Fritz Riehle

Frequency Standards

Basics and Applications

WILEY-
VCH

WILEY-VCH Verlag GmbH Co. KGaA

Fritz Riehle

Frequency Standards

Basics and Applications

Fritz Riehle

Frequency Standards

Basics and Applications

WILEY-VCH

WILEY-VCH Verlag GmbH Co. KGaA

Dr. Fritz Riehle
Physikalisch-Technische Bundesanstalt, Braunschweig
Germany
e-mail: fritz.riehle@ptb.de

Cover picture: Magneto-optical trap with 10 millions laser cooled calcium atoms in an optical frequency standard. Physikalisch-Technische Bundesanstalt, Braunschweig

To the memory of my parents

and to Hildegard and Ruth

Contents

Preface

The contributions of accurate time and frequency measurements to global trade, traffic and most sub-fields of technology and science, can hardly be overestimated. The availability of stable sources with accurately known frequencies is prerequisite to the operation of world-wide digital data networks and to accurate satellite positioning, to name only two examples. Accurate frequency measurements currently give the strongest bounds on the validity of fundamental theories. Frequency standards are intimately connected with developments in all of these and many other fields as they allow one to build the most accurate clocks and to combine the measurements, taken at different times and in different locations, into a common system.

The rapid development in these fields produces new knowledge and insight with breathtaking speed. This book is devoted to the basics and applications of frequency standards. Most of the material relevant to frequency standards is scattered in excellent books, review articles, or in scientific journals for use in the fields of electrical engineering, physics, metrology, astronomy, or others. In most cases such a treatise focusses on the specific applications, needs, and notations of the particular sub-field and often it is written for specialists. The present book is meant to serve a broader community of readers. It addresses both graduate students and practising engineers or physicists interested in a general and introductory actual view of a rapidly evolving field. The volume evolved from courses for graduate students given by the author at the universities of Hannover and Konstanz. In particular, the monograph aims to serve several purposes.

First, the book reviews the basic concepts of frequency standards from the microwave to the optical regime in a unified picture to be applied to the different areas. It includes selected topics from mechanics, atomic and solid state physics, optics, and methods of servo control. If possible, the topics which are commonly regarded as complicated, e.g., the principles and consequences of the theory of relativity, start with a simple physical description. The subject is then developed to the required level for an adequate understanding within the scope of this book.

Second, the realisation of commonly used components like oscillators or macroscopic and atomic frequency references, is discussed. Emphasis is laid not only on the understanding of basic principles and their applications but also on practical examples. Some of the subjects treated here may be of interest primarily to the more specialised reader. In these cases, for the sake of conciseness, the reader is supplied with an evaluated list of references addressing the subject in necessary detail.

Third, the book should provide the reader with a sufficiently detailed description of the most important frequency standards such as, e.g., the rubidium clock, the hydrogen maser, the caesium atomic clock, ion traps or frequency-stabilised lasers. The criteria for the "impor-

tance" of a frequency standard include their previous, current, and future impact on science and technology. Apart from record-breaking primary clocks our interest also focusses on tiny, cheap, and easy-to-handle standards as well as on systems that utilise synchronised clocks, e.g., in Global Navigation Satellite Systems.

Fourth, the book presents various applications of frequency standards in contemporary high-technology areas, at the forefront of basic research, in metrology, or for the quest for most accurate clocks. Even though it is possible only to a limited extent to predict future technical evolution on larger time scales, some likely developments will be outlined. The principal limits set by fundamental principles will be explored to enable the reader to understand the concepts now discussed and to reach or circumvent these limitations. Finally, apart from the aspect of providing a reference for students, engineers, and researchers the book is also meant to allow the reader to have intellectual fun and enjoyment on this guided walk through physics and technology.

Chapter 1 reviews the basic glossary and gives a brief history of the development of clocks. Chapters 2 and 3 deal with the characterisation of ideal and real oscillators. In Chapter 4 the properties of macroscopic and in Chapter 5 that of microscopic, i.e., atomic and molecular frequency references, are investigated. The most important methods for preparation and interrogation of the latter are given in Chapter 6. Particular examples of frequency standards from the microwave to the optical domain are treated in Chapters 7 to 10, emphasising their peculiarities and different working areas together with their main applications. Chapter 11 addresses selected principles and methods of measuring optical frequencies relevant for the most evolved current and future frequency standards. The measurement of time as a particular application of frequency standards is treated in Chapter 12. The remainder of the book is devoted to special applications and to the basic limits.

I would like to thank all colleagues for continuous help with useful discussions and for supporting me with all kinds of information and figures. I am thankful to the team of Wiley–VCH for their patience and help and to Hildegard for her permanent encouragement and for helping me with the figures and references. I am particularly grateful to A. Bauch, T. Binnewies, C. Degenhardt, J. Helmcke, P. Hetzel, H. Knöckel, E. Peik, D. Piester, J. Stenger, U. Sterr, Ch. Tamm, H. Telle, S. Weyers, and R. Wynands for careful reading parts of the manuscript. These colleagues are, however, not responsible for any deficiencies or the fact that particular topics in this book may require more patience and labour as adequate in order to be understood. Furthermore, as in any frequency standard, feedback is necessary and highly welcome to eliminate errors or to suggest better approaches for the benefit of future readers.

<div align="right">

Fritz Riehle
(fritz.riehle@ptb.de)
Braunschweig
June 2004

</div>

1 Introduction

1.1 Features of Frequency Standards and Clocks

Of all measurement quantities, frequency represents the one that can be determined with by far the highest degree of accuracy. The progress in frequency measurements achieved in the past allowed one to perform measurements of other physical and technical quantities with unprecedented precision, whenever they could be traced back to a frequency measurement. It is now possible to measure frequencies that are accurate to better than 1 part in 10^{15}. In order to compare and link the results to those that are obtained in different fields, at different locations, or at different times, a common base for the frequency measurements is necessary. Frequency standards are devices which are capable of producing stable and well known frequencies with a given accuracy and, hence, provide the necessary references over the huge range of frequencies (Fig. 1.1) of interest for science and technology. Frequency standards link the different areas by using a common unit, the hertz. As an example, consider two identical clocks whose
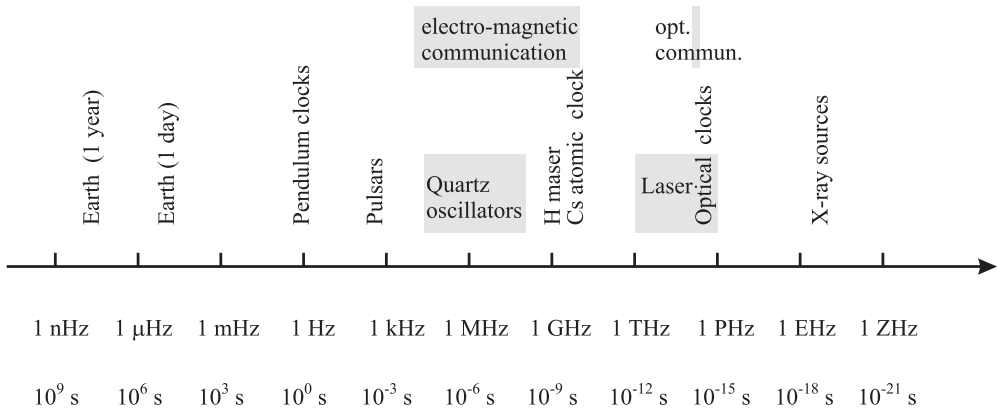
Figure 1.1: Frequency and corresponding time scale with clocks and relevant technical areas.

relative frequencies differ by 1×10^{-15}. Their readings would disagree by one second only after thirty million years. Apart from the important application to realise accurate clocks and time scales, frequency standards offer a wide range of applications due to the fact that numerous physical quantities can be determined very accurately from measurements of related frequencies. A prominent example of this is the measurement of the quantity *length*. Large distances are readily measured to a very high degree of accuracy by measurement of the time interval that a pulse of electromagnetic waves takes to traverse this distance. Radar guns used

by the police represent another example where the quantity of interest, i.e. the speed of a vehicle is determined by a time or frequency measurement. Other quantities like magnetic fields or electric voltages can be related directly to a frequency measurement using the field-dependent precession frequency of protons or using the Josephson effect, allowing for exceptionally high accuracies for the measurement of these quantities.

The progress in understanding and handling the results and inter-relationships of celestial mechanics, mechanics, solid-state physics and electronics, atomic physics, and optics has allowed one to master steadily increasing frequencies (Fig. 1.1) with correspondingly higher accuracy (Fig. 1.2). This evolution can be traced from the mechanical clocks (of resonant
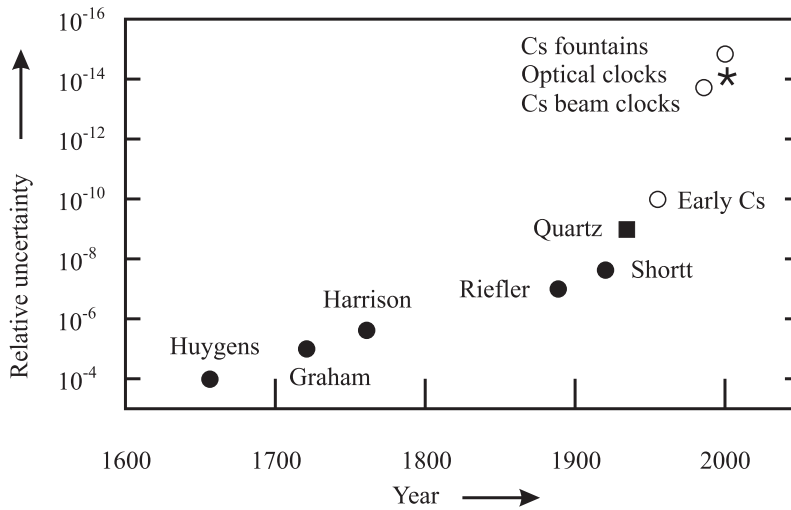
Figure 1.2: Relative uncertainty of different clocks. Mechanical pendulum clocks (full circles); quartz clock (full square); Cs atomic clocks (open circles); optical clocks (asterisk). For more details see Section 1.2.

frequencies $\nu_0 \approx 10^0$ Hz) via the quartz and radio transmitter technology (10^3 Hz $\leq \nu_0 \leq 10^8$ Hz), the microwave atomic clocks (10^8 Hz $\leq \nu_0 \leq 10^{10}$ Hz) to today's first optical clocks based on lasers ($\nu_0 \lesssim 10^{15}$ Hz). In parallel, present-day manufacturing technology with the development of smaller, more reliable, more powerful, and at the same time much cheaper electronic components, has extended the applications of frequency technology. The increasing use of quartz and radio controlled clocks, satellite based navigation for ships, aircraft and cars as well as the implementation of high-speed data networks would not have been possible without the parallel development of the corresponding oscillators, frequency standards, and synchronisation techniques.

Frequency standards are often characterised as active or passive devices. A "passive" frequency standard comprises a device or a material of particular sensitivity to a single frequency or a group of well defined frequencies (Fig. 1.3). Such a frequency reference may be based on macroscopic resonant devices like resonators (Section 4) or on microscopic quantum systems (Section 5) like an ensemble of atoms in an absorber cell. When interrogated by a suitable oscillator, the frequency dependence of the frequency reference may result in an absorption
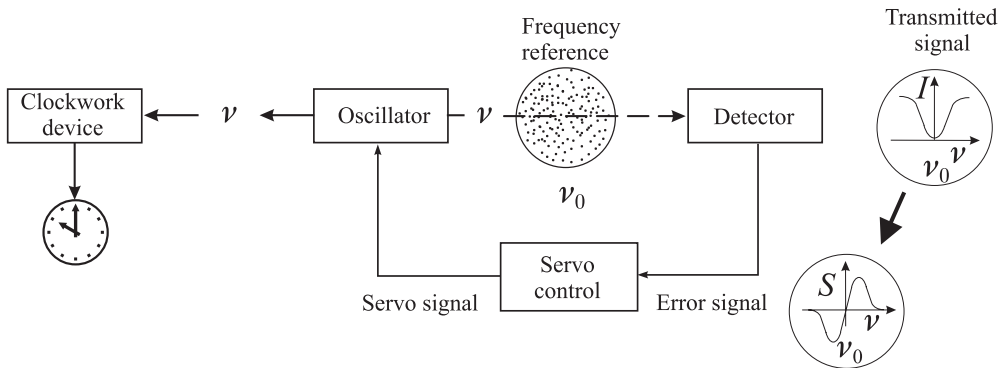
Figure 1.3: Schematics of frequency standard and clock.

line with a minimum of the transmission at the resonance frequency ν_0. From a symmetric absorption signal I an anti-symmetric error signal S may be derived that can be used in the servo-control system to generate a servo signal. The servo signal acting on the servo input of the oscillator is supposed to tune the frequency ν of the oscillator as close as possible to the frequency ν_0 of the reference. With a closed servo loop the frequency ν of the oscillator is "stabilised" or "locked" close to the reference frequency ν_0 and the device can be used as a frequency standard provided that ν is adequately known and stable.

In contrast to the passive standard an "active" standard is understood as a device where, e.g., an ensemble of excited atomic oscillators directly produces a signal with a given frequency determined by the properties of the atoms. The signal is highly coherent if a fraction of the emitted radiation is used to stimulate the emission of other excited atoms. Examples of active frequency standards include the active hydrogen maser (Section 8.1) or a gas laser like the He-Ne laser (Section 9.1).

A frequency standard can be used as a clock (Fig. 1.3) if the frequency is suitably divided in a clockwork device and displayed. As an example consider the case of a wrist watch where a quartz resonator (Section 4.1) defines the frequency of the oscillator at $32\,768\,\text{Hz} = 2^{15}\,\text{Hz}$ that is used with a divider to generate the pulses for a stepping motor that drives the second hand of the watch.

The specific requirements in different areas lead to a variety of different devices that are utilised as frequency standards. Despite the various different realisations of frequency standards for these different applications, two requirements are indispensable for any one of these devices. First, the frequency generated by the device has to be stable in time. The frequency, however, that is produced by a real device will in general vary to some extent. The variation may depend, e.g., on fluctuations of the ambient temperature, humidity, pressure, or on the operational conditions. We value a "good" standard by its capability to produce a stable frequency with only small variations.

A stable frequency source on its own, however, does not yet represent a frequency standard. It is furthermore necessary that the frequency ν is known in terms of absolute units. In the internationally adopted system of units (Systéme International: SI) the frequency is mea-

sured in units of Hertz representing the number of cycles in one second (1 Hz = 1/s). If the frequency of a particular stable device has been measured by comparing it to the frequency of another source that can be traced back to the frequency of a primary standard [1] used to *realise* the SI unit, our stable device then – and only then – represents a frequency standard.

After having fulfilled these two prerequisites, the device can be used to calibrate other stable oscillators as further secondary standards.
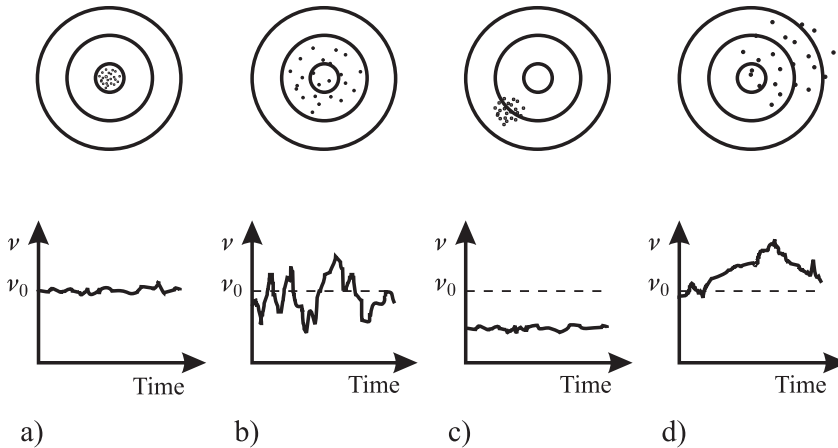
Figure 1.4: Bullet holes on a target (upper row) show four different patterns that are precise and accurate (a), not precise but accurate (b), precise but not accurate (c), not precise and not accurate (d). Correspondingly a frequency source (lower row) shows a frequency output that is stable and accurate (a), not stable but accurate (b), stable but not accurate (c), and not stable and not accurate (d).

There are certain terms like stability, precision, and accuracy that are often used to characterise the quality of a frequency standard. Some of those are nicely visualised in a picture used by Vig [2] who compared the temporal output of an oscillator with a marksman's sequence of bullet holes on a target (Fig. 1.4). The first figure from the left shows the results of a highly skilled marksman having a good gun at his disposal. All holes are positioned accurately in the centre with high precision from shot to shot. In a frequency source the sequence of firing bullets is replaced by consecutive measurements of the frequency ν, where the deviation of the frequency from the centre frequency ν_0 corresponds to the distance of each bullet hole from the centre of the target.[2] Such a stable and accurate frequency source may be used as a frequency standard. In the second picture of Fig. 1.4 the marks are scattered with lower precision but enclosing the centre accurately. The corresponding frequency source would suffer from reduced temporal stability but the mean frequency averaged over a longer period would be accurate. In the third picture all bullet holes are precisely located at a position off the centre. The corresponding frequency source would have a frequency offset from the desired

---

[1] A primary frequency standard is a frequency standard whose frequency correponds to the adopted definition of the second , with its specified accuracy achieved without external calibration of the device [1].

[2] The distances of bullet holes in the lower half plane are counted negative.

frequency ν_0. If this offset is stable in time the source can be used as a frequency standard provided that the offset is determined and subsequently corrected for. In the fourth picture most bullet holes are located to the right of the centre, maybe due to reduced mental concentration of the marksman. The corresponding oscillator produces a frequency being neither stable nor accurate and, hence, cannot be used as a frequency standard.

The accuracy and stability of the frequency source depicted in the third picture of Fig. 1.4 can be quantified by giving the deviation from the centre frequency and the scatter of the frequencies, respectively, in hertz. To compare completely different frequency standards the relative quantities "relative accuracy" ("relative stability", etc.) are used where the corresponding frequency deviation (frequency scatter) is divided by the centre frequency. As well as the terms accuracy, stability and precision the terms inaccuracy, instability and imprecision are also in current use and these allow one to characterise, e.g., a good standard with *low* inaccuracy by a *small* number corresponding to the *small* frequency deviation, whereas a *high* accuracy corresponds to a *small* frequency deviation.

The simple picture of a target of a marksman (Fig. 1.4) used to characterise the quality of a frequency standard is not adequate, however, in a number of very important cases. Consider, e.g., a standard which is believed to outperform all other available standards. Hence, there is no direct means to determine the accuracy with respect to a superior reference. This situation is equivalent to a plain target having neither a marked centre nor concentric rings. Shooting at the target, the precision of a gun or the marksman can still be determined but the accuracy cannot. It is, however, possible to "estimate" the uncertainty of a frequency standard similarly as it is done by measuring an *a priori* unknown measurand. There are now generally agreed procedures to determine the uncertainty in the Guide to the Expression of Uncertainty in Measurement (GUM) [3]. The specified uncertainty hence represents the "limits of the confidence interval of a measured or calculated quantity" [1] where the probability of the confidence limits should be specified. If the probability distribution is a Gaussian this is usually done by the standard deviation (1σ value) [3] corresponding to a confidence level of 68 %. For clarity we repeat here also the more exact definitions of accuracy as "the degree of conformity of a measured or calculated value to its definition" and precision as "the degree of mutual agreement among a series of individual measurements; often but not necessarily expressed by the standard deviation" [1].

1.2 Historical Perspective of Clocks and Frequency Standards

1.2.1 Nature's Clocks

The periodicity of the apparent movement of celestial bodies and the associated variations in daylight, seasons, or the tides at the seashore has governed all life on Earth from the very beginning. It seemed therefore obvious for mankind to group the relevant events and dates in chronological order by using the time intervals found in these periodicities as natural measures

---

[3] In cases where this confidence level is too low, expanded uncertainties with $k\,\sigma$ can be given, with, e.g., 95.5% ($k = 2$) or 99.7% ($k = 3$).

of time. Hence, the corresponding early calendars were based on days, months and years related to the standard frequencies of Earth's rotation around its polar axis (once a day), Earth's revolution around the Sun (once a year) and the monthly revolution of the moon around the Earth (once a month), respectively. The communication of a time interval between two or more parties had no ambiguity if all members referred to the same unit of time, e.g., the day, which then served as a natural standard of time. Similarly, a natural standard of frequency (one cycle per day) can be derived from such a natural clock. The calendar therefore allowed one to set up a time scale based on an agreed starting point and on the scale unit.[4] The establishment of a calendar was somewhat complicated by the fact that the ratios of the three above mentioned standard frequencies of revolution are not integers, as presently the tropical year [5] comprises 365.2422 days and the synodical month 29.5306 days.[6] Today's solar calendar with 365 days a year and a leap year with 366 days occurring every fourth year dates back to a Roman calendar introduced by Julius Caesar in the year 45 B.C.[7]

The use of Nature's clocks based on the movement of celestial bodies has two disadvantages. First, a good time scale requires that the scale unit must not vary with time. Arguments delivered by astronomy and geochronometry show that the ratio of Earth's orbital angular frequency around the sun and the angular frequency around its polar axis is not constant in time.[8] Second, as a result of the low revolution frequency of macroscopic celestial bodies the scale unit is in general too large for technical applications.[9]

1.2.2 Man-made Clocks and Frequency Standards

Consequently, during the time of the great civilisations of the Sumerians in the valley of Tigris and Euphrates and of the Egyptians, the time of the day was already divided into shorter sections and the calendars were supplemented by man-made clocks. A clock is a device that indicates equal increments of elapsed time. In the long time till the end of the Middle Ages the precursors of today's clocks included sundials, water clocks, or sand glasses with a variety of modifications. The latter clocks use water or sand flowing at a more or less constant rate and use the integrated quantity of moved substance to approximate a constant flow of time. Progress in clock making arose when oscillatory systems were employed that

---

[4] The set-up of a time scale, however, is by no means exclusively related to cyclic events. In particular, for larger periods of time, the exponential decay of some radioactive substances, e.g., of the carbon isotope ^{14}C allows one to infer the duration of an elapsed time interval from the determination of a continuously decreasing ratio ^{14}C/^{12}C.

[5] The tropical year is the time interval between two successive passages of the sun through the vernal equinox, i.e. the beginning of spring on the northern hemisphere.

[6] The synodical month is the time interval between two successive new moon events. The term "synode" meaning "gathering" refers to the new moon, when moon and sun gather together as viewed from the earth.

[7] The rule for the leap year was modified by Pope Gregor XIII in the year 1582 so that for year numbers being an integer multiple of 100, there is no leap year except for those years being an integer multiple of 400. According to this, the mean year in the Gregorian Calendar has 365.2425 days, close to its value given above.

[8] The growth of reef corals shows ridges comparable to the tree rings that have been interpreted as variations in the rate of carbonate secretion both with a daily and annual variation. The corresponding ratios of the ridges are explained by the fact that the year in the Jurassic (135 million years ago) had about 377 days [4].

[9] Rapidly spinning millisecond pulsars can represent "Nature's most stable clocks" [5], but their frequency is still too low for a number of today's requirements.

operate at a specific resonance frequency defined by the properties of the oscillatory system. If the oscillation frequency ν_0 of this system is known, its reciprocal defines a time increment $T = 1/\nu_0$. Hence, any time interval can be measured by counting the number of elapsed cycles and multiplying this number with the period of time T. Any device that produces a known frequency is called a frequency standard and, hence, can be used to set up a clock. To produce a good clock requires the design of a system where the oscillation frequency is not perturbed either by changes in the environment, by the operating conditions or by the clockwork.

1.2.2.1   Mechanical Clocks

In mechanical clocks, the clockwork fulfils two different tasks. Its first function is to measure and to display the frequency of the oscillator or the elapsed time. Secondly, it feeds back to the oscillator the energy that is required to sustain the oscillation. This energy from an external source is needed since any freely oscillating system is coupled to the environment and the dissipated energy will eventually cause the oscillating system to come to rest. In mechanical devices the energy flow is regulated by a so-called escapement whose function is to steer the clockwork with as little as possible back action onto the oscillator. From the early fourteenth century large mechanical clocks based on oscillating systems were used in the clock towers of Italian cathedrals. The energy for the clockwork was provided by weights that lose potential energy while descending in the gravitational potential of the Earth. These clocks were regulated by a so-called verge-and-foliot escapement which was based on a kind of torsion pendulum. Even though these clocks rested essentially on the same principles (later successfully used for much higher accuracies) their actual realisation made them very susceptible to friction in the clockwork and to the driving force. They are believed to have been accurate to about a quarter of an hour a day. The relative uncertainty of the frequency of the oscillator steering these clocks hence can be described by a fractional uncertainty of $\Delta T/T = \Delta\nu/\nu \approx 1\,\%$. The starting point of high-quality pendulum clocks is often traced back to an observation of the Italian researcher Galileo Galilei (1564 – 1642). Galilei found that the oscillation period of a pendulum for not too large excursions virtually does not depend on the excursion but rather is a function of the length of the pendulum. The first workable pendulum clock, however, was invented in 1656 by the Dutch physicist Christian Huygens. This clock is reported to have been accurate to a minute per day and later to better than ten seconds per day corresponding to $\Delta T/T \approx 10^{-4}$ (see Fig. 1.2). Huygens is also credited with the development of a balance-wheel-and-spring assembly. The pendulum clock was further improved by George Graham (1721) who used a compensation technique for the temperature dependent length of the pendulum arriving at an accuracy of one second per day ($\Delta T/T \approx 10^{-5}$).

The contribution of accurate clocks to the progress in traffic and traffic safety can be exemplified from the development of a marine chronometer by John Harrison in the year 1761. Based on a spring-and-balance-wheel escapement the clock was accurate to 0.2 seconds per day ($\Delta T/T \approx 2 - 3 \times 10^{-6}$) even in a rolling marine vessel. Harrison's chronometer for the first time solved the problem of how to accurately determine longitude during a journey [6]. Continuous improvements culminated in very stable pendulum clocks like the ones manufactured by Riefler in Germany at the end of the nineteenth century. Riefler clocks were stable to a hundredth of a second a day ($\Delta T/T \approx 10^{-7}$) and served as time-interval standards in

the newly established National Standards Institutes until about the twenties of the past century before being replaced by the Shortt clock. William H. Shortt in 1920 developed a clock with two synchronised pendulums. One pendulum, the master, swung as unperturbed as possible in an evacuated housing. The slave pendulum driving the clockwork device was synchronised via an electromagnetic linkage and in turn, every half a minute, initialised a gentle push to the master pendulum to compensate for the dissipated energy. The Shortt clocks kept time better than 2 milliseconds a day ($\Delta T/T \approx 2 \times 10^{-8}$) and to better than a second per year ($\Delta T/T \approx 3 \times 10^{-8}$).

1.2.2.2   Quartz Clocks

Around 1930 quartz oscillators (Section 4.1) oscillating at frequencies around $100\,\text{kHz}$, with auxiliary circuitry and temperature-control equipment, were used as standards of radio frequency and later replaced mechanical clocks for time measurement. The frequency of quartz clocks depends on the period of a suitable elastic oscillation of a carefully cut and prepared quartz crystal. The mechanical oscillation is coupled to electronically generated electric oscillation via the piezoelectric effect. Quartz oscillators drifted in frequency about 1 ms per day ($\Delta \nu/\nu \approx 10^{-8}$) [7] and, hence, did not represent a frequency standard unless calibrated. At this time, frequency calibration was derived from the difference in accurate measurements of mean solar time determined from astronomical observations.

The quartz oscillators (denoted as "Quartz" in Fig. 1.2) proved their superiority with respect to mechanical clocks and the rotating Earth at the latest when Scheibe and Adelsberger showed [7] that from the beginning of 1934 till mid 1935, the three quartz clocks of the Physikalisch-Technische Reichsanstalt, Germany all showed the same deviation from the siderial day. The researchers concluded that the apparent deviations resulted from a systematic error with the time determination of the astronomical institutes as a result of the variation of Earth's angular velocity.[10] Today, quartz oscillators are used in numerous applications and virtually all battery operated watches are based on quartz oscillators.

1.2.2.3   Microwave Atomic Clocks

Atomic clocks differ from mechanical clocks in such a way that they employ a quantum mechanical system as a "pendulum" where the oscillation frequency is related to the energy difference between two quantum states. These oscillators could be interrogated, i.e. coupled to a clockwork device only after coherent electromagnetic waves could be produced. Consequently, this development took place shortly after the development of the suitable radar and microwave technology in the 1940s. Detailed descriptions of the early history that led to the invention of atomic clocks are available from the researchers of that period (see e.g. [9–12]) and we can restrict ourselves here to briefly highlighting some of the breakthroughs. One of the earliest suggestions to build an atomic clock using magnetic resonance in an atomic beam was given by Isidor Rabi who received the Nobel prize in 1944 for the invention of this spectroscopic technique. The successful story of the Cs atomic clocks began between 1948 and

---

[10] T. Jones [8] points out that "The first indications of seasonal variations in the Earth's rotation were gleaned by the use of Shortt clocks."

1955 when several teams in the USA including the National Bureau of Standards (NBS, now National Institute of Standards and Technology, NIST) and in England at the National Physical Laboratory (NPL) developed atomic beam machines. They relied on Norman Ramsey's idea of using separated field excitation (Section 6.6) to achieve the desired small linewidth of the resonance. Essen and Parry at NPL (denoted as "Early Cs" in Fig. 1.2) operated the first laboratory Cs atomic frequency standard and measured the frequency of the Cs ground-state hyperfine transition [13, 14]. Soon after (1958) the first commercial Cs atomic clocks became available [15]. In the following decades a number of Cs laboratory frequency standards were developed all over the world with the accuracy of the best clocks improving roughly by an order of magnitude per decade. This development led to the re-definition of the second in 1967 when the 13th General Conference on Weights and Measures (CGPM) defined the second 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 caesium 133 atom". Two decades later the relative uncertainty of a caesium beam clock (e.g., CS2 in 1986 at the Physikalisch-Technische Bundesanstalt (PTB), Germany, denoted as "Cs beam cock" in Fig. 1.2) was already as low as 2.2×10^{-14} [16].

A new era of caesium clocks began when the prototype of an atomic Cs fountain was set up [17] at the Laboratoire Primaire du Temps and Fréquences (LPTF; now BNM–SYRTE) in Paris. In such clocks Cs atoms are laser cooled and follow a ballistic flight in the gravitational field for about one second. The long interaction time made possible by the methods of laser cooling (Section 6.3.1) leads to a reduced linewidth of the resonance curve. The low velocities of the caesium atoms allowed one to reduce several contributions that shift the frequency of the clock. Less than a decade after the first implementation, the relative uncertainty of fountain clocks was about 1×10^{-15} [18–20] (see "Cs fountain clock" in Fig. 1.2).

1.2.2.4   Optical Clocks and Outlook to the Future

As a conclusion of the historical overview one finds that the development of increasingly more accurate frequency standards was paralleled by an increased frequency of the employed oscillator. From the hertz regime of pendulum clocks via the megahertz regime of quartz oscillators to the gigahertz regime of microwave atomic clocks, the frequency of the oscillators has been increased by ten orders of magnitude. The higher frequency has several advantages. First, for a given linewidth $\Delta\nu$ of the absorption feature, the reciprocal of the relative linewidth, often referred to as the line quality factor

$$Q \equiv \nu_0/\Delta\nu, \tag{1.1}$$

increases. For a given capability to "split the line", i.e. to locate the centre of a resonance line, the frequency uncertainty is proportional to Q and hence, to the frequency of the interrogating oscillator. The second advantage of higher frequencies becomes clear if one considers two of the best pendulum clocks with the same frequency of about 1 Hz which differ by a second after a year ($\Delta\nu/\nu_0 \approx 3 \times 10^{-8}$). If both pendulums are swinging in phase it takes about half a year to detect the pendulums of the two clocks being out of phase by 180°. With two clocks operating at a frequency near 10 GHz the same difference would show up after 1.6 ms. Hence, the investigation and the suppression of systematic effects that shift the frequency of a standard is greatly facilitated by the use of higher frequencies. One can therefore expect

further improvements by the use of optical frequency standards by as much as five orders of magnitude higher frequencies, compared with that of the microwave standards. The recent development of frequency dividers from the optical to the microwave domain (Section 11) also makes them available for optical clocks [21] which become competitive with the best microwave clocks (see "Optical clocks" in Fig. 1.2).

It can now be foreseen that several (mainly optical) frequency standards might be realised whose reproducibilities are superior to the best clocks based on Caesium. As long as the definition of the unit of time is based on the hyperfine transition in caesium, these standards will not be capable to realise the second or the hertz better than the best caesium clocks. However, they will serve as secondary standards and will allow more accurate frequency ratios and eventually may lead to a new definition of the unit of time.

2 Basics of Frequency Standards

2.1 Mathematical Description of Oscillations

A great variety of processes in nature and technology are each unique in the sense that the same event occurs periodically after a well defined time interval T. The height of the sea level shows a maximum roughly every twelve hours ($T \approx 12.4\,\text{h}$). Similarly, the swing of a pendulum ($T \lesssim 1\,\text{s}$), the electric voltage available at the wall socket ($T \approx 0.02\,\text{s}$), the electric field strength of an FM radio transmitter ($T \approx 10^{-8}\,\text{s}$) or of a light wave emitted by an atom ($T \approx 2 \times 10^{-15}\,\text{s}$) represent periodic events. In each case a particular physical quantity $U(t)$, e.g., the height of the water above mean sea level or the voltage of the power line, performs oscillations.

2.1.1 Ideal and Real Harmonic Oscillators

Even though the time interval T and the corresponding frequency $\nu_0 \equiv 1/T$ differ markedly in the examples given, their oscillations are often described by an (ideal) harmonic oscillation

$$U(t) = U_0 \cos(\omega_0 t + \phi). \tag{2.1}$$

Given the amplitude U_0, the frequency

$$\nu_0 = \frac{\omega_0}{2\pi} \tag{2.2}$$

and the initial phase ϕ, the instantaneous value of the quantity of interest $U(t)$ of the oscillator is known at any time t. ω_0 is referred to as the angular frequency and $\varphi \equiv \omega_0 t + \phi$ as the instantaneous phase of the harmonic oscillator. The initial phase determines $U(t)$ for the (arbitrarily chosen) starting time at $t = 0$.

The harmonic oscillation (2.1) is the solution of a differential equation describing an ideal harmonic oscillator. As an example, consider the mechanical oscillator where a massive body is connected to a steel spring. If the spring is elongated by U from the equilibrium position there is a force trying to pull back the mass m. For a number of materials the restoring force $F(t)$ is to a good approximation proportional to the elongation

$$F(t) = -DU(t). \qquad \text{(Hooke's law)} \tag{2.3}$$

The constant D in Hooke's law (2.3) is determined by the stiffness of the spring which depends on the material and the dimensions of the spring. This force, on the other hand, accelerates

the mass with an acceleration $a(t) = d^2U(t)/dt^2 = F/m$. Equating both conditions for any instant of time t leads to the differential equation

$$\frac{d^2U(t)}{dt^2} + \omega_0^2\, U(t) = 0 \qquad \text{with} \qquad \omega_0 \equiv \sqrt{\frac{D}{m}}. \tag{2.4}$$

(2.1) is a solution of (2.4) as can be readily checked. The angular frequency ω_0 of the oscillator is determined by the material properties of the oscillator. In the case of the oscillating mass, the angular frequency ω_0 is given according to (2.4) by the mass m and the spring constant D.

If we had chosen the example of an electrical resonant circuit, comprising a capacitor of capacitance C and a coil of inductance L the frequency angular would be $\omega_0 = 1/(\sqrt{LC})$. In contrast, for an atomic oscillator the resonant frequency is determined by atomic properties. In the remainder of this chapter and in the next one we will not specify the properties of particular oscillators but rather deal with a more general description.

It is common to all oscillators that a certain amount of energy is needed to start the oscillation. In the case of a spring system potential energy is stored in the compressed spring elongated from equilibrium by U_0. When the system is left on its own, the spring will exert a force to the massive body and accelerate it. The velocity $v = dU(t)/dt$ of the body will increase and it will gain the kinetic energy

$$E_{\text{kin}}(t) = \frac{1}{2}mv^2 = \frac{1}{2}m\left[\frac{d\,U(t)}{dt}\right]^2 = \frac{1}{2}m\omega_0^2 U_0^2 \sin^2(\omega_0 t + \phi) \tag{2.5}$$

where we have made use of (2.1). The kinetic energy of the oscillating system increases to a maximum value as long as there is a force acting on the body. This force vanishes at equilibrium, i.e. when $\sin^2(\omega_0 t + \phi) = 1$ and the total energy equals the maximum kinetic (or maximum potential) energy

$$E_{\text{tot}} = \frac{1}{2}m\omega_0^2 U_0^2 = \frac{1}{2}DU_0^2. \tag{2.6}$$

The proportionality between the energy [1] stored in the oscillatory motion and the square of the amplitude is a feature which is common to all oscillators.

Rather than using a cosine function to describe the harmonic oscillation of (2.1) we could also use a sine function. As is evident from $\cos\varphi = \sin(\varphi + \frac{\pi}{2})$ only the starting phase ϕ would change by $\pi/2$. More generally, each harmonic oscillation can be described as a

---

[1] The energy discussed here is the energy stored in the oscillation of an oscillator that has been switched on and that would be oscillating forever if no dissipative process would reduce this energy. It must not be mixed with the energy that can be extracted from a technical oscillator which uses another source of energy to sustain the oscillation. The voltage $U(t)$, for instance, present at the terminals of such an oscillator is capable of supplying a current $I(t)$ to a device of input resistance R. This current $I(t) = U(t)/R$ produces a temporally varying electrical power $P(t) = U(t)I(t) = U^2(t)/R = U_0^2/R\cos^2(\omega_0 t + \phi)$ at the external device. The mean power $\bar{P}$, i.e. the power integrated for one period $\int_0^T U_0^2/R\cos^2(\omega_0 t + \phi)dt = U_0^2/(2R)$ is also proportional to U_0^2, as well as the energy $E(t') = \int_0^{t'} P(t)dt = U_0^2/R \int_0^{t'} \cos^2(\omega_0 t + \phi)dt$ delivered by the oscillator within the time t'. In contrast to the energy stored in an undamped oscillator this energy $\int_0^{t'} P(t)dt$ increases linearly with time t'.

superposition of a sine function and a cosine function having the same frequency as follows

$$
\begin{aligned}
U(t) = U_0 \cos(\omega_0 t + \phi) &= U_0 \cos(\omega_0 t) \cos(\phi) - U_0 \sin(\omega_0 t) \sin(\phi) \\
&= U_{01} \cos(\omega_0 t) - U_{02} \sin(\omega_0 t)
\end{aligned}
\tag{2.7}
$$

where we have used $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$. The two quantities $U_{01} = U_0 \cos\phi$ and $U_{02} = U_0 \sin\phi$ are termed quadrature amplitudes of the oscillation. As computations including sine and cosine functions can sometimes become awkward it is more convenient to describe the harmonic oscillation by a complex exponential using Euler's formula $\exp i\varphi = \cos\varphi + i\sin\varphi$. Then (2.1) can be replaced by

$$
\begin{aligned}
U(t) &= \Re\left\{U_0 e^{i(\omega_0 t + \phi)}\right\} = \Re\left\{\widetilde{U}_0 e^{i\omega_0 t}\right\} = \frac{\widetilde{U}_0 e^{i\omega_0 t} + \widetilde{U}_0^* e^{-i\omega_0 t}}{2} \\
&= \frac{1}{2}\left\{\widetilde{U}_0 e^{i\omega_0 t} + c.c.\right\},
\end{aligned}
\tag{2.8}
$$

with the complex phasor

$$
\widetilde{U}_0 = U_0\, e^{i\phi} = U_{01} + i\,U_{02}.
\tag{2.9}
$$

The phasor $\widetilde{U}_0$ contains the modulus $U_0 = |U(t)|$ and the starting phase angle in a single complex number. Calculations using the complex representations of the oscillation take advantage of the simple rules for dealing with complex exponentials. Having obtained the final (complex) result one keeps only the real part.[2] Accordingly, there are different ways to represent the ideal harmonic oscillation of (2.8) graphically. To depict the oscillation in the time

a)   b)   c)
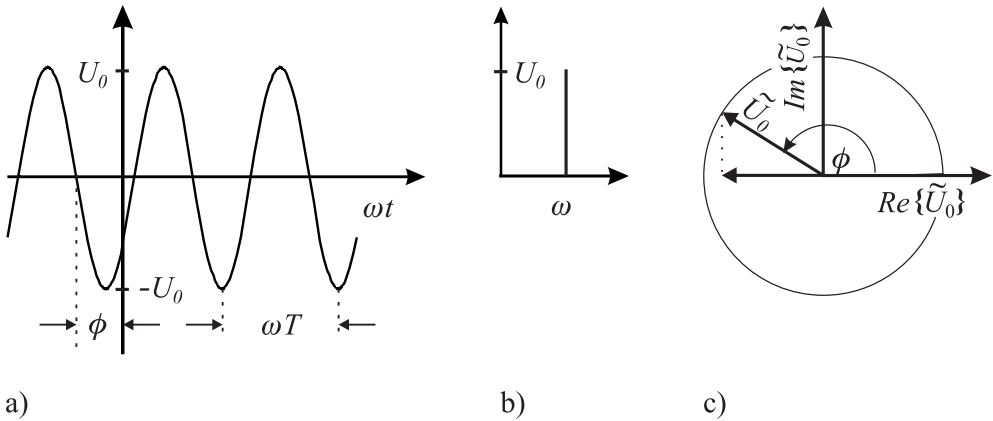
Figure 2.1: Ideal harmonic oscillator. a) Time-domain representation. b) Frequency-domain representation. c) Phasor representation.

---

[2] For simplicity, the operator $\Re\mathrm{e}$ is often not written in the course of the complex computations and the real part is taken only at the final result. Notice, however, that this procedure is only applicable in the case of linear operations as e.g. addition, multiplication with a number, integration or differentiation, but not in the case of non-linear operations. This can be seen in the case of the product of two complex numbers where obviously in general $\Re\mathrm{e}\,(A^2) \neq [\Re\mathrm{e}\,(A)]^2$.

domain (Fig. 2.1 a) it is necessary to know the initial phase ϕ, the amplitude U_0, and frequency $\nu_0 = 1/T$ of the oscillation. The oscillation in the frequency domain (Fig. 2.1 b) does not contain any information on the phase of the oscillator. If represented by a complex phasor, $\widetilde{U}_0 = U_0 \exp(i\phi)$ may be visualised in the complex plane (Argand diagram; Fig. 2.1 c) by a pointer of length U_0 that can be represented either in polar coordinates or in Cartesian coordinates. The initial phase is depicted as the angle ϕ between the real coordinate axis and the pointer. The phasor must not be mixed with the complex pointer $U_0 \exp[i(\omega_0 t + \phi)]$ that rotates counterclockwise [3] at constant angular velocity ω_0.

A specific property of the ideal harmonic oscillator is that we can predict its phase according to (2.1) starting from the initial conditions (phase, amplitude, and frequency) at any instant with any desired accuracy. For real oscillators used as examples above, these properties can be predicted only with an inherent uncertainty. For instance, the tidewaters do not always rise and fall to the same levels, but also show from time to time exceptionally high spring tides. In this case the amplitude of the oscillation resulting from the attraction of the moon is also "modulated" by the gravitational influence of the sun. In the example of the swinging pendulum the amplitude is constant only if the energy dissipated by friction is compensated for. Otherwise the amplitude of the swinging pendulum will die away similar to the amplitude of an oscillating atom emitting a wave train. In reality neither the amplitude nor the frequency of a real oscillator are truly constant. The long-term frequency variation may be very small as in the case of the ocean tides, where the angular velocity of the earth is decreasing gradually by friction processes induced by the tides of the waters and the solid earth but will become important after a large number of oscillations (see footnote 8 in Chapter 1). Apart from the natural modulations encountered in these two examples, the frequency of an oscillator may also be modulated on purpose. The frequency of the electromagnetic field produced by a FM (frequency modulated) transmitter is modulated to transmit speech and music. Basically, one refers to any temporal variation in the amplitude of an oscillator as amplitude modulation and the variation of its phase or frequency as phase modulation or frequency modulation, respectively. In the following we will investigate the processes of amplitude and phase modulation of an oscillator in more detail and we will develop the methods for the description.

For the oscillators relevant for frequency standards one may assume that the modulation represents only a small perturbation of the constant amplitude U_0 and of the phase $\omega_0 t$. An amplitude-modulated signal can then be written as

$$U(t) = U_0(t) \cos\varphi(t) = [U_0 + \Delta U_0(t)] \cos[\omega_0 t + \phi(t)]. \tag{2.10}$$

The instantaneous frequency

$$\nu(t) \equiv \frac{1}{2\pi} \frac{d\varphi(t)}{dt} = \frac{1}{2\pi} \frac{d}{dt}[2\pi\nu_0 t + \phi(t)] = \nu_0 + \frac{1}{2\pi} \frac{d\phi(t)}{dt} \tag{2.11}$$

differs from the frequency ν_0 of the ideal oscillator by

$$\Delta\nu(t) \equiv \frac{1}{2\pi} \frac{d\phi(t)}{dt}. \tag{2.12}$$

---

[3] Actually, there is an equivalent way to describe the oscillation mathematically by choosing a negative phase in (2.8), i.e. writing $U(t) = \Re\left\{\widetilde{U}_0 e^{-i\omega_0 t}\right\}$. From $e^{-i\phi} = \cos\phi - i\sin\phi$ it is clear that the pointer then rotates clockwise in the Argand diagram. As a consequence, one would have to change the sign of the imaginary part of several quantities. In the cases when this might lead to confusion we shall explicitly refer to this point again.