The Measurement of Time

Time, Frequency and the Atomic Clock

Claude Audoin
CNRS, Orsay, France

Bernard Guinot
Paris Observatory, France

Translated by Stephen Lyle

CAMBRIDGE
UNIVERSITY PRESS

4

The evolution of time measurement

When measuring time, just as when measuring any other physical quantity, it would seem logical to begin by evaluating metrological needs in view of applications, and then to seek ways of satisfying them. This approach was indeed attempted for astronomical dynamics. But more often than not, standards were realised before considering the use to which they would be put; and then those uses themselves preceded official recognition of the standards by organisations set up to manage metrology on a worldwide scale. In this sense, metrology is a strange combination of pragmatism and rigour.

The main part of this book is devoted to presenting a snapshot of time measurement as it is practised at the time of writing. This contrasts with the present chapter, in which we describe how such measurements have evolved, in terms of both techniques and ideas. It is a story punctuated by hesitation, doubt and sometimes even inconsistency.

4.1 Date, calendar and hour

The date of an event in the scientific sense is a whole set of data attributing a time label to it relative to some specified time scale. Traditionally, it comprises some way of identifying the day, known as a calendar, together with a subdivision of the day which we commonly call hours.

The various calendars (Gregorian, Jewish, Islamic, etc.) are ingenious schemes for making use of the natural cycles that have precise astronomical definitions:

- the apparent *solar day* is the duration between two successive transits of the Sun at the local meridian,
- the *tropical year* is the duration between two transits of the Sun through the vernal equinox (the spring equinox),

- the *lunation* (or lunar month) is the duration between two successive new moons.

We shall restrict ourselves mainly to the well-known Gregorian calendar, which respects the tropical year to a good approximation.†

The day is subdivided into hours, minutes and seconds in a way inherited from the Babylonians [4.1]. It has resisted any attempts at decimalisation. In France, a decree was issued on 24 November 1793 with the intention of imposing a decimal division of the day and a *decimal second*, but it was suspended on 7 April 1795. It is only for intervals shorter than one second that decimal submultiples are used, i.e., the millisecond, microsecond, etc.

In this complex system, a whimsical pocket of resistance to technocratic intervention, the units do not even bear constant relation to one another. Added to the fact that a year may contain either 365 or 366 days, a further claim to originality was to come in 1971. The day, normally lasting 86 400 s, can sometimes include one second more or less in the *Coordinated Universal Time* (UTC) that governs our lives.

In order to simplify their work, astronomers sometimes use the *Julian Date* (JD). This is based on a continuous count of the days since 4713 BC. The Julian Date can be complemented by giving also the decimal fraction of the day. In this system, days are counted from midday. For example, 1 January 2000 at midday corresponds to JD = 2 451 545.0.

In time metrology, the study of the Earth's rotation, and space science, the *Modified Julian Date* (MJD) is often used. This is defined by

$$\mathrm{MJD} = \mathrm{JD} - 2\,400\,000.5\,. \tag{4.1}$$

This means that MJD = 0.0... corresponds to 17 November 1858 at time 0 h. (Note that we are following international usage as regards acronyms. MJD comes from English, whilst International Atomic Time TAI comes from the French *Temps atomique international*.)

By their definitions, the calendar and the JD and MJD systems are closely linked to alternation between day and night. They are nevertheless used with time scales based upon other phenomena, such as International Atomic Time, when these systems drift only slightly relative to the true solar day. No ambiguity can result, provided that the time scale is indicated in the record of the date.

---

† Between the Gregorian reform of 1582 and about the year 3200, the Gregorian calendar will gain roughly one day over the tropical calendar. The uncertainty in this gain is mainly due to unpredictable irregularities in the Earth's rotation.

4.2 Time measurement based on alternation of day and night

4.2.1 Mean solar time

In ancient times, the day, from sunrise to sunset, and the night, from sunset to sunrise, were almost invariably divided into 12 hours each. Naturally, these hours did not have the same duration in the day and at night, except at the equinoxes or on the equator, and they varied with season and latitude. These variable hours, or *seasonal hours*, were still being used in the fifteenth century, even though ancient astronomers had already invented and used *equinoctial hours* long before, dividing the apparent solar day between successive passages of the Sun across the local meridian into 24 equal periods.

More precisely, the *apparent solar time* (or apparent time) is defined as the hour angle of the Sun, i.e., the angle between the half-plane of the observer's meridian and the half-plane defined by the Earth's axis of rotation and the Sun. This angle is then counted in 'hours' of 15°. (Astronomers do not always clearly distinguish between angles and times, as is revealed by the fact that the astronomical ephemerides of the Bureau des longitudes in Paris are called *Connaissance des temps*.)

Apparent solar time is directly observable and was still used in country areas up until the beginning of the twentieth century. However, it is not a 'convenient' time in the sense that H. Poincaré would have understood the term (see Section 2.3). We may say that it is not a uniform time. Its irregularities were already known to Ptolemy (*circa* AD 150), and are mainly due to the elliptical shape of the Earth's orbit around the Sun and the tilting of the Earth's axis of rotation relative to the ecliptic (the plane of its orbit). They have a total amplitude of 30 minutes and are reproduced cyclically in an identical manner each year, as described by the so-called *equation of time* (see Figure 4.1). Astronomers took these variations into account by using a regularised time known as *mean solar time*. This time scale is, like apparent time, associated with the local meridian. However, attempts to perfect the system did not stop there. The ultimate aim was to guarantee strict proportionality between mean solar time and the angle through which the Earth turns about its axis. In Chapter 8, we shall discover some of the subtleties of this requirement.

Despite the advantages brought about by its regularity, mean solar time was long considered a mere tool in the astronomer's panoply. Once their work was done, they would carry out conversions so that the astronomical ephemerides could be expressed in terms of apparent time. As an example, the mean solar time at Greenwich, England, was only introduced in 1834, in the *Nautical Almanac and Astronomical Ephemeris*. The following year, in France, the *Connaissance des temps* adopted the mean solar time of Paris.

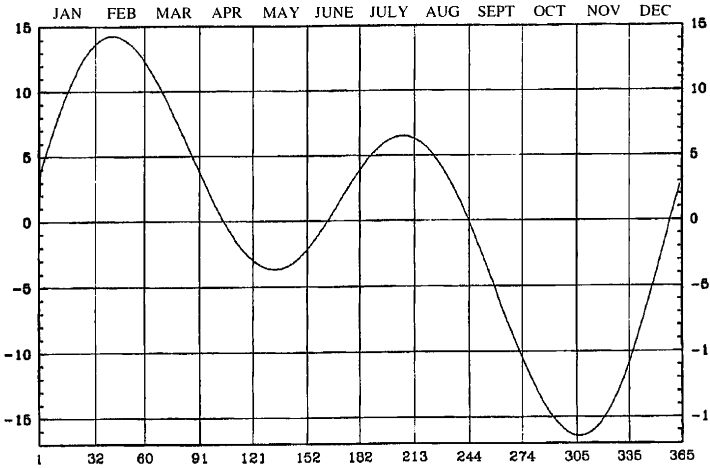

Fig. 4.1. The equation of time. Mean solar time minus apparent solar time in minutes. (Kindly communicated by the Bureau des longitudes, Paris.)

4.2.2 Universal time scales and time zones

During the second part of the nineteenth century, the introduction of rail networks made it essential to use a single definition of the hour, at least on the national level. Many countries then adopted the mean time of the meridian through their capital city, increased by 12 hours. Hence, in France, an Act of Parliament voted on 14 March 1891 imposed the hour of the Paris meridian.

However, this solution proved inadequate in certain countries which extended over a very wide range of longitudes. The idea of dividing such countries into time zones, the time differing from one to the next by a whole number of hours, in such a way that solar midday always occurs at around 12 h, is apparently due to Ch. Dowd in the USA in 1870 [4.1]. It was applied shortly afterwards by S. Fleming, an engineer on the Canadian railroads.

The use of mean solar time thus began to spread, whilst still lacking a single time standard the world over. Agreement for a worldwide unification of time came in October 1884 at an international conference held in Washington 'for the adoption of a single prime meridian and a universal time' [4.2]. The Greenwich meridian stood out as a natural choice since it was already taken as the origin for longitudes on most maritime charts, and indeed it was almost unanimously adopted by national representatives. The universal time was therefore the one determined at this meridian. Moreover, it was stipulated at the conference that the universal day should begin at midnight on the prime meridian, in contrast to the common astronomical practice of counting days from midday, and that hours should be counted from 0 to 24.

During the following two decades, the time zone system was related to universal time and extended around the Earth. Initially, the planet was cut into 24 time zones, each occupying 15° of longitude, with the first centred on Greenwich. Each zone was centred on its *standard meridian* and *standard time* was thereby established within the zone. Any country covering a reasonably small range of longitudes was to adopt the time zone containing its capital city. Little by little, countries began to implement the new system. Since 1971, it has been associated with Coordinated Universal Time (UTC), to be defined in Section 4.5 and discussed in greater depth in Chapter 7. For example, on 9 March 1911, a French Act of Parliament introduced the time of the Greenwich meridian to France in terms intended to spare national sensitivities: 'Official time in France ... is the Paris mean time delayed by nine minutes and twenty-one seconds.' (It is interesting to note the inappropriacy of the terms, since the mean time at midday is 0 h.) This Act remained in force until 9 August 1978, when a government decree associated official time with UTC.

The standard time system has since lost something of its original simplicity through the use of permanent shifts between official time and the time of the relevant zone, and through the use of summer time. For example, the latter appeared in France in 1916 'in order to counter an annoying tendency of a great number of town dwellers who get up and go to bed too late' [4.3]. It is extremely difficult to find out the difference between UTC and official time or the commonly used time in various countries, since no organisation is responsible for collating such information.

Astronomers were slower to accept the universal time of the 1884 conference. The *Connaissance des temps*, for example, was aligned with Greenwich Mean Time in 1916, but not with universal time which, as we have observed, differs from GMT by 12 hours.

It was only in 1925 that the ephemerides used by navigators and astronomers introduced the day which began at midnight rather than at midday. The new time scale, deduced from mean time by adding 12 hours, was referred to as *civil time*. For example, the civil time at Greenwich, which was nothing other than the universal time specified in 1884, was taken as the time argument for the *Connaissance des temps* from 1925 to 1950. However, it is commonplace, even in astronomy, to keep the appellation *Greenwich Mean Time* for this new scale, and it has been the source of much confusion. Although the name *Universal Time* arose straightforwardly from the terms used in the 1884 conference, it was not until 1948 that the International Astronomical Union, after a good deal of hesitation, finally made firm recommendations in its favour. This has not prevented continued incorrect use of the acronym GMT, where UT should be used.

Even today, there still remains a vestige of the system whereby days are counted from midday, viz., the Julian Date mentioned in Section 4.1.

4.2.3 Towards a unique realisation of Universal Time

Let us begin with a convention aimed at simplifying language. We shall use the name *universal time* to speak of any time scale based on the Earth's rotation and referred to the prime meridian. This therefore includes Universal Time (UT) with its modern definition, as well as GMT, shifted by 12 hours relative to UT, and also *Greenwich Sidereal Time*. The latter is the hour angle of the vernal equinox measured from the prime meridian and it is mathematically related to GMT and UT.

From the middle of the nineteenth century up until around 1970, national timekeeping organisations operated along more or less unchanged lines. The basic equipment included clocks that were as stable as possible, together with refracting telescopes that could be used to observe the transits of stars across the meridian, or occasionally at equal altitudes. This was complemented by devices for emitting time signals that were originally electrical, optical or sound signals. Then, in about 1910, emitters and receivers of radio time signals were used. In periods of clear weather, astronomers spent their evenings, or even the whole night, timing star transits by means of one of the local clocks taken as master clock. They would then go to bed before setting about 'reducing their observations', which basically meant deducing *clock corrections*. These corrections were to be added to the readings of the master clock in order to obtain, at the instant of astronomical observation, the local time values or, after adding the longitude, the universal time. The purpose of the clock was therefore to provide a sort of average for observations, smoothing them out, so to speak, and to keep time between evenings when observations were made. By extrapolation, the clock thus gave an approximation to universal time in real time. With the help of this extrapolation, time signals could be emitted in the form of pulses at agreed nominal times. Figure 4.2 shows how this works, highlighting some of the uncertainties associated with the method. The point in having a good *timekeeper* is clear. However, by pushing forward progress in clock timekeeping, astronomers have finally lost their hold upon the management of time.

When radio time signals began to cross the Atlantic in around 1910, it was found that they could differ by as much as one or two seconds. This discovery came at a time when random uncertainties in local measurements of universal

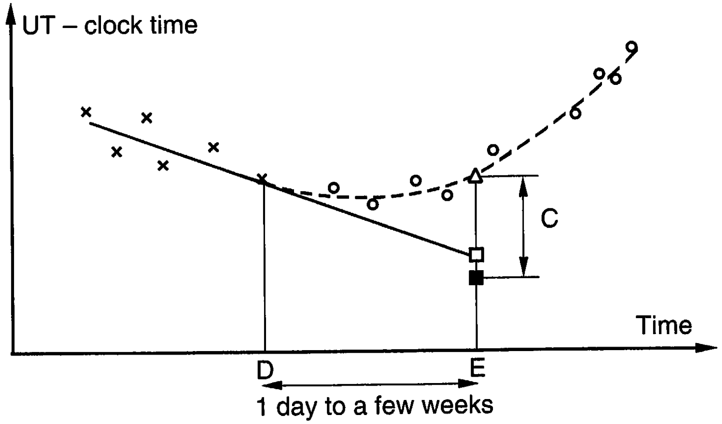

Fig. 4.2. Emission of time signals in UT. The astronomical observation × of UT minus clock time, available at the time E when the time signal is emitted, only extends up to the time D because of computation and transmission delays, and bad weather. A linear extrapolation is used to predict the value (marked □) at E. Subsequent observations (marked o) belie this prediction. In addition, the emission time ■ may not actually coincide with the prediction. The total error in the emission time is C.

time were as low as a few hundredths of a second. The main causes of systematic error, to be combined with random errors, were errors in observed star positions, instrumental delays and, above all, errors in longitudes, the three causes adding together in unknown measures. Universal time is indeed unique by its definition, but its realisations could diverge in a quite unacceptable manner.

At the Bureau des longitudes in Paris, it was clear that this situation would have to be remedied. Therefore, in 1911, on the initiative of Ferrié, they suggested that the French government convene an international conference to create a Bureau international de l'heure (BIH). Its basic task would be to unify world time. The meeting took place in 1912, but war was to intervene before the articles of the BIH could be established. The BIH nevertheless began to carry out its functions immediately, in a semi-official manner, with assistance from the Paris Observatory. It acquired official status in 1919, when it was placed under the responsibility of the International Astronomical Union, and carried out its mandate until 1988. At this point its activities were split into two parts. Atomic time measurement was transferred to the International Bureau of Weights and Measures (Bureau international des poids et mesures, BIPM), whilst its astronomical and geodetic activities were reorganised and

extended to form the International Earth Rotation Service (IERS). The IERS Central Bureau was set up in the Paris Observatory until 2001, before moving to the Institute of Cartography and Geodesy in Frankfurt, Germany. We shall comment further on the activities of the IERS later in the book. The references [4.4, 4.5, 4.6] give a historical review of the BIH and summarise its activities.

The aim of the BIH was to provide a single approximation to the theoretical universal time, known as the *definitive time* and then to give the discrepancy between this definitive time and the nominal emission times of time signals. Calculations were extremely complex, although based on simple enough principles. We may outline them in the following way. Each time service referred astronomical observations and time signal emissions to its master clock. Then by exchanging time signals, the BIH could put together a unique worldwide master clock, referring the whole set of observations and emissions to it. It was then a simple matter to construct the definitive time, by averaging astronomical measurements, and refer time signal emissions to it. The results were published in the *Bulletin horaire* every month or every other month, depending on the period.

Let us consider the results for October 1936 as an example, published in April 1937. The definitive time of fifty-six daily emissions is given to the nearest millisecond. We may expect a fuzziness to the extent of a few milliseconds in the definitive time, due to uncertainties over the estimated propagation delays of the radio signals. With regard to this fuzziness, the offsets of emission times with respect to their nominal values as estimated by the BIH were considerable, reaching as much as 0.2 s.

We may well wonder to what extent the BIH's definitive time represented the theoretical universal time defined by astronomers. It is hardly possible to give a precise answer to this question because of the systematic errors already mentioned, but it would be wise to expect uncertainties of up to several hundredths of a second. This shows how important it is to distinguish synchronisations which can be very accurately guaranteed on some conventionally agreed basis, and the realisation of a theoretical time scale using this basis. We shall see that this distinction explains a certain reluctance towards setting up an atomic measure of time.

The work carried out at the BIH barely changed until around 1960, although technical improvement of the clocks used gradually made it possible to reduce synchronisation errors in time signal emissions. We may say that, by this time, the unification of time could be ensured to within one or two milliseconds, with a time lag of about one year before results were announced. In real time, by extrapolating clock readings and using local observations, the discrepancies in the unification of time were of the order of 10 ms.

4.2.4 *Definition of the second before 1960: the second of mean solar time*

The measurement of time based on the Earth's rotation originally rested upon the reproducibility postulate, as discussed in Chapter 2. This was first applied to the mean duration of the day, and then to the period of rotation. When Euler showed in 1737 that the Earth, considered as an undeformable solid, was spinning at a uniform rate, the mean solar time took on the character of a Newtonian dynamical time. However, two centuries later, when it was realised that the Earth's rotation was not so regular after all, it turned out to be much more difficult to establish a mathematical model, and thereby define a time system. Only a small empirical correction was applied from 1955, in order to remove an annual fluctuation (the corrected scale being denoted UT2). It had been realised that mean solar time, and subsequently, Universal Time, were not a good measure of time, and a better clock was sought to replace the Earth's rotation.

The second, unit of time, or rather, duration, was tacitly and universally defined as the duration of 1/86 400 of a mean solar day. It is surprising that this definition of the second of mean time, which remained in force until as late as 1960, had never been officially ratified by the organisations responsible for world metrology since the 1875 Metre Convention.

4.3 Time based on Solar System dynamics: Ephemeris Time

4.3.1 *First doubts on the uniformity of the Earth's rotation*

Whilst Copernicus accepted the ancient Greek dogma concerning the uniformity of the Earth's rotation, Kepler mentioned the possibility of some irregularities. The first attempt to demonstrate the existence of such phenomena appears to be due to Flamsteed. Shortly after the foundation of the Greenwich Observatory in 1677, he set up gigantic pendulums in the hope of revealing some anomaly, but found none whatever [4.7]. Later, in 1752, the Berlin Academy of Arts and Sciences, presided by Maupertuis, raised the following question: 'Has the daily motion of the Earth always occurred at the same rate or not? What means do we have at our disposal for answering this question? And if there were some irregularity, what might be its cause?' In his reply, Kant said that there could be a slowing down effect due to dissipation of energy in tidal movements of the oceans. He was right, but the idea was only confirmed by observation two centuries later.

The conviction that the Earth's rotation must be perfectly uniform was deeply rooted, and these first doubts were barely heard. In 1825, Laplace wrote: 'It is therefore certain that, since Hipparcos, the day's duration has not changed by as much as one hundredth of a [decimal] second [i.e., 0.008 64 ordinary seconds].' The claim is unfounded, as we shall see. Whilst Ferrel in 1864 and then Delaunay in 1865 both asserted that discrepancies observed between the lunar ephemeris and observation were due to a lengthening of the day, Fleming declared in 1864 that there was no motion more uniform than that of the Earth about its axis.

4.3.2 *Acceptance of irregularities in the Earth's rotation*

Proof of irregularities in the Earth's rotation first came from a study of the orbital motions of the planets and the Moon. To the degree of accuracy required by observations, the orbital motions of the planets can be studied under the assumption that both they and the Sun are point objects. The purely gravitational interactions between them are simple enough for their motions to be worked out to a high level of precision, using for example the model provided by Newtonian mechanics which was common practice at the time. (Relativistic theories have been in use for the past few years, but the differences will not be relevant in the context of this discussion.) Generally speaking, satellite motions about a central body are more complex to treat. This is due to the fact that the central object may not be exactly spherical and also to the presence of phenomena leading to energy dissipation. Such effects are nevertheless small in the case of the Moon, which we shall be concerned with now, and we shall ignore them for the moment. (Further explanations of this point will be given in Chapter 8.) The next step is to fit the theory to observations. In other words, numerical values must be attributed to unknown parameters so that an ephemeris can be established, giving the positions of the bodies in tabular form as a function of time, e.g., in terms of the mean solar time. But what happens if the time used to date observations differs from the uniform time of Newtonian theory? Then discrepancies arise between observed positions and those given by the ephemeris at the instant of observation. Such discrepancies, if large enough, cannot be explained by the random errors of observation. Clearly, they are minimised to a certain degree by the fitting of orbital parameters, which are therefore biased. Even then, they still remain to some extent, and because of the errors in the orbital parameters, they are liable to increase rapidly with time when the ephemeris is used to predict what will happen.

This is precisely what Newcomb found when using the lunar ephemeris established by Hansen in 1857. Having checked that the theory had been used correctly, Newcomb considered the possibility of irregularities in the mean solar time. In other words, he put the Earth's rotation to question. However, it had first to be checked that the ephemerides of the other planets revealed the concomitant discrepancies, and in a way compatible with the hypothesis of an irregularity in the Earth's rotation. This was a difficult task, since the effects are much smaller for the planets than for the Moon. Indeed, they are proportional to the apparent speed of the orbital motion. In 1927, following research by Brown and de Sitter [4.8], no doubt could remain. The Earth's rotation was not a good clock.

Almost immediately, in 1929, Danjon suggested using the dynamics of orbital motions to measure time, in an article that proved remarkable for its clarity and the correctness of its predictions [4.9]. Let us quote some short extracts (translation).

> ... it is legitimate to consider the [rotation of the] Earth as the sole cause for the apparent disorder that still reigned in the Solar System. Although Newton's law has been saved, it is experiencing a quite extraordinary adventure: henceforth called upon to gauge the passage of time, it becomes in part unverifiable and ceases to be what could strictly be termed a law. [...] Since we would ask these [Kepler's] laws to provide a measure for the passage of time, we could no longer subject them to experimental control without entering into a vicious circle. [...] Let us simply hope that we shall one day discover a good terrestrial time standard, so that we may leave these purely logical difficulties behind us.

The good terrestrial time standard appeared in 1955 in the form of the atomic time standard. Danjon may well have been thinking of it.

Unfortunately, Danjon's article was published in an amateur astronomical review and did not receive the attention of his peers. It was not until 1950, during an international colloquium on astronomical constants held in Paris [4.10], that Clemence made a precise proposal for defining a dynamical time. This was christened *Ephemeris Time*, following a suggestion by Brouwer. The definition of Ephemeris Time, to be discussed in Chapter 8, was ratified by the IAU in 1952.

We shall see in particular that Ephemeris Time suffers from a serious drawback, the wretched quality of its readings, with uncertainties of the order of 0.1 s. It may be compared with certain luxury wrist watches which, although provided with an excellent mechanism, have no graduations marked on the face to help us gauge the position of the hands. It is thus only possible to exhibit the large scale and long term variations in the Earth's rotation. In 1936, an irregularity in the Earth's rotation was revealed for the first time and independently in two laboratories, relative to artificial clocks, including both pendulum clocks and quartz clocks. The annual total amplitude was later found to be 60 ms [4.11, 4.12]. However, as can be seen from Figure 4.3, the annual irregularity was at the limit of what could be revealed by the clocks in use at the time. Later, atomic time proved adequate to demonstrating a great many other irregularities (see Chapter 8).

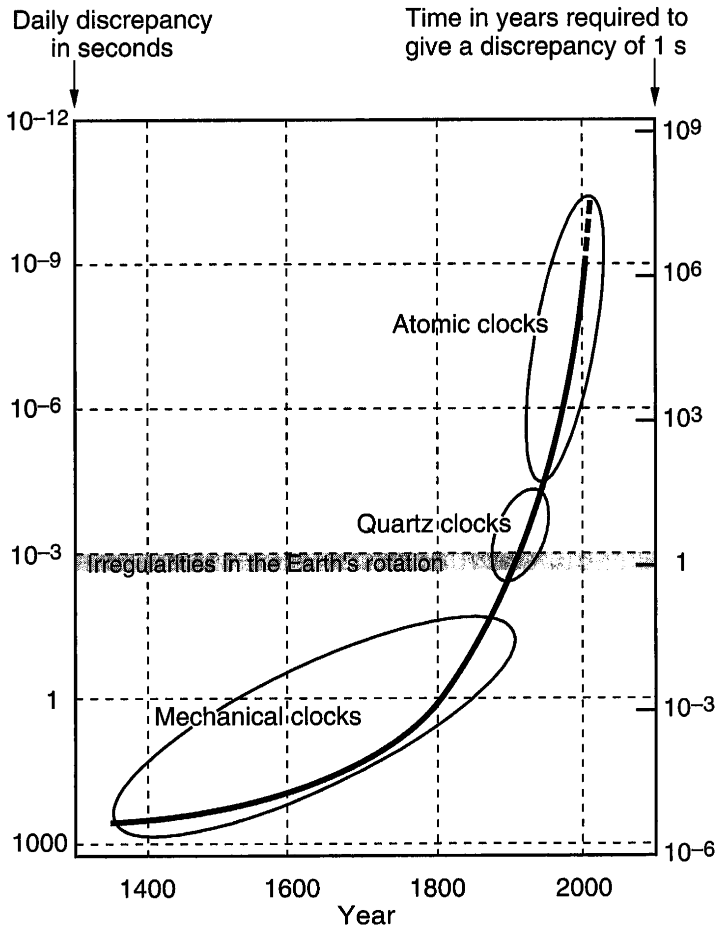

Fig. 4.3. Improvement in the quality of artificial clocks and comparison with the clock provided by terrestrial rotation.

All these irregularities lead to variations in the duration of the second of mean time. In order to eliminate some short term fluctuations and the annual irregularity, whilst also reducing the role of measurement uncertainties, we may consider the average annual value of this second. Unpredictable variations nevertheless remain, reaching almost 10^{-7} s.

4.3.3 *Definition of the second from 1960 to 1967: the ephemeris second*

As explained in Section 2.4.2, if we aim to base the definition of the second on dynamical considerations, it suffices to fix the duration of some particular phenomenon numerically. The 1950 colloquium recommended basing this numerical value on Newcomb's *Tables of the Sun*, a proposal which raised no objections. However, the choice of reference phenomenon did give rise to some hesitation. The one chosen by the International Astronomical Union at its 1952 General Assembly was the sidereal year of 1900, the duration measured between two solar transits of a fixed equinox (i.e., without taking the precession of the equinoxes into account), because this duration is virtually constant. It was soon realised, however, that the more directly observable tropical year was preferable, despite its slow variation. For this reason, and quite exceptionally, the IAU resolution was corrected by correspondence.

In 1956, the International Committee for Weights and Measures used the powers conferred upon it by the 10th General Conference on Weights and Measures (CGPM) in 1954 to decide upon the following definition:

> The second is the fraction 1/31 556 925.9747 of the tropical year for 1900 January 0 at 12 h Ephemeris Time.

This definition was ratified by the 11th CGPM in 1960. Shortlived, it was abrogated in 1967 in favour of an atomic definition for the second.

The definition of the *ephemeris second* proved puzzling to those who were not familiar with astronomy. Its strange definition arises from the fact that the mean longitude of the Sun is, according to Newcomb, a quadratic function of Ephemeris Time (ET). The tropical year for 1900 January 0 at 12 h ET is a fictitious year corresponding to the speed of the mean longitude at the given date.

Let us look more closely at the numerical value used in the definition of the ephemeris second. Newcomb's tables were based upon astronomical observations from the nineteenth century. Naturally, these were dated using mean solar time. The duration of the ephemeris second is therefore roughly equal to the average duration of the second of mean time over this century. However, in 1960, when the ephemeris second was adopted as the unit of time for SI units, the new second was shorter than the currently used second of mean time (averaged over the year 1960) by 1.4×10^{-8} s, and this offset was roughly known at the time of the decision. This goes against common metrological practice of adjusting successive definitions of a unit in such a way that they lead to an invariable quantity during changes of definition, at least, within the limits of experimental error. It could not have been predicted in 1960 that this would have unfortunate consequences for the Coordinated Universal Time system that is now in use.

The ephemeris second is uniquely defined, but it is rooted in the past, in 1900. Its physical realisation at a given date depends on the astronomical ephemerides. It is estimated that it could be achieved to within $\pm 2 \times 10^{-9}$ s around 1960 by the analysis of astronomical observations extending over several years. Compared with the second of mean time, the gain in reproducibility was by a factor of 50.

4.3.4 *Ephemeris Time: a scale reserved for astronomers*

The beauty and simplicity of Ephemeris Time raised some enthusiasm. In 1966, when atomic clocks had already been running for eleven years, Woolard and Clemence wrote [4.13]: 'The traditional practice of measuring time by the apparent motions of celestial bodies may now be based upon an exact dynamical foundation by adopting as a primary standard the measure of time implicitly defined by the laws of motion.' The story of Ephemeris Time is nevertheless plagued by a long series of difficulties and misunderstandings, as we shall see in Chapter 8.

Ephemeris Time was never made available, nor used, in everyday life. In fact, its application was limited to the needs of astronomical dynamics. There has never been an official organisation expressly set up to centralise and process its measurements so as to produce a unique realisation that could have been adopted by convention.

Let us emphasise that, until 1960, the official time scale, Universal Time, and the definition of the second (of mean time) were both based in a consistent way upon the Earth's rotation. Hence, the dissemination of the time scale effectively provided the SI second. With the adoption of ET began a period of incompatibility between the time scale, still UT, and the unit of time. The atomic measurement of time was to put an end to this confusion in 1971.

4.4 Atomic time measurement

4.4.1 *Atomic frequency standards and the atomic definition of the second (1967)*

Science has gradually accepted that the various chemical species are composed of molecules which are themselves made from atoms. A relation was eventually established between the structure of emission spectra and the atomic and molecular composition of excited gases. The assumed universality of atomic properties could then be transposed to these spectra. Maxwell in 1873 and Kelvin in 1879 both suggested that the wavelength of a spectral line and the period of the corresponding radiation could be used to define the unit of length and the unit of time, respectively [4.14]. These proposals were truly prophetic if we remember that the definition of the metre was only based upon the wavelength of a spectral line from 1960 (and until 1983). Moreover, optical emissions can now be used in this way by highly specialised laboratories, and optical time and frequency standards may soon compete with current atomic standards using transitions at centimetre wavelengths.

As in many other areas, the understanding we now have of electromagnetism, quantum physics, atomic physics and spectroscopy, which have led to the invention and development of time and frequency standards, was acquired at the end of the nineteenth century and in the first half of the twentieth century. We shall summarise these years of progress by citing those physicists who made key contributions. Planck laid the foundations for quantum theory, whilst Einstein introduced the idea of the photon and stimulated emission, amongst many other significant contributions, including the theory of relativity. Bohr applied quantum theory to the explanation of atomic structure and introduced the idea of energy levels. Hertz was the first to produce and detect radio waves. De Broglie, Heisenberg and Schrödinger created and developed wave mechanics. Stern contributed to the kinetic theory of gases and discovered with Gerlach the magnetism of atoms and its spatial quantisation.

Technological advances which began in the 1930s and were pushed ahead by the need for radio communication and radar detection during the Second World War also played a key role. In 1945, it became possible to produce radio waves with frequencies up to 30 GHz and to measure this frequency. These developments were the starting point for a major leap forward in precision Hertzian spectroscopy, led by Townes and Pound.

The first atomic clock was made in 1948 by Lyons at the National Bureau of Standards in the USA, now called the National Institute of Standards and Technology [4.15]. The reference was an intense absorption line of the ammonia molecule, lying at around 24 GHz. The molecular resonance controlled the frequency of a quartz oscillator which then provided pulses to mark time. However, the resonance was broadened by the Doppler effect and collisions. Its long term stability was no better than that of a quartz clock and this approach was dropped.

The caesium clock was the result of work carried out by Rabi and his students [4.16, 4.17]. At the end of the 1930s, Rabi was working on the magnetic resonance technique for atomic or molecular beams. This method, originally aimed at measuring magnetic moments, was soon adapted to the study of the radio frequency spectra of atoms and molecules. The first experimental evidence for the hyperfine transition in the caesium atom was obtained in 1940. The possibility of using the magnetic resonance method with an atomic beam to make an atomic clock had already been discussed within the Rabi group in 1939. It seems well-established that Rabi presented this idea publicly at a conference in 1945, during a meeting of the American Physical Society. In fact, he suggested using the caesium atom. In 1950, Ramsey improved conditions for the interaction between the electromagnetic wave and the atoms by introducing two separate oscillating fields. Putting together the experimental methods developed by Rabi and Ramsey, Essen and Parry built the first reliable caesium beam clock in 1955 at the National Physical Laboratory in the United Kingdom [4.18]. It was regularly used to calibrate the frequency of quartz oscillators. At the same time, one of Rabi's students, Zacharias, developed a prototype commercial caesium clock at the Massachusetts Institute of Technology, in which the atomic resonance controlled a quartz oscillator through a servo loop. The first caesium clocks based on this prototype were commercialised in 1956 by the National Company under the name of Atomichron [4.14].

Experimental confirmation of the principle of amplification by stimulated emission of radiation was first obtained in 1954–55 by Townes in the USA and by Basov and Prokhorov in the USSR. They built the first self-oscillating masers (Microwave Amplification by Stimulated Emission of Radiation), using the ammonia molecular transition at 24 GHz. They were considered for use in atomic clocks but their long term instability of the order of 10^{-10} was too great and the idea was not taken further [4.19]. These achievements were nevertheless useful in showing the way to a solid state maser amplifier devised by Bloembergen in 1956, and then the laser, developed by Schawlow and Townes in 1958 [4.17].

The hydrogen maser, elaborated by Ramsey, results from an evolution of Rabi's magnetic resonance method. The original aim was to narrow the resonance line of the caesium atom by increasing dwell time between the two oscillating fields [4.17]. To this end, the caesium atoms were held back for a certain time, allowing collisions to occur in a cell coated with paraffin wax and equipped with baffles. It was then realised that a Teflon coating was better. The idea was put to the test with hydrogen atoms, allowing them to undergo a great many collisions without significant perturbation. The detection of hydrogen atoms by stimulated emission was also achieved and the hydrogen maser was thereby invented in 1959–60. Its success rests upon the very long dwell time of the atoms in the Teflon coated cell (about 1 s), and hence upon the very long period of interaction with the electromagnetic field in a resonant cavity, which favours stability of the frequency. A commercial version was soon developed by Vessot at the manufacturer Varian Associates.

Since many applications required compact atomic time and frequency standards, the possibility of using optical pumping was studied from the moment of its discovery. In 1950, Kastler suggested replacing the intense inhomogeneous magnetic field, until then used to select and detect the atomic states, and involving bulky magnets in those days, by an optical method. Dicke showed in 1953 that Doppler broadening of microwave resonance lines could be practically eliminated by mixing the atoms with a neutral gas that limited their scattering speeds. In 1956, Dehmelt discovered that the microwave resonance was manifested by a reduction in intensity of the pumping light transmitted by the cell containing the atoms and the buffer gas [4.15, 4.16]. It thus became possible to build compact rubidium clocks, and they were soon being produced by various manufacturers. These are still the most widespread atomic clocks. From the 1980s, optical pumping methods have been applied to caesium beam clocks, thanks to the development of semiconductor lasers.

The two- or three-dimensional confinement of ions in a non-uniform alternating electric field was developed by Paul from 1955. The long term trapping obtained in this way is favourable for observing narrow lines in the microwave region. Dehmelt was the first, in 1965, to use such traps in the precise measurement of a hyperfine transition frequency, namely that of $^3\mathrm{He}^+$. In 1969, Major suggested combining the optical pumping and ion confinement methods in a time and frequency standard using $^{199}\mathrm{Hg}^+$. Such a standard was first achieved in 1979 at the Laboratoire de l'horloge atomique in France. Several improved operating models of this standard were produced by Hewlett-Packard under the guidance of Cutler in 1983. Studies of trapped-ion frequency standards continue, using transitions in the microwave and optical regions. Some of these involve the laser cooling method proposed by Wineland and Dehmelt in 1975.

In the same year, Hänsch and Schawlow described an efficient way of slowing down and reducing the thermal motions of neutral atoms. The method was first used to slow down atom beams. Then, in 1989, Wieman and his team at the University of Colorado succeeded in applying it to caesium atoms held at low pressures in a cell. A ball of atoms with a kinetic energy of thermal excitation corresponding to a temperature of the order of 1 microkelvin was trapped at the crossing point of six light beams. Several teams in the USA and France, among others, have made experimental and theoretical contributions towards improving techniques for cooling, trapping and manipulating cold atoms. It is now possible to launch cooled caesium atoms with a speed of the order of $1\ \mathrm{m\,s^{-1}}$. It has thus been possible to use them in caesium clocks, where the advantages of slow atoms are decisive. The first primary standard based on cooled caesium atoms was built in France in 1996 by Clairon at the Laboratoire primaire du temps et des fréquences, Salomon at the Ecole normale supérieure, and their respective research teams.

The Nobel committee has rewarded some of those physicists who directly contributed to the invention and development of atomic clocks: I. Rabi (1944), C. Townes, N. Basov and A. Prokhorov (1964), A. Kastler (1966), N. Ramsey, W. Paul and H. Dehmelt (1989), S. Chu, C. Cohen-Tannoudji and W. Philips (1997), the dates corresponding to the attribution of the prize.

Chapter 6 explains how the various atomic frequency standards work. We shall see that it is the caesium standard that presently leads to the greatest accuracy. Accuracy is a key idea in metrology. It refers here to the capacity to deliver a frequency bearing a perfectly known relationship to the transition frequency of the unperturbed atom. It excludes all frequency drift. At the time of writing, the caesium standard thus plays a privileged role in fundamental time measurement. Let us now review how its accuracy has evolved and outline the consequences of this evolution for the definition of the second.

The first caesium standards involved frequency inaccuracies with a relative value of the order of 10^{-9}. (Uncertainties and small variations are generally expressed as relative values. We shall no longer specify this in the following, since an absence of physical unit will be a sufficient indicator.) Progress was made in leaps and bounds and, as we shall see, standards could be compared with one another in order to assess their agreement. Clearly, the transition frequency ν_{Cs} had to be expressed in terms of the unit of time then available. In their early research work, Essen and Parry used the second of mean time which could be supplied in real time. The ephemeris second was already defined, although not yet the recognised SI unit. However, as we have already noted, it could only be realised as a result of long analysis. This analysis was undertaken by Markowitz *et al.* [4.20], giving in 1958 the frequency value

$$\nu_{\mathrm{Cs}} = 9\,192\,631\,770 \pm 20\ \mathrm{Hz}\,. \tag{4.2}$$

The unit Hz mentioned here is based on the ephemeris second. The uncertainty of ±20 Hz was almost entirely due to uncertainty in the realisation of the ephemeris second.

Naturally the various laboratories studying atomic frequency standards were keen to agree upon a single frequency value for caesium. The value due to Markowitz *et al.* was immediately adopted by convention, in the sense that it was regarded as a strictly accurate reference for expressing frequencies. This implicitly defined an unofficial atomic second. Indeed, it was recognised as such by the 12th General Conference on Weights and Measures (CGPM) in

1964, although it was felt that the moment had not yet come to adopt a new official definition of the second in view of further significant progress that might soon be made. However, aware of the urgency of the situation, the hyperfine transition of caesium 133 was designated for use as standard and the above value was assigned to its frequency.

In 1967, the frequency inaccuracies of the atomic standards were reduced to 10^{-12} and it was decided to move ahead with an atomic definition for the unit of time. This was accomplished by the 13th General Conference on Weights and Measures in 1967:

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

The definition of the so-called ephemeris second was thereby abrogated. The atomic second was fitted as far as possible with the ephemeris second as it stood at the time. Subsequent measurements of Ephemeris Time confirmed this fitting. The current second thus has a duration that corresponds to that of the second of mean time as it was averaged over the nineteenth century.

The improvement of caesium standards continued. In 1976, inaccuracies in the best standard, then in Germany, were reduced to a few multiples of 10^{-14}. The situation remained at this point until 1995 when a new stage in the ascension towards improved accuracy was set in motion by the discovery that atoms could be cooled by lasers. The technique has already been integrated into some caesium clocks, as we shall see in Chapter 6.

Applications of time measurement are on the increase because industry is able to manufacture high quality timing equipment. In particular, caesium standards are commercially available and, although they are less accurate than the best laboratory instruments, they are less bulky, have high frequency stability and can operate continuously with great reliability. Strictly speaking, these are caesium clocks, stringing together the seconds. The existence of such clocks has made it feasible to base the world time scale on atomic transitions rather than the motions of celestial bodies. This brings us to the birth of International Atomic Time.

4.4.2 From Universal Time to International Atomic Time

(a) The age of frequency comparisons and integrated atomic time

The first caesium frequency standards operated sporadically. They were used from time to time to calibrate the frequency of independent quartz clocks which, between such measurements, retained a memory of the atomic frequency. It was then possible to establish a time scale based upon the atomic transition frequency, that is, an atomic time. Let us discuss some of the details of this operation. We may thereby bring out the properties of the atomic time realised by this method and also introduce useful definitions in what follows.

An oscillator is characterised by a *nominal frequency* ν_0, stated or claimed by its manufacturer and referred to the definition of the second. Its actual frequency is slightly different and varies. This is denoted $\nu(\theta)$, where θ is the date in a time scale Θ taken as reference. We shall see that few metrological qualities are required of Θ which may, for example, be taken as Universal Time. Since the nominal frequencies of various oscillators are not necessarily the same, it is useful to define:

- the *normalised frequency* $\Phi(\theta)$ by
$$\Phi(\theta) = \frac{\nu(\theta)}{\nu_0}\,, \tag{4.3}$$
which is close to unity;
- the *relative frequency offset* $y(\theta)$ by
$$y(\theta) = \Phi(\theta) - 1\,, \tag{4.4}$$
which is small compared to unity.

In order to construct an atomic time scale, the value of the nominal frequency of the transition $\nu_{\mathrm{Cs},\,0}$ must first be fixed. This can be the value that has been used to define the second, but in the beginning this was not known and different values were used. If the frequency of the external quartz clock is ν_{Q}, the episodic comparisons of frequencies at dates θ_i can be expressed in terms of the difference between normalised frequencies, viz.,

$$y_{\mathrm{Cs}}(\theta_i) - y_{\mathrm{Q}}(\theta_i) \equiv [y_{\mathrm{Cs}} - y_{\mathrm{Q}}]\,(\theta_i)\,. \tag{4.5}$$

These are in fact mean values around the date θ_i. Starting from these measured values, the function $[y_{\mathrm{Cs}} - y_{\mathrm{Q}}]\,(\theta)$ must be estimated as accurately as possible. The atomic time τ_{A} is then obtained by correcting the readings τ_{Q} of the quartz clock:

$$[\tau_{\mathrm{A}} - \tau_{\mathrm{Q}}]\,(\theta) = [\tau_{\mathrm{A}} - \tau_{\mathrm{Q}}]\,(\theta_0) + \int_{\theta_0}^{\theta} [y_{\mathrm{Cs}} - y_{\mathrm{Q}}]\,(\theta)\,\mathrm{d}\theta\,. \tag{4.6}$$

This explains why time scales produced in this way are often referred to as *integrated time scales*. The origin of τ_{A} is arbitrary. Whilst it was at first chosen differently by the various laboratories, it was eventually fixed in 1960 by the condition that τ_A should equal the Universal Time UT2 on 1 January 1958 at 0 h UT2. As far as the value of $\nu_{Cs,0}$ was concerned, it was universally agreed to use the value obtained by Markowitz *et al.* It is clear that the atomic time scale depends on what assumptions are made to interpolate the frequency, and also on the numerical method adopted to carry out the integral. In other words, it is not uniquely determined.

When the first atomic time scales were being built up in this way in the years 1955 to 1960, very low frequency (VLF, 10–30 kHz) radio broadcasts were already being used for long distance radio communications and as navigational aids. The carrier frequencies were stabilised by quartz oscillators. These frequencies could be very accurately received by phase tracking. They thus constituted a good reference to which frequencies available in different laboratories could be compared. Frequency comparisons thus became possible over intercontinental distances. Time comparisons remained uncertain, however, since, even though an initial time comparison was available, it was virtually impossible to avoid losing some periods of the carrier frequency, so that anything from a few microseconds to a few tens of microseconds might slip away each year.

The pioneers of atomic time therefore only had available frequency comparisons between standards, tainted with a good many uncertainties. These included weaknesses in their local mode of frequency interpolation, a poor knowledge of radio wave propagation, and phase losses. Through computations, they were nevertheless able to construct a mean atomic frequency standard that was probably better than each of the individual standards, and produce mean integrated atomic times that were materialised locally by correcting readings from one of their clocks. This method was used in particular by the Bureau international de l'heure until 1969 [4.21].

Such integrated atomic times had originally been set up to study irregularities in Universal Time, that is, in the Earth's rotation. They were also used to stabilise time signal broadcasts, allowing as they did a better extrapolation of Universal Time than earlier clocks. In contrast, their application to other fields of activity, such as celestial dynamics in particular, raised little enthusiasm. The general opinion of astronomers is well expressed in a work published in 1966 [4.13]:

> An atomic clock provides only a standard of *frequency*; it determines a unit of time, but not the continuous count of units that is necessary to determine the interval elapsed since any initial epoch in the past. Astronomical time determinations are essential for defining an epoch and referring instants of time to it, as no artificial clock can be indefinitely sustained in continuous operation in the manner of the celestial motions.

(b) The advent of modern atomic time

It is true to say that the rise of atomic time is primarily due to technical advances made in the design of atomic clocks and their general proliferation. But it would not have taken place at all without precise methods for comparing their readings, for *time comparisons*, and without being able to transfer times in such a way as to carry out *synchronisation*. Conventional radio time signals as they appeared at the beginning of the twentieth century, and which are still used, are subject to some uncertainty as regards the time required for their propagation. This uncertainty, estimated to be around 1 ms, is quite unacceptable if we hope to exploit the qualities of atomic clocks over large distances. For example, at the time when the atomic definition was adopted (1967), thirty years of averaging would have been required to benefit from afar from the full accuracy of the best caesium clocks by the use of such time signals.

In order to fix the orders of magnitude relevant to the metrological problems treated in this work, we may reasonably assume that the frequency degradation introduced by remote time comparisons over one day of measurements should not exceed the inaccuracies in the frequency standards. For standards accurate to 10^{-12}, uncertainties in comparisons should be less then 0.1 μs, whilst at 10^{-14}, an uncertainty of 1 ns should be achieved. These are truly demanding requirements. It has almost always been impossible to harmonise the accuracy of time comparisons with the stability and accuracy of atomic clocks.

Nevertheless, in 1962, an experimental transmission of signals by the telecommunications satellite Telstar provided a time transfer link accurate to 1 μs (for the spatial segment) between the United Kingdom and the USA [4.22]. Although this experiment was followed by many others, the exercise remained at the experimental stage. As recalled in the quotation at the end of the last section, one feature of the quantity time is that its measurement requires a continued effort. Methods were therefore needed to compare times that were simple and cheap enough to apply on a day-to-day basis.

A practical solution was found around 1968 using the navigation system Loran-C. The signals travel by ground waves and thus have highly stable propagation delays (with possible variations of ±1 μs), although they are not known a priori, and a range of 1000–2000 km. Locally, the simultaneous reception of ordinary television signals was also used successfully, to even greater accuracy. The idea was therefore to begin by determining the propagation delays and then check them episodically. This was done by transporting operating caesium clocks, by road and on commercial flights. These methods will be described in Chapter 5. For the moment, let us just say that large regions of the globe could be covered in this way, including North America and

Europe, with a routinely operating network of time comparisons accurate to ±1 μs.

From this moment on, mean atomic time scales could be constructed from clock readings. These atomic times were realised by calculating corrections to the readings of the participating clocks. This led to a new problem. Optimal algorithms had to be established in order to build time scales from data of varying quality (see Chapter 7). The Bureau international de l'heure replaced its mean frequency standard by a mean atomic clock on 1 January 1969, without introducing any discontinuity into its atomic time scale. In 1973, it set up an algorithm and a world coordination that have survived to this day with only minor modifications (see Chapter 7) in order to keep up with technical progress in standards and time transfers. A considerable improvement in remote clock comparisons was provided in 1983 by the satellite system known as the Global Positioning System (GPS). The GPS has gradually superseded Loran-C in timing laboratories and reduced uncertainties to a few nanoseconds. Recently, time comparisons by exchange of signals relayed by telecommunications satellites began to be used operationally, with improved results, especially for frequency comparisons.

National organisations, in particular the US Naval Observatory in Washington, were also producing atomic time scales. The need for agreement upon a single reference led several key international organisations to recommend the use of the BIH scale, including the International Astronomical Union (IAU) in 1967, the International Union of Radio Science (URSI, Union radio-scientifique internationale) in 1969, and the International Radio Consultative Committee (CCIR) in 1970. Ultimate consecration came from the organisations in the Metre Convention, with the definition of *International Atomic Time* (TAI, Temps atomique international) formulated in 1970 by the International Committee for Weights and Measures (CIPM) in the following terms:

> International Atomic Time (TAI) is the time reference coordinate established by the Bureau international de l'heure on the basis of the readings of atomic clocks operating in various establishments in accordance with the definition of the second, the unit of time of the International System of units.

This definition, or rather, this official recognition, was ratified by the 14th General Conference on Weights and Measures in 1971. After a transition period that had extended from 1955 to 1971, time measurement had finally become coherent again, with the unit of time and the world time scale based on the same physical phenomenon, a specified atomic transition. Time, like all other fundamental physical quantities, was now measured by laboratory standards. Nothing could prevent it from taking its place amongst those other quantities, within the International Bureau of Weights and Measures at Sèvres in France. This happened in 1988.

Since the atomic time scale at the BIH had been maintained continuously since 1955, it was retroactively called TAI, even though computational methods had evolved.

Whilst the accuracy and stability of atomic time standards and the TAI were improving, it became clear that a fuller definition within the context of general relativity would soon be needed. We should mention the extremely apposite ideas put forward on this subject by Becker as early as 1967 [4.23]. But it was not until 1980 that the Consultative Committee for the Definition of the Second (CCDS, a CIPM committee which became in 1997 the Consultative Committee for Time and Frequency) gave a relativistic definition of the TAI, a definition that was completed by the IAU in its resolution A4 in 1991, and which should be slightly modified following the redefinition of TT in 2000 (see Section 3.3.2e). An important consequence of progress in time measurement for the history of science is that general relativity has become an essential tool in metrology and in practical applications. Indeed the topic of 'metrology and relativity' is the subject of a report prepared under the auspices of the CCDS [4.24].

4.5 Coordinated Universal Time

During the period when International Atomic Time was being set up, the idea of coordinating the broadcast of radio time signals (which had continuously sent out Universal Time) was gradually maturing.

The initiative was taken by Great Britain and the United States. In each of these countries, from August 1959, signals were broadcast to mark locally obtained atomic time scales. At this time, the value of the caesium transition frequency that would later serve to define the second was already in general use. As noted, this value had been related to Ephemeris Time. Consequently, atomic time scales were observed to run fast relative to Universal Time. In order to build an atomic time in approximate agreement with UT, a relative frequency offset y_U had to be introduced. The same value of y_U was chosen in both countries and time signals initially synchronised. Since UT involves unpredictable variations, time adjustments had to be made in order to avoid changing y_U too often. These adjustments were made in steps of 50 ms at agreed times in the light of observations of UT.

As many time signals had their second markers strictly linked to the carrier frequency (e.g., for an emission at 10 MHz, the second marker begins every 10^7 cycles), the atomic frequency could be found from y_U, either by using the spacing of the markers, or by measuring the carrier frequency.

This coordinated effort was soon joined by other countries and as early as 1960 the International Union of Radio Science asked the BIH, which was responsible for unifying UT, to fix the offset y_U each year. Its value would then remain constant throughout the year. Subsequently, the BIH also decided the dates when time steps should be introduced.

This system evolved fairly rapidly. In 1963, the time steps went from 50 ms to 100 ms. Until 1965, the time scale that was more or less common to all the coordinated signals, and which had been spontaneously christened *Coordinated Universal Time* (UTC), still had not been given a rigorous definition. For example, the BIH considered UTC as an average of the emission times of the coordinated signals. An important step in the development of UTC was the BIH's decision to define it, from 1 January 1965, by a mathematical relation with the atomic time of the BIH, which later became the TAI. This relation was

$$UTC - TAI = y_U(TAI - TAI_0) + B \ , \tag{4.7}$$

where TAI_0 is an arbitrarily fixed date origin and B a constant modified by steps so as to maintain the condition

$$|UTC - UT2| < \varepsilon \ , \tag{4.8}$$

with UT2 being Universal Time corrected for its quasi-periodic annual irregularity and ε the agreed tolerance.†

This definition of UTC was used until 1972. However, the need for the frequency offset y_U, sometimes changed at the beginning of the year, became more and more burdensome. (The largest modification was made on 1 January 1966, when it went from -1.5×10^{-8} to -3.0×10^{-8}.) One of the drawbacks was that certain emissions had their carrier frequency linked to the shifted frequency of UTC, so that technical alterations had to be carried out on the emitters. For this reason, the International Radio Consultative Committee had accepted in 1966 that certain time signals would be broadcast with $y_U = 0$, at the expense of frequent time steps. The size of these steps was fixed at 200 ms and they were only imposed at the beginning of a month when needed to keep ε smaller than 200 ms. This *Stepped Atomic Time* was only adopted by the German signals (Federal Republic of Germany) and one emitter in the USA. It is significant mainly because it prepared the way for the currently used UTC system, adopted in 1972.

---

† Note that a time scale such as UT1 is designated by its acronym written in upright capitals. Its reading at an instant specified in terms of a time scale θ is written in italics, e.g., $UT1(\theta)$. The argument θ, value of θ, is not always shown explicitly.

We may wonder why time signals, and hence world time, were not adjusted to International Atomic Time when this time scale was officially recognised in 1971. Of course, TAI runs fast relative to UT. In the very long term, 12 o'clock would have occurred well before the Sun had transited the Greenwich meridian. However, this deviation is of the order of one hour every thousand years. When we realise that in Britain (and other countries), we accept changing the clocks by one hour twice a year and hence living through the summer an hour or so ahead of solar time, the idea of putting the clocks back one hour every ten centuries would appear a rather minor adjustment.

Objections to a general use of TAI are partly sentimental, harking back to the cherished belief that time is determined by the celestial motions, and partly technical. Sailors who use astronomical means of navigation were firmly opposed to the idea of no longer broadcasting Universal Time, which they needed for their fix. The correction required to obtain UT from TAI could be predicted to the nearest second several years ahead, and this is accurate enough for astronomical positioning at sea. However, maritime organisations considered that the risk of error (e.g., a correction made with the wrong sign) was too great. It was also risky to broadcast two types of time signal, one in UT and the other in TAI. In 1970, after long debate and much disagreement [4.25], the CCIR was ready to recommend that UTC be defined with $y_U = 0$ and B equal to a whole number of seconds, modified by leap seconds. At first, the tolerance ε (now referred to UT1 and not UT2) was fixed at 0.7 s, before being set at 0.9 s in 1974. Moreover, time signals had to carry coded information that could be used to obtain UT1 to within 0.1 s.

The new UTC was established according to rules whose details were laid out by the CCIR (see Chapter 7), and was officially recognised by the 14th CGPM in 1971. It was put into practice on 1 January 1972, requiring a time step that was not a round fraction of 1 s. It has been in general use ever since. The system is being increasingly criticised, but attempts to form a continuous time scale like TAI have failed. The leap seconds introduced at intervals of one to a few years, depending on the whims of our planet's rotation, do indeed have their drawbacks, but they must be accepted if we wish to have a single world time. Humankind is perhaps not ready to move to a purely atomic time, and so the Earth's rotation continues to regulate world affairs, at least as far as common usage is concerned.

4.6 Final comments

Since 1972, the measurement of time has officially rested upon the hyperfine transition of caesium, even though the Earth's rotation is still followed in an approximate way. Some believed that astronomical measures of time were to make a comeback when the first millisecond pulsar was discovered in 1982. We shall see in Chapter 8 that such bodies cannot provide a definition of the second, no matter how regular their rotation may be.

It will be noted how one definition of the second succeeds another. Under the pressure of scientific and technical requirements, the use of a new definition may even precede its official adoption.

The situation for time scales is quite different. New ones are created, associated with the current definition of the second, but formerly used scales do not disappear. This can be explained to some extent by their historical role. They have been used to date events and consequently remain in the archives. Naturally, there is every reason to maintain such archives in their original form. This is true for astronomical observations, dated in the apparent time of the place in which they were made. If we wish to make use of them, the dates can be converted into Universal Time or Ephemeris Time, although this involves some degree of interpretation and the result is not therefore unique, whereas the original document remains authentic. As far as atomic time is concerned, it has only existed since 1955.

Another reason is that the various time scales, based on different physical phenomena, retain an intrinsic meaning. Universal Time represents the Earth's rotation. Ephemeris Time and the various coordinate times defined in the context of general relativity are dynamical times, associated with dynamical theories. Who knows whether their long term comparison with atomic time might not lead to crucial data? In the same spirit, a *Pulsar Time* is now under construction.

The survival of formerly used time scales is not a manifestation of conservatism. Neither is the definition of new scales a sign of perfectionism, for it is required by applications. We must accept that we have moved into an era of multiple time scales. Unfortunately for the reader, we shall be unable to simplify without being guilty of omission. At least, it will be appreciated that the idea of a universal time has been safeguarded for public use.

5

Clock time

5.1 Introduction

The base unit of the International System of Units (SI units) in the time domain is a unit of duration, the second. It would therefore have been appropriate to refer to the atomic standards which provide this unit as time standards, or time interval standards. In practice, however, these terms are rarely used, with the term *frequency standard* being preferred, for it is indeed the frequency that they actually supply.

According to the official prescription [5.1], we must distinguish a *primary standard* 'that is designated or widely acknowledged as having the highest metrological quality and whose value is accepted without reference to other standards of the same quantity' from a secondary standard 'whose value is assigned by comparison with a primary standard of the same quantity'. We shall follow this prescription even though the qualitative aspects in the definition of primary standards are somewhat embarrassing. Indeed there exist frequency standards that are certainly not secondary, and yet which could in no sense be considered as primary.

Anyone building a frequency standard generally attempts to produce a wave as close as possible to a sine wave, thereby aiming to ensure spectral purity, and with frequency as close as possible to the nominal value. But the distinct periods of this wave are not identified and numbered. By definition, a clock supplies further information, precisely by identifying these periods. This is usually achieved by numbering pulses at a frequency of 1 Hz which are called *second markers*, or more familiarly, *second pulses*. The word 'clock' also implies the idea of continuous operation over long periods, e.g., of several years. A clock therefore produces a time scale (its *proper time*, in relativistic terminology).