future development, it is expected that progress in both small-current measurement technology and in the further development of self-referenced SET current sources will contribute markedly to the future realization of the "quantum ampere". Even today, PTB is a leader in both fields.

5. Outlook

In the near future, it is expected that advanced SET pumps with controllable potential barriers will supply currents into the nA range with relative uncertainties of better than 10^{-7}. PTB has set itself the objective of realizing the re-defined ampere by means of a single-electron current source for currents of around 100 pA with a relative uncertainty of smaller than $1 \cdot 10^{-7}$. In this way, together with the new current-amplifying instruments developed at PTB, the current range between 1 fA and 1 µA can be accessed with superior accuracy. Thus, even the accuracy of instrument calibrations in the small-current range is expected to be improved by up to two orders of magnitude. This is relevant not only for research on single-electron circuits, but also for industrial applications and for environmental and medical metrology. The modern semiconductor industry (microelectronics and nanoelectronics) as well as environmental and medical metrology (such as dosimetry and emission protection measurements) increasingly require methods for precisely measuring small currents.

Concerning higher currents, realizations of the ampere by means of the quantum Hall effect and the Josephson effect (via Ohm's law) will continue to be better suited. In fact, these two quantum effects, both of which are long-established in electrical metrology, will become even more important due to the re-definition of the SI: Since the value for h will be established in the new SI in addition to the value for e, Josephson voltage standards will realize the SI volt (based on $K_J = 2e/h$) and quantum Hall resistors will realize the SI ohm (based on $R_K = h/e^2$). When this happens, both the direct realization of the ampere by means of SET pump circuits (via $I = n \cdot e \cdot f$) and the indirect realization by means of a Josephson voltage standard and a quantum Hall resistor will conform to the SI.

Concerning electrical metrology in general, it should be noted that the use of the Josephson and quantum Hall effects will be further developed based on the advanced level that has already been achieved – in particular for AC applications. The further development of electrical metrology will not only improve practical electrical measurement technology, but will also allow new, fundamental experiments to take place that are designed to

acquire a better understanding of how quantized charges are transported in solids. Such experiments include the "quantum metrology triangle". Detailed representations of this aspect can be found in references [1], [5] and [11].

In the new SI, the connection between electrical and mechanical units will be established by means of an additional, important metrological experiment that has been developed and optimized over the past few decades: The so-called "watt balance" (another electromechanical apparatus) is used to establish the connection between mass (as a mechanical quantity) and the electrical quantities by comparing mechanical and electrical power (or energy) [1].

6. Conclusion

By establishing the numerical values of the fundamental constants of *elementary charge* and P*lanck's constant*, the importance of electrical quantum effects will continue to increase, both for fundamental metrological applications and for metrological practice. The most direct realization of the ampere can take place by means of electronic circuits that "count" single electrons. The redefinition of the SI will bring the quantum effects of the volt and the ohm (Josephson effect and quantum Hall effect), which have long become established in the field of metrology, into conformity with the SI; as a result, it will become possible to use these quantum effects for the realization of the future SI ampere. However, this latter realization is more indirect, as it is based on the combination of voltage and resistance.

While establishing the numerical values of the defining constants, attention is devoted to ensuring that the quantitative differences between the "new" units and those of the "old" SI are as small as possible. For the electrical units, the redefinition of the SI will lead to very small changes (if any) of approximately one part in 10 million; consequently, the link to the "old" units will not involve any large steps. It is therefore certain that our electricity bills will in no way be affected by the redefinition of the ampere.

Literature

[1] *E. O. Göbel and U. Siegner*; Quantum Metrology: Foundation of Units and Measurements, Wiley-VCH, Weinheim (2015), ISBN 978-3-527-41265-5

[2] *M. Planck*; Über irreversible Strahlungsvorgänge, in Sitzungsbericht der Königlich Preußischen Akademie der Wissenschaften, erster Halbband, (1899) pp. 479–480

[3] Single Charge Tunneling – Coulomb blockade phenomena in nanostructures, in NATO ASI Series B, Vol. 294, Editoren *H. Grabert and M. H. Devoret*, Plenum Press, New York (1992), ISBN 0-306-44229-9

[4] *B. Kästner and V. Kashcheyevs*; Non-adiabatic quantized charge pumping with tunable-barrier quantum dots: a review of current progress, Reports on Progress in Physics **78** (2015) 103901

[5] *J. P. Pekola, O.-P. Saira, V. F. Maisi, A. Kemppinen, M. Möttönen, Y. Pashkin, D. V. Averin*; Single-electron current sources: toward a refined definition of the ampere, Review of Modern Physics **85** (2013) 1421

[6] *D. Drung, C. Krause, U. Becker, H. Scherer, F. J. Ahlers*; Ultrastable low-noise current amplifier: a novel device for measuring small electric currents with high accuracy, Review of Scientific Instruments **86** (2015) 024703

[7] *D. Drung, M. Götz, E. Pesel, H. Scherer*; Improving the traceable measurement and generation of small direct currents, IEEE Transactions on Instrumentation and Measurement **64** (2015) 3021

[8] *L. Fricke, R. Dolata, B. Kästner, F. Hohls, H.-W. Schumacher*; Die Kunst des Elektronenzählens, Physik in unserer Zeit **46** (2015) pp. 70–76

[9] *L. Fricke, M. Wulf, B. Kästner, V. Kashcheyevs, J. Timoshenko, P. Nazarov, F. Hohls, P. Mirovsky, B. Mackrodt, R. Dolata, T. Weimann, K. Pierz, H.-W. Schumacher*; Counting statistics for electron capture in a dynamic quantum dot, Physical Review Letters **110** (2013) 126803

[10] *L. Fricke, M. Wulf, B. Kaestner, F. Hohls, P. Mirovsky, B. Mackrodt, R. Dolata, T. Weimann, K. Pierz, U. Siegner, H.-W. Schumacher*; Self-referenced single-electron quantized current source, Physical Review Letters **112** (2014) 226803

[11] *H. Scherer and B. Camarota*; Quantum metrology triangle experiments: a status review, Measurement Science and Technology **23** (2012) 124010

Counting Atoms for Mass and Amount of Substance

Peter Becker*, Horst Bettin**

Summary

In addition to Planck's constant, the Avogadro constant is an important input quantity for the international efforts to redefine the unit of the mass – the kilogram – with the aid of fundamental constants. Currently, it is proposed to refer the unit of amount of substance, the mole, directly to the Avogadro constant N_A and to use Planck's constant h for the redefinition of the mass unit (in addition to two other constants, i.e. the speed of light c and the hyperfine transition frequency of the cesium atom $\Delta\nu$). At present, the definition of the kilogram is still based on a prototype which is more than 100 years old: the international prototype of the kilogram in Paris. Yet the relations between this kilogram piece and all other national prototypes vary. The masses of these artifacts change (as is, in principle, to be expected for macroscopic objects in interaction with the environment). This is, in the long run, an intolerable state for the definition of the mass unit. However, before a redefinition is possible, the experimenters have to face considerable challenges: They must determine the constants in question with sufficiently small measurement uncertainties.

Until 2003, attempts had been made to determine the Avogadro constant with the aid of natural silicon (which consists of three isotopes). The task to measure the molar mass with a relative measurement uncertainty of smaller than 10^{-7} proved, however, to be impossible to achieve. In the case of highly enriched silicon, the heavy isotopes ^{29}Si and ^{30}Si, however, quasi represent only an impurity which can be determined more exactly by mass spectrometry. Therefore, an international consortium (International Avogadro Coordination, IAC) started, in 2003, a collaboration with research institutes in Russia, in order to have approx. 5 kg of highly enriched silicon (99.99 % ^{28}Si) manufactured as a single crystal and to determine the Avogadro constant by 2010 with a relative measurement uncertainty of approx. $2 \cdot 10^{-8}$.

In 2011, the international consortium composed of metrology institutes and the *Bureau International des Poids et Mesures* (BIPM) under the auspices of PTB succeeded – in spite of extremely high experimental challenges – in "counting" the silicon atoms in a 1 kg silicon sphere of enriched ^{28}Si for the first time with high accuracy. As a result, this experiment furnished a value for the Avogadro constant and – without a loss in accuracy – a value for Planck's constant, because both are coupled with each other via a very exactly known relation [1]. In 2015, a measurement uncertainty of $2 \cdot 10^{-8}$ was reached under improved conditions [2], whereby a not sufficiently perfect spherical shape turned out to be a limiting factor. The whole experiment is described in detail in a special issue of the journal *Metrologia* [3].

According to the requirements of committees of the Metre Convention, at least three independent measurement results – among them results of Planck's constant via watt balance experiments – must be available for the planned redefinition. The value published by the *National Research Council of Canada* (NRC) in 2014 is in very good agreement with the Avogadro value, so that the metrologists decided to start the countdown for the redefinition of the mass unit in 2018.

Introduction

Since the first General Conference on Weights and Measures in 1889, the international prototype of the kilogram has embodied the unit of mass.

For years, experts at the BIPM in Paris have worried about the long-term stability of the international prototype of the kilogram as embodiment of the mass unit. Compared to the mass standards of the national metrology institutes, it has, in the course of the decades, apparently lost approx. 50 µg of its mass. The reasons for this loss are unknown. What is also not known is whether the mass change is mainly due to the international prototype of the kilogram itself or whether these mass differences are due to changes of the national mass standards.

Within the scope of the International System of Units (SI), the kilogram is connected with the definitions of the SI units ampere, candela and mole. This is why it has so far been impossible to realize

* Dr. Peter Becker, former head of the international Avogadro project, e-mail: peter.becker@ptb.de

** Dr. Horst Bettin, Department "Mass - Realization of the Unit", e-mail: horst.bettin@ptb.de

the electric units more exactly in the SI on the basis of the Josephson constant and on the basis of the von Klitzing constant. A definition of the mass unit on the basis of a fundamental quantity of physics would solve these problems [4, 5].

In 1900, Planck's constant was introduced by Max Planck when he described the fundamentals of the black-body radiator, and thus the energy of the harmonic oscillator as quantized, in units of $E = h \cdot \nu$. h has the unit of an action, i.e. kg m^2 s^{-1}, and, thus, a connection to the mass. h, as well as the speed of light c – which is assigned to the unit of length, the meter – are fundamental constants of physics. Two experiments are available to relate the kilogram to natural constants: the so-called watt balance experiment, which links the mass unit up with Planck's constant h, and the Avogadro experiment, which represents a link-up to the atomic mass constant m_u. Both fundamental constants are linked with each other via a relation which is very exactly known, the so-called molar Planck constant ($N_A\ h$ = 3.990 312 7110(18) · 10^{-10} Js/mol, with a relative measurement uncertainty of 4.5 · 10^{-10} (for details, see info box), so that both experiments can ultimately contribute to a redefinition "on a partnership basis". A prerequisite is, however, that they can achieve a relative measurement uncertainty of 2 · 10^{-8}. This condition follows from the minimum requirements for the development of the measurement uncertainties during the metrological dissemination of the mass unit. According to the proposal of an advisory committee of the Metre Convention (CCM: Consultative Committee for Mass and Related Quantities), the Avogadro constant would therefore play a central role both for the redefinition of the kilogram and also for the redefinition of the mole.

For the mole definition, the connection to the Avogadro constant is obvious: An amount of substance of one mole contains exactly N_A particles. Today, the definition says that the molar mass of ^{12}C amounts to exactly 12 g/mol. However, as there is an obvious connection to the mass unit, the molar mass of ^{12}C will, after the redefinition, be affected by an uncertainty.

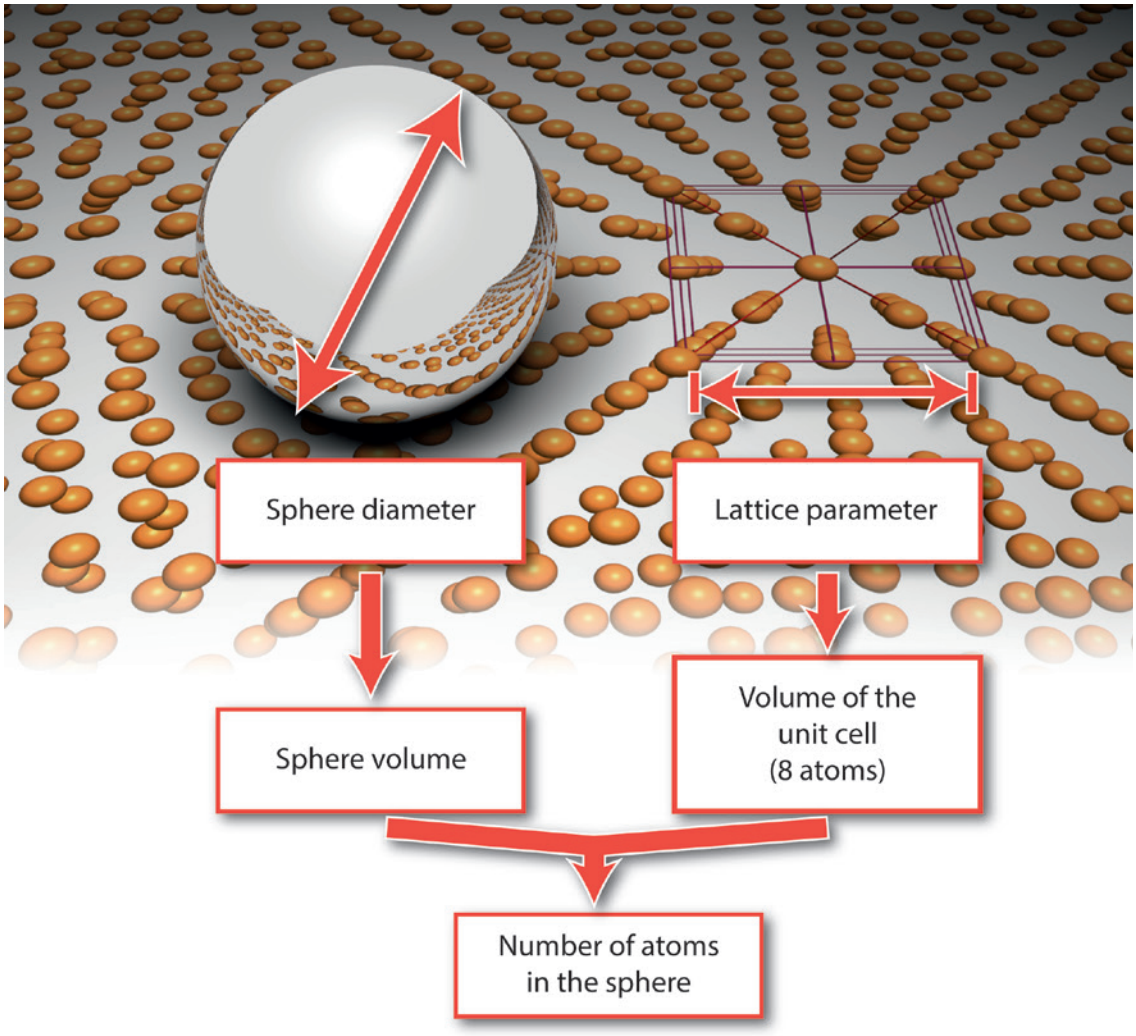


Diagram:
The measurement principle for the Avogadro experiment is quite simple. To determine the number of atoms in the silicon sphere, only the diameter of the sphere and the elementary distance of the atoms in the crystal, the lattice parameter, must be determined. Therefore, the number of the atoms in the whole crystal can be directly determined from the sphere volume and the volume of the unit cell (in the case of the diamond lattice of silicon, it contains eight atoms). Then the sphere is weighed once (i.e. it is compared to the international prototype of the kilogram), thereby determining the mass of a silicon atom. Consequently, the experiment also gives an answer to the question of how many silicon atoms make up a kilogram. The relation of the experiment to the Avogadro constant is quite direct: In addition to the mass of a silicon atom, the chemists also determine, in the case of the Avogadro experiment, the molar mass of the silicon used. The quotient from both is the Avogadro constant. The relation of the experiment to Planck's constant is indirect (see info box): This relation is given via the relation between the atomic mass and the electron mass, which can be measured with great accuracy, for example with the aid of Penning traps. As a result, the Avogadro experiment thus furnishes two natural constants: the Avogadro constant and (indirectly calculated from it) Planck's constant.

The molar Planck constant $N_A h$

The molar Planck constant designates the product of Avogadro constant N_A and Planck constant h. What makes this remarkable is the fact that the value of this product is more exactly known than any of the two constants individually. The reason for this is that there is a fundamental link between the two constants which is known with high accuracy. The **Avogadro constant** is defined as the number of particles which are contained in the amount of substance of one mole. One mole, in turn, consists of as many identical single particles as there are atoms in 12 g ^{12}C, so that the following is valid for the Avogadro constant:

$$N_A \equiv \frac{12\ \mathrm{g}}{m_{^{12}C}}\mathrm{mol}^{-1} \qquad (1)$$

The **atomic mass** constant is also defined via the mass $m_{^{12}C}$ of the carbon isotope ^{12}C:

$$m_u \equiv \frac{m_{^{12}C}}{12} \qquad (2)$$

Thus, a relation between atomic mass constant and Avogadro constant can be established:

$$N_A = \frac{1}{m_u}\frac{1}{1000}\mathrm{kg\ mol}^{-1} \qquad (3)$$

$$M_u \equiv \frac{1}{1000}\mathrm{kg\ mol}^{-1} \qquad (4)$$

is referred to as the **molar mass constant.**

It is required because, in the definition of the Avogadro constant, the unit "gram" is used, whereas the unit "kilogram" is used in the definition of the atomic mass unit. This allows equation (3) to be simplified as follows:

$$N_A = \frac{M_u}{m_u} \qquad (5)$$

However, the connection between Avogadro and Planck constant is ultimately established by the **rest mass of the electron** m_e, as the same can be calculated from other constants on the basis of quantum physics:

$$m_e = \frac{2R_\infty h}{c\alpha^2} \qquad (6)$$

One of these constants is Planck's constant. The other constants are the Rydberg constant R_∞, the speed of light c and the fine-structure constant α. All of these three constants are known with a much higher accuracy than Planck's constant. As long as this is the case, the accuracy with which the numerical values of Planck's constant and of the electron mass can be converted into each other is greater than the accuracy with which they are known themselves. The mass of the electron can be well compared with the atomic mass constant. With this so-called **relative atomic mass of the electron**

$$A_r^e = \frac{m_e}{m_u} \qquad (7)$$

the cycle closes due to the fact that Planck's constant can now not only be converted into the mass of the electron, but also into the atomic mass unit. According to equations (3) and (5), the numerical value of the Avogadro constant can then also be calculated without an additional loss in accuracy.

This can be summarized in the following equation:

$$N_A h = \frac{M_u\, c\alpha^2}{2R_\infty} A_r^e \qquad (8)$$

In the old SI, the molar mass constant M_u and the speed of light c have fixed numerical values, i.e. they are exactly known. The Rydberg constant R_∞ and the fine-structure constant α are, as explained above, known with a much higher accuracy than Planck's constant or the Avogadro constant. As the measurements of the relative atomic mass of the electron are also clearly more exact, the value of the product is also known with a much greater accuracy than the two factors themselves.

Historical facts

The Avogadro constant is named after the Italian earl Amedeo Avogadro who – at the beginning of the 19th century – dealt with atomism, which had been established by Dalton. Dalton had demonstrated that all gases expand in the same way when they are heated. This, concluded Avogadro, could be explained only by the fact that the number of gas particles involved is also always the same: Identical volumes of all gases had to contain – in the case of identical external conditions – the same number of the smallest particles. However, Avogadro did not know at that time how this number could be determined.

Only in 1865 did Josef Loschmidt calculate the particle number per volume with the aid of the mean free path of gas molecules. At that time, this value showed only a small deviation from the value best known today. Forty years later, Albert Einstein tried to indicate the Avogadro constant more precisely on the basis of a novel diffusion equation from measurements carried out on a sugar solution: his values deviated, however, considerably due to a calculation error. In 1917, Robert Millikan reduced the "error" to less than one percent by distinctly improving his famous oil droplet test for the determination of the elementary charge.

Approximately 100 years ago, X-ray diffraction on crystals opened up new experimental procedures for the determination of the Avogadro constant. In the end, however, all attempts to further reduce the accuracy failed due to the number of defects in the crystals as found in nature, as this number was too large. Only with the beginning of the "silicon era" have almost perfect industrially manufactured single crystals been available. The breakthrough came about in 1965 with the development of an X-ray interferometer by Bonse and Hart [6] which allowed lattice distances to be measured with great accuracy in the unit "meter". Thus, the X-ray wavelengths, which were mostly indicated in the so-called X unit and which were known with only low accuracy, were no longer required as a measure. On the basis of this invention, PTB started a project in 1971 for the determination of the lattice plane distance in a silicon single crystal. This project was successfully concluded in 1981 with a relative measurement uncertainty of $6 \cdot 10^{-8}$ [7]. This laid the foundation for an exact determination of the Avogadro constant. In addition to the lattice parameter, the density of this material and its molar mass then had to be determined from the knowledge of the isotopic ratio. In 1992, PTB achieved for the first time a relative measurement uncertainty of $1.1 \cdot 10^{-6}$ for a crystal with natural isotopic composition [8].

Other metrology institutes followed the proposal of Bonse and also started projects for the determination of the Si lattice parameter and the Avogadro constant: In the USA, in Great Britain, in Japan, in Italy and in Russia, similar projects – of different duration – were launched. The National Bureau of Standards, later the National Institute of Standards and Technology (NIST, USA), for example, had already discontinued its experiments by the end of the 1970s, although the results obtained were trend-setting for other metrology institutes. Without having achieved results worth mentioning, the National Physical Laboratory (NPL) in England and the metrology institute Mendeleyev Institute for Metrology (VNIIM) in Russia also abandoned this field of research. To this day, however, the following institutes are involved in the measurements for the determination of the Avogadro constant:

- Physikalisch-Technische Bundesanstalt (PTB),
- National Metrology Institute of Japan (NMIJ),
- National Institute of Metrology (NIM, PR China),
- Istituto Nazionale di Ricerca Metrologica (INRIM, Italy),
- National Institute of Standards and Technology (NIST, USA),
- National Research Council (NRC, Canada),
- Bureau International des Poids et Mesures (BIPM).

However, for a possible redefinition of the kilogram on the basis of these measurements, the measurement uncertainties achieved until then were much too large. A potential for improvement was expected from the use of silicon spheres for density determination and from their optical measurement in special interferometer arrangements. Also Si single crystals, which were specially manufactured for the PTB project by industry (Wacker-Siltronic Burghausen) in a close collaboration, were suggested to further reduce the influences of crystal defects. To improve the coordination of the national efforts, the institutes involved cooperated from 1995 on under the leadership of PTB in a joint working group within the scope of the CCM. Whereas at least two independent determinations were available for the most important physical quantities, only the results of one laboratory (Institute for Reference Materials and Measurements, IRMM, of the European Commission) were available for the determination of the molar mass. In 2003, a relative measurement uncertainty of $3 \cdot 10^{-7}$ was achieved which could not be reduced any further: The determination of the Si isotopic ratios had reached its technical limits [9]. Therefore, the investigations on natural

silicon were terminated. An overview of the determination of the Avogadro constant until 2000 can be found in [10].

Unfortunately, besides the insufficient accuracy for a redefinition of the kilogram, the observed deviation of the measurement value of approx. $1 \cdot 10^{-6}$ relative to the results obtained with other methods also caused quite a headache for the researchers. Was the deviation due to miscounting? Or were there unrecognized inconsistencies in the system of the physical constants? The possibility of using enriched silicon-28 on a large scale for sample preparation gave new impetus to the Avogadro project. Estimates showed that the uncertainties for the measurement of the isotope composition would be considerably reduced, but the researchers were forced to also repeat all other investigations on this material. It was not for the first time that the possibility of using enriched silicon was discussed. Up to that point, however, this had always failed due to the insufficient perfection of the single crystal [11].

In 2004, scientists from the metrology institutes of the USA, Great Britain, Australia, Italy, Japan and the European Commission again joined forces under the leadership of PTB to jointly determine the Avogadro constant as exactly as never before with the aid of an almost perfect mono-isotopic silicon single crystal. A first milestone was reached in 2011 with a relative measurement uncertainty of $3 \cdot 10^{-8}$: All required measurement methods were installed, analyzed and tested. Only the insufficient perfection of the silicon spheres reduced the success. Four years later, the objective $2 \cdot 10^{-8}$ was achieved after extensive improvements had been made to the silicon surfaces.

How are atoms exactly counted?

As the individual counting of 10^x atoms ($x \approx 23$) is unrealistic, the perfection of a single crystal is made use of. Such single crystals can be manufactured from silicon with only a very small number of impurities and lattice defects, as the manufacturing technology in the semiconductor industry is very advanced. Consequently, silicon proves to be an ideal candidate for this experiment. For the determination of the Avogadro constant, i.e. of the number of atoms in a mole, first the volume V_{sphere} and, with an X-ray interferometer, the volume V_{atom} of an atom are determined on a silicon sphere (see Figure 1). This provides knowledge of the number of atoms in the sphere. If the molar mass M_{mol} and the mass of the sphere m_{sphere} are determined in addition according to equation (1), the number of atoms per mole is obtained. The determination of the number of the atoms in exactly one kilogram of the same substance then only requires a small calculation step.

$$N_A = \frac{V_{sphere}}{V_{atom}} \cdot \frac{M_{mol}}{m_{sphere}} = \frac{V_{sphere} \cdot M_{mol}}{m_{sphere} \cdot \sqrt{8} \cdot d^3} \qquad (1)$$

The *d* in equation 1 is the spacing of the (220) lattice planes in the Si crystal

All measured quantities must, of course, be traced back to the SI base units. Therefore, the volume of the silicon sphere is determined by interferometric measurements of the sphere diameter (see Figure 3). This provides a complete surface topography of the sphere. In the first step, the spheres were polished at the Australian Centre for Precision Optics (ACPO) and exhibited, at best, a deviation of only a few 10 nanometers from a perfect shape.

The silicon sphere is covered with a thin oxide layer which is formed during the polishing process. This layer, which is only a few nanometers in thickness, is nevertheless "disturbing". It consists almost exclusively of silicon dioxide, SiO_2. It must not be taken into account in the "counting", and its geometry and mass must, therefore, be characterized. The mass of the silicon sphere is linked up to the national prototype of the kilogram by weighing. This process is necessary for the seamless transition from the old to the new definition. As the two masses (or material measures) have different densities and surfaces, extensive buoyancy and sorption corrections are required. Volume and mass determinations of the Si material must then – in addition – be corrected with respect to the oxide layer.

The atomic volume is determined from the lattice plane spacings of the crystal lattice with the aid of a combined optical and X-ray interferometer. All mass and length determinations are carried out in vacuum at 20 °C because the crystal lattice depends on the pressure and on the temperature. Due to the Si expansion coefficient of approx. $2.6 \cdot 10^{-6}$ / K, the crystal temperature must be known with an accuracy of better than 0.001 K and must be the same in any laboratory involved worldwide.

The molar mass is calculated from the atomic masses of the three isotopes ^{28}Si, ^{29}Si and ^{30}Si which occur in natural silicon, and from their relative frequencies which have before been determined by mass spectrometry. The atomic masses are linked up with the mass of the carbon isotope ^{12}C, the reference value for all atomic masses, by means of Penning traps.

The experiment

Not all measurands were determined by all parties involved in the experiment. Instead, there was a division of labor: The sphere diameters were measured at the NMIJ, at the NMI-A (National Measurement Institute, Australia) and at PTB with the aid of specially developed optical interferometers, using the wavelength of an iodine-stabi-

lized helium-neon laser as a measure. The lattice parameter was determined by the INRIM with an X-ray interferometer in the unit "meter". In a separate comparison experiment[1] at NIST, the lattice parameters of natural silicon and of enriched silicon were compared and verified with calculated data. Due to the optical wavelength definition, this experiment was carried out in vacuum, and due to the temperature-dependent expansion of Si, related to 20 °C. The sphere mass was compared in mass comparators with the national or international mass standards at the BIPM, at PTB and at the NMIJ. The molar mass was determined at PTB with the aid of a mass spectrometer with inductively coupled plasma (ICP-MS) and calibrated by means of synthetically manufactured sample mixtures. As the determination of the molar mass has so far represented the largest uncertainty factor, this experiment was also repeated in other laboratories within the scope of an interlaboratory comparison. In addition, a great number of spectrometric investigations were performed for the determination of the crystal defects.

[1] In the experiment at NIST, only the lattice parameters were compared. The differences are largely independent of temperature. The measurements at NIST were carried out at 21 °C, the measurements at INRIM at 20 °C.

The material

The highly enriched material used here for the measurements was manufactured in cooperation with a Russian consortium under the leadership of the Development Office for Centrifuges (CENTROTECH) of the Nuclear Ministry in St. Petersburg and at the Institute of the Chemistry of High-Purity Materials (IChHPS) in Nishni Novgorod. Gaseous silicon tetrafluoride, SiF_4, was enriched in several cascades of centrifuges, after which the gas was converted into silane, SiH_4, chemically purified and separated as polycrystalline silicon. Finally, in 2007, a monocrystal was manufactured at the Leibniz Institute for Crystal Growth (IKZ) in Berlin (see Figure 1). In order to determine the molar mass with sufficient accuracy, the enrichment of the ^{28}Si isotope was to reach at least 99.985 %. For this purpose, all required technological steps were taken, and the mandatory analytical investigation methods were described for all partners in a so-called "road map". The degree of enrichment of the single crystal of more than 99.995 % and its chemical purity of approx. 10^{15} cm^{-3} impurity atoms exceeded the required specifications by far [12]. For the measurement of the required crystal parameters, two spheres of approx. 1 kg each, an X-ray interferometer and a great number of special crystal samples were manufactured from the crystal which had a mass of 4.7 kg.

In 2008, two polished 1 kg spheres (with the designations AVO28-S5 and AVO28-S8) were handed over to the Avogadro project in Sydney. They had

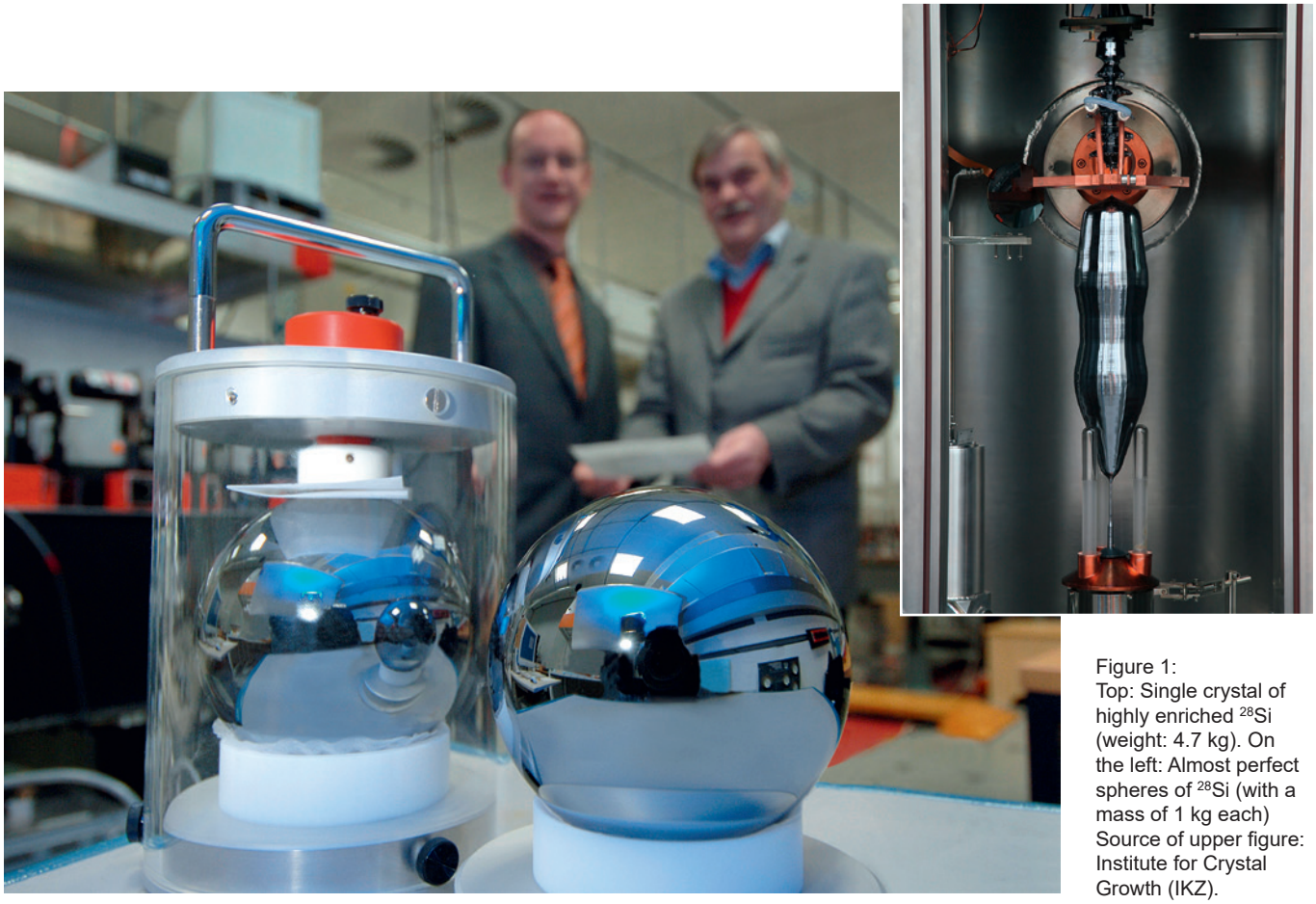


Figure 1:
Top: Single crystal of highly enriched ^{28}Si (weight: 4.7 kg). On the left: Almost perfect spheres of ^{28}Si (with a mass of 1 kg each)
Source of upper figure: Institute for Crystal Growth (IKZ).

been precision-engineered, mainly in manual operation, at the ACPO in Sydney. By the middle of 2010, the first measurement cycle had been completed, and a relative measurement uncertainty of $3 \cdot 10^{-8}$ was achieved for the determination of the Avogadro constant which, after a correction of the surface quality of the two spheres, could be reduced to $2 \cdot 10^{-8}$ by 2015.

The crystal contaminations (see Table 1) such as carbon, oxygen and boron, were measured at PTB with Fourier infrared spectrometers. Infrared spectrometry is a fast and uncomplicated method for the determination of impurity atoms in crystalline silicon. Impurities from the fourth and sixth main group, such as carbon and oxygen, can be measured jointly at cryogenic temperatures (7 K) [1–3]. Whereas oxygen on interstitial sites leads to an enlargement of the crystal lattice, carbon occupies the places of silicon atoms in the crystal lattice and contracts the lattice structure. As calibration samples of known content are not available for mono-isotopically enriched silicon, a model has been developed with the aid of which the calibration factors of silicon can be transmitted to a highly enriched material with natural isotopic distribution. Vacancies were investigated by positron annihilation at Halle University, hydrogen with Deep Level Transient Spectroscopy (DLTS) at the Technical University of Dresden, and possibly occurring lattice deformations with high-resolution X-ray topography at the Photon Factory KEK in Japan. The lattice deformations caused by defects were taken into account in the measurements of the lattice parameter, the sphere mass and the sphere volume. The maximum carbon concentration of $2 \cdot 10^{15}$ cm^{-3} which is required for the envisaged relative measurement uncertainty of $2 \cdot 10^{-8}$ was not exceeded.

Sphere surface*

On the silicon spheres, an oxide layer is formed (as on every silicon surface), and a carbonaceous contamination layer and adsorbed water layers cannot be avoided (Figure 2). The whole layer system on the surface can reach a thickness of up to 3 nm. In investigations performed on other silicon spheres, an even slightly thicker SiO_2 layer was generated by thermal oxidation, so that its thickness could be determined from oscillations observed in X-ray reflectometry (XRR). However, the thickness of the thermal oxide layer depends strongly on the crystal orientation. For the spheres of highly enriched silicon, thermal oxidation was therefore omitted and the XRR measurements were carried out on SiO_2 reference layers on silicon wafers to exactly determine their thicknesses in the range around 10 nm.

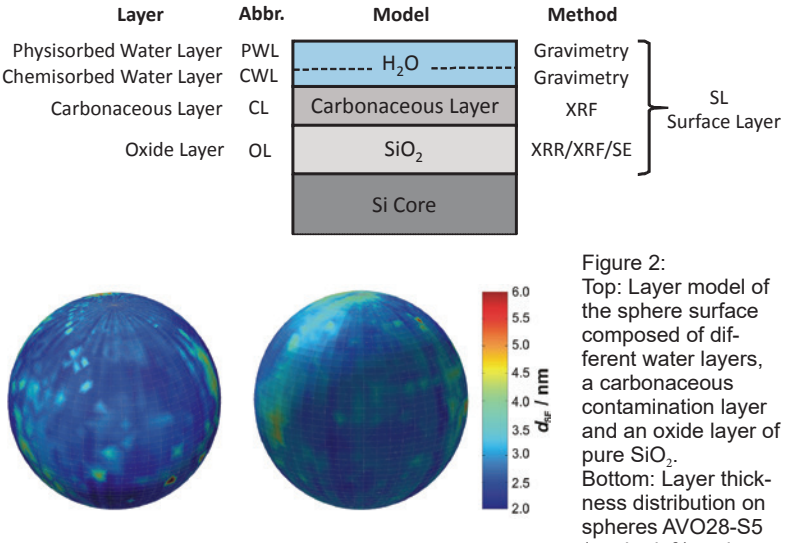


Figure 2:
Top: Layer model of the sphere surface composed of different water layers, a carbonaceous contamination layer and an oxide layer of pure SiO_2.
Bottom: Layer thickness distribution on spheres AVO28-S5 (on the left) and AVO28-S8 (on the right) measured with spectral ellipsometry [13]. The values indicated are uncorrected values.

For the measurements with mono-chromatized synchrotron radiation at BESSY II, several energies in the range of the oxygen-K-absorption edge were selected so that the optical constants of silicon and oxide differ significantly and the contributions of the oxide layer and of the contamination layer can be separated. At the same time, the intensity of the oxygen-fluorescence line was measured on these layers with X-ray fluorescence analysis (XRF) at an excitation energy of 680 eV.

In the case of XRF measurements on three marked points on the spheres, the intensity of this line was also measured, so that the oxide layer on the sphere follows from the intensity ratio and the thickness of the reference layer determined by XRR. Corresponding measurements were carried out with a carbon reference layer at an excitation energy of 480 eV, in order to also determine the mass per area of the carbon-containing contamination layer. Based on the layer thicknesses measured at the three points, the whole layer thickness topography of the spheres was recorded with spectral ellipsometers at PTB (Figure 2) and at the NMIJ.

* Co-author of this chapter:
Dr. Michael Krumrey, Working Group "X-ray Radiometry", e-mail: michael.krumrey@ptb.de

Defect	Unit	AVO28-S5	AVO28-S8	XINT
Carbon	10^{15} cm^{-3}	0.40(5)	1.93(19)	1.07(10)
Oxygen	10^{15} cm^{-3}	0.283(63)	0.415(91)	0.369(33)
Boron	10^{15} cm^{-3}	0.011(4)	0.031(18)	0.004(1)
Vacancies	10^{15} cm^{-3}	0.33(11)	0.33(11)	0.33(11)

Table 1:
Point defect density in the two spheres (AVO28-S5, AVO28- S8) and in the X-ray interferometer crystal (XINT).

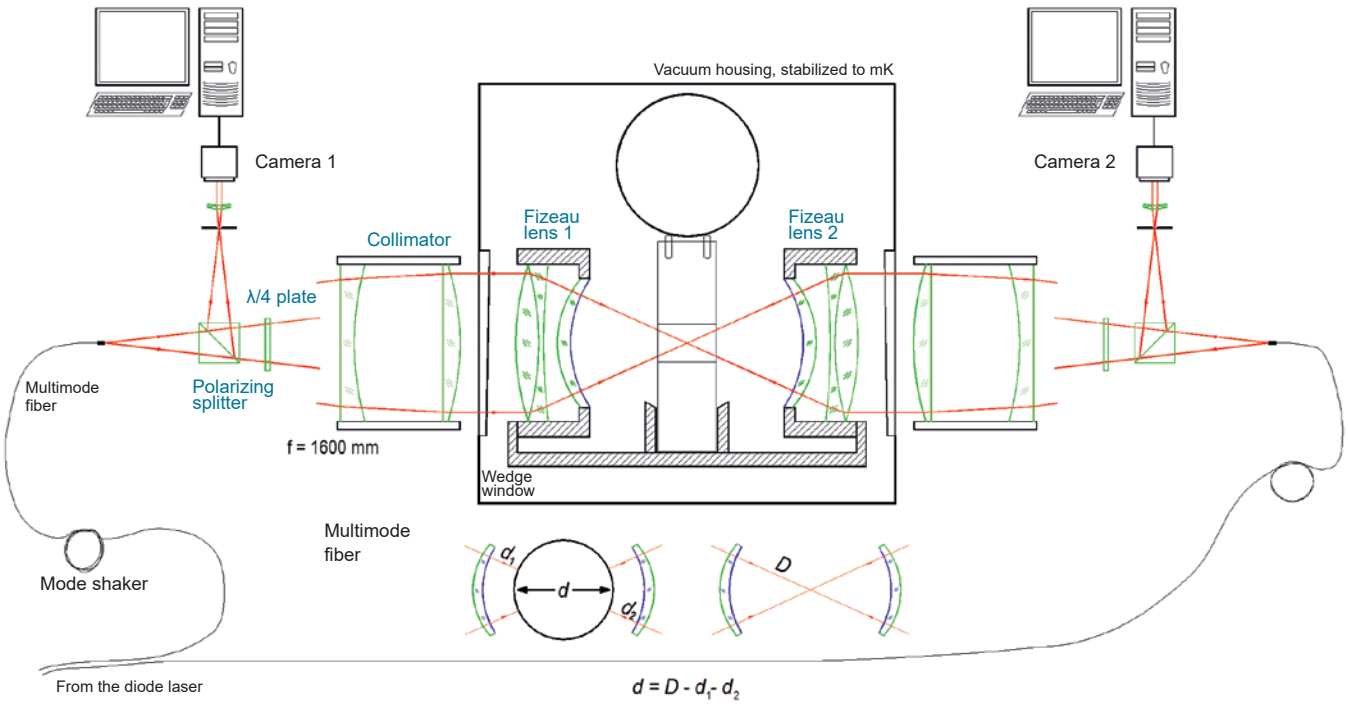


Figure 3: Schematic set-up of PTB's sphere interferometer. The laser light is coupled from both sides into the two objectives (composed of collimator and Fizeau lenses) via optical fibers. A lifting mechanism moves and rotates the spheres in the interferometer. The sphere diameter is determined by distance measurements between the objectives (with and without sphere).

Additional measurements with X-ray photoelectron spectroscopy (XPS) were performed at the Swiss metrology institute METAS in Bern, in order to determine the stoichiometry and possible contaminations of the surface layer. It turned out that during the first polishing process at the ACPO, the metals nickel and copper from the polishing tools had been bound as silicides and polished into silicon regions close to the surface. Signs of lattice deformations close to the surface were not found, however.

After completion of the first measurement campaign, the two spheres were liberated from their surface layers by a so-called Freckle etching process and polished again without additional contamination. For this purpose, PTB has manufactured special machines for polishing the spheres to avoid the influence of manual work. For the newly polished spheres AVO28-S5c and AVO28-S8c, layer thicknesses of 1.76 nm and 1.83 nm and a layer mass of only 78 µg and 86 µg were obtained compared to more than 200 µg before the new polishing process. The measurement uncertainties could not yet be considerably reduced compared to the uncertainties obtained before (see Table 2).

As the data published in 2011 and 2015 show, the measurement uncertainty which follows from these considerations still makes the second largest contribution to the total uncertainty budget of the Avogadro constant. Observations of the layer thickness over many months did not indicate any significant time-dependent changes of the layer.

Currently, a new apparatus has been put into operation at PTB which allows any point on the sphere surface to be investigated with XRF and XPS.

Sphere volume*

The volume of the silicon spheres is derived from the diameter measurements by means of optical interferometry. For this purpose, the spheres are placed inside the mirrors of an interferometer.

Table 2: Thickness and mass of the layers on the sphere surfaces.

Sphere	Lab.	Layer thickness / nm	Layer mass / µg
AVO28-S5c	PTB	1.79(24)	79.5(10.9)
	NMIJ	1.64(33)	70.4(17.7)
	mean value	1.76(23)	77.7(10.0)
AVO28-S8c	PTB	1.94(22)	92.2(10.2)
	NMIJ	1.41(31)	60.0(16.3)
	mean value	1.83(28)	85.5(14.8)

* Co-author of this chapter: Dr. Arnold Nicolaus, Working Group "Interferometry on Spheres", e-mail: arnold.nicolaus@ptb.de

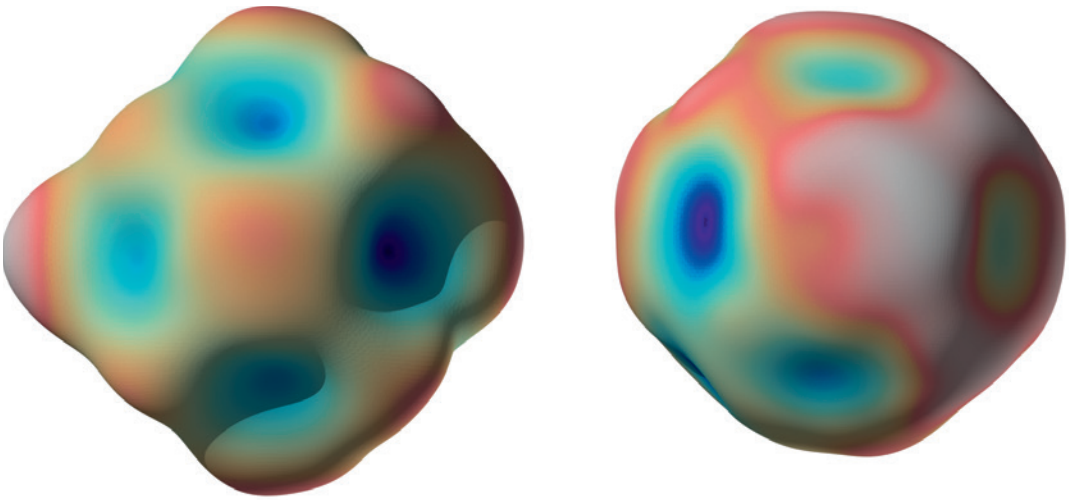


Figure 4:
Surface topography of the ^{28}Si-spheres AVO28-S5c, on the left, and AVO28-S8c, on the right. The deviations from an ideal sphere are shown (in a strongly increased way).

Two different interferometer types are used: In the case of the PTB set-up, spherical surfaces are used, whereas in the case of the interferometers at the NMIJ, plane surfaces are used (and at the NMI-A, *were* used). Figure 3 schematically shows the PTB set-up and the measurement method.

This interferometer uses spherical waves and can therefore determine the topography of the diameters with great accuracy (a function which is the only one of its kind in the world). If the results are applied to the so-called stitching calculation method, one obtains the "true" shape of the sphere from the measured diameters (Figure 4) [14].

Temperature and temperature constancy as well as the shape of the wavefronts are important influence quantities. A temperature constancy of a few millikelvins and a temperature measurement uncertainty of less than 1 mK are achieved. The spheres are covered with an oxide layer which influences the optical phase shift of the light wave on the surface and the diameter measurement, so that only apparent diameters are measured. To take

Material	Layer	*n*	*k*
H_2O	CWL, PWL	1.332(10)	$1.54(1.00)\cdot10^{-8}$
C_mH_n	CL	1.45(10)	0(0.1)
SiO_2	OL	1.457(10)	0
Si	Solid	3.881(1)	0.019(1)

Table 3:
The optical constants *n* (refractive index) and *k* (absorption index) of the sphere surface layers (CWL, PWL, CL, OL – see Figure 2).

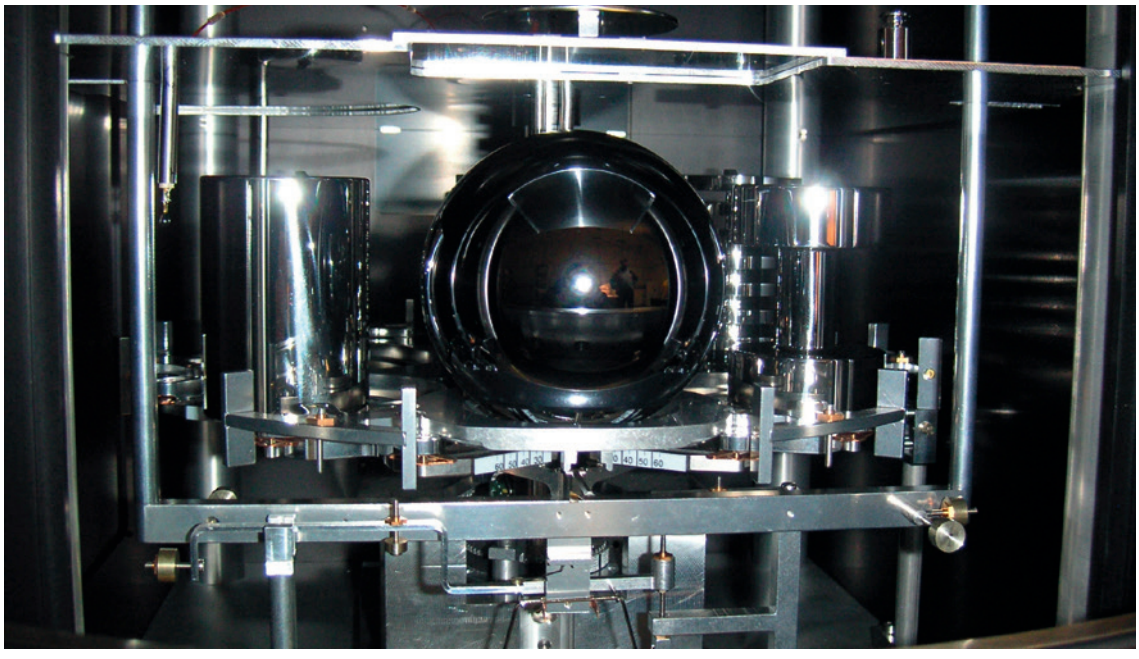


Figure 5:
View into the weighing space of a mass comparator with buoyancy artifacts (left and right), Si-sphere (centre) and sorption bodies.

this phase shift into account, the optical constants, i.e. the refractive index n and the absorption factor k of the layer model, must be known (see Table 3).

An overall measurement uncertainty of 0.007 mm^3 was obtained for the AVO28-S8c sphere volume. Figure 4 shows the surface topography of the two ^{28}Si spheres: At a diameter of 93.7 mm, the deviation from the spherical shape amounts to approx. 70 nm for the AVO28-S5c and to only 40 nm for sphere S8c. The "cube shape" which can still weakly be recognized is caused by the cubic shape of the crystal lattice.

Comprehensive simulations of the wavefronts and their phases at the outlet of the sphere interferometer were performed for both the plane-wave interferometers and for the Fizeau interferometer. The respective phase corrections and the influence of maladjustments in the beam path on the measurement result were determined in addition [15]. Measurements of selected sphere diameters with all interferometer types showed very good agreement within the scope of the measurement uncertainty. The number of selected diameters which is – at least – required for the volume calculation was also investigated: From a number of 150 randomly selected diameters onwards, there were no more significant deviations. If the uncertainty budget of the volume measurements is taken into account, the deviations from the ideal spherical shape make, at more than 80 %, the largest contribution.

Sphere mass*

* Co-author of this chapter: Dr. Michael Borys, Working Group "Realization of Mass", e-mail: michael.borys@ptb.de

Both for the present determination of the Avogadro constant and of Planck's constant, and for the future realization and dissemination of the unit, mass comparisons between silicon spheres and national and international mass standards with uncertainties in the range of a few micrograms (relative 10^{-9}) are required.

Due to the great density differences between silicon spheres, kilogram prototypes of platinum-iridium and conventional steel standards, the smallest uncertainties have so far been obtained under vacuum conditions, i.e. without the buoyancy corrections required in air.

For the mass determinations in air and under vacuum conditions, high-resolution 1-kg mass comparators are used which are accommodated in a housing which can be evacuated. Compared to their maximum capacity, these comparators have only a small weighing range between 1 g and 2 g which is, however, at 0.1 µg, very large and can be resolved with linearity deviations of ≤ 2 µg. To keep the weighing differences to be determined as small as possible, smaller mass standards (derived from a kilogram prototype) are used as auxiliary weights.

The measurements are carried out in vacuum in a pressure range between 10^{-4} Pa and 0.1 Pa and in air (10^5 Pa) under pressure-stable conditions and temperature variations of a few millikelvins. Hereby, relative standard deviations of $\leq 2 \cdot 10^{-10}$ can be achieved.

As transfer standards between the mass standards in air and the silicon spheres in vacuum, special sorption artifacts are used [16–18]. Sorption artifacts consist of two bodies of the same mass which have been manufactured from the same material with identical surface properties, but which show a large surface difference. From the change of the mass difference between the bodies due to the transition between air and vacuum, the change of the sorption coefficient and, thus, the mass change of the standards can – if the surface difference is known – be determined accurately to a few micrograms.

However, for the most exact mass determinations in air, buoyancy artifacts are used [19]. Buoyancy artifacts are composed of two bodies of the same mass which have been manufactured from the same material with identical surface properties and which show a large volume difference. From the change of the mass difference between the bodies in air and in vacuum, the air density required for the buoyancy correction can – if the volume difference is known (which is usually determined by hydrostatic weighing) – be determined with relative standard uncertainties in the range of $2 \cdot 10^{-5}$. Currently, this influence limits the uncertainty with which mass comparisons between silicon spheres and kilogram prototypes can be performed in air to approx. 10 µg (relative $1 \cdot 10^{-8}$).

For the dissemination of the future definition of the kilogram realized with the aid of silicon spheres in vacuum, mass comparisons between primary standards of silicon and secondary standards of platinum-iridium or steel in air are, in principle, not required.

If sorption artifacts of steel and platinum-iridium are used as transfer standards between the silicon spheres in vacuum and secondary standards of platinum-iridium or steel in air, the influence of the buoyancy correction is negligible if mass comparisons between sorption artifacts and mass standards of the same density are performed in air, and the influence of the sorption correction is limited to a few micrograms. Thus, the unit realized with the aid of silicon spheres in vacuum can be disseminated to kilogram prototypes of platinum-iridium and to conventional steel standards in air with an almost identical uncertainty.

Figure 5 shows the weighing room of a mass comparator with a silicon sphere, buoyancy artifacts and sorption artifacts. International mass comparisons in air and in vacuum showed very good agreement. From measurements at the BIPM,

the NMIJ and at PTB, the mass of two ^{28}Si spheres in vacuum could be determined with a standard uncertainty of approx. 4 µg (relative $4 \cdot 10^{-9}$) [2, 20].

Before a mass determination is carried out, the spheres are subjected to a prescribed cleaning procedure. Due to different sorption layers, the difference between the sphere mass in air and the sphere mass in vacuum amounts to approx. 10 µg. After deduction of the mass of the surface layers and correction of the vacancies and contaminations of the crystal, the mass of the pure silicon sphere searched for can finally be determined with a relative standard uncertainty of $\leq 1.6 \cdot 10^{-8}$ (see Table 6).

Molar mass – isotopic composition*

The method of gas mass spectrometry used so far for the determination of the molar mass of natural silicon requires sophisticated sample preparation in order to convert the crystalline silicon into a SiF_4 gas. In addition, it turned out that, in the preparation of measuring and calibration samples, an unexpectedly high contamination of natural silicon significantly falsified the measurement result.

The measurement of the molar mass of the silicon requires the determination of the fractions of the three isotopes ^{28}Si, ^{29}Si and ^{30}Si. As approx. a million times more ^{28}Si than ^{30}Si is contained in the silicon used, the metrological obstacles which had to be overcome to achieve the required measurement uncertainty have long been insurmountable. In 2008, PTB therefore developed the concept of isotope dilution mass spectrometry with a "virtual element" (VE-IDMS) [21]. Here, the sum of ^{29}Si and ^{30}Si is regarded as contamination in the total silicon and determined without measuring ^{28}Si. For this purpose, the real silicon sample is mixed with practically pure ^{30}Si in such a way that a ^{30}Si/^{29}Si isotopic ratio close to 1 is adjusted which can, in turn, be determined with a small measurement uncertainty. The measurements of the isotopic ratios were corrected with the aid of a novel experimental and mathematical calibration scheme [22] so that – for the first time – a closed calculation of the measurement uncertainty was possible by using a complete analytical model equation. The first use of multicollector ICP mass spectrometry – instead of the gas mass spectrometry which had been used for decades – has brought about essential metrological advantages: In addition to the advantage of a reduced measurement uncertainty, the ubiquitous and interfering natural silicon can, for the first time, be measured and thus be reasonably corrected in the mass spectrometer and in the chemicals used. Also for the selection of the solvent, the change from the acid media which had so far been used (hydrofluoric acid) to alkaline media (first: caustic soda, now: tetramethylammonium hydroxide, TMAH) [23] has made a considerable contribution to the reduction in the measurement uncertainty.

In the past several years, several metrology institutes (NRC, NIST, NMIJ and NIM) have theoretically tested, experimentally reproduced, and successfully applied this procedure (developed by PTB) for the determination of the molar mass, whereby consistent results were obtained (see Table 4). In the past several months, the isotopic homogeneity of the whole crystal has also been confirmed at PTB with this procedure, by measuring a total of 14 single samples which were systematically distributed over the crystal. From these measurements, the currently most reliable molar mass of $M = 27.976\,970\,12(12)$ g/mol has been determined (with an associated combined uncertainty of $u_{rel}(M) = 4.4 \cdot 10^{-9}$) [24]. This implies a reduction of the measurement uncertainty by the PTB procedure by a factor of 100 in the past 10 years.

Measurements at different crystal places confirmed an excellent homogeneity of the crystal in the isotopic composition (see Table 4).

Lattice parameters**

Figure 6 shows the principle of the X-ray scanning interferometer. It consists of three equidistant crystal lamellas (approx. 1 mm in thickness) S, M and A. By Bragg reflection of the X-rays on lamellas S and M, a standing wavefield with periodicity of the lattice plane spacing is formed at the place of lamella A. If lamella A is shifted through this wavefield as indicated, the lattice planes can be counted due to the occurring Moiré effect between wavefield and lattice.

Knowledge of the X-ray wavelength is not required. The displacement is measured optically with a laser interferometer. The displacement length of the lamella amounts to approx. 1 mm, without tilting in the subnanorad range. Auxiliary interferometers and capacitive sensors ensure – fed back online – an alignment that is always perfect. For a good interference contrast, the three lamella thicknesses must agree very well in the µm range.

* Co-authors of this chapter:
Dr. Axel Pramann, Working Group "Inorganic Analysis", e-mail: axel.pramann@ptb.de
Dr.-Ing. Olaf Rienitz, Working Group "Inorganic Analysis", e-mail: olaf.rienitz@ptb.de

** Co-author of this chapter:
Dr. Ulrich Kuetgens, Working Group "X-ray Optics" e-mail: ulrich.kuetgens@ptb.de

Table 4:
Molar mass of the enriched Si material, measured by means of a modified IDMS at different laboratories (PTB: Physikalisch-Technische Bundesanstalt, NIST: National Institute of Standards and Technology - USA, NMIJ: National Metrology Institute of Japan).

Sample	Year	M/(g/mol)
PTB-4-1 (S5)	2015	27.97697029(14)
NIST-5-1 (S5)	2014	27.976969842(93)
NMIJ-5-1 (S5)	2014	27.97697010(22)
PTB-8 (S8)	2014	27.97697020(17)
NIST-8-1 (S8)	2014	27.976969745(57)
NMIJ-8-1 (S8)	2014	27.97697014(21)
PTB-9-1 (S8)	2015	27.97697008(11)

For this purpose, PTB has established an ultra-precision laboratory for silicon processing within the scope of the Avogadro project. The interferometer crystal was manufactured with the aid of diamond grinding tools on a precision 3-coordinate milling machine; the manufacturing tolerances were in the range of a few µm. The lattice parameter was measured absolutely at the INRIM. Comparison measurements of lattice plane spacings of different ^{28}Si samples and a natural Si-crystal at the X-ray crystal diffractometer of NIST showed the following results: The different crystal ranges are very homogeneous in the lattice parameter and differ only by the influence of the different defect densities. The difference of the lattice plane spacings of enriched and natural crystal corresponds very well to the values calculated from the isotopic ratios [25]. Compared to the first measurement campaign in 2011, the following aspects of the measuring apparatus have been improved: The 633 nm diode laser so far used was replaced by a 532 nm Nd:YAG laser, the temperature measurement was recalibrated, and the interferometer crystals were cleaned from surface contaminations. Lamella A was measured at both its front side and its rear side in order to recognize possible crystal deformations due to fastening or gravity. Influences of temperature and contaminations were also corrected. An overall measurement uncertainty of $0.34 \cdot 10^{-18}$ m was achieved.

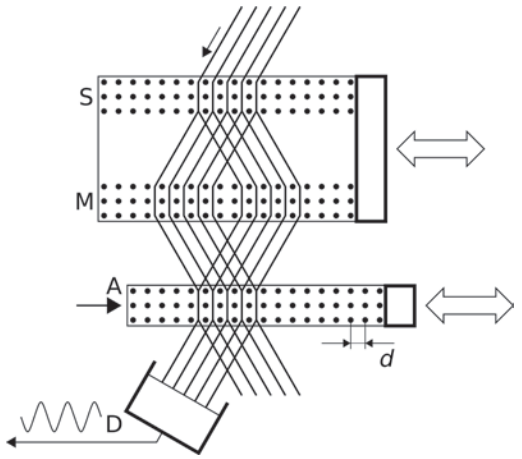


Figure 6:
Principle of the X-ray scanning interferometer. The beam path corresponds to that of an optical Mach-Zehnder interferometer.

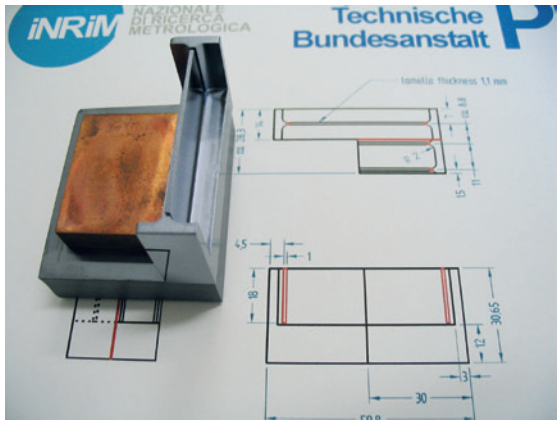


Figure 7:
Analyzer lamella A of the X-ray scanning interferometer with copper block for temperature determination.
Dimensions:
Length: 50 mm
Width: 15 mm
Height: 30 mm.

Table 5:
Determination of the Avogardo constant with enriched Si: simplified uncertainty budget for the AVO28-S5c sphere.

Quantity	Relative uncertainty/10^{-9}	Contribution/%
Molar mass	5	6
Lattice parameter	5	6
Surface layer	10	23
Sphere volume	16	59
Sphere mass	4	4
Crystal defects	3	2
Total	21	100

The latest "counting result"

For all measured parameters and for the total result, the measurement uncertainties were determined with the aid of the Guide to the Expression of Uncertainty in Measurement (GUM). For the most important parameters, at least two independent measurements were carried out for reasons of redundancy. Independent determinations of the molar mass were concluded at the American NIST, the Canadian NRC and in other laboratories, and they support the PTB results.

In a further experiment, the density differences of the two spheres derived from volume and mass measurements were checked relative $1 \cdot 10^{-8}$: When immersing the Si spheres into a liquid of the same density, they float. Smallest changes of the density cause an upward or downward movement of the spheres, which can be compensated for by a change of the hydrostatic pressure on the liquid. This flotation device achieves a relative measurement uncertainty of $1 \cdot 10^{-8}$ and serves to check the density homogeneity of the ^{28}Si crystal [26]. Table 5 contains a list of the contributions to the uncertainty budget. The largest contributions are caused by the characterization of the surface and the determination of the sphere volume. Table 6 shows the final results of molar mass, density and lattice parameter, derived separately for the two spheres. Within the scope of the redefinition of the kilogram, the weighted mean value $N_A = 6.022\ 140\ 76(12) \cdot 10^{23}\ \mathrm{mol}^{-1}$ represents one of the currently most exact input values. Figure 8 shows a comparison of the most exact measurements of the Avogadro constant which are at present available, whereby the results of the watt balances were converted with the aid of the relation $N_A\, h = 3.990\ 312\ 7110(18) \cdot 10^{-10}$ Js/mol. The most exact value of the Avogadro constant obtained from watt balance measurements agrees very well with the value published here.

Outlook

In addition to the metrological applications of the determination of the Avogadro constant for the realization and dissemination of the units kg and mol, silicon spheres as a single-crystalline material are particularly well suited for stability investigations of the international prototype of the kilogram: It is planned to include several Si spheres into the so-called *Ensemble of Reference Mass Standards* at the BIPM, to link them up to primary standards of different realizations (Avogadro and watt balance experiments) and to use them for the dissemination of the unit.

Other possible applications of the enriched ^{28}Si material are based on the almost complete absence of the isotope ^{29}Si with its nuclear spin and, due to the isotopic purity, the lacking broadening of spectral lines. For example, on ^{28}Si crystals doped with Cu, the valence states of Cu in Si were successfully re-described [27], and ^{28}Si was successfully used as a basic material for the development of novel quantum information systems [28]. In addition, ^{28}Si is being considered for use as a support for graphene [29] in this field of application. Further potential applications of the material are based on the reduced thermal conductivity, in particular in the case of low temperatures [30].

Experiments for the reduction of the relative measurement uncertainty to $1 \cdot 10^{-8}$ and below by manufacturing improved surface qualities of spheres and sphere objectives are to be concluded in the medium term. For this purpose, further silicon spheres of enriched silicon are to be manufactured in the next few years with a degree of enrichment of 99.998 %.

Based on the measurement uncertainty of $2 \cdot 10^{-8}$ which has now been achieved within the scope of the Avogadro experiment, the dissemination of the kilogram would, in the case of a redefinition, also start with a measurement uncertainty of 20 µg. This value complies with the minimum requirements of the *Consultative Committee for Mass and Related Quantities* (CCM) [31]. Before a redefinition, these requirements must be met to avoid negative impacts on the realization and dissemination of the mass unit [32].

The results described here are an essential part of a project which is very important for metrology, i.e. for the redefinitions of the units. They are, therefore, already a milestone for stability investigations of the current mass standard: As a drift of the mass scale of at least 0.5 µg per year is assumed, this drift could, with the aid of Si spheres, for the first time be observed within a period of 10 years. For a redefinition of the mass unit in 2018, the CCM demands three independent determinations of Planck's constant. This condition is fulfilled with the results of the Avogadro project, of the NRC watt balance and the new watt balance of NIST (NIST-4). Therefore, from PTB's point of view, nothing would stand in the way of a redefinition of the mass unit "kilogram".

Table 6:
Molar mass, lattice parameter, mass and volume of the two spheres AVO28-S5c and AVO28-S8c and the Avogadro constant calculated from them.

Quantity	Unit	AVO28-S5c	AVO28-S8c
M	g/mol	27.97697009(15)	27.97697009(15)
a	pm	543.0996219(10)	543.0996168(11)
V	cm^3	430.8912891(69)	430.7632225(65)
m	g	999.698359(11)	999.401250(16)
$\rho=m/V$	kg/m^3	2320.070943(46)	2320.070976(51)
N_A	10^{23} mol^{-1}	6.022 140 72(13)	6.022 140 80(14)

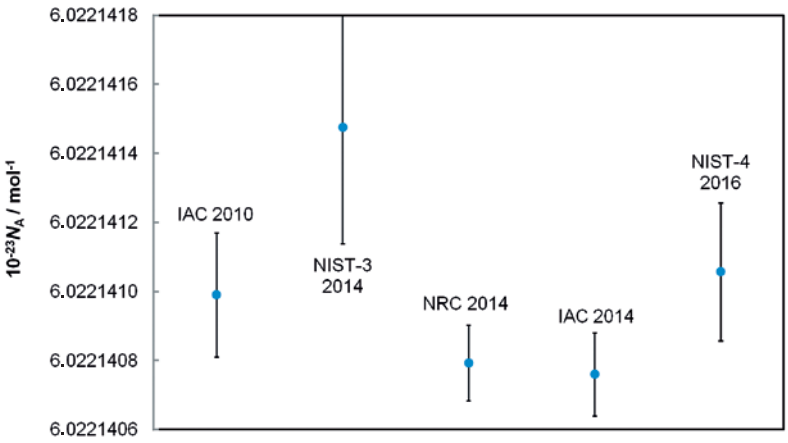


Figure 8:
Survey of the currently most exact determination of the Avogadro constant (IAC: International Avogadro Coordination, NIST: National Institute of Standards and Technology – USA, NRC: National Research Council of Canada).

Acknowledgements

In addition to their own scientific tasks, the authors of this article had (P. B.) and have (H. B.) a coordinating function as project manager in the Avogadro project. The authors therefore wish to thank all of their colleagues in the Avogadro project for their technical production steps, metrological services, mathematical evaluations and (not to forget!) science policy decisions. All these individual services provide the basis for the success of the Avogadro project. Special thanks are due to some colleagues who have particularly contributed to the described activities: Birk Andreas, Guido Bartl, Michael Borys, Ingo Busch, Bernd Güttler, Michael Krumrey, Ulrich Kuetgens, Arnold Nicolaus, Axel Pramann, Olaf Rienitz and Detlef Schiel.

Literature

[1] *B. Andreas, Y. Azuma, G. Bartl, P. Becker, H. Bettin, M. Borys, I. Busch, M. Gray, P. Fuchs, K. Fujii, H. Fujimoto, E. Kessler, M. Krumrey, U. Kuetgens, N. Kuramoto, G. Mana, P. Manson, E. Massa, S. Mizushima, A. Nicolaus, A. Picard, A. Pramann, O. Rienitz, D. Schiel, S. Valkiers, A. Waseda*; Determination of the Avogadro constant by counting the atoms in a ^{28}Si crystal, Phys. Rev. Lett. **106** (2011) 030801.

[2] *Y. Azuma, P. Barat, G. Bartl, H. Bettin, M. Borys, I. Busch, L. Cibik, G. D'Agostino, K. Fujii, H. Fujimoto, A. Hioki, M. Krumrey, U. Kuetgens, N. Kuramoto, G. Mana, E. Massa, R. Meeß, S. Mizushima, T. Narukawa, A. Nicolaus, A. Pramann, S. A. Rabb, O. Rienitz, C. Sasso, M. Stock, R. D. Vocke Jr, A. Waseda, S. Wundrack, S. Zakel*; Improved measurement results for the Avogadro constant using a ^{28}Si-enriched crystal, Metrologia **52** (2015) 360–375.

[3] International determination of the Avogadro constant, special issue, ed. by *E. Massa and A. Nicolaus*, Metrologia **48** (2011) 1–119.

[4] *I. M. 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 (CI-2005) Metrologia **43** (2006) 227–246.

[5] *P. Becker, P. De Bièvre, K. Fujii, M. Gläser, B. Inglis, H. Lübbig, G. Mana*; Consideration on future redefinitions of the kilogram, the mole and of other units, Metrologia **44** (2007) 1–14.

[6] *U. Bonse, M. Hart*; An x-ray interferometer. Appl. Phys. Lett. **6** (1965) 155–156.

[7] *P. Becker, K. Dorenwendt, G. Ebeling, R. Lauer, W. Lucas, R. Probst, H.-J. Rademacher, G. Reim, P. Seyfried, H. Siegert*; Absolute measurement of the (220) lattice plane spacing in a silicon crystal, Phys. Rev. Lett. **46** (1981) 1540–1543.

[8] *P. Seyfried, P. Becker, A. Kozdon, F. Lüdicke, F. Spieweck, J. Stümpel, H. Wagenbreth, D. Windisch, P. De Bièvre, H. H. Ku, G. Lenaers, T. J. Murphy, H. J. Peiser, S. Valkiers*; A determination of the Avogadro Constant, Z. Phys. B Condensed Matter **87** (1992) 289–298.

[9] *P. Becker, H. Bettin, H.-U. Danzebrink, M. Gläser, U. Kuetgens, A. Nicolaus, D. Schiel, P. De Bièvre, S. Valkiers, P. Taylor*; Determination of the Avogadro constant via the silicon route, Metrologia **40** (2003) 271–287.

[10] *P. Becker*; History and progress in the accurate determination of the Avogadro constant, Rep. Prog. Phys. **64** (2001) 1945–2008.

[11] *Yu. V. Tarbeyev, A. K. Kaliteyevsky, V. I. Sergeyev, R. D. Smirnov, O. N. Godisov*; Scientific, Engineering and Metrological Problems in Producing Pure ^{28}Si and Growing Single Crystals, Metrologia **31** (1994) 269–273.

[12] *P. Becker, D. Schiel, H.-J. Pohl, A. K. Kaliteyevsky, O. N. Godisov, G. G. Devyatykh, A. V. Gusev, A. D. Bulanov, S. A. Adamchik, V. A. Gavva, I. D. Kovalev, N. V. Abrosimov, B. Hallmann-Seiffert, H. Riemann, S. Valkiers, P. Taylor, P. De Bièvre, E. M. Dianov*; Large-scale production of highly enriched ^{28}Si for the precise determination of the Avogadro constant, Meas. Sci. Technol. **17** (2006) 1854–1860.

[13] *I. Busch, Y. Azuma, H. Bettin, L. Cibik, P. Fuchs, K. Fujii, M. Krumrey, U. Kuetgens, N. Kuramoto, S. Mizushima*; Surface layer determination for the Si spheres of the Avogadro project, Metrologia **48** (2011) 62–82.

[14] *G. Bartl, M. Krystek, A. Nicolaus, W. Giardini*; Interferometric determination of the topographies of absolute sphere radii using the sphere interferometer of PTB. Meas. Sci. Technol. **21** (2010) 115101.

[15] *B. Andreas, I. Ferroglio, K. Fujii, N. Kuramoto, G. Mana*; Phase corrections in the optical interferometer for Si sphere volume measurements at NMIJ, Metrologia **48** (2011) 104–111.

[16] *R. Schwartz*; Precision determination of adsorption layers on stainless steel mass standards by mass comparison and ellipsometry: Part II. Sorption phenomena in vacuum, Metrologia **31** (1994) 129–136.

[17] *A. Picard, H. Fang*; Methods to determine water vapour sorption on mass standards, Metrologia **41** (2004) 333–339.

[18] *A. Picard, N. Bignell, M. Borys, S. Downes, S. Mizushima*; Mass comparison of the 1 kg silicon sphere AVO#3 traceable to the International Prototype K, Metrologia **46** (2009) 1–10.

[19] *M. Gläser, R. Schwartz, M. Mecke*; Experimental Determination of Air Densitiy Using a 1 kg Mass Comparator in Vacuum, Metrologia **28** (1991) 45–50.

[20] *A. Picard, P. Barat, M. Borys, M. Firlus, S. Mizushima*; State-of-the-art mass determination of ^{28}Si spheres for the Avogadro project, Metrologia **48** (2011) 112–119.

[21] *O. Rienitz, A. Pramann, D. Schiel*; Novel concept for the mass spectrometric determination of absolute isotopic abundances with improved measurement uncertainty: Part 1 – theoretical derivation and feasibility study, Int. J. Mass Spectrom. **289** (2010) 47–53.

[22] *G. Mana, O.Rienitz*; The calibration of Si isotope ratio measurements, Int. J. Mass Spectrom. **291** (2010) 55–60.

[23] *A. Pramann, O. Rienitz, D. Schiel*; Silicon Isotope Ratios Affected by Sodium-Induced Broadband Interference in High Resolution Multicollector-ICPMS, Analytical Chemistry **84** (2012) 10175–10179.

[24] *A. Pramann, K-S. Lee, J. Noordmann, O. Rienitz*; Probing the homogeneity of the isotopic composition and molar mass of the 'Avogadro'-crystal, Metrologia **52** (2015) 800–810.

[25] *S. W. Biernacki*; The Influence of Temperature, Isotope Composition and Impurities on the Lattice Parameter of Si, Metrologia **31** (1994) 251–254

[26] *H. Bettin, H. Toth*; Solid density determination by the pressure-of-flotation method, Meas. Sci. Technol. **17** (2006) 2567–2573.

[27] *M.L.W. Thewalt, M. Steger, A. Yang, N. Stavrias, M. Cardona, H. Riemann, N.V. Abrosimov, M.F. Churbanov, A.V. Gusev, A.D. Bulanov, I.D. Kovalev, A.K. Kaliteevskii, O.N. Godisov, P. Becker, H.-J. Pohl, J.W. Ager III, E.E. Haller*; Can highly enriched ^{28}Si reveal new things about old defects? Physica B **401–402** (2007) 587–592.

[28] *S. Simmons, R.M. Brown, H. Riemann, N.V. Abrosimov, P. Becker, H.-J. Pohl, M.L.W. Thewalt, K.M. Itoh, J.J.L. Morton*; Entanglement in a solid-state spin ensemble, Nature **470** (2011) 69–72.

[29] *C. Williams*; private note 2011.

[30] *P. Becker, H.-J. Pohl, H. Riemann, N. Abrosimov*; Enrichment of silicon for a better kilogram, feature article, Phys. Stat. Sol. A **207** (2010) 49–66.

[31] [CCM 2013] Consultative Committee for Mass and Related Quantities (CCM), Report of the 14th meeting (2013) to the International Committee for Weights and Measures, http://www.bipm.org/utils/common/pdf/CC/CCM/CCM14.pdf (last accessed on 13 June 2016).

[32] *M. Gläser, M. Borys, D. Ratschko, R. Schwartz*; Redefinition of the kilogram and the impact on its future dissemination, Metrologia **47** (2010) 419–428.

Electro-Mechanical Balance – the Watt Balance

Michael Gläser*

Abstract

In the introduction below, the history of the development of the watt balance is explained as well as its significance for the new International System of Units (SI). Mention is made of the motivation behind the idea of the watt balance – as an experimental method of linking electric quantities to the kilogram, the meter and the second, which promises a higher accuracy than has been achieved in earlier tests. The principle of the watt balance and the representation of current and voltage via the macroscopic quantum effects of the Josephson effect and the quantum Hall effect by means of Planck's constant and the elementary charge are described. The modes of operation of the currently most significant watt balances are presented and their results are shown in a diagram in comparison to the result of the Avogadro project and in comparison to the CODATA value based on a least-squares adjustment. Finally, the dissemination of the unit of the kilogram is described – ranging from the mass standards used in the watt balance to the weights used in industry, on the markets or for the verification of weighing instruments.

1. Introduction

History

André-Marie Ampère was the first to develop a current balance after he had discovered the magnetic forces of current-carrying conductors. This current balance can be imagined as two parallel conductors, one of which is movable and linked to a rocking bar whose equilibrium position can be adjusted by means of a traveling weight (Figure 1). If electric currents flow in the same direction in both conductors, they attract each other and the weight will be adjusted to the equilibrium position on the side of the rocking bar which is opposite to the conductor. If the currents flow in opposite directions, the weight on the side of the conductor is adjusted. Later on, Lord Kelvin developed a weighing instrument which could be used to attribute a known electric current to a measurable mechanical force (Figure 2). In terms of the present units, this means that the unit of ampere is attributed to a force in the units of the meter, the kilogram and the second. Finally, in 1948, the unit of the ampere was defined according to an idealized arrangement corresponding to the ampere balance. However, the arrangement of conductors of infinite length, which is described in

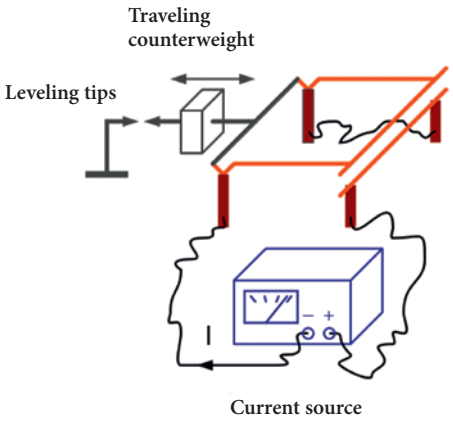


Figure 1:
Scheme of a current balance according to Ampère (see Wikipedia).

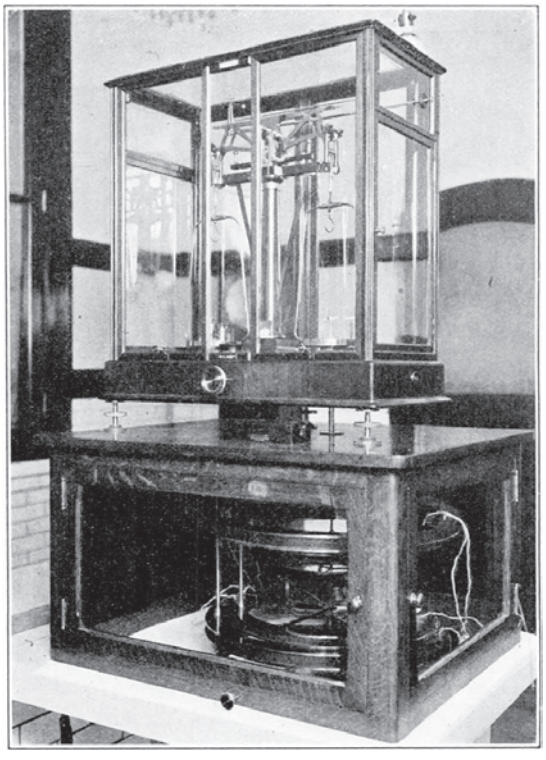


Figure 2:
Current balance at NIST in 1912. In the upper box, there is an equal-armed beam balance, in the lower box, there is the solenoid suspended at the right balance arm [4].

* Dr. Michael Gläser, former head of PTB's laboratory "Unit of Mass", and later of PTB's Department "Solid Mechanics", e-mail: glaeser-braunschweig@t-online.de

this definition, cannot be realized so that different experimental designs with coils were selected in order to realize the unit of the ampere. The difference between the two geometries can be exactly calculated. However, the current density in the coil wires is not constant across its cross section, so that the geometrical distance of the coils does not agree with the distance of the electric currents [1]. Yet the relative uncertainty of some 10^{-6} of such measurements [2] could not be reduced to the extent one had hoped for at the beginning of the 1970s. The critical quantity was the magnetic induction which required an exact knowledge of the current distribution also inside the coil wires and of the magnetic field distribution in space. Brian Kibble was the first to suggest a measurement method, in 1975, which no longer required such measurements [3] (see Section 4). Thus, in the first test made with a conventional current balance, the force which acts on a current-carrying conductor in a magnetic field was to be compared to the weight force of a mass standard. In the second test with the same arrangement, the electric conductor was to be moved in the magnetic field and thus a voltage between its ends was to be generated (see Section 2). If the equations of these two tests are combined in such a way that the magnetic induction is eliminated, a relationship between current, voltage, mass, gravitational acceleration and speed is created – quantities which can be measured with much higher accuracy than the magnetic induction. Soon after Kibble had published this proposal, he himself at the *National Physical Laboratory* (NPL, Great Britain) – but also a team at the *National Bureau of Standards* (NBS, today: *National Institute of Standards and Technology*, NIST, USA) – began to set up the respective apparatuses which later on were called the "watt balance". The two institutes pursued different concepts. Whereas the NPL used an existing equal-armed beam balance and a permanent magnet, the NBS used a cable pulley instead of a balance beam and a superconducting coil for the generation of the magnetic field. In 1998, the Swiss Federal Office of Metrology (OFMET, today METAS, Switzerland) also began to set up a watt balance. Its concept differed from that of the two other institutes. A commercial mass comparator is used to compare the forces, the vertical movement is carried out with a double seesaw, and the two test modes are mechanically separated. Also the *Bureau International des Poids et Mesures* (BIPM, France) and other national metrological institutes started to set up a watt balance, e.g. the *Laboratoire national de métrologie et d'essais* (LNE, France), the *Korea Research Institute of Standards and Science* (KRISS, South Korea) and the *Measurement Standards Laboratory* (MSL, New Zealand). The *National Metrology Institute of Japan* (NMIJ, Japan), the *Ulusal Metroloji Enstitüsü* (UME, Turkey) and the *D.I. Mendeleyew Institute for Metrology* (VNIIM, Russian Federation) plan to set up a watt balance in the near future.

The *National Institute of Metrology* (NIM, China) began to set up a weighing instrument according to a slightly different principle: a joule balance, which equates mechanical and electric energies.

Significance of the watt balance for the new SI

For the new *Système international d'unités* (SI) envisaged for 2018, the values of several fundamental constants will be fixed. These values consist of a number and a dimension, the latter being the product of units. In this way, the units used in these values, e.g. kilogram, meter, second, ampere and kelvin, are indirectly defined. In the case of the watt balance, the product of current and voltage will be replaced by two frequencies and Planck's constant (see Section 3), whereby two macroscopic quantum effects are exploited: the Josephson effect for voltage and the quantum Hall effect for a resistance. The watt balance then creates a relationship between a mass, the gravitational acceleration, a velocity, two frequencies and Planck's constant. Thus, the kilogram will be defined by the meter, the second and Planck's constant. The meter is already defined – and will continue to be defined in future – via the value of the speed of light (see the article *Interferometry – How Do I Coax a Length Out of Light* in this publication), and the second via an atomic constant, the period duration of a hyperfine splitting transition of the ^{133}Cs nuclide (see the article *How Does an Atomic Clock tick* in this publication). Thus, the watt balance in the new SI will be a possible realization of the unit of kilogram due to the fixation of Planck's constant. There will also be other realizations, as far as they relate to the fixed fundamental constants, such as the Avogadro experiment (see the article *Counting Atoms for Mass and Amount of Substance* in this publication).

2. From the kilogram to the watt

The watt balance can be realized in different arrangements (see Section 4). Here, a cable pulley serves to better illustrate the principle and takes on the function of the equal-armed beam balance (Figure 3). One of the two tests (modes) operating the watt balance is the static mode. In this mode, a mass standard (a weight) is applied on one of the two cable ends on a pan. To the other end of the cable, a current-carrying coil is attached which is positioned in the field of a magnet in such a way that a vertical force is generated downwards. If the gravitational force of

the mass standard and the electromagnetic force acting on the coil are balanced, the product from the mass of the mass standard m and the local gravitational acceleration g is equal to the product of the current in the coil I, the magnetic flux density B and the length of the coil wire L:

$$mg = I\,B\,L \qquad (1)$$

Equation (1) applies in vacuum. In air, the gravitational force must be completed by the correction for air buoyancy. The current I can now be calculated by m, g, B and L. As, however, the magnetic flux density cannot be determined precisely enough, the balance is operated in a second test – the moving-coil mode. Here, a voltmeter is connected to the ends of the coil instead of a current source. At the other end of the cable, a driver is connected instead of a mass standard, which sets the pulley in motion in such a way that the coil moves at constant speed vertically inside the magnetic field. Thereby, a voltage is induced in the coil which is recorded by means of the voltmeter. This effect is known from the dynamo; it is based on Maxwell's equations. The speed of the motion is measured with a laser interferometer. The generated voltage U is then equal to the product from B, L and the velocity v:

$$U = B\,L\,v \qquad (2)$$

If equations (1) and (2) are written as vectors, the direction of the force can also be recognized. In practice, the right-hand rule becomes useful, e.g. for equation (1): the thumb shows the direction of the current, the index finger shows the direction of the magnetic field lines, and the 90°-curled middle finger shows the direction of the (Lorentz) force. However, equation (2) is only valid if the coil moves along a distance on which the magnetic field is constant. In practice, however, this is not the case.

As the velocity can, however, be adjusted in a way that is sufficiently constant, and as the position on the path is known at which the measuring takes place in the static mode, the inhomogeneity of the magnetic field showing in the measured voltage can be modeled and taken into account correspondingly by means of a correction. Equations (1) and (2) can then be combined in such a way that the product BL will be eliminated:

$$\frac{U}{v} = \frac{mg}{I}$$

or

$$UI = mgv \qquad (3)$$

On each side of equation (3), there is now a power whose unit is the watt. Therefore, this balance is called the "watt balance". However, this equation cannot be illustrated in a simple way because these powers have been derived from two different tests. In other words: neither the electric (left side) nor the mechanical (right side) power becomes evident in the experiment. These are fictive powers which result from the combination of two equations. Nevertheless, the equation is physically correct. The quantity which is common to both tests is the product from magnetic induction and a length which has been eliminated by calculation.

3. From the watt to Planck's constant. How does *h* fix the watt and ultimately the kilogram?

Current and voltage in equation (3) can be measured in units of the previous SI. However, to trace them back to a fundamental constant – Planck's constant – the Josephson effect and the quantum Hall effect are used (see the article *Counting Electrons to Measure Current* in this publication). Without referring to details, only the relationships

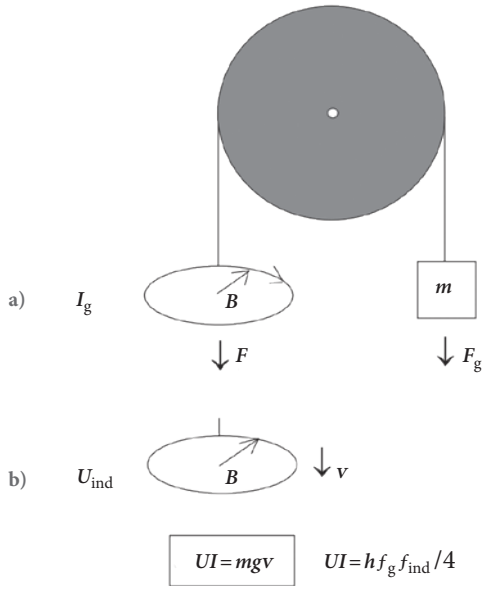


Figure 3:
Principle of the watt balance:
I_g current in the coil,
B radial magnetic field,
F electro-magnetic force,
m mass of a weight,
F_g weight force,
U_{ind} induced voltage,
v velocity in the moving coil mode,
g gravitational acceleration,
h Planck's constant,
f_g, f_{ind} microwave frequencies of the Josephson voltage [5].

between the decisive quantities will be specified here. Thus, the Josephson voltage U_J depends on a microwave radiation of the frequency f, on an integral number n_1, on Planck's constant h and on the elementary charge e according to equation (4).

$$\frac{U_J}{f} = \frac{n_1 h}{2e} \qquad (4)$$

For n_1=1, the reciprocal value on the right side of equation (4) is named after the discoverer of the

effect – the "Josephson constant". In 1973, the Briton Brian D. Josephson was awarded the Nobel Prize of Physics for his discovery. The quantum Hall resistance R_H depends on Planck's constant h, another integral number n_2 and the square of the elementary charge e in accordance with equation (5).

$$R_H = \frac{h}{n_2 e^2} \tag{5}$$

For n_2=1, the right side of equation (5) is named after the discoverer of the effect called the "von Klitzing constant". For the discovery of this effect, the German physicist Klaus von Klitzing was awarded the Nobel Prize of Physics in 1985.

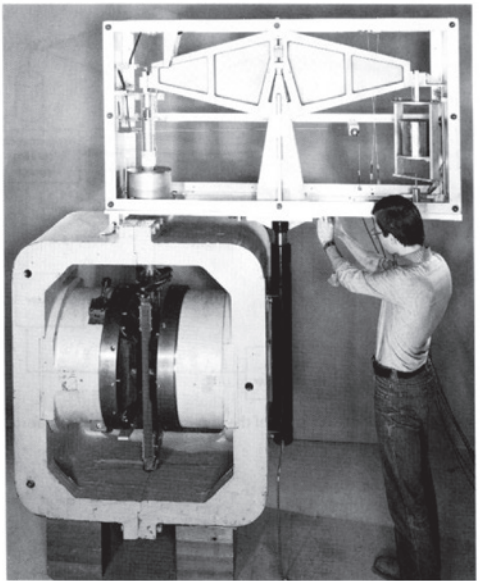


Figure 4:
NPL watt balance MARK 1. At the top, the balance beam, on the left below, the magnetic poles with the moving coil which is suspended at the left balance beam [6].

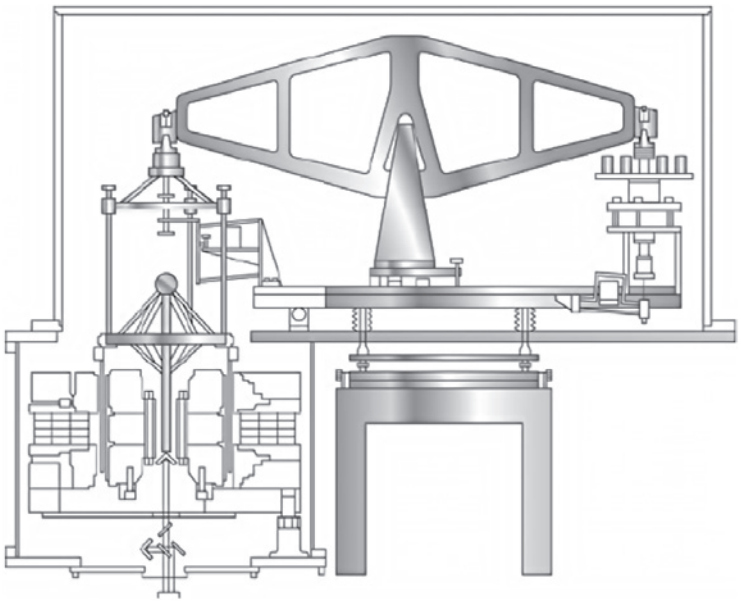


Figure 5:
Scheme of the NPL watt balance MARK 2. At the left beam arm, two vertically staggered coils are suspended in axial magnetic fields; on top of it a device for applying and lifting a 1 kg weight. At the right beam arm, there is a counterweight of 500 g [7, 16].

Current I in equation (3) is measured according to Ohm's law, equation (6) by means of a resistance R and a voltage U.

$$I = \frac{U}{R} \tag{6}$$

For the left side of equation (3), the voltage U_m – in the moving coil mode –, the voltage U_g and the resistance R_g – both in the static mode – are measured. The following is then obtained for the electric watt:

$$U_J I = \frac{1}{4} n_{1m} n_{1g} n_2 f_m f_g h \tag{7}$$

whereby the index m refers to the moving-coil mode and the index g refers to the static mode. Equation (7) now shows how h determines the watt. After equating it with the mechanical watt according to equation (3) and the appropriate transformation, one obtains:

$$m = \frac{1}{4} n_{1m} n_{1g} n_2 f_m f_g \frac{h}{gv} \tag{8}$$

Equation (8) shows how h determines the kilogram – or any other mass. After the kilogram has been redefined, the mass m of a mass standard can thus, in future, be determined by means of the watt balance using the fixed value of Planck's constant (unit Js = kg m²/s), whereby two frequencies (unit 1/s²), the local gravitational acceleration (unit m/s²) and a velocity (unit m/s) are measured.

4. The two most important watt balances

The NRC/NPL watt balance

The watt balance which today is located at the *National Research Council* (NRC, Canada) represents a further development of the watt balance of the *National Physical Laboratory* (NPL, Great Britain).

In 1975, Brian Kibble (NPL) made the proposal at the *Fifth International Conference on Atomic Masses and Fundamental Constants* in Paris to represent the unit of the ampere in a new way. His speech was published in 1976 in the subsequent conference volume [1]. In the first part of his speech, he described a measurement of the proton's gyromagnetic ratio in a strong magnetic field. The apparatus used consisted of an equal-armed beam balance for a maximum of 44 kg where a weight was suspended at one beam arm, and at the other arm a rectangular coil was suspended whose horizontal windings were arranged perpendicularly in a horizontal magnetic field (Figure 4). The magnetic field was generated between the two poles of a permanent magnet. In the magnetic field, the test sample was also positioned at which the preces-

sion frequency of the proton was measured. In the second part of his speech, Kibble suggested to use the same apparatus to compare the gravitational force of the applied weight in the first step with the electromagnetic force applied on the current-carrying coil, and to move the coil through the magnetic field and thus measure the induced voltage in the second step (see Section 2). The objective of such a measurement was to achieve the relation between the ampere – in accordance with his definition in the SI – and the ampere as quotient from the units of voltage and resistance maintained in the respective laboratory.

Not long after that, this weighing instrument was modified accordingly at the NPL in order to carry out the planned measurements. Initially, it was named "moving-coil apparatus" and later on "watt balance", as the watt is the unit of the quantity which is used to compare a mechanical power with an electric power.

Around 1990, the arrangement of magnet and coil was modified. A (short) cylindrical coil with a vertical axis was suspended in a horizontally arranged radial magnetic field (Figure 5). Actually, there are two axially staggered coils with a winding in directions opposite to each other which are suspended in radial magnetic fields poled in the opposite direction. A radial magnetic field is realized by two circular magnets whose identical poles are opposite to each other. As both poles repel each other, the magnetic field lines emanating from them are diverted and radially guided to the outside, parallel to the gap between the poles. A scheme of the NIST watt balance shows the shape of the field lines, see Figure 6. In order to reduce systematic errors, the static mode applies a procedure that differs from the above-mentioned procedure (Section 2). The static mode is carried out in two steps. The arm of the balance opposite of the coil is loaded with 500 g. In the first step, a current is conducted through the coils that balances the weight force of 500 g. In the second step, the suspension with the movable coil is loaded with a 1 kg weight by means of a remote-controlled mechanism, and a current with a reversed sign is conducted through the coils and lifts up the coil suspension (which is pulling down with 500 g) into the equilibrium position. Around 1992, the whole apparatus was accommodated in a vacuum housing in order to eliminate the influence of air buoyancy and barometric variation on the gravitational force. In order to be able to measure the velocity of the moving coil, a laser interferometer was used. For the measurement of the gravitational acceleration, an absolute gravimeter was used.

Between 2009 and 2011, this watt balance was transported from the NPL to the NRC and reestablished (Figure 7). At the NRC, further improvements were made and new measuring instruments were used. In the newly established underground measuring room, the prevailing environmental conditions are more stable, e.g. fewer mechanical underground vibrations than at the NPL. Thus, the measurement uncertainty was considerably reduced (see Section 5).

The NIST watt balance

Whereas a balance beam is used at the NRC, NIST uses a cable pulley (Fig. 8). The advantage of a cable pulley is that – in the moving- coil mode – the coil moves in an exactly vertical direction and does not

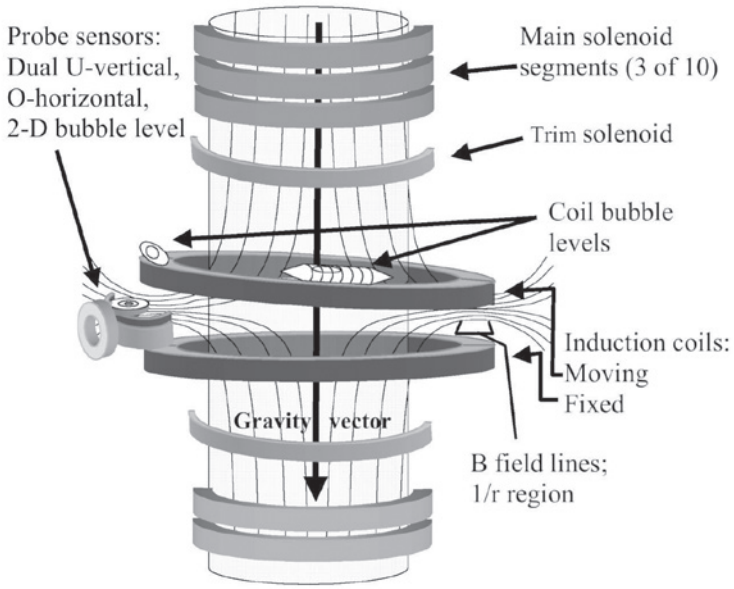


Figure 6:
Scheme of the magnetic field lines and arrangement of the superconducting coils as well as the moving and the fixed induction coils of the watt balance at NIST. The tilted induction coils show a typical misalignment. The sensors and levels are used to identify the alignment of the coils [8].

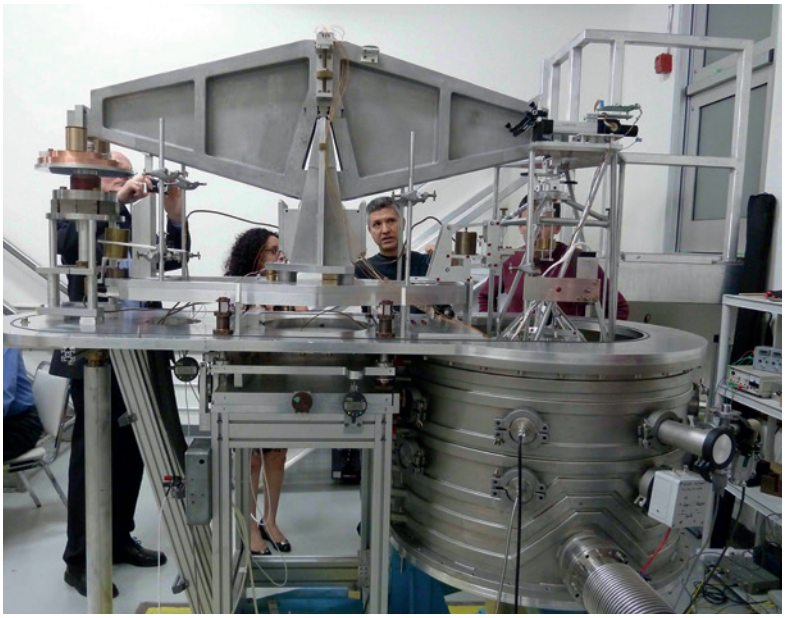


Figure 7:
The NRC watt balance, vacuum lid raised. (Photo: by the author)

have a horizontal component like the balance beam. However, a cable pulley also has disadvantages: e.g. the cable can twist and lead to undesired torsions of the attached coil. At NIST, too, the coil is suspended in a radial magnetic field, but the magnetic field is generated by two coils here. Initially, the coils were normal conducting coils at ambient temperature. Later on, superconducting coils were used. The advantage of a magnetic field generated by coils is that measurements can be carried out at differently strong magnetic fields as compared to a permanent magnet and that thus the reliability of the measurements can be verified. Whereas the coil-magnet system of the NRC watt balance is relatively compact – the coil plunges into a cylindrical slit of the complex magnet – the system of NIST has considerably larger dimensions (Figure 9 and 10). For reasons of control, the NIST watt balance has a second coil, besides the moving coil, which is similar to the moving coil, but stationary. To be able to better distinguish these two coils, we will call them "induction coils" in the following. The two induction coils surround the superconducting coils. Therefore, there is a strong magnetic stray field towards the outside. Anyone who enters the room must remove their credit cards and wrist watches to avoid damage. In order to separate the area of the pulley and the weights from the influence zone of the magnetic field, the whole apparatus is located on two levels. After the first measurements had been carried out in ambient air, the NIST watt balance was accommodated in a vacuum housing. To avoid the above-mentioned torsions, the cable was manufactured of 50 parallel platinum-tungsten wires and, in addition, the suspension was electrostatically fixed by three wings staggered by 120°.

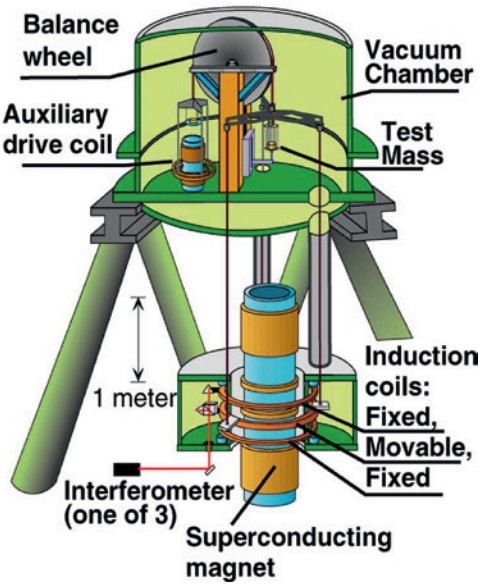


Figure 8:
Scheme of the NIST watt balance. The upper part with the cable pulley (auxiliary drive coil) and the weight is located far enough from the superconducting coils (bottom), so that their magnetic field does not influence the equilibrium position [9].

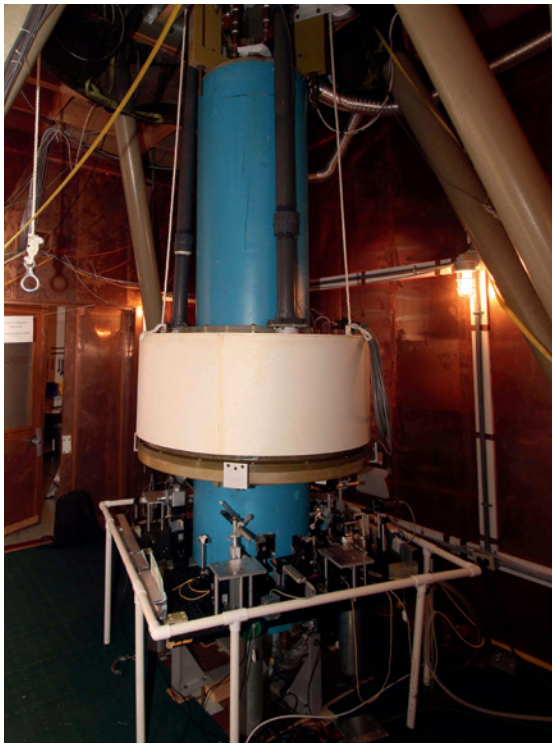


Figure 9:
The NIST watt balance, lower level. In the blue cylinders, there are the super-conducting magnetic coils, in the white ring there are the induction coils, the latter being in vacuum. (Photo: by the author)

Disturbing effects

Apart from the above-mentioned effects, e.g. air buoyancy, inhomogeneity of the magnetic field and possible torsions, the following effects are of great importance for the accuracy of measurement:

- errors which occur during the vertical alignment of the moving coil, or during the parallel alignment of the forces;
- possible horizontal oscillations which are caused in the moving coil mode or also when changing the load conditions;
- vibrations of the underground;
- the temperature sensitivity of the permanent magnet, but also that of the resistance used;
- lack of shielding against electromagnetic disturbances from outside;
- errors which occur during the determination of the gravitational acceleration at the location of the weight;
- errors occurring when taking into account the temporal changes of the gravitational acceleration;
- inadequate characteristics of the mass standard used (magnetic, alterable).

Suitable measures – including modeling – helped to reduce these disturbing effects in the respective laboratories as far as possible.

5 Measurement results

The first measurements carried out by means of the watt balances at the NPL and at NIST achieved relative uncertainties of approximately 10^{-6} for the determination of Planck's constant – comparable to the previous current measurements by means of the current balance. The experiments to determine the Avogadro constant, e.g. at PTB, also are appropriate to determine Planck's constant by means of a conversion with other fundamental constants. Very soon, an unacceptable discrepancy showed as compared to the results of the watt balances. It is only within the past few years that a tolerable agreement has been reached due to improvements of both experiment types. The relative uncertainty of $2 \cdot 10^{-8}$ required by the international committees has meanwhile been achieved for at least one of the experiments. Figure 11 shows the last results for the determination of Planck's constant.

The relative uncertainties of the results of the NRC and of the *International Avogadro Coordination* (IAC) are around $2 \cdot 10^{-8}$ or below and thereby fulfill part of the requirements of the *Comité Consultatif pour la masse et les grandeurs apparantées* (CCM); at $5.6 \cdot 10^{-8}$, the relative uncertainty of NIST is only insignificantly higher. However, these results did not meet the consistency requirement as the result of NIST differs too much from those of the NRC and the IAC.

It remains to be seen whether these three results will be consistent with each other after the future new measurements. There is still some time left until 2018. Although the result of the LNE is consistent with the three other results, its relative uncertainty of $3 \cdot 10^{-7}$ is too large to satisfy the preconditions for a redefinition.

6 Dissemination of the unit of mass

The weight used in the watt balance – here called "A" – whose exact mass was determined according to equation (8), will subsequently be used to determine the mass of other weights via a chain of comparison measurements, ultimately for weights which are used in practice for the weighing of goods or for the verification of weighing instruments. Weight A consists of a high-quality material which is particularly non-magnetic, e.g. gold, a platinum-iridium alloy or silicon. The other weights in the chain consist of stainless steel or cast iron which are considerably cheaper.

Currently, the mass standards used in watt balances and for the Avogadro experiment are compared with the international kilogram prototype within the scope of a pilot study. This comparison is made via transfer standards and an ensemble of mass standards maintained at the BIPM. Even after the redefinition, this ensemble is intended to serve to disseminate the kilogram as an embodiment of the kilogram weighted and averaged from individual later realizations (watt balance, Avogadro experiment).

In the first step, the evacuated watt balance is ventilated. After that, a film made of water molecules will accumulate on the surface of weight A at normal humid ambient air. Here, its mass increases depending on the relative humidity of the environment. This change in mass will be determined in vacuum and in air with a known relative humidity by comparison weighing of a similar weight B and a differently shaped weight C with a larger surface but with (almost) identical mass. Weight C can also consist of a number

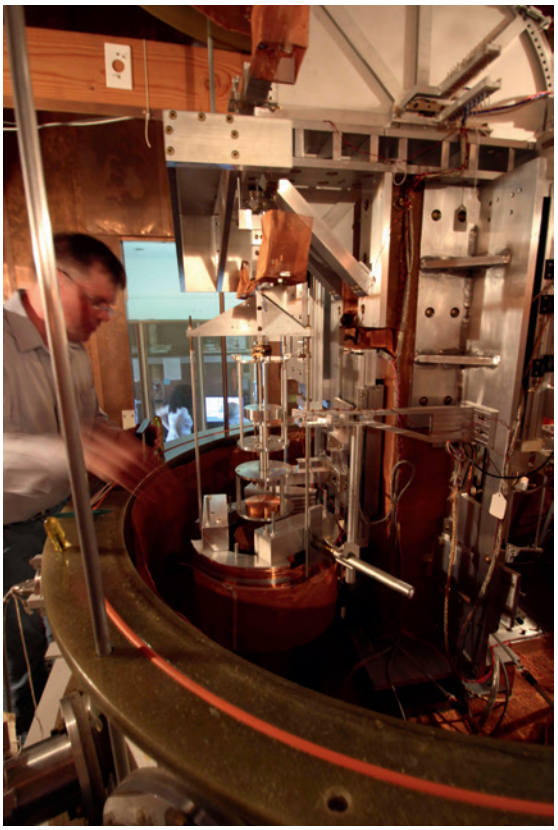


Figure 10:
The NIST watt balance, upper level with open vacuum bell jar, in the middle there is the suspension with the weight exchange facility for the weight. The cable pulley is located further up. (Photo: by the author)

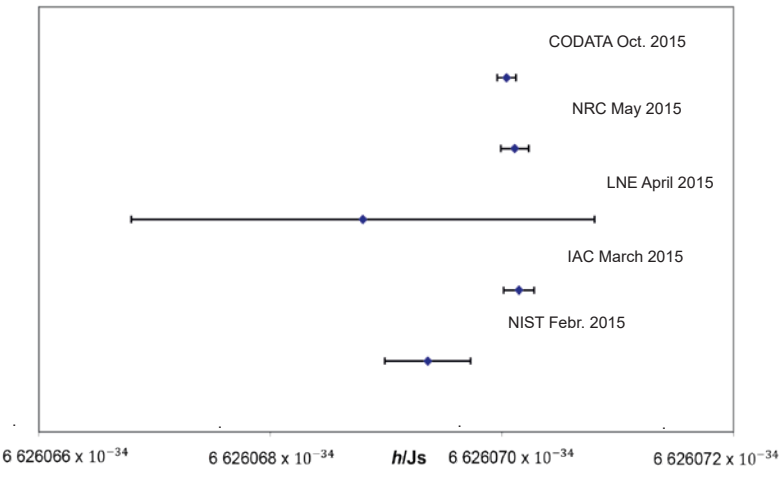


Figure 11:
Planck's constant – measurement results from 2015: NIST [10], IAC [11], LNE [12], NRC [13], CODATA [14].

of disks. The relationships determined – from change in mass to surface at different relative humidities – serve to correct for the mass of weight A. Weight A is then used to calibrate the weights D of national metrology institutes, which then calibrate the weights E of verification authorities or calibration laboratories. Weights E are then used to verify weights in industry or on markets, or also weighing instruments which are used for trade or other areas regulated by law. Taking into account the humidity is no longer necessary for weights of the verification authorities and calibration laboratories, as such modifications lie within the maximum permissible errors.

In addition, weights change over the months and years due to the absorption of aerosols, particularly hydrocarbons and integrated dust particles. In order to obtain the mass of the calibration again, the weights must be cleaned, just like windows or mirrors must be cleaned of their film from time to time. As weights which are subject to legal surveillance (verification) must be verified again after a fixed time, such cleaning is generally not necessary for them.

7 Outlook

The work which has been carried out on the watt balance for about 40 years – lately in order to achieve a higher accuracy for the determination of the value of Planck's constant – has reached a level which will allow the kilogram to be redefined in the next few years by means of fixing these fundamental constants. The international committees in charge of a redefinition have already envisaged the year 2018 for this purpose. As can be seen in Figure 11, compliance of the NIST value with the other results still has to be improved.

Literature

[1] *F.J. Ahlers*; private communication.

[2] *P. Vigoureux*; 1965, A Determination of the Ampere, Metrologia **1**, 3–7.

[3] *B.P. Kibble*; 1976, A measurement of the gyromagnetic ratio of the proton by the strong field method, in: Sanders, J.H. and Wapstra, A. H., Atomic Masses and Fundamental Constants 5, Plenum New York, pp. 545–551.

[4] *E.B. Rosa, N. E. Dorsey and J. M. Miller*; 1912, A determination of the international ampere in absolute measure, *Bull. Bur. Stand.* **8,** pp. 269–393.

[5] *M. Gläser and M. Borys*; 2009, Precision mass measurements, Rep. Prog. Phys. **72** (32 pp.), doi:10.1088/0034-4885/72/12/126101.

[6] *B.P. Kibble, R.C. Smith and I.A. Robinson*; 1982, A moving coil apparatus for realising the S.I. ampere, Proc. Conf. on Precision Electromagnetic Measurements CPEM 1982, pp. H3–H4.

[7] *B.P. Kibble, I.A. Robinson and J.H. Belliss*; 1990, Re-defining the kilogram via moving-coil apparatus, Proc. Conf. on Precision Electromagnetic Measurements CPEM 1990, pp. 178–179.

[8] *R. Steiner*; *D. Newell and E. Williams*; 2005, Details of the 1998 Watt Balance Experiment Determining the Planck Constant, J. Res. NIST, **110,** pp. 1–26.

[9] *R.L Steiner, D.B. Newell, E.R. Williams, R. Liu and P. Gournay*; 2005, The NIST project for the electronic realization of the kilogram, IEEE Trans. Instrum. Meas. **54**, pp. 846–849.

[10] *S. Schlamminger et al.*; 2015, A summary of the Planck constant measurements using a watt balance with a superconducting solenoid at NIST, Metrologia **52**, pp. L5–L8.

[11] *Y. Azuma et al.*; 2015, Improved measurement results for the Avogadro constant using a ^{28}Si-enriched crystal, Metrologia **52**, pp. 350–375.

[12] *M. Thomas et al.*, 2015, First determination of the Planck constant using the LNE watt balance, Metrologia **52**, pp. 433–443.

[13] *C.A. Sanchez et al.*; 2015, Corrigendum to the 2014 NRC determination of Planck's constant, Metrologia **52**, pp. L23.

[14] http://physics.nist.gov/constants.

Review articles

[15] *R. Steiner*; 2013, History and progress on accurate measurements of the Planck constant, Rep. Prog. Phys. **76** (46 p.) doi: 10.1088/0034-4885/76/1/016101.

[16] *M. Stock*; 2013, Watt balance experiments for the determination of the Planck constant and the redefinition of the kilogram, Metrologia **50**, pp. R1–R16.

[17] *A. Eichenberger, G. Genevès and P. Gournay*; 2009, Determination of the Planck constant by means of a watt balance, Eur. Phys. J. Special Topics **172**, pp. 363–383.

[18] *M. Gläser*; 2008, Neudefinition des Kilogramm, PTB-Mitteilungen 118, issue 2, pp. 65–70.

How Much Energy Does Temperature Contain? Determination of the Boltzmann Constant

Joachim Fischer*, Bernd Fellmuth, Christof Gaiser*****

Determination of the Boltzmann constant

In 1827, the Scottish botanist Robert Brown noticed that pollen move in a strange zigzag way in a glass of water. What was the reason for this? All attempts to explain this effect – which was called "Brownian motion" later on – initially failed. It was only Albert Einstein who realized that the movement of the small particles in the liquid was caused by continual collisions of the water molecules. In 1905, he submitted his work to the "Annalen der Physik" [1], in which he explained Brownian motion. At that time, this was a substantial argument for the existence of atoms and molecules – as in the 19th century, this existence was still utterly controversial. And at the same time, Einstein's description matched the molecular theory of heat. The warmer the water, for example, the greater the mean velocity by which the water molecules move in an unordered way and thus can cause collisions (Fig. 1, left). Thus, the term "thermodynamics" can be explained: heat is something which is dynamic.

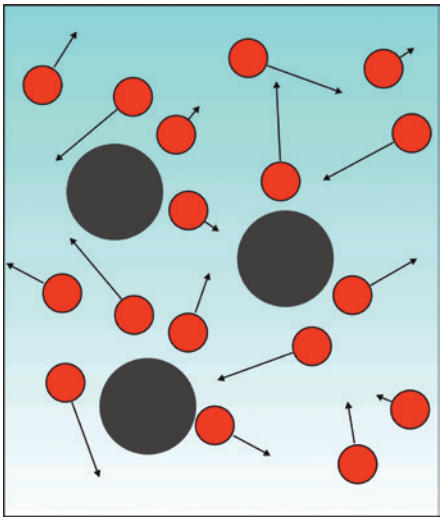
Nearly at the same time as Albert Einstein, the Polish physicist Marian Smoluchowski also provided a correct explanation of Brownian motion. However, it was then the French physicist Jean-Baptiste Perrin who confirmed Brownian motion experimentally with great precision. On the basis of Einstein's model concepts, he was one of the first, in 1908, to be able to experimentally determine the Boltzmann constant k and thus also the Avogadro constant N_A. The fact that the values determined in this way showed an error of less than 1 % for N_A and k can be rated as quantitative proof for the correctness of the kinetic approach and thus also as a further indicator for the atomic structure of matter. In 1926, Perrin was awarded the Nobel Prize in Physics for this.

But how, in detail, does the velocity of the microscopic particles depend on temperature? As one of the founders of the kinetic gas theory, Ludwig Boltzmann derived the Maxwell-Boltzmann velocity distribution (see Fig. 1, right). Therein, the quantity which is characteristic of the distribution is the mean microscopic thermal energy kT which increases linearly with the temperature with the proportionality constant k. By fixing its value, the kelvin will, in future, be directly linked to the unit of energy, the joule. Strictly speaking, a unit of temperature of its own would no longer be necessary; however, all thermometers would then have to indicate the joule. For many reasons, this is not feasible and

* Dr. Joachim Fischer, Department "Temperature", e-mail: joachim.fischer@ptb.de

** Dr. Bernd Fellmuth, Working Group "Fundamentals of Thermometry", e-mail: bernd.fellmuth@ptb.de

*** Dr. Christof Gaiser, Working Group "Fundamentals of Thermometry", e-mail: christof.gaiser@ptb.de



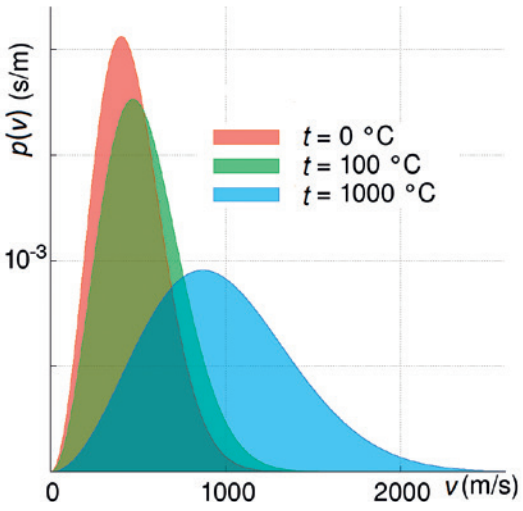


Fig. 1:
On the left: Model of Brownian motion. The water molecules (shown in red) of the surrounding medium collide with the suspended particles due to their thermal energy, so that they move on completely irregular paths. On the right: Maxwell-Boltzmann velocity distribution for nitrogen molecules at three different temperatures ($p(v)$ probability density).

enforceable.

As the microscopic thermal energy kT is experimentally not directly accessible, macroscopic quantities which are unambiguously correlated with the thermal energy [2] must be measured for the determination of k at a known temperature. Table 1 gives an overview of such dependencies in which kT is only related to other *measurable* quantities and known constants. The thermometers used for this purpose are called "primary thermometers", as they do not require any calibration. To attain the smallest possible uncertainties, the experiments are carried out at the triple point of water. As the kelvin – the base unit of thermodynamic temperature – is currently defined via this fixed point, this temperature can be realized with the greatest precision.

Boltzmann noticed that in the case of an ideal gas, the thermal energy increases proportionally to the mean kinetic energy of the gas particles. In a closed volume, this energy is directly measurable via the gas pressure and the number of particles. The pressure p is described – with the interaction between the particles being negligible – by means of the equation of state of the ideal gas. The thermometer based on this law is the traditional gas thermometer whose uncertainty, however, is too large for the determination of the Boltzmann constant.

In the case of the acoustic gas thermometer, the imprecise density determination of the classical gas thermometer is replaced by measuring the speed of sound. Furthermore, the density, which changes with temperature at a constant pressure, can be determined via the dielectric constant or the refractive index. If you use the conducting electrons of a metallic resistance material as a "measuring gas", the electrical Johnson noise according to the Nyquist formula is suited for thermometry. Laser spectroscopy provides the kinetic energy of the gas particles from the Doppler broadening of absorption lines of a gas. Eventually, an isothermal cavity, in which the light quanta are in thermal equilibrium with the walls, emits radiation. This can be used with the spectral radiation thermometer according to Planck's radiation law. In Table 1, the respective measurement uncertainties to be expected are listed for the determination of k, which is a compact overview of all current methods [2].

When redefining the kelvin, the measurement uncertainty of the value of the Boltzmann constant k should be comparable to the uncertainty of the realization used so far. The value which is at present accepted internationally [3] of $1.380\,648\,52 \cdot 10^{-23}$ JK[1] has a sufficient relative standard uncertainty of $5.7 \cdot 10^{-7}$, but is essentially based on results of acoustic thermometry only. However, measurements taken with only one method are not considered to be a sufficient basis for fixing the numerical value. They need to be confirmed by other independent methods to be able to detect and correct hidden errors inherent in the system. Therefore, experts from all the

Thermometer	Law of Physics	Relative Standard Uncertainty
Acoustic gas thermometer	$u_0 = \sqrt{\frac{\gamma_0 RT}{M}}$	$1 \cdot 10^{-6}$
Dielectric-constant gas thermometer	$p = kT\frac{(\varepsilon - \varepsilon_0)}{\alpha_0}$	$2 \cdot 10^{-6}$
Refractive index gas thermometer	$p = kT\frac{(n^2 - 1)\varepsilon_0}{\alpha_0}$	$10 \cdot 10^{-6}$
Johnson noise thermometer	$\langle U^2 \rangle = 4kTR_{el}\Delta\nu$	$2 \cdot 10^{-6}$
Doppler broadening thermometer	$\Delta\nu_D = \sqrt{\frac{2kT}{mc_0^2}}\nu_0$	$10 \cdot 10^{-6}$
Spectral radiation thermometer	$L_\lambda = \frac{2hc_0^2}{\lambda^5}\left[\exp\left(\frac{hc_0}{\lambda kT}\right) - 1\right]^{-1}$	$50 \cdot 10^{-6}$

Table 1: Overview of the primary thermometers which are suitable for the determination of the Boltzmann constant k, with the relative measurement uncertainties to be expected. T thermodynamic temperature, u_0 speed of sound in the limiting case of zero pressure and of very low frequency, $\gamma_0 = c_p/c_V$ ratio of specific heat capacities at constant pressure and constant volume, M molar mass of the gas, $R = N_A k$ molar gas constant, (N_A Avogadro constant), p pressure, ε dielectric constant of the gas, ε_0 electric constant, α_0 electric polarizability of the gas, n refractive index, $\langle U^2 \rangle$ mean square noise voltage, R_{el} electric resistance, ν frequency, $\Delta\nu_D$ Doppler width of a spectral line of frequency ν_0, m atomic mass, L_λ spectral radiance, h Planck's constant, c_0 speed of light in vacuum, λ wavelength.

metrology institutes dealing with fundamental investigations elaborated a concept and a schedule at two workshops at the Physikalisch-Technische Bundesanstalt (PTB) in Berlin in order to achieve a value of k which is based on several different methods [4]. In the following, the primary thermometric methods envisaged for this purpose will be presented, together with the uncertainties to be expected [5].

Acoustic gas thermometer

In a gas, a number of measurands depend on temperature. Usually, the describing equations have been derived from the ideal gas. As the interaction between the gas particles is not sufficiently known for the determination of k, the measurements are carried out by means of extrapolation to zero pressure in such a way that the approximation of the ideal gas is sufficient. Nowadays, the speed of sound u_0 of the noble gases of argon or helium at the temperature of the triple point of water is measured via a spherical resonator (Fig. 2) in order to determine k. Thus, k can then be calculated from the respective formula in Table 1, whereby R is replaced there by kN_A.

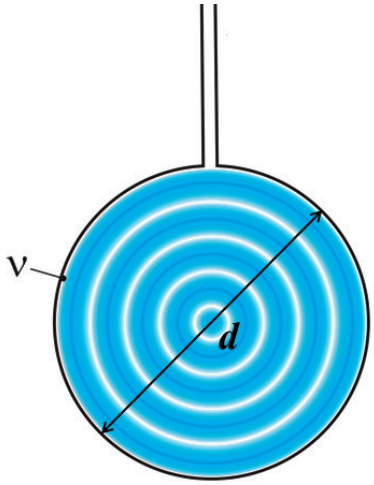
In 1988, a group at the US metrology institute NIST (*National Institute of Standards and Technology*) determined the dimensions of its resonator by filling it with mercury, whose density is exactly known. Today, the resonator is measured by means of microwave resonances and is thus traced to a frequency measurement. It is, in particular, the British national institute NPL (*National Physical Laboratory*), the French LNE (*Laboratoire National de Métrologie et d'Essais*), the Italian INRIM (*Istituto Nazionale di Ricerca Metrologica*) and the Chinese NIM (*National Institute of Metrology*) which apply this method [6]. From the determination of the dimensions and the connection to the triple point temperature, essential uncertainty components result. Components from the dependency of the molar mass on the isotopic composition, the purity of the measuring gas, the extrapolation to zero pressure and the position of the acoustic transducers and receivers are added. With the current measuring technique, an uncertainty of $1 \cdot 10^{-6}$ is attained [5].

Dielectric-constant gas thermometer

The determination of the dielectric constant of helium has been used at low temperature thermometry for quite some time. Compared to the measurement of the refractive index, the determination of the dielectric constant is the more mature method with considerably smaller achievable uncertainties.

For an ideal gas, the dielectric constant is obtained via the electric polarizability of the gas particles and the density of their number of particles. By combining it with the equation of state, the relation between the pressure p and the dielectric constant ε indicated in Table 1 is obtained. Considerable progress made with the *ab initio* calculation of the polarizability of helium, whose relative uncertainty could be reduced to significantly less than $1 \cdot 10^{-6}$ in the past few years, has made this method competitive. For the measurement of ε, the measuring gas is filled into suitable capacitors. However, due to the very low polarizability of helium, absolute measurements are not possible. Therefore, the measuring capacitor is alternately filled with helium up to a pressure of 7 Mpa (the 70-fold atmospheric pressure of the Earth) and evacuated, and ε is derived from the relative change in capacitance (Fig. 3).

During the measurement, pairs of values of the pressure p and of the dielectric constant ε are recorded at a constant temperature (isotherm measurement). From the linear component of



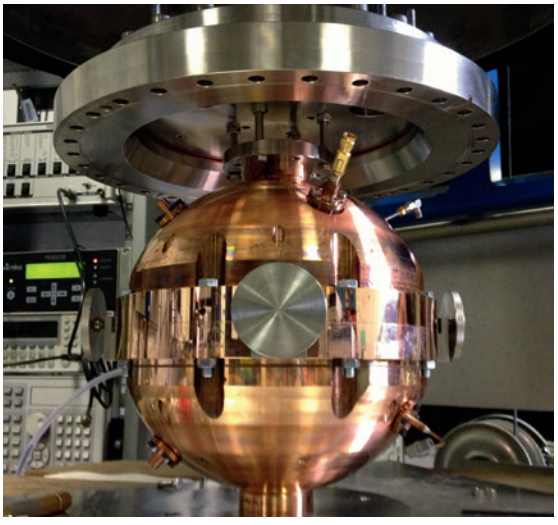
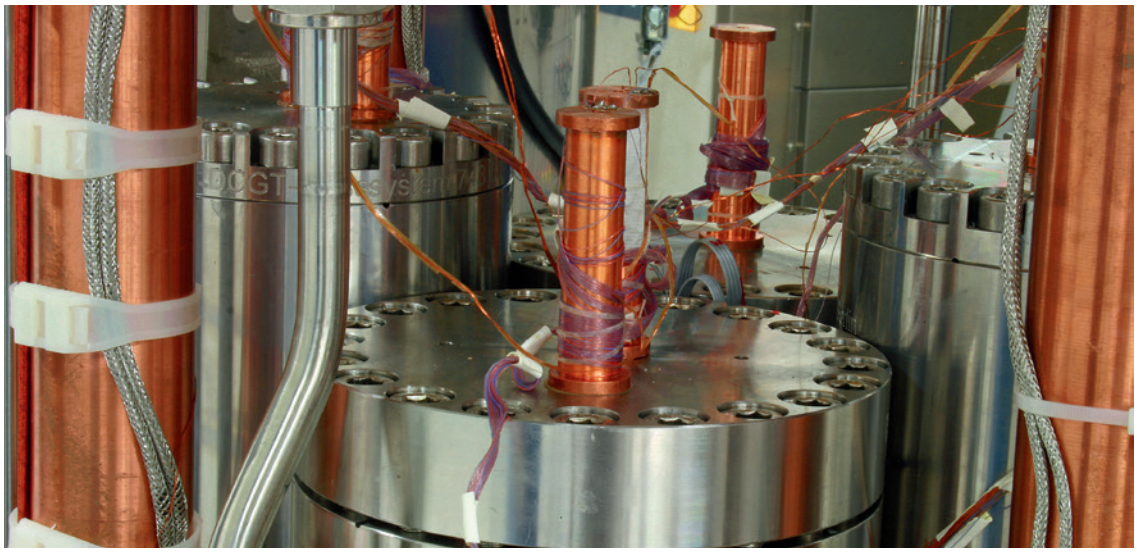


Fig. 2:
Acoustic gas thermometer with spherical resonator. The speed of sound u_0 is derived from the resonance frequencies ν and the diameter d.



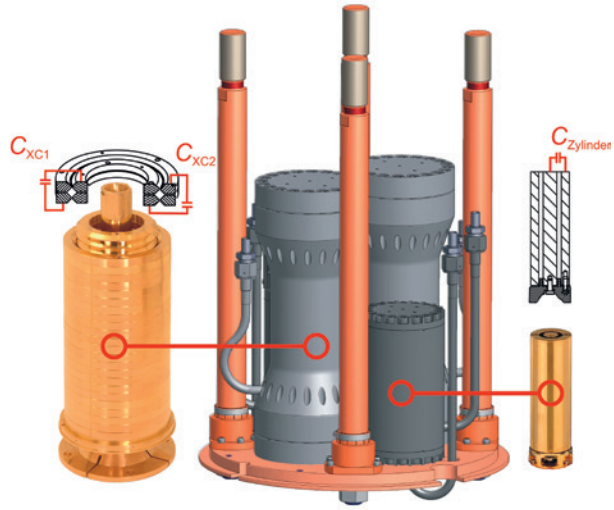


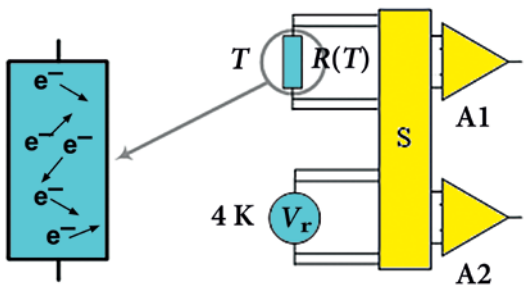
Fig. 3:
The central element of the dielectric-constant gas thermometer used at PTB includes cylindrical and ring-shaped measuring capacitors which are filled with helium gas at pressure p and whose relative capacitance change $[C(p) - C(0)]/C(0)$ is measured. The capacitors are available as pairs in the symmetrical set-up, as the capacitance bridge can only measure ratios with the highest precision. A capacitor thereby serves as a comparison capacitor. The figure shows the gold-plated capacitors which are mounted into bolted stainless-steel pressure vessels.

the function of $\varepsilon(p)$, k is eventually determined. A considerable source of error when using this method results from the deformation of the measuring capacitors due to the gas pressure. The pressure and capacitance measurements must also be improved up to the feasible limits. The estimation of all contributions allows an overall uncertainty comparable to acoustic gas thermometry of about $2 \cdot 10^{-6}$ to be expected. This estimation is supported by the results which have recently been achieved [7].

Johnson noise thermometer

The temperature of an object can also be derived from the statistical movement of the charge carriers in an ohmic resistor; this treatment generates a noise voltage (white noise). This noise thermometry has already been applied for many years. It determines the temperature from the mean square of the noise voltage by means of the Nyquist formula (line 4 in Table 1). As parameters, the electric resistance value of the test sample at the triple point of water as well as the amplification and bandwidth of the electronic measuring system have to be precisely determined.

So far, noise thermometry has been well established at very low temperatures below 5 K and at very high temperatures. At high temperatures above approximately 1000 K, it uses the larger measurement signals and is thus less sensitive to disturbances. At low temperatures, highly sensitive superconducting detectors can be used and the small signals can be detected with adequate dynamics. In the range around the triple point of water, the effective noise voltage lies – with the realizable parameters of the measurement technology – on the order of 1 µV. To measure these extremely small signals, you have to use special electronic circuits and carry out an in situ comparison with a reference noise source that is traced to the quantum standard of the voltage (see the article *Counting Electrons to Measure Current* in this publication) (Fig. 4). This is the only way to keep the amplification and the bandwidth stable and to eliminate the influence of drifts. To eliminate amplifier and line noise, the cross correlation of two measurement channels is evaluated. Thereby, only the time-correlated signal of the noise sources which is measured in both detection channels is detected. Related projects are currently carried out by NIST and NIM. The main problem is the long measuring time that is required for low uncertainties. To achieve an uncertainty in the range of $1 \cdot 10^{-5}$ for a bandwidth of 20 kHz, a measuring time of about five weeks is necessary. NIM currently achieves a relative uncertainty of $4 \cdot 10^{-6}$ [5] with measurements ranging over an extended



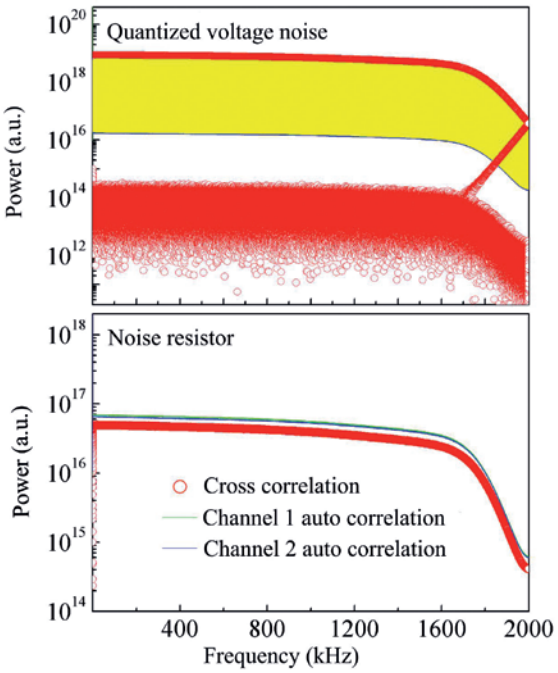


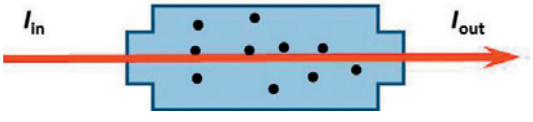
Fig 4:
The noise thermometer compares the thermal noise source $R(T)$ with the voltage reference V_r traced to the voltage standard by means of switch S. To eliminate amplifier and line noise, the cross correlation of the two measurement channels is evaluated via the amplifiers A1 and A2 (left).
Spectra of the quantized voltage noise source and the noise spectra of the measuring resistor as well as the result of the cross correlation (on the right).

bandwidth.

Doppler broadening thermometer

A further method to determine k by means of laser spectroscopy has been proposed only recently [4]. It measures the Doppler broadening of an absorption line in a gas cell at a homogeneous temperature, and thus the mean movement of the gas particles (Fig. 5). Thereby, the absorption profile is sampled by means of tunable laser radiation. The temperature is obtained from its width (see the equation in line 5 of Table 1). The advantage of this method is that no complicated absolute radiation thermometry is required (as the radiant power – if it is clearly below the saturation intensity – does not influence the full width at half maximum, except by means of heating-up effects).

At the *Université Paris 13 Nord*, experiments have so far been carried out with a CO_2 laser at an ammonia line at 30 THz. Evaluating 2000 individual absorption profiles resulted in a relative standard uncertainty for k of approximately $50 \cdot 10^{-6}$. In a second project, the *Second University of Naples* and the *Polytechnic of Milan* are cooperating with each other. They are using high-resolution spectroscopy at water vapor with a diode laser in the near infrared range. At present, the relative uncertainty is $24 \cdot 10^{-6}$. However, distinguishing Doppler broadening from other line shape modifications, which are caused by the interaction of the particles, is causing enormous difficulties. A reduction of the uncertainty to below $1 \cdot 10^{-5}$ is improbable [5].



Radiation thermometer

Radiation thermometry is based on the emitted radiant power of a blackbody cavity with an absorptivity of 1, approximated by an isothermal cavity in which the light quanta are in thermal equilibrium with the walls. The radiant power is independent of the material and form of the cavity and only depends on the temperature (and the wavelength) and on fundamental constants. Primary thermometry assumes an absolute determination of the radiant power (Fig. 6). Planck's radiation law is used to describe the spectral radiant power (Table 1, last equation).

The responsivity of the detector used in Fig. 6 is determined by an electrical substitution radiometer. A cavity receiver thereby absorbs radiation which increases the temperature of the cavity. In a second measurement, the same temperature increase is generated – with the input aperture

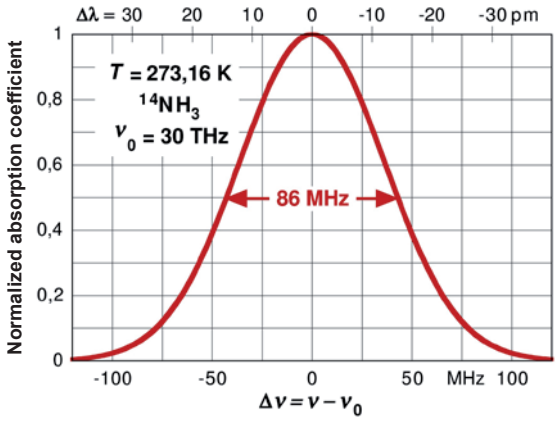
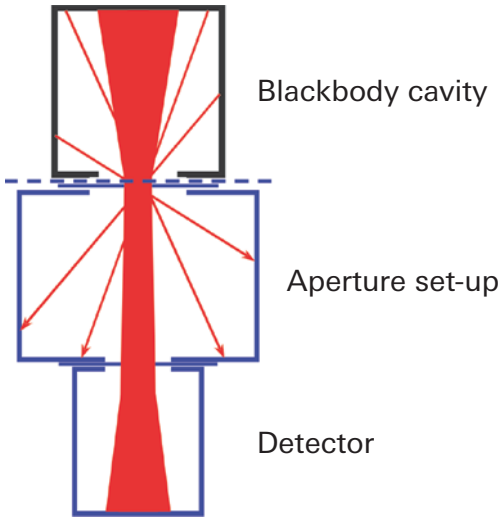


Fig. 5:
The core piece of the Doppler broadening thermometer is the gas cell (left) in which the broadening of an absorption line (right) is measured by means of tunable laser radiation.



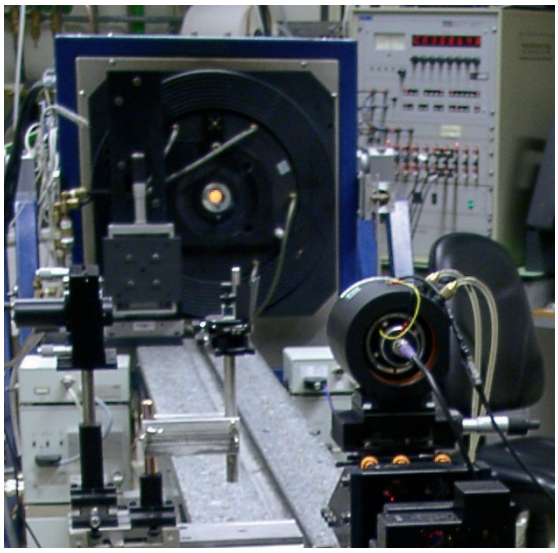


Fig. 6:
The radiation thermometer is based on the relation between the emitted radiant power of a blackbody cavity and its temperature.
An aperture set-up between the blackbody cavity and the detector defines the solid angle of the radiation.

closed – by means of electrical heating power, which can be determined very precisely. To attain the required detection sensitivity, the thermal capacity of the cavity must be as small as possible. This can be achieved by cooling down to temperatures near the boiling point of liquid helium (4.2 K). These absolute detectors are therefore called "cryogenic radiometers". It is only by using them that the – currently – low uncertainties can be achieved. Furthermore, the precise values of the areas of the apertures and their distance must be experimentally determined, as well as the emissivity (equal to absorptivity) of the blackbody cavity.

The spectral radiation thermometer has the advantage that you can choose a suitable spectral range via a filter. In this spectral range, the optical characteristics of the detector can be determined very precisely. However, as only part of the emitted spectrum is measured, the radiant power at the detector is low. Besides this, the transmission curve of the filter must be measured with extreme precision. Therefore, it is not to be expected that it will be possible to reduce the uncertainty further down than to some 10^{-5}, not even by bestowing the greatest care on the experiment [2]. However, after the redefinition of the kelvin, radiation thermometry will continue to play an important role in the measurement of high temperatures.

Consequences of the redefinition

The definition valid since 1954 determines the kelvin via a material property of a special substance. The kelvin is the 273.16th part of the thermodynamic temperature of the triple point of water [8]. Thus, influences of the isotope composition and the purity of the water used are of essential importance for their practical realization. Thereby, the long-term stability is jeopardized over space and time. By determining the Boltzmann constant, this defect is remedied.

A new definition of the kelvin based on the methods described above will be as follows [9]: The kelvin, which is the unit of the thermodynamic temperature T, is defined in such a way that the Boltzmann constant has exactly a value of $1.380\,648\,52 \cdot 10^{-23}$ J/K (explicit definition of a fundamental constant). One kelvin is thus the change of the thermodynamic temperature T which corresponds to a change of the thermal energy kT of exactly $1.380\,648\,52 \cdot 10^{-23}$ joule. This formulation is in analogy to the current definition of the meter and can be considered as an explicit definition of the unit itself. In effect, both formulations are equivalent. But what are the consequences of such a new definition?

Initially, the consequences will only be of importance for precision metrology. As has already been discussed, they must not even be felt in everyday use. Only in this way can international metrology applied so far function in an undisturbed way, and the world economy will not be affected. To achieve this, the Consultative Committee for Thermometry of the CIPM (*Comité International des Poids et Mesures*) is already working on a recommendation for implementation (*mise en pratique*). Anything else would create enormous costs. The *mise en pratique* will include recommendations for the direct measurement of the thermodynamic temperature T. Texts defining the International Temperature Scales ITS-90 and PLTS-2000, which will remain valid, [10] will add to this (see info box). The CIPM recommendation for the implementation will also discuss the differences $T–T_{90}$ and $T–T_{2000}$ of these two scales with the respective uncertainties. The temperature values of T_{90} and T_{2000} will thereby be measured in accordance with the requirements of ITS-90 and PLTS-2000.

This approach allows direct thermodynamic temperature measurements which are far away

from the triple point of water. These are, for example, high temperatures where the radiation thermometer can be used as an interpolation instrument of the ITS-90, but in future also as a primary thermometer. At the highest fixed points of the ITS-90, at 1300 K, for example, the uncertainties are about one thousand times larger than the reproducibility of the triple point of water of approximately 30 µK. These uncertainties can be considerably reduced in future by means of primary radiation thermometers.

In the temperature range around the triple point of water (which is important in practice), the ITS-90 will keep its right to exist – as it will, also in future, be of great importance for the worldwide harmonization of temperature measurement. The uncertainty of its realization is currently still up to one order of magnitude lower than the uncertainty of the thermodynamic temperature T. However, the triple point of water, which is currently – by definition – provided with an exact temperature, will lose its outstanding position. It will then be a temperature fixed point like any other, which has exactly the same uncertainty as the Boltzmann constant at the time of its fixing. A relative uncertainty of $5 \cdot 10^{-7}$ will then correspond to 0.14 mK. We are optimistic that this target will be achieved by 2018. Then the kelvin, together with three other units – the kilogram, the mole and the ampere – will be redefined by the General Conference on Weights and Measures [12].

Conclusion

Our short summary of the methods reveals that the development of the primary thermometers is making considerable progress. Therefore, we may expect that a value is achievable for the Boltzmann constant k with a relative uncertainty of $5 \cdot 10^{-7}$ which is based on various experiments with presumably three different methods. The acoustic gas thermometer (AGT), the dielectric-constant gas thermometer (DCGT) and the noise thermometer (noise) are particularly promising for this purpose (Fig. 7). Besides this, the Doppler broadening thermometer could be able to provide additional confirmation, though with larger uncertainties.

By 2018, we will be able to determine the Boltzmann constant via the primary thermometers so precisely that by fixing its numerical value once for all, the redefinition of the kelvin will become possible. In this way, a fundamental constant will be taken as a scale instead of an embodied standard (water sample). The substantial consequences are of a long-term nature, as the measuring system for the temperature would thus be infinitely stable. This objective is well worth the effort being made worldwide.

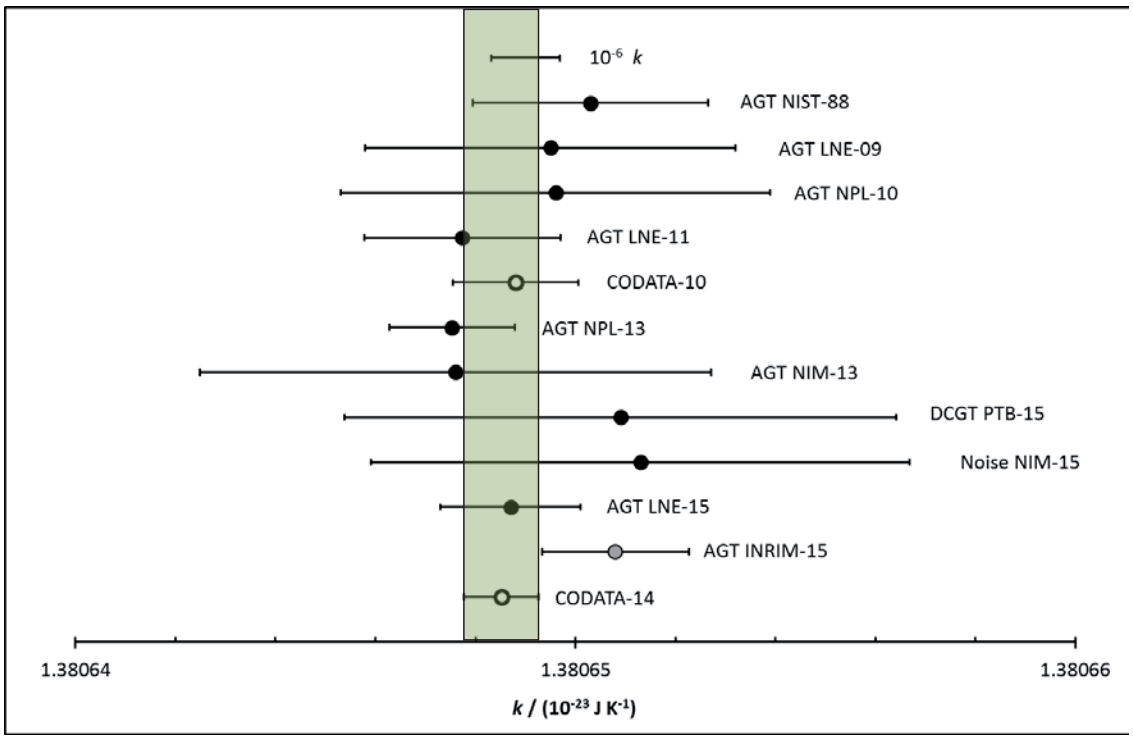


Fig. 7:
All measurements of the Boltzmann constant which contributed to the CODATA adjustment of 2014 [5], as well as the adjusted CODATA values of 2010 and 2014 and the INRIM-15 result which – due to a delayed submission – was not taken into account. The error bars specify the standard uncertainty. The green area specifies the standard uncertainty of the CODATA value of 2014.

The International Temperature Scale of 1990

The international temperature scales pursue the principle of imaging the complex measurements of thermodynamic temperatures to phase transitions of pure substances. These phase transitions can be more easily realized in the laboratory. The International Temperature Scale of 1990 (ITS-90) which is currently valid extends from 0.65 K to the highest temperatures practically measurable with the aid of Planck's radiation law. It is based on 17 well reproducible thermodynamic states of equilibrium, e.g. between the solid and the liquid phases of a pure metal [11]. Specific temperature values are assigned to these defining fixed points which – at the time of assigning – were considered as the best available approximations of the thermodynamic values. The ITS-90 prescribes for several temperature ranges various interpolation procedures between the fixed points. This gives the user as much choice in the realization of the scale as is compatible with the requirements placed on a high reproducibility. Figure 8 gives a schematic overview of the temperature ranges and the respective interpolation instruments of the ITS-90.

Since 2000, the low-temperature scale *Provisional Low Temperature Scale* PLTS-2000 has complemented the ITS-90 by three decades down to low temperatures. The temperature T_{2000} is measured by means of a ^{3}He melting pressure thermometer.

Between 0.65 K and 5 K, T_{90} is determined from the vapor pressure of ^{3}He (0.65 K to 3.2 K) or ^{4}He (1.25 K to 5 K) using a vapor pressure thermometer. In the temperature range from 3.0 K to 24.5561 K, a special gas thermometer with ^{3}He or ^{4}He serves for interpolation. In the range from 13.8 K to 1234.93 K (961.78 °C), special types of platinum resistance thermometers are interpolation instruments. The temperature of these interpolation instruments is calculated from the resistance ratio $W(T_{90}) = R(T_{90})/R(273.16\ \mathrm{K})$ by using detailed specifications. $R(T_{90})$ is the electric resistance measured at the temperature T_{90}, and $R(273.16\ \mathrm{K})$ is the electric resistance measured at the triple point of water. Temperatures above 1234.93 K (961.78 °C) are determined by means of the spectral radiation thermometer. For this purpose, the spectral radiances $L_\lambda(\lambda, T_{90})$ of a blackbody cavity of the unknown temperature T_{90} are compared – ideally at a fixed wavelength λ – with $L_\lambda(\lambda, T_{90,\mathrm{ref}})$ of a fixed point blackbody cavity whose temperature $T_{90,\mathrm{ref}}$ is known. T_{90} is calculated from the radiance ratio.

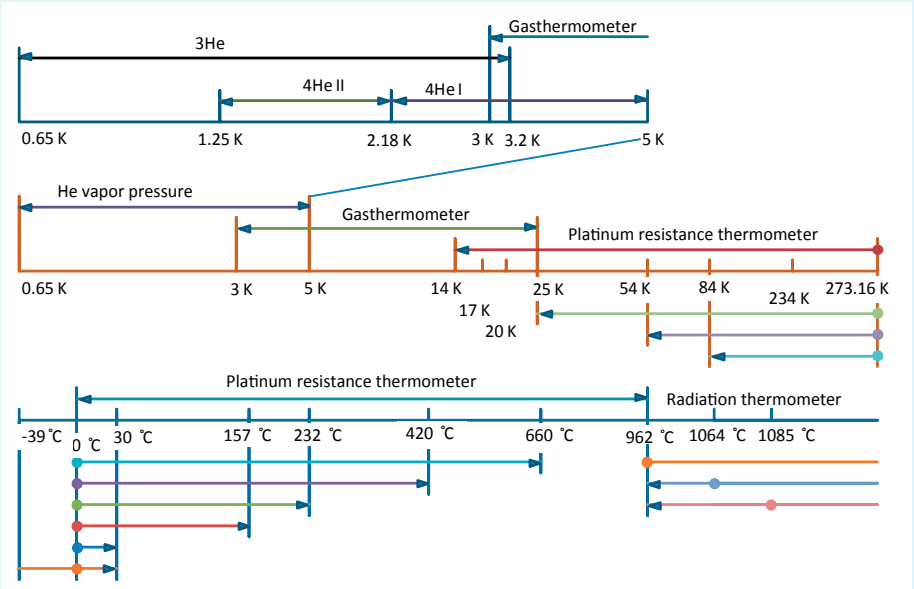


Fig. 8:
Schematic overview of the temperature ranges of the International Temperature Scale ITS-90 and the respective interpolation instruments. The temperature values of the defining fixed points are stated in rounded values only.

Literature

[1] *A. Einstein*; Über die von der molekularkinetischen Theorie der Wärme geforderten Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Ann. Phys. **17** (1905) 549–560.

[2] *B. Fellmuth, Ch. Gaiser, J. Fischer*; Determination of the Boltzmann Constant – Status and Prospects, Meas. Sci. Technol. **17** (2006) R145–R159.

[3] *P. J. Mohr, D. B. Newell, B. N. Taylor*; CODATA Recommended Values of the Fundamental Physical Constants: 2014. arXiv:1507.07956 [physics.atom-ph] (2015).

[4] *B. Fellmuth, J. Fischer* (Eds.); Talks of the 221th PTB-Seminar "Workshop on Progress in Determining the Boltzmann Constant", Report PTB-Th-3, ISBN 978-3-86509-684-5, Braunschweig (2007).

[5] *J. Fischer*; Progress towards a new definition of the kelvin. Metrologia **52** (2015) 364–375.

[6] *M. Moldover, R. M. Gavioso, J. B. Mehl, L. Pitre, M. de Podesta, J. T. Zhang*; Acoustic gas thermometry, Metrologia **51** (2014) R1–R19.

[7] *C. Gaiser, T. Zandt, B. Fellmuth*; Dielectric-constant gas thermometry, Metrologia **52** (2015) 217–226.

[8] *H. Preston-Thomas*; The International Temperature Scale of 1990 (ITS-90), Metrologia **27** (1990) 3–10, www.bipm.org/en/publications/its-90.html (last accessed on 13 June 2016).

[9] The International Systems of Units, Draft 9th Brochure, BIPM Cons. Com. Units (2013), www.bipm.org/en/committees/cc/ccu/publications-cc.html (last accessed on 13 June 2016).

[10] Mise en pratique for the definition of the kelvin, BIPM (2011), www.bipm.org/utils/en/pdf/MeP_K.pdf (last accessed on 13 June 2016).

[11] *J. Fischer, J. Hollandt*; Temperatur – Die SI-Basiseinheit Kelvin. PTB-Mitteilungen **122** (2012) 69–81.

[12] *J. Fischer, B. Fellmuth, C. Gaiser, T. Zandt*; Naturkonstanten als solide Basis, Phys. Unserer Zeit **42** (2011) 118–124.

A Measure for Visible Light – Development and Importance in the International System of Units

Armin Sperling*, Stefan Kück**

1. Introduction

The light visible to us as human beings affects every aspect of our daily lives: our behaviour, our daily rhythm, our metabolism, our communication, our well-being and our productivity. Light serves as a tool, as information and as medicine all at the same time. The advent of industrialization and the accompanying means of creating artificial illumination for workplaces and other environments made the quantitative and qualitative description of light a necessity, as insufficient light limits all of the "natural" functions described above. For these reasons, the objective physical evaluation and measurability of light became one of the central tasks of the *Physikalisch-Technische Reichsanstalt* (formerly PTR – Physical-Technical Imperial Institute, today PTB) shortly after it was founded in 1887.

The scientific discipline concerned with the measurement of light is called *photometry*. In order to explain the photometric units, we will first address the different meanings of the terms "light" and "radiation". We will then describe the past and present approaches to integrating light-measuring technology into the International System of Units (SI), shedding special light on the role of the spectral luminous efficiency of a human observer. We will close by briefly looking at possible future developments.

2. Light

The visually perceivable spectral range of electromagnetic radiation that we refer to as "light" is determined by the sensitivity of the human eye. This spectral range extends from 360 nm to 830 nm (i.e. from the ultraviolet (UV) to the near infrared (NIR) range). Thus, light comprises only a small segment of electromagnetic radiation. The spectral curve of the luminous efficiency of the eye was determined at the beginning of the 20th century from the mean value of measurements conducted on a series of (European) test subjects. The result of these measurements is the luminous efficiency function of the human eye. It is represented by the so-called $V(\lambda)$-function, which is normalized to "unity", and is shown in Fig. 1 together with the terrestrial solar spectrum. The maximum of the $V(\lambda)$-function is at exactly 555 nm (i.e. about where the spectrum of the sun available on Earth is at its maximum).

However, the actual, individual photosensitivity of the human eye varies from person to person, as it is dependent not only on one's age and ethnicity but also on the given lighting. For this reason, in the transition from visual photometry to physical photometry, so-called standard observers with defined spectral sensitivity distributions were determined. Currently, the International Commission on Illumination (CIE, *Commission Internationale de l´Éclairage*) determines the spectral luminous efficiency $V(\lambda)$ for photopic vision (i.e. for an eye adapted to brightness), $V'(\lambda)$ for scotopic vision (for an eye adapted to darkness) and $V_{mes}(\lambda)$ for mesopic vision (in the transitional area). These functions are used for the physical evaluation of light – independently of the actual, individual visual sensitivity of a particular human being.

* Dr. Armin Sperling, Working Group "Photometry", e-mail: armin.sperling@ptb.de

** Prof. Dr. Stefan Kück, Department "Photometry and Applied Radiometry", e-mail: stefan.kueck@ptb.de

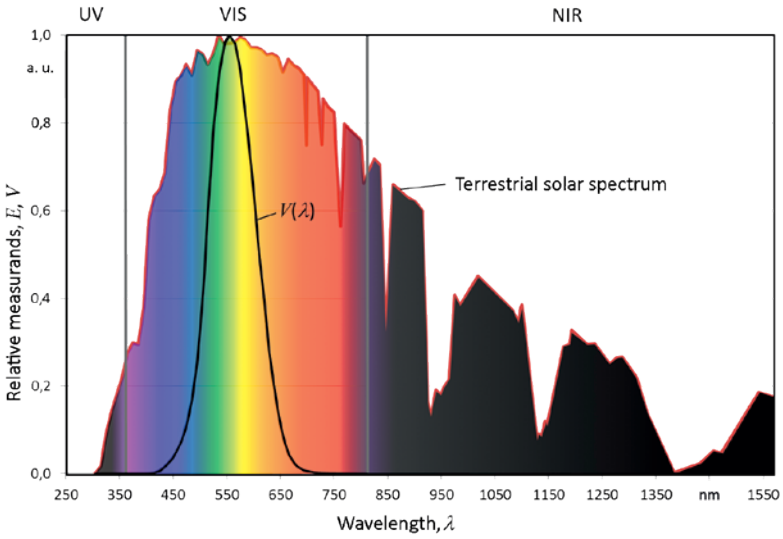


Fig. 1:
Relative progression of the terrestrial solar spectrum in the ultraviolet (UV), visual (VIS) and near-infrared spectral range with the absorption bands of the atmosphere. Within the visual spectral range, the progression of the $V(\lambda)$-function is plotted.

3. Link to the SI

Until the mid-20th century, before the invention of the first electrical detectors that were sufficiently sensitive, it was the human eye that was used in photometry as a *visual comparator*. In this process, the human eye, as an imaging system, measures and evaluates (in principle) the luminance of objects, as the light of an observed object is collected via the eye lens and reproduced onto a surface element of the retina. For this reason, the main focus of photometry was not only the total amount of the weighted electromagnetic radiation to be evaluated (luminous flux), but also especially the spatial radiation characteristic of the light sources that emitted or reflected the observed light.

The elementary physical quantity that describes the emission of light from a source is luminous intensity I_v. It is the photometric equivalent of the radiometric quantity of radiant intensity I_e, evaluated with the luminous efficiency function. Here, the "v" index stands for "visual" and the "e" index stands for "energy-related"; these indices clearly define the relation to photometry and radiometry. Expressed more simply, luminous intensity describes the amount of luminous flux emitted by a light source in one direction in a sufficiently narrow beam of light. A physically correct description would be that luminous intensity is the density (relative to the solid angle Ω) of the luminous flux Φ_v emitted in a particular direction, (i.e. $I_v = \mathrm{d}\Phi_v/\mathrm{d}\Omega$).

Even before the International Metre Convention had been signed, luminous intensity was already being used as a base quantity for the evaluation of light and for the clear description of light sources. Luminous intensity was also chosen because it was measurable using simple means (see Fig. 2) via a visual comparison of illuminance incorporating distance and solid angle. In Fig. 2, the distance d_2 of a test lamp is altered until the corresponding translucent paper windows are equally bright (i.e. the paper illuminated by the test lamp demonstrates the same luminous intensity as the paper illuminated by the standard lamp). The luminous intensity is then yielded via the relation of the distance squares as follows:

$$I_2 = I_1 \frac{d_2^2}{d_1^2} \quad . \tag{1}$$

Thus, the unit of luminous intensity was originally traced to the directed light emission of a standard light source, as this was considered a natural approach. Here, in the early phases of photometry, the value of the unit was dependent on the type of the standard light source used. The most important representatives of these standard light sources were the Hefner lamp in German-speaking Europe, the Pentan lamp in England and the Carcel lamp in France [2]. In 1937, an agreement was reached within the *Comité Consultatif de Photométrie* (CCP, the predecessor to the current CCPR, *Comité Consultatif de Photométrie et Radiométrie*, prior to the incorporation of radiometry) to create an internationally uniform, source-based definition for the unit of luminous intensity [3]. To this end, a cavity radiator was chosen as the most suitable light source because the radiation from a confined cavity in thermal equilibrium can be physically described by Planck's law of radiation. In a vacuum, the radiation of such a cavity depends only on its temperature, apart from its dependence on fundamental constants. For this reason, a cavity radiator (also called a "black-body radiator") is a light source that can be calculated.

Because white light requires a high cavity temperature, a platinum cavity radiator operated at the temperature of freezing platinum was chosen as the new standard light source. The new definition of the unit of luminous intensity was originally planned to be introduced worldwide beginning in 1942; however, this was delayed due to the outbreak of World War II. Not until 1948 did all Member Countries to the Metre Convention convert to the new unit of luminous intensity, which bore the name *candela* (cd). The 1967 version [4] defined the candela as follows:

"The candela is the luminous intensity, in the perpendicular direction, of a surface of 1/600 000 square meter of a black body at the temperature of freezing platinum under a pressure of 101 325 newtons per square meter."

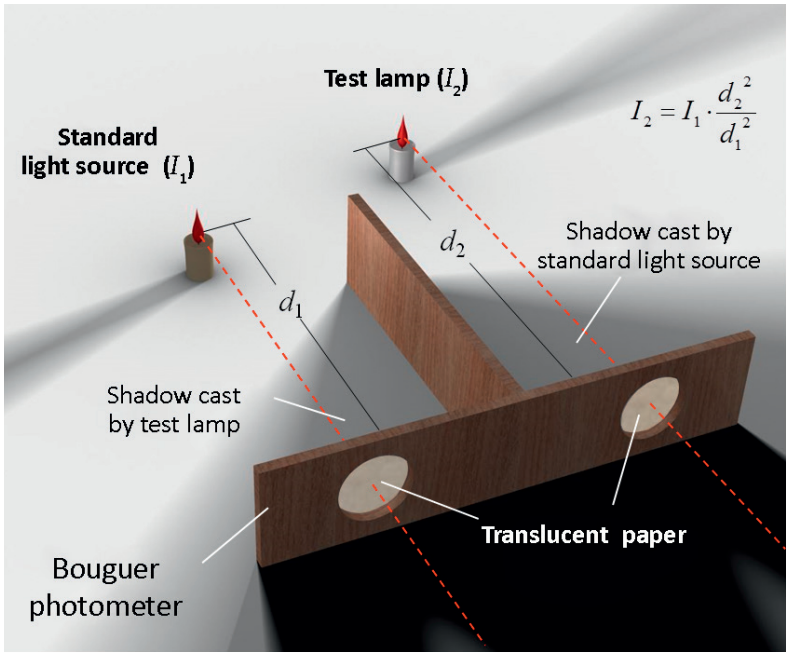


Fig. 2:
Principle of the first photometer described by Bouguer (1698–1758) around 1725 [1]. See text for explanation.

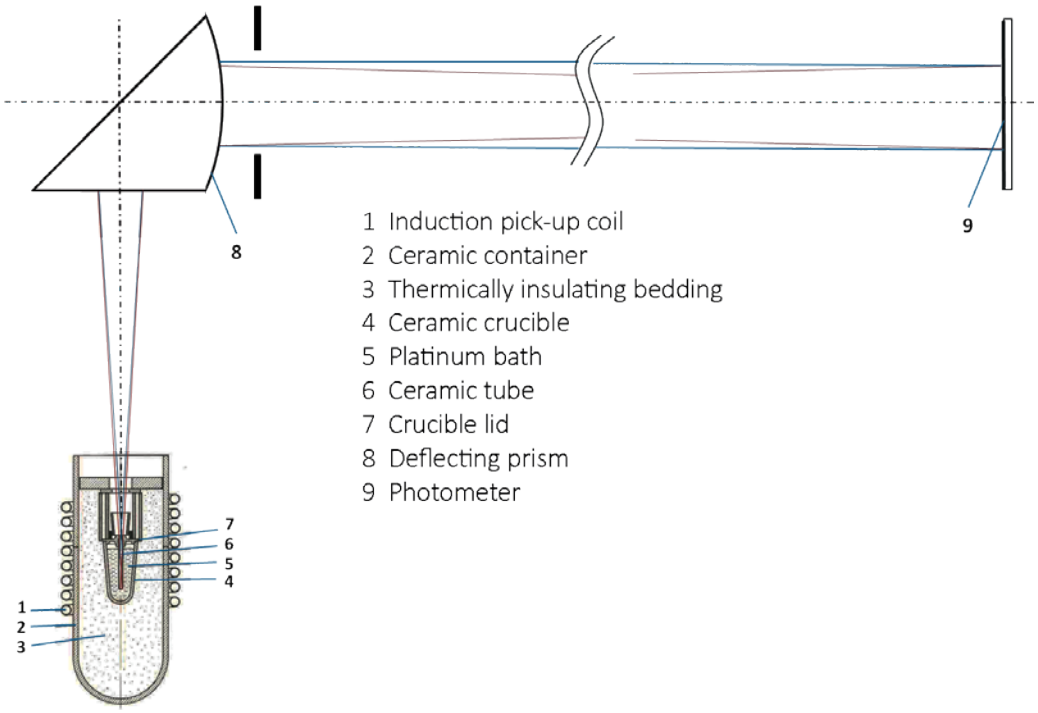


Fig. 3: Measuring set-up with a platinum point black-body radiator [5] for the realization of the candela.

The size of the surface segment was dimensioned in such a way that the numerical value of the unit of the candela was in the same order of magnitude as the unit of the "international candle" (IK), which had previously been used in England, France and the United States. Fig. 3 shows a typical measuring set-up for the realization of the candela.

With this definition, which is linked to a material property, the unit of luminous intensity was established in the SI system in such a way that it was independent of radiometric units and dependent only on the units of the quantities of surface area and air pressure. Via a direct comparison on an optical bench (which effectively replaced the Bouguer photometer shown in Fig. 2), incandescent lamps were compared with this black-body radiator by means of equation (1), thereby determining their luminous intensity absolutely.

If the temperature of the freezing point T_{Pt} of platinum is known, the spectral radiance $L_{e,\lambda}(\lambda,T_P)$ of the black-body radiator is known as well via Planck's radiation law. The spectral radiance weighted with the $V(\lambda)$-function (i.e. $\int L_{e,\lambda}(\lambda,T_P)\cdot V(\lambda)\,d\lambda$) then indicates the radiometric radiance emitted in the visible range expressed in the unit $Wm^{-2}sr^{-1}$. However, the luminous intensity in the far field divided by the luminescent surface of the black-body radiator (i.e. $\Delta I/\Delta A$) is just equal to the photometric luminance L_v of the source, expressed in the unit cdm^{-2}. The quotient of this luminance to the weighted radiance described above is thus a measure of the equivalence between photometric and radiometric quantities. This equivalence has general validity for the conversion of radiometric quantities X_e into photometric quantities X_v, and can be described via the relation $X_v = K_m \int X_{e,\lambda}(\lambda)\,V(\lambda)\,d\lambda$. If the surface area and the magnitude of the luminous intensity from the old definition of the candela are entered into the equation, the following is yielded:

$$K_{m,old} = \frac{600000\ \frac{cd}{m^2}}{\int_0^\infty L_{e,\lambda}\left(\lambda,T_{Pt}\right)\cdot V\left(\lambda\right)d\lambda}\quad , \qquad (2)$$

As stated in literature of the time, the value of the freezing temperature of platinum (according to the International Practical Temperature Scale IPTS-48) was a temperature of T_{Pt} = 2042 K [6], which led to a K_m of 682.1 lm/W [7]. The currently valid values of the fundamental constants and the freezing temperature of platinum, 2041.4 K, would result in a value of 685.06 lm/W [8]. Similarly, using the scotopic luminous efficiency function $V'(\lambda)$, T_{Pt} = 2042 K produced the spectral luminous efficacy of K'_m = 1752 lmW^{-1} for an eye adapted to darkness.

The product of K_m and $V(\lambda)$ is called the spectral luminous efficacy of radiation $K(\lambda)$, which – as $V(\lambda) \leq 1$ – has a maximum value of K_m.

Because the radiation of a perfect black-body radiator is not dependent on the material used, it was also possible (in principle) to avoid using expensive, high-purity platinum in favour of less-expensive black-body radiators (such as those based on graphite). However, it was necessary to ensure that the black-body radiator that was then used for the traceability had the same temperature as freezing platinum at the defined atmospheric

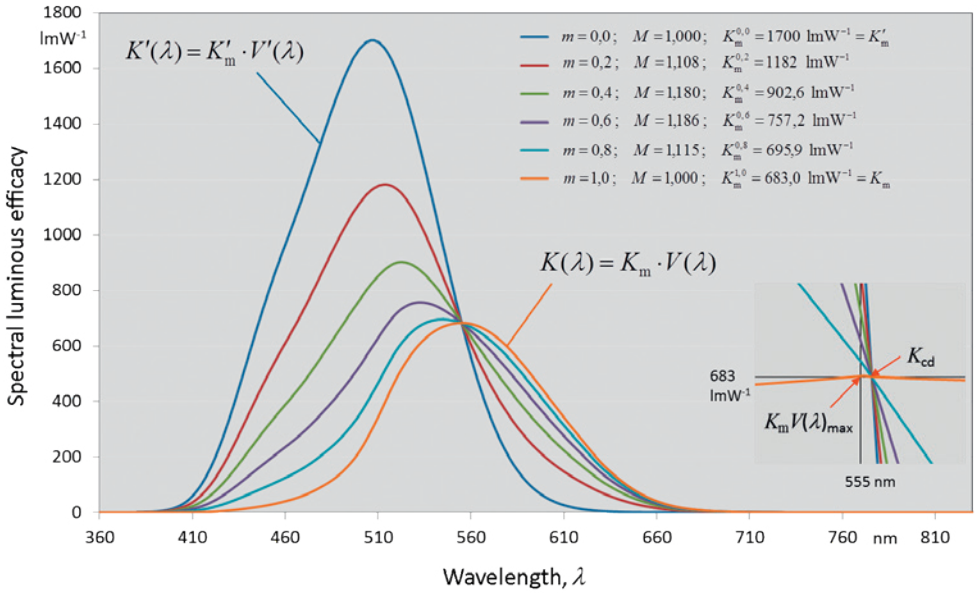


Fig. 4:
Spectral luminous efficacy $K(\lambda)$ for the different photopic, scotopic and mesopic spectral efficiency functions, and a selection of mesopic adaptation levels with their maximum efficacies K_m^m; in K_m^m, the top m index describes the adaptation level at mesopic spectral efficiency.

conditions. This method was used by many national metrology institutes. However, when the freezing temperature of platinum was assigned the value of T_{Pt} = 2045 K [6] in the late 1960s via measurements for the International Practical Temperature Scale (IPTS-68), a great deal of uncertainty was generated within the field of photometry concerning the realization of this unit. Ultimately, in the 1970s, this led to a search for a new definition for the SI base unit of the candela that was independent of artefacts and material properties. After lengthy discussions, an agreement was reached in 1979 to abandon the independence of the photometric quantities and to link luminous intensity to the radiometric unit of radiant power by means of a defined spectral luminous efficacy. This allowed the unit of luminous intensity to be realized by applying radiometric procedures.

The following definition was devised [9]:

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

A close look at this definition will reveal that it was not devised to serve as instructions for the direct realization of the luminous intensity of a candela, for the following reasons:

- Every real radiation source has a finite bandwidth.
- Luminous intensity is directly related to the sensitivity of the human eye, which is sensitive not only to a single wavelength, but also in a spectral range from 360 nm to 830 nm.
- The definition contains no statement on the quantitative assessment of radiation of other frequencies/wavelengths.

Instead, the definition merely establishes the equivalence of 1 cd $\triangleq$ 1/683 W sr^{-1} for the defined frequency of 540 THz. Here, due to $n\lambda = c/f$, the wavelength which is to be assigned depends on the value of the refractive index of air. Often, n = 1.00028 [10] is taken as the refractive index of air, although its value depends on the chemical composition and temperature of the air. In practice, the corresponding wavelength to the defined frequency will typically be in the range of (555.02 ±0.01) nm, which is very close to 555 nm.

From this equivalence, the photometric spectral luminous efficacy K_{cd} yields $K_{cd} = K(\lambda =$ 555,016 nm) = 683 cdsrW^{-1} = 683 lmW^{-1}.

In order to establish the numerical value for the maximum spectral luminous efficacy K_m, extensive measurements were performed between 1971 and 1979 – both for the determination of the freezing temperature of platinum and for the realization of the candela by means of the platinum cavity radiator [7]. Ultimately, based on these measurements, an integer frequency value in the terahertz range was agreed upon, as well as an integer spectral luminous efficacy value K_{cd}, which is very close to the maximum of the photopic spectral luminous efficacy K_m.

One reason why K_m was not simply chosen for K_{cd} is the universal validity of the candela definition for all photometric luminous efficiencies:

The definition of the unit of the candela for luminous intensity implies that the photometric spectral luminous efficacies of all existing and future photometric luminous efficiencies must intersect exactly in the point ($540 \cdot 10^{12}$ Hz, 683 lmW^{-1}).

Only via this further restraint does the candela become equally applicable for all efficiency functions. By determining this intersection, new maximum values resulted for the photometric spectral luminous efficacies of different luminous efficiencies (see also Fig. 4). Thus, the statement above results in K_m being assigned a value of 683.002 lmW^{-1} (for n = 1.00028) for photopic light (see detail enlargement in Fig. 4). However, this value is so close to the value of K_{cd} that a distinction between K_m and K_{cd} is rarely made in practice.

The changes were clearer under scotopic marginal conditions, where K'_m decreased by 3 % from 1752 lmW^{-1} to 1700 lmW^{-1}. However, this difference was tolerated in the new definition, as $V'(\lambda)$ was hardly ever used in practice at that time.

In 2010, an algorithm for the determination of mesopic luminous efficiencies, $V_{mes}(\lambda)$, was defined by the CIE [11]. Based on the fact that the eye "measures" luminances, both the photopic and scotopic luminous efficiencies have to be considered as limit functions for high and low luminances. The challenge therefore lay in finding a correlation for luminances within the intermediate range of 0.005 cdm$^{-2} \leq L \leq$ 5 cdm^{-2} that would transform the existing relative spectral efficiency functions $V(\lambda)$ and $V'(\lambda)$ as a function of the level of luminance. An agreement was reached to use a linear combination of the photopic and scotopic luminous efficiencies, controlled via the adaptation level m:

$$V_{mes}(\lambda) = \frac{1}{M(m)}(mV(\lambda) + (1-m)V'(\lambda)) \quad (3)$$

Here, $M(m)$ is a scaling factor that ensures that the maximum of $V_{mes}(\lambda)$ is always normalized to one. Fig. 4 shows the curves of several selected adaptation levels. In [11], the relation between the adaptation level m and the luminance L is explained in greater detail.

At this point, it should be pointed out that, in addition to the functions given above that describe the so-called "CIE 2° observer", additional CIE observers exist (such as the 10° observer) that can be used for the evaluation of light. The 10° observer is based on the fact that the retina of the human eye does not have an equal distribution of photosensitive cone and rod cells. The penetration point of the optical axis of the eye with the retina, the fovea, contains the maximum concentration of cones. The cones responsible for colour vision are described, for the most part, by the $V(\lambda)$-function. The highest concentration of rods sensitive to black and white, whose sensitivity is best described by the $V'(\lambda)$-function, is located at an angle between 15° and 20° from the optical axis. For this reason, the spectral sensitivity of the eye changes depending on the angle of observation, thus also changing the sensory impression of brightness and colour. Because these mechanisms play a minor role in the objective assessment of direct or indirect light sources, we have restricted ourselves here to the angular section of the CIE 2° observer that is important for focussed observing. However, in order to address sight and visual perception, it is necessary to devote attention to these and other characteristics of the human visual apparatus.

Having described the special relevance of luminous efficiency for understanding the definition of the unit of the candela, we will address the actual realization of this unit in the next section.

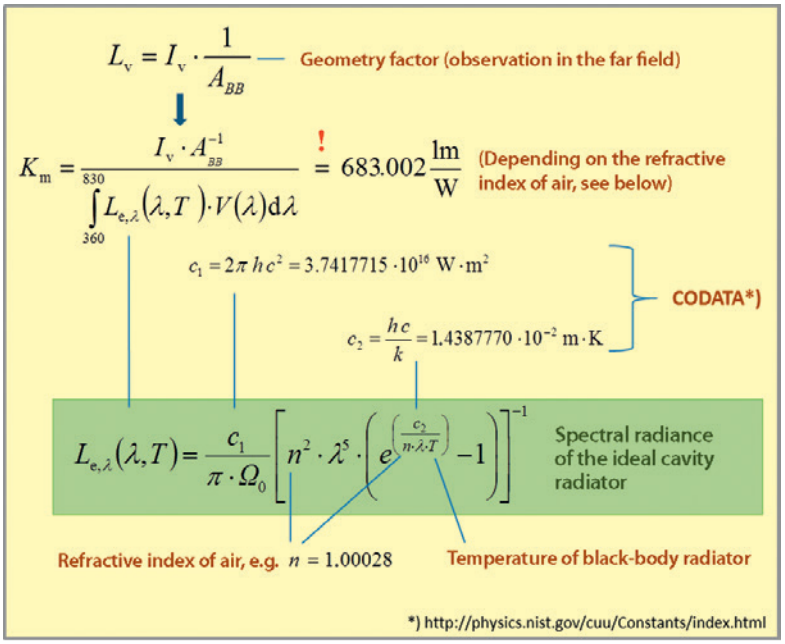


Fig. 5:
Influence quantities for the radiator-based realization of the unit of the candela.

4. Metrological realization

A realization that would naturally suggest itself would be possible by using any black-body radiator as a standard light source if it emits sufficient radiation power in the visual spectral range. If the temperature of the black-body radiator is known (for example, by using fixed-point cells with a known freezing temperature), the radiance of the black-body radiator can be determined by means of Planck's radiation equation, applying the constants $c_1 = 2\pi hc^2$ and $c_2 = hc/k$ (see Fig. 5). In accordance with equation (2), a geometry factor would have to be theoretically determined and experimentally maintained if the black-body radiator is applied; this geometry factor ensures that the maximum value of the spectral luminous efficacy of the source is at K_m = 683.002 lmW^{-1} (for n = 1.00028).

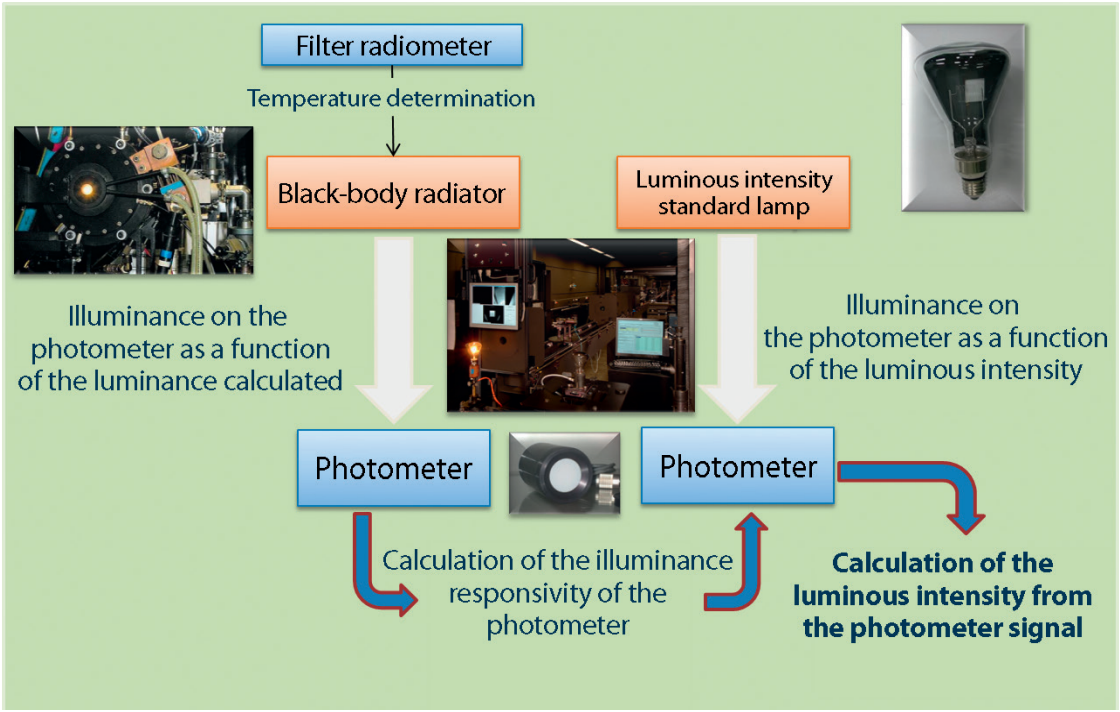


Fig. 6: Traceability chain for the radiator-aided realization of the unit of the candela. In practice, black-body radiators made of pyrolitic graphite are normally used. Long-lasting OSRAM WI41/G lamps are used as standard lamps.

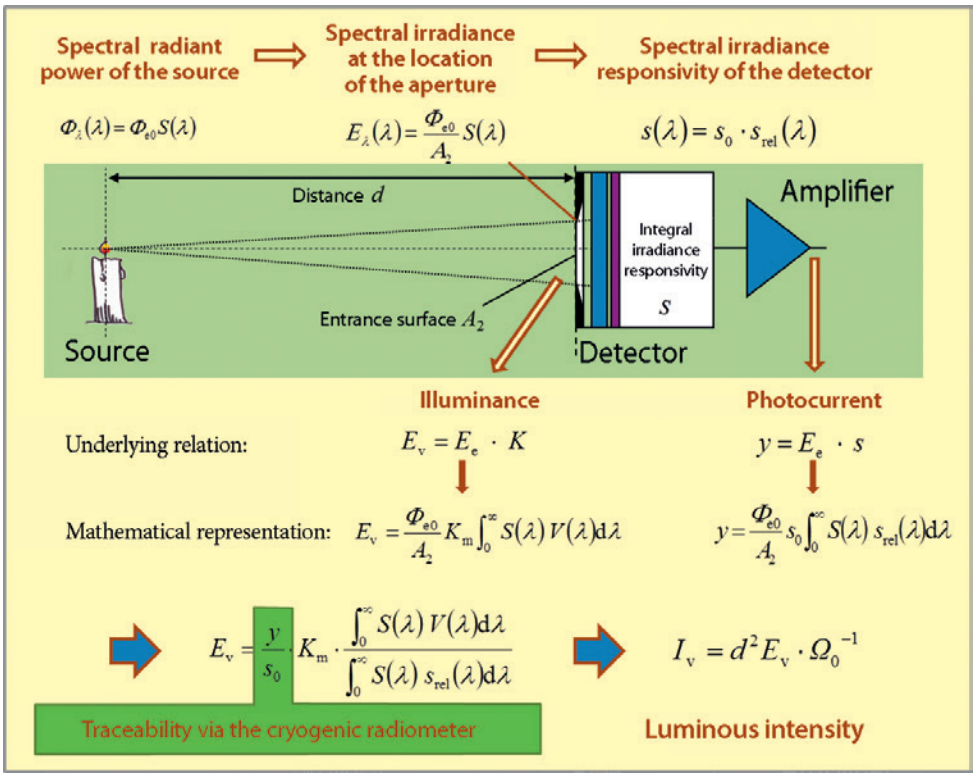


Fig. 7: Detector-aided realization of the unit for the quantity of luminous intensity. Here, the quantity Ω_0 describes the unity angle of a steradian [sr]. The quantities E_e, K and s are the values integrated across the entire spectrum for irradiance, luminous efficacy and responsivity, respectively.

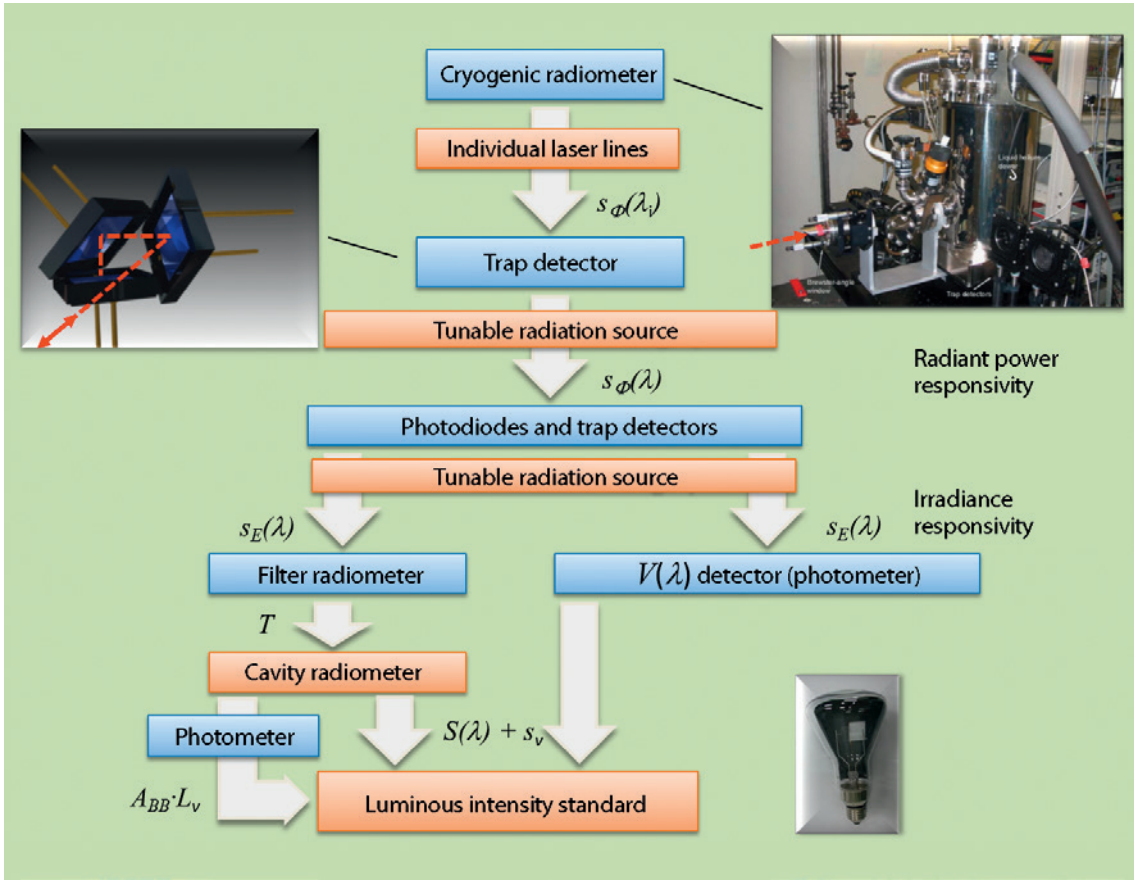


Fig. 8:
Traceability chain for the detector-aided traceability to the cryogenic radiometer.

Once a black-body radiator that fulfils the above measurement conditions has been found, other light sources can be connected to this primary standard via a direct comparison using a photometer (Fig. 6). The uncertainty of this realization depends primarily on the uncertainty of the geometry factor, as well as on the uncertainty when determining the cavity temperature.

Instead of a source-based realization of this nature, in which the unit of the quantity of luminous intensity is determined directly by means of the calculated radiation of a source, a detector-aided realization is also possible. A prerequisite for this is for the irradiance responsivity of the radiation detector used (i.e. the relation of the output signal measured to the irradiance detected) to be known precisely.

Since all known detectors have spectral dependencies, both the absolute spectral irradiance responsivity of the detector and the relative spectral distribution of the light source measured using the detector are required in order to calculate the illuminance from the spectral distribution weighted with $K_m \cdot V(\lambda)$ and the spectral irradiance measured. Via the squared distance, the luminous intensity can then be determined easily at great distances from the source (far field) (see Fig. 7).

In order to trace the spectral irradiance responsivity to the SI, a cryogenic radiometer is used: first, by means of this radiometer, the optical radiant power of a monochromatic source is compared to an electrical power. To do so, the optical radiation is fully absorbed in a black cavity, and the resultant heating of the cavity (after shielding the source) is substituted with an equally strong heating by adding an electrical heating power. In this way, radiant power can be expressed via electrical power.

Due to the material characteristics, the highest levels of sensitivity and the smallest uncertainties are reached at a cavity temperature of around 6 K, from which the name "cryogenic radiometer" is derived [12, 13].

After the radiant power of the monochromatic source has been precisely determined via this method, the source can be used to calibrate the spectral radiant power responsivity $s_\Phi(\lambda_i)$ of (for example) semiconductor detectors at a given wavelength (see Fig. 8). Here, so-called trap detectors are particularly well-suited; these detectors are made up of several single series-connected photodiodes (see Fig. 8 top left and [12]) that are very homogeneous across their detector area, and have a relative spectral responsivity distribution that can be calculated very well with a few additional assumptions. Thus, with a single wavelength of the cryogenic radiometer (or with only a very small number of wavelengths), an absolute link is obtained between the scale for the spectral radiant power responsivity $s_\Phi(\lambda)$ and the other SI units.

If the irradiated area of the detector is known in addition (to this end, a precisely calibrated aperture is positioned in front of the detector opening), the spectral irradiance responsivity $s_E(\lambda)$ can be derived from the spectral radiant power responsivity. In the next step, via comparison with the trap detector, other detectors (such as radiometers and pyrometers, as well as photometers) can be calibrated in terms of their spectral irradiance responsivity.

It is here that we come full circle: To determine the temperature of the black-body radiator for source-based traceability (see Fig. 6), a pyrometer is usually used whose spectral responsivity is determined beforehand via the calibration chain using the cryogenic radiometer.

The two methods specified here for the realization of this unit are also described in the *mise en pratique* of the BIPM for the candela [14].

5. Dissemination of the unit

Despite the above considerations, we have not yet reached our goal within the realm of photometry. If the relative spectral responsivity of the detector used (in this case, the photometer) were identical to the $V(\lambda)$-function, a single calibration of the photometer using the cyrogenic radiometer would be enough to determine correct illuminances for any spectral distributions (i.e. for any light source). However, in practice, the responsivity of a photometer deviates from the spectral efficiency function established. As a result, this spectral mismatch, whose effect always depends on the given spectral distribution of the source, creates incorrect measurement results. For this reason, in order to calibrate a light source using a calibrated photometer, the relative spectral distribution of the source is needed as well. This spectral distribution can be determined (for example) from the spectral comparison of the light source with a black-body radiator of a known temperature.

In order to gain better control of this problem, and in order to ensure comparability of measuring instruments, so-called illuminants (standard light types) are specified in photometry with defined spectral distributions [15]. Here, two illuminants, Illuminant A and Illuminant D65, are of special significance. Illuminant A in particular, which is the spectrum of a black-body radiator with a temperature of 2856 K, serves as a standard spectrum for the determination of the integral photometric responsivity s_v. In practice, all characteristics of photometric detectors are stated in relation to this spectral distribution. This is aided by the fact that even normal incandescent lamps have a spectral distribution that is very close to Planck's spectrum. Thus, in order to keep the uncertainty associated with the dissemination of the candela as small as possible, incandescent lamps set via their lamp current to a distribution temperature of around 2856 K are used as transfer standard lamps. (The distribution temperature is the temperature of a black-body radiator at which the spectral distribution is equal or almost equal to that of the temperature radiator observed.) This ensures that the correction factors designed to take spectral mismatches of the photometer into account during its calibration remain small. Here, it is possible to maintain and reproduce the lamp current (as a nominal value) with very high precision.

Illuminant D65 has a spectrum similar to that of the sun; this spectrum is mainly used to assess applications that use daylight. Here, D65 represents a set spectral distribution whose luminous colour is closest to that of a black-body radiator with a temperature of 6500 K.

6. Possible future developments

With the new wording of the candela definition in the sense of an "explicit constant definition", a definition has been established that eliminates the ambiguity of parts of the old wording.
According to BIPM, the current concept reads as follows [16].

> *The candela, symbol cd, is the SI unit of luminous intensity in a given direction. It is defined by taking the fixed numerical value of the luminous efficacy of monochromatic radiation of frequency 540 · 10^{12} Hz, K_{cd}, to be 683 when expressed in the unit cd sr W^{-1}, which is equal to lm W^{-1}, or kg^{-1} m^{-2} s^3 cd sr, where the kilogram, meter and second are defined in terms of h, c and $\Delta\nu_{Cs}$.*

Frequently, the very existence of the candela is called into question. However, the form currently chosen is the sole possibility of creating a coherent representation in the SI of the complex photometric measurement technique with its linked action spectra.

To date, maintaining and disseminating the unit of the candela via the luminous intensity of standard lamps specially designed for this purpose (see inset images, Fig. 8) has been the method that can be realized with the smallest uncertainties. The international ban on incandescent lamps, which has also resulted in manufacturing requirements for the production of incandescent lamps (including those used for scientific purposes) ceasing

to apply, means that measurement laboratories around the world are now increasingly lacking traceability artefacts.

Thus, it is often considered inevitable that a transition to spectroradiometric measurements will take place on the one hand, and that the unit will be disseminated via (for example) LED-based light sources on the other hand. However, the measurement uncertainties of a traceability chain based on this are often underestimated; suitable framework conditions for traceability will have to be found concerning the spectral distribution of LEDs.

Literature

[1] *J.W.T Walsh:* Photometry; Constable & Company LTD (1958).

[2] *H. Lux;* Das moderne Beleuchtungswesen, Verlag von B.G. Teubner in Leipzig (1914)

[3] Procés-verbaux des séances, Comité International des Poids et Mesures, 18, Paris, Gauthier–Villars (1937) 236.

[4] Bureau International des Poids et Mesures (ed.); Comptes Rendus des Séances de la 13e Conférence Générale des Poids et Mesures, Annex II: Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92310 Sèvres (1968).

[5] *H. Willenberg;* Die Lichtstärke-Einheit Candela, Technisches Messen, Volume 349, Issue 2, *pp. 21–22.*

[6] *C.R. Barber et al;* The Platinum Metals in the Measurement of Temperature, Platinum Metals Rev. 13 (2), (1968) 65–67.

[7] *G. Bauer;* Zusammenfassung der Messergebnisse der PTB zur Bestimmung des Strahlungsäquivalents $K_{m,77}$ mit dem Platin-Hohlraumstrahler für die 9. Sitzung des CCPR (7.–9. Sept. 1977) [Summary of the Measurement Results of PTB for the Determination of the Spectral Luminous Efficacy $K_{m,77}$ with the Platinum Cavity Radiator for the 9th Meeting of the CCPR (7.–9. Sept. 1977)]

[8] *G. Sauter;* Die Candela: Erläuterungen zum Verständnis der Definition und der Realisierung, PTB-Mitteilungen 107, 6/97, 397ff

[9] 16th General Conference on Weights and Measures (1979).

[10] *W.R. Blevin;* Corrections in Optical Pyrometry and Photometry for the Refractive Index of Air, Metrologia 8 (1972) 146–147.

[11] CIE 191:2010; Recommended System for Mesopic Photometry Based on Visual Performance, 73 pages.

[12] *K.D. Stock et al;* Trap-Empfänger und Dünnschicht-Thermosäulen als Bindeglieder zwischen Kryoradiometer und Gebrauchsempfänger. Laser in Forschung und Technik: Lectures of the 12th International Congress on Lasers 95, (1996), DOI 10.1007/978-3-642-802638_175.

[13] *J. E Martin, N. P. Fox, P. J. Key;* A cryogenic radiometer for absolute radiometric measurements, Metrologia 21 (1985) No 3, 147–155.

[14] *J. Zwinkels, A. Sperling, T. Goodmann, J. Campos Acosta, Y. Ohno, M. L. Rastello, M. Stock, E. Woolliams;* Mise en pratique for the definition of the candela and associated derived units for photometric and radiometric quantities in the International System of Units (SI), Metrologia 53, No. 3.

[15] ISO 11664-2:2007/CIE S 014–2:2006, Joint ISO/CIE Standard: Colorimetry-Part 2: Standard Illuminants for Colorimetry.

[16] BIPM: Draft text for the 9th edition of the SI Brochure, http://www.bipm.org/utils/common/pdf/si-brochure-draft-2016.pdf (last accessed on 22 June 2016).

PTB mitteilungen

Imprint

The *PTB-Mitteilungen* are the metrological specialist journal and the official information bulletin of the Physikalisch-Technische Bundesanstalt. As a specialist journal, the *PTB-Mitteilungen* publish original scientific contributions and general articles on metrological subjects from the areas of activities of the PTB. The individual volumes are focused on one subject. The *Mitteilungen* have a long tradition dating back to the beginnings of the Physikalisch-Technische Reichsanstalt (founded in 1887).

Editor
Physikalisch-Technische Bundesanstalt (PTB)
ISNI: 0000 0001 2186 1887
Braunschweig and Berlin
Postal address:
POB 33 45,
38023 Braunschweig
Delivery address:
Bundesallee 100,
38116 Braunschweig

Editorial Staff/Layout
Press and Information Office, PTB
Sabine Siems
Dr. Dr. Jens Simon
(Editor in Chief)
Dr. Thomas Middelmann
(Scientific Editor)
Phone: +49 531 592-82 02
Fax: +49 531 592-30 08
E-mail: sabine.siems@ptb.de

Translation (PTB Translation Office)
Undine Baier-Blott
Cécile Charvieux
Ella Jones
Charles Minnick
Kathleen Spalinger

Printed in Germany ISSN 0030-834X

The technical articles from this issue of the PTB-Mitteilungen are also available online at:
doi: 10.7795/310.20160299en02

Bundesministerium
für Wirtschaft
und Energie

The Physikalisch-Technische Bundesanstalt, Germany's national metrology institute, is a scientific and technical higher federal authority falling within the competence of the Federal Ministry for Economic Affairs and Energy.

The Physikalisch-Technische Bundesanstalt, Germany's national metrology institute, is a scientific and technical higher federal authority falling within the competence of the Federal Ministry for Economic Affairs and Energy.

Physikalisch-Technische Bundesanstalt
Braunschweig and Berlin
National Metrology Institute

Bundesallee 100
38116 Braunschweig, Germany

Press and Information Office

Phone: +49 531 592-30 06
Fax: +49 531 592-30 08
E-mail: press@ptb.de
www.ptb.de