Eur. Phys. J. H
DOI: 10.1140/epjh/e2016-70004-3

THE EUROPEAN PHYSICAL JOURNAL H

The history of time and frequency from antiquity to the present day

Judah Levine[a]

Time and Frequency Division and JILA, National Institute of Standards and Technology and the University of Colorado, Boulder, CO 80305, USA

Received 21 January 2016 / Received in final form 22 January 2016
Published online 21 March 2016

Abstract. I will discuss the evolution of the definitions of time, time interval, and frequency from antiquity to the present day. The earliest definitions of these parameters were based on a time interval defined by widely observed apparent astronomical phenomena, so that techniques of time distribution were not necessary. With this definition, both time, as measured by clocks, and frequency, as realized by some device, were derived quantities. On the other hand, the fundamental parameter today is a frequency based on the properties of atoms, so that the situation is reversed and time and time interval are now derived quantities. I will discuss the evolution of this transition and its consequences. In addition, the international standards of both time and frequency are currently realized by combining the data from a large number of devices located at many different laboratories, and this combination depends on (and is often limited by) measurements of the times of clocks located at widely-separated laboratories. I will discuss how these measurements are performed and how the techniques have evolved over time.

1 Introduction

Time and time interval have played important roles in all societies. The earliest definitions of these parameters were based on apparent astronomical phenomena, so that there is a long-standing strong link between time and astronomical observations. This link still plays an important role in timekeeping. Various types of clocks were developed to measure these parameters, but, initially, they were used to measure relatively short intervals or to interpolate between the consecutive astronomical observations that defined time and time interval.

The earliest clocks, which were used in Egypt, India, China, and Babylonia before 1500 BCE, measured time intervals by measuring the accumulation of a controlled, constant flow of water or sand [1]. The sand clocks of antiquity were very similar to the contemporary hour-glass shaped design in which the sand flows from an upper chamber to a lower one through a very small hole. The clock was (and is) useful primarily for making a one-time measurement of a fixed time interval determined by

[a] e-mail: judah.levine@colorado.edu

the size of the hole and the amount of sand. The design of these clocks is a compromise between resolution, which would favor a rapid flow and therefore potentially a large amount of material, and the maximum time interval that could be measured, which would favor a slower flow, and which would require less material.

There were a number of different designs of water clocks. In the outflow design, water flows out from a small hole near the bottom of a large container. The marks on the inside of the container show the time interval since the container was filled as a function of the height of the water remaining. The flow rate depends on the height of the liquid so that the marks are not exactly equally spaced. In the inflow design, a bowl with a small hole in the bottom is floated on the top of water in a larger vessel. The water enters the bowl through the small hole in the bottom, and the bowl gradually sinks as it is filled. The level of the liquid in the bowl at any instant is a measure of the time interval, and the maximum time interval has been reached when the bowl is completely filled and sinks.

The tine interval measured by these clocks was stable and reproducible, but did not necessarily correspond to any definition of a standard time interval. In contemporary terms we would characterize these clocks as stable but not necessarily accurate without some form of calibration.

The Chinese developed a water clock in which the flow of water turned a mechanical wheel that acted as a counter, which effectively de-coupled the time resolution of the clock and the maximum time it could record. However, this advantage was offset by the need to maintain a constant water flow for long periods of time, so that it was difficult to exploit the ability of the clock to measure long time intervals reproducibly. Without a careful design, the clock was neither stable nor accurate.

In the 16th century, Galileo developed the idea of measuring time interval by means of a pendulum that generated periodic "ticks". He had discovered that the period of a pendulum was a function only its length provided that the amplitude was kept small, so that it could be used as the reference period for a clock [1]. It was probably Huygens who built the first pendulum clock in about 1656. Starting from that time, clocks were constructed from two components. The first component was a device or natural phenomenon. The device was usually a pendulum whose length was carefully controlled (in later designs, the simple pendulum was replaced by more complicated designs that compensated for the variation in the length of the pendulum with temperature. The compensation was realized by replacing the single pendulum rod with a number of rods of different lengths that had different coefficients of expansion). The pendulum generated nominally equally-spaced time intervals – what we would call a frequency standard. The period of a simple pendulum depends on the square root of its length. If the time of a pendulum clock is to be accurate to within 1 s per day, which is a fractional accuracy of about 1.2×10^{-5}, then the fractional length of the pendulum must be held constant to twice this value, or about 2.4×10^{-5}. This value is comparable to the thermal expansion coefficient (fractional change in length per Celsius degree) of many common metals, so that the temperature of the pendulum must be held constant to about 1 degree Celsius to realize the accuracy of 1 s per day.

The second component is some method of counting the number of intervals that had elapsed since some origin. The counter is often implemented by gears driven by an escapement that moved by a fixed angle for every swing of the pendulum. The escapement also provided energy to replace the energy lost to friction. Pendulum clocks continued to be improved over the next centuries; in 1921 William H. Shortt invented a new pendulum clock that used two pendulums [2] – the master pendulum in an evacuated enclosure that actually kept the time and a slave pendulum that provided periodic impulses to the master to replace the energy lost to friction. The time accuracy was about 1 ms per day, or a fractional accuracy of about 10^{-8} (the meaning

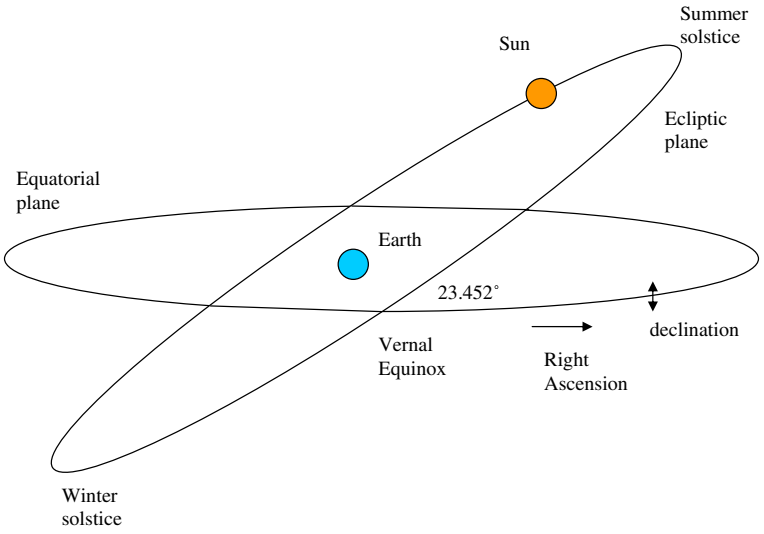

Fig. 1. The astronomical coordinate system. The coordinate system is defined with the Sun in an orbit around the Earth at the origin. The equatorial plane is the projection of the equator onto the sky and the ecliptic plane is the apparent orbit of the sun relative to the Earth. The ecliptic plane is tipped at an angle of 23.452° relative to the equatorial plane and it crosses the equatorial plane at the Vernal (or Spring) and Autumnal equinoxes. The dates of the equinoxes are approximately 21 March and 23 September, respectively. Astronomical positions are measured as a right ascension, which is measured Eastward along the equator from the Vernal equinox and declination, which is measured North and South of the equatorial plane. The angle of right ascension is typically measured in hours rather than degrees, where the conversion is 360 degrees = 24 h or 15 degrees = 1 h. Thus, the coordinates of the summer solstice are right ascension = 6 h and declination = 23.452 degrees North.

of the *accuracy* of a clock or frequency standard is discussed in more detail in the glossary at the end of the document). This was a very significant development because it was sufficiently stable to measure the variations in the length of the astronomical day, and this was the beginning of the end of time scales based purely on astronomical observations.

The combination of the clock and the time origin is often referred to as a time scale, and I will discuss the evolution of the concept of a time scale from antiquity to the present day. An important aspect of the discussion is how a standard time scale is defined and realized and how the time of an individual clock is set to realize the standard. This will naturally lead to a discussion of how clocks are compared and especially how the comparison is realized when the clocks are far apart.

2 Early astronomical time

The time intervals most commonly used in antiquity were the lengths of the apparent solar day (from sunset to sunset, for example), the lunar month (from the observation of one first crescent to observation of the next one), and the solar year, which was important for scheduling both religious rituals and secular agricultural events.

The solar year was often determined as the interval between consecutive Vernal (or Spring) equinoxes (when the Sun is directly over the equator as shown in Fig. 1), but observations of the solstices (where the Sun appears to "stand still") or of the first

appearance above the horizon at sunrise of a star such as Sirius (the "heliacal" rising) were also used. Determining the date of the solstice has the advantage that it does not require knowing the latitude of the observation, but it has the disadvantage that the position of the Sun (and therefore the position of the shadow of an object or opening) changes very slowly at that time of year, so that a precise determination is difficult. It is possible that an important function of Stonehenge was to determine the date of the summer solstice, when the Sun rises over the heel stone as viewed from the center of the stone circle [3]. In Babylonian times, the Spring and Fall equinoxes were the dates of the heliacal rising (the "first" point) of the stars in the constellations Aries and Libra, respectively, but this alignment is no longer correct because of the precession of the equinoxes (the rotation of the line joining the two equinox points with respect to the distant stars), and the Sun now appears to be "in" the constellations Pisces and Virgo, respectively, at these instants [4].

Time was reckoned by counting the number of intervals that had elapsed from some origin; the origin was often chosen to be sufficiently far in the past so that almost all times were positive. For example, the origin of the Roman calendar was set to the founding of Rome, which was taken (perhaps incorrectly) to have occurred in 753 BCE. The Jewish calendar counts years from the creation described in the Bible. The calculation in the Talmud puts this event in the year 1, which corresponds to 3760 BCE. The Julian cycle, which is used by astronomers and (in a truncated version) by the modern time and frequency community, defines the year 1 as equivalent to 4713 BCE. All of these values are ambiguous because they are all based on backward extrapolations computed long after the origin epoch, and all use different definitions for the length of a year [5].

Many societies made use of all of these time intervals for different purposes. The reckoning was complicated in practice because the apparent solar day, the lunar month, and the solar year are not commensurate; there are not an integral number of solar days in a lunar month or lunar months in a solar year. Every group produced a calendar that addressed these complexities in some way, and the resulting calendars were often quite complex. I will limit myself to describing the definition of the day and how it is subdivided, so that the peculiarities and complexities of the many different calendars are outside of the scope of this discussion (the solar day and the solar year continue to play an important role in timekeeping. I will discuss these intervals from an astronomical perspective, which is independent of any particular calendar).

The Babylonians were probably the first to use a base-60 numbering system, perhaps because 60 has many integer factors, and we still combine counting time intervals in base 60 with the Egyptian system of dividing the day into 24 h, each of which has 60 minutes, with each minute having 60 seconds. The definition of the length of the day therefore implicitly defines the length of the second, and vice versa. This is not merely an academic consideration. The linkage between the lengths of the day and the second was an important consideration in the definition of the international time scale, UTC (Coordinated Universal Time), and there has been a lengthy debate on the question of modifying this relationship.

Suppose that I wish to construct a clock based on a device that will produce a signal every second, and these periodic events will be counted to compute elapsed time. The periodic events can also be considered as a realization of time interval, which would have units of seconds per event and as a realization of a frequency, which would have units of events per second. Depending on the application, the same device would provide a realization of time, time interval, and frequency, so that these three parameters are tightly coupled in the sense that a realization of any one of them implicitly realizes the others. This coupling has played an increasingly important role in modern times, and has significantly contributed to the complexity of the definition

of modern time scales, which attempt to satisfy the often incompatible requirements of the different user communities.

If the definition of time interval is based on an astronomical observation, then the reference signal that drives the clock is not a primary frequency standard, but rather is an interpolator between consecutive astronomical observations, and the interval between its "ticks" must be adjusted so that a time interval measured by the clock agrees with the time interval defined by astronomy. The most extreme version of this principle is the definition of canonical or ecclesiastical "hours", which divide the interval between sunrise and sunset into twelve equal parts, with special prayers or rituals associated with the third (Terce), sixth (Sext), and ninth (Nones) hours of the day [6]. These "hours" are obviously longer in summer than in winter and are also a function of latitude, and it would be difficult to design a clock whose frequency reference reproduced this variation.

Defining the length of the day as exactly twelve canonical hours had mystical significance in identifying especially propitious times for prayers and rituals, and made it easier to divide the day into quarters for these purposes. However, it compromised the simplicity of telling time without sophisticated instruments and without the need for a formal method of the transfer of time information, which were two of the advantages of all of the time scales based on apparent astronomical phenomena. The tension between apparent astronomical phenomena, which play a central role in everyday life, and the definitions of time that are derived from other considerations will appear several times in the subsequent sections.

3 Mean solar time

Even without the variation that results from ecclesiastical hours, the elliptical shape of the orbit of the Earth introduces an annual variation in the length of the apparent solar day, measured as the time interval between two consecutive solar noons, for example (apparent solar noon is the instant when the sun crosses the local meridian and is therefore precisely north or south of the observer. The elevation of the sun is a maximum at that time, so that the shadow cast by an object has the minimum length. The shadow will point directly north at most locations in the Northern hemisphere).

In order to appreciate the problem, it is useful to recall the fiction that the Earth is fixed and that the Sun is in an orbit around it (see Fig. 1). The path of the orbit is the ecliptic, and the requirement of conservation of orbital angular momentum requires that the ecliptic be very nearly a plane. The ecliptic plane is tilted about 23.452 degrees with respect to the equatorial plane projected onto the sky, and the two planes intersect in a line that points in the direction of the Vernal and Autumnal equinoxes. It is common in astronomical usage to measure angles (either along ecliptic or equatorial planes) from this line as the reference direction. The angle from the Vernal equinox along the equatorial plane towards the East is called the right ascension, and is typically given in hours rather than degrees, where 1 h corresponds to 15 degrees.

During the interval when the Earth has made one revolution about its axis, the Sun has also moved along the ecliptic, so that the time between consecutive apparent solar noons, for example, is somewhat longer than the time it takes the Earth to revolve by 360 degrees (in principle, this rotation period could be measured easily with respect to the very distant fixed stars, but these measurements are complicated by the small motion of the equinox, which is the reference for the tabulated positions of the stars and by the precession and nutation of the Earth itself. At least initially, it is convenient to ignore all of these complexities). Since the orbital period of the Sun around the Earth is very nearly 365.25 days, in one day the Sun has moved approximately 360/365.25 degrees along the ecliptic. If the period of rotation of the

Earth is taken as 24 h (86 400 s), the motion of the Sun increases the length of the apparent solar day by the product of the seconds of elapsed time for each degree of rotation of the Earth multiplied by the advance of the sun (in degrees) in one day, or $(86\,400/360) \times (360/365.25) = 236$ s or about 4 min. However, the actual value varies through the solar year as I will show in the next section.

The principle of conservation of angular momentum requires that the vector between the Sun and the Earth sweep out equal areas in equal times. Since the orbit is an ellipse with the Sun at the focus, the length of the vector varies through the year, and the angular speed measured along the ecliptic with respect to the equinox must vary as well to satisfy the requirement of conservation of angular momentum. The orbital angular momentum is proportional to $r^2\omega$, where r is the length of the vector from the Sun to the Earth and ω is the angular velocity of this vector with respect to the equinox. The annual variation in the terms that make up this product results in an annual variation in the contribution of the orbital motion of the Sun to length of the apparent solar day. The radius vector is shortest in the winter in the Northern Hemisphere (October, November, and December), the angular motion of the Sun is correspondingly faster during this time, and the apparent solar day is longest. The opposite effect happens during the months of June, July, August.

The variation in the length of the apparent solar day was already known to ancient Babylonian astronomers, and Ptolemy worked to construct a uniform time scale that he could use for tables of the position of the Sun in about the year 100 CE. The Earth is the center of the solar system (and the entire Universe) in his model, and all orbits are perfectly circular. He modeled the apparent eastward motion of the Sun with respect to the distant fixed stars, and the variation in the length of the apparent solar day, as a result of the fact that although the orbit of the Sun was a circle, the Earth was not exactly at the center, but was offset by a quantity called the "eccentric". With some adjustment, this model can look very much like the configuration in Figure 1, and it is not surprising that the model of Ptolemy was the accepted picture of the solar system for over 1300 years.

The solution to the variation in the length of the apparent solar day, which was already being used by the Babylonian astronomers in the first century before the common era, was to define *mean solar time*, which imagines a fictitious sun that moves at a uniform rate on the equator (not the ecliptic) and which agrees as closely as possible with the actual motion of the sun along the ecliptic averaged over a year [7].

This definition of mean solar time models two effects of the motion of the real sun and the resulting annual variation in the length of the apparent solar day: (1) the variation in the angular speed of the sun described above and (2) the apparent North-South motion of the Sun because its actual motion is along the tilted ecliptic and not along the equator as in the model.

The difference between mean and apparent solar time is called the equation of time, and is often displayed on sundials (see Fig. 2). As discussed above, the apparent solar day is longest in the winter months of the northern hemisphere, so that the apparent sun is increasingly behind the mean sun during this time. Each apparent solar day is almost 30 seconds longer than the average during this period, and the minimum integrated time difference is about –14 min and is reached early in February. The maximum integrated time difference is about 16 min and is reached early in November. Each apparent solar day is about 20 seconds shorter than the average during this period.

As a practical matter, mean solar time was determined by observing the distant stars rather than the sun itself; these observations were then converted to the position of the fictitious sun. For example, a clock that was synchronized to apparent solar time (by using a sundial to detect consecutive apparent solar Noons, for example) could be used to record the time of meridian transit of a particular constellation at

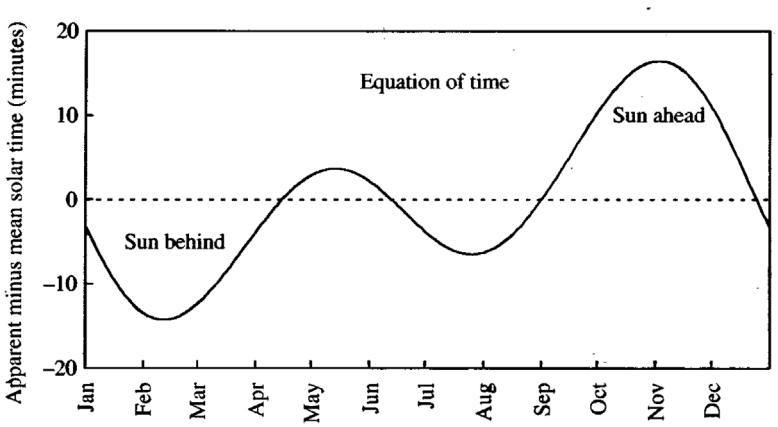

Fig. 2. The equation of time giving the difference between apparent and mean solar time as a function of the day of the year.

midnight apparent solar time. The Sun would be "in" the constellation on the opposite side of the Earth at that instant. The apparent angular motion of the constellation on consecutive nights would give a measure of the length of the apparent solar day. An accurate conversion is somewhat complicated by the motions of the direction of the equinox, because it is typically used as the reference direction for these angular positions.

Mean solar time was measured from Noon prior to 1925, and the day beginning at noon was the "astronomical day." Greenwich Mean Time was defined as mean solar time (starting from noon) and measured on the Greenwich meridian. The start of each day was changed to midnight on 1 January 1925, and Greenwich Mean Time (GMT) was used to identify the time from this new origin. The time referenced to noon was referred to as Greenwich Mean Astronomical Time (GMAT). To avoid the confusion of the change in the origin, in 1928 the International Astronomical Union recommended that the term Greenwich Mean Time be replaced by Universal Time, which was defined as mean solar time on the Greenwich meridian with the day starting at midnight. However, the name Greenwich Mean Time continues to be used today, especially in the United Kingdom where it is the official time scale. In addition, the name GMT is often used (incorrectly, in principle) to refer to Coordinated Universal Time (UTC), which I will discuss below.

The definition of mean solar time is the first of many compromises intended to maintain the linkage between the technical time scale, which emphasized uniform time intervals, and the everyday notion of time based on the length of the apparent solar day. Even when the difference between mean solar time and apparent solar time is largest, it is still smaller than the width of a contemporary time zone (± 30 min in time or 15 degrees in latitude) and the hour offset introduced by daylight saving time, so that it is not significant in everyday timekeeping.

4 Universal time

Although Universal Time was defined as mean solar time, it was actually determined by astronomical observations of the Moon or the stars. Observations of the rotation angle of the Earth were made by timing the meridian transit of a selected group of stars, which is really local sidereal time, and this was combined with the longitude of the station to determine Greenwich sidereal time. The conversion between Greenwich

mean sidereal time (GMST) and Universal time (mean solar time) was calculated based on the position of the Sun as specified by a mathematical expression derived from Newcomb's Tables of the Sun:

$$\begin{aligned} UT &= GMST - (18{:}38{:}45.836 + 8\,640\,184.542T + 0.0929T^2) - 12:00:00 \\ UT &= GMST - (6{:}38{:}45.836 + 8\,640\,184.542T + 0.0929T^2) \end{aligned} \quad (1)$$

where T is the number of Julian centuries of 36 525 days that have elapsed since the origin time, which is 12:00:00 Universal Time (Greenwich mean noon) on 1900 January 0. The coefficients of the time-dependent terms are in units of seconds of universal time. The length of the Universal day is $24 \times 60 \times 60 = 86\,400$ s. The difference in the lengths of the Sidereal day and the Universal day is approximately given by the second term in equation (1). If we convert T from Julian centuries to days, the difference is $8\,640\,184.542/36\,525 = 236.56$ s, a value that is consistent with the value derived above based on a qualitative argument. The term proportional to T^2 implies that there is an increase in the rate as well. The magnitude of the *average* acceleration, which is twice the coefficient of the T^2 term, is approximately 0.186 s/(julian century)^2. The length of the year 1900 in seconds can be computed as the time needed for the value in the parentheses to increase by 24 h or 86 400 s. Since T is in Julian centuries, this value is $86\,400 \times 36\,525 \times 86\,400/8640184.542 = 31\,556\,925.9747$ s, and this value was adopted as the definition of the ephemeris second as I will discuss below. The time origin in equation (1) can be converted to an angle: $(18{:}38{:}45.386/24{:}00{:}00) \times 360 = (67\,125.386/86\,400) \times 360 = 279°\ 41'\ 20.79"$, which is the position of the Sun at the origin time 1900 January 0, 12 h Universal time. The position of the Sun at 1900 January 0 12 h Ephemeris time is 279° 41' 48.04", a difference of 27.25" or about 1.8 ms in time.

The raw data (called UT0) from several stations at approximately the same latitude were compared to estimate the motion of the pole of the rotation axis of the Earth, and the resulting time scale, which was independent of the data from any single station in principle, was called UT1.

The simplest method of observing the meridian transits is a specialized telescope called a transit circle. This is a telescope that is aligned so that it can move only exactly North-South along the local meridian. Therefore, it can be adjusted to match the elevation of the star whose time of meridian transit is to be measured. The uncertainty in the determination of the time of meridian transit is typically a few milliseconds (a fractional uncertainty in the length of the day of order 10^{-8}).

The photographic zenith tube (PZT) is a more sophisticated device. It is a tube mounted vertically, and an image of the sky is reflected in a pool of mercury (see Fig. 3). The image of the sky is recorded photographically at several consecutive times, typically two before meridian transit and two afterwards. The photographic plate and the lens are rotated 180 degrees between exposures. The time of meridian transit of each star is then found by interpolation. In some designs the photographic plate could be moved continuously so as to compensate for the apparent motion of the stellar image caused by the rotation of the Earth.

Very long baseline interferometry (VLBI) is currently used for these observations. This method computes the cross-correlation between signals received from distant radio sources at widely-separated antennas. The time delay that maximizes the cross-correlation in the received signals gives the difference in the distances between the distant radio sources and the antennas, and the evolution of this time difference can be used to estimate the length of the sidereal day (and therefore UT1) and the motion of the pole of rotation the Earth with respect to the distant radio source.

By the 1930s, clocks had improved enough to detect an annual variation in the position of the Earth as predicted by the UT1 time scale. The annual variation had an amplitude of tens of milliseconds, and was ascribed to the annual variation in the

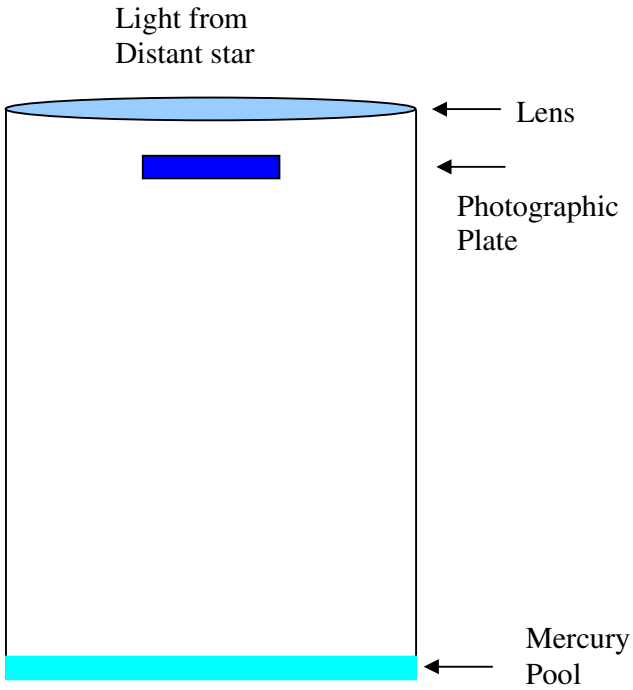

Fig. 3. Schematic of a Photographic Zenith Tube (PZT). The image of a star passes through a lens with a long focal length, is reflected by a pool of mercury and is imaged on a photographic plate located just below the lens. At the US Naval Observatory, the diameter of the lens was about 66 cm, its focal length was about 10 m; the drawing is not to scale. If the photographic plate is rotated about a vertical axis through the center of telescope, the image of the star on the plate is not displaced at meridian transit. In practice two observations were generally made before meridian transit and two afterwards. The instant of meridian transit was computed by interpolation between these images.

moment of inertia of the Earth produced by effects such as the seasonal movement of water from the oceans to the mountains in the Northern hemisphere. The UT2 time scale was defined to correct for this annual variation and was considered to be the most stable astronomical time scale in the 1950s and 1960s.

5 The secular variation in the length of the day

I will discuss atomic clocks in some detail in a following section, but I will briefly discuss the cesium frequency standard here because of its significance in the historical development of time scales.

A cesium frequency standard locks an electronic oscillator to the hyperfine transition in the ground state of cesium (a hyperfine transition is a change in the total spin angular momentum of the nucleus and the outermost valence electron. The ground state of cesium is an S state, so that there is no net orbital angular momentum). The currently defined value for the frequency of this transition is 9 192 631 770 Hz, but the important point for the current discussion is that the frequency of the device, and time intervals that were determined by counting the cycles of this frequency, was much more stable than anything that had come before.

The first operational cesium standard was built by Louis Essen and Jack Parry at the National Physical Laboratory in the UK. The description of its operation was published in August, 1955 [8]. The initial fractional frequency stability was about 10^{-9}, and this was subsequently improved by about an order of magnitude.

Starting in late 1955 and continuing for about 3 years, William Markowitz at the US Naval Observatory compared the UT2 one-second time interval with a time interval derived from the frequency of the cesium standard [9]. The comparison was realized by means of a variation of the common-view time transfer method that I will describe below.

The common-view frequency method used the time signals that were broadcast from radio station WWV, which was operated by the National Bureau of Standards (NBS). The station was located at Greenbelt, Maryland (originally named Beltsville, Maryland) at that time. Essen and Parry periodically measured the number of cycles of the oscillator locked to the cesium reference between two pulses transmitted from WWV that were one second apart as determined by the clock at the radio station. This calibrated the oscillator at WWV to the cesium frequency. At the same time, William Markowitz measured the time interval between the two pulses by means of the oscillators at the Naval Observatory. A comparison of the two measurements effectively transferred the cesium frequency in the UK to the oscillators at the Naval Observatory in Washington. The method is not sensitive to the propagation delays along either path, provided only that the propagation delay (and the properties of the oscillator at each of the end points) does not change during the transmission time. This was a reasonable assumption, even for the relatively unstable oscillators of that era because the propagation delay was only on the order of milliseconds.

The data showed that the length of the UT2 day was increasing by about 1.3 ms per year (a fractional increase in the length of the day of about 1.5×10^{-8} per year), or a total increase in the length of the day of somewhat less than 4 ms over the duration of the experiment. To put these values in perspective, a constant increase in the length of the day by 1.3 milliseconds per year would result in a time dispersion of about $0.5 \times 365 \times 1.3 \times 10^{-3} = 0.24$ s after the first year (note that this is much larger than the contribution of the acceleration term in equation (1) above, which does not include any contribution due to the change in the period of rotation of the Earth).

6 Ephemeris time

The variation in the length of the UT2 day turned out to be irregular and not completely predictable, and the next proposal was to define a time scale based on the ephemerides of the Moon and the planets, which amounts to a definition based on the year rather than the day. The time defined in this way would be the independent variable of the equations of motion, and would be defined so that the observed position would match the position predicted by the equation for some value of the independent time parameter. The initial proposal was to use the equation for the position of the Sun given by Newcomb and to define the second based on the length of the sidereal year 1900. This proposal was adopted by the International Astronomical Union in 1952, and the definition was later change to tropical year (the year measured by the periodic motion of the sun, that is the time interval between consecutive passages of the sun through the same point on the ecliptic). The length of the tropical year 1900 was defined to be 31 556 925.9747 ephemeris seconds, and the origin of the time scale was chosen as that instant when the mean longitude of the sun was the value given by Newcomb's equation: 279 degrees, 41 minutes, 48.04 seconds. The ephemeris time at that instant was 0 January 12 hours ephemeris time precisely. Since astronomical days begin at Noon on the same date as the civil date that started at the previous

midnight, this time is equivalent to Greenwich Mean Noon. The choice of the tropical year meant that astronomical events (such as the spring equinox) would occur on approximately the same calendar date every year, and the choice of the length of ephemeris second resulted in a definition of the second that was very close to the previous value of the second defined by mean solar time.

Ephemeris time was completely uniform in principle and was independent of the irregular variation in the rotation of the Earth. However, it had the great disadvantage that measuring it required lengthy observations often spread over several years. Furthermore, ephemeris time was defined as the independent argument of the equations of motion of the Moon or the Sun, and there was no clock that realized ephemeris time.

The lengthy observation intervals that were needed to realize an accuracy of milliseconds was only part of the problem. A definition of the length of the second based on the length of the year 1900 may have given standards committees a warm feeling, but it provided only headaches for practical metrology. It had the same basic difficulty as the definition of the length of the meter as a fraction of the circumference of the Earth. In both cases, the primary standard of the definition was not really observable, and practical metrology had to be based on an artifact derived from the standard (as for the meter) or on some extrapolation of the standard to a contemporary observable (in the case of the second). This was particularly important for ephemeris time, since, as I have discussed above, it had no physical realization. In the case of the definition of the second, the contemporary observable was the frequency of an atomic clock, and the first question was how to transfer the length of the ephemeris second from an astronomical observable to one based on the frequency of an atomic transition.

7 The cesium second

The goal of the comparison between the cesium frequency and the ephemeris second was to transfer the definition of the length of the ephemeris second to a value determined by counting cycles of an oscillator that was locked to the hyperfine transition in the ground state of cesium with the smallest possible effect on practical metrology (I will discuss the motivation for choosing a transition in cesium below. From the perspective of the current discussion, the transition in cesium was chosen because an oscillator stabilized to that transition was much more stable than the astronomical observations).

The ephemeris second could be determined from the observation of any astronomical object in principle, and the comparison was performed using the Moon because it has the shortest period of any object in the solar system. Since the Moon is a large object, its position is typically determined by occultation – observing the time when it passes in front of a given star so that the star disappears from view.

Although the method of occultation is simple in principle, there are a number of practical difficulties. The Moon is much brighter than the distant star, so that it is difficult to find a photographic exposure that can record both images simultaneously. A much more serious problem is that the apparent monthly period of the Moon is much faster than the apparent annual motion of the distant star – that is why the Moon was chosen to start with. However, this disparity means that a telescope that is driven to compensate for the rotation of the Earth and keep the star at a fixed position in the field of view (by rotating to the West by approximately $360/1440 = 0.25$ degrees per minute) will not move at the correct rate to keep the image of the Moon from becoming blurred due to its apparent orbital motion. The orbital period of the Moon is approximately 29.5 days, so that the apparent angular motion of the Moon has an additional term of somewhat more than $360/(29.5 \times 24) = 0.5$ degrees per hour

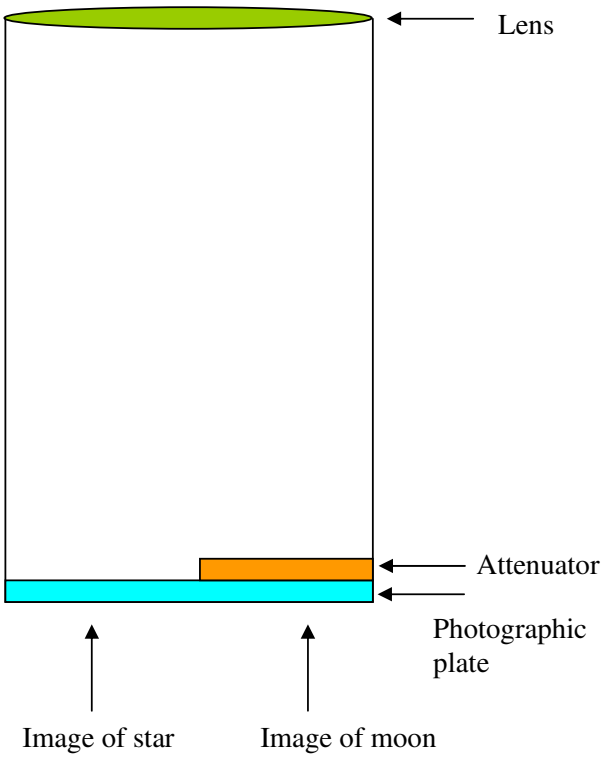

Fig. 4. Schematic of the Markowitz Moon Camera. The attenuator compensates for the difference in the brightness of the moon and the much fainter background stars. The attenuator and telescope are movable to compensate for the difference in the motions of the two images as the Earth rotates.

(or 0.5 seconds of arc per second of time) to the East. In other words, if the telescope is rotated so as to compensate for the apparent motion of the instrument with respect to the distant stars then the image of the moon will appear to move Eastward with an angular speed of approximately 0.5 degrees per hour, or approximately one moon-diameter per hour. Conversely, if the telescope is adjusted to track the apparent motion of the moon, then the image of the distant stars will move Westward at the same rate.

Markowitz's solution was the moon camera, which he designed in 1951 and which was used to calibrate the length of the cesium second starting in 1952 [10]. The moon camera compensates for the great difference in brightness between the moon and the background stars by means of a relatively thick glass-plate attenuator placed only in front of the image of the moon (see Fig. 4). It compensates for the difference in the apparent motion of the moon and the background stars by slowly tilting the plate so as to move the part of the image behind the plate relative to the remainder of the image. This method can compensate for the differential apparent motion between the moon in one part of the image and the background stars in another part.

The data were acquired in the same way as described above from the middle of 1955 to 1958. For each data point, a photograph was taken of the Moon at a known universal time, and the corresponding ephemeris time was determined based on its observed position with respect to the distant stars. Based on these measurements, the length of the ephemeris second was measured to be equal to $9\,192\,631\,770 \pm 20$ cycles of the cesium frequency at epoch 1957.0, which was the midpoint of the observation

period. The fractional uncertainty of about 2×10^{-9} was limited primarily by the difficulty of determining the position of the moon and thus ephemeris time at the instant of the observation. The accuracy claimed by Markowitz et al. was much better than would be expected based on the experimental limitations of that time [11].

Markowitz et al. also published values for the time difference between ephemeris time and universal time as a function of epoch. The time difference was 30.5835 s at the epoch 1957.0, and the data showed an increase in this difference by about 1 s by the end of the experiment in mid-1958. The variation in the time difference was quadratic, implying a linear increase in the rate of the evolution of the time difference. The rate at 1950.0 was estimated as 0.469 s per year. If the rate was zero in 1900, the change in the rate was about $0.469/50 = 9.4 \times 10^{-3}$ s per year^2, which is a fractional change in the length of the day of about 3×10^{-10} per year, assuming that the change in the rate was linear over that time period. The quadratic coefficient in Markowitz's model was 0.0615 s per year^2, which is a much larger fractional change in the length of the day of about 2×10^{-9} per year. The significant difference in these estimates illustrates the difficulty in predicting the length of the universal-time day in cesium seconds.

I will discuss the historical development of International time and frequency standards in the next section, but I will note here that the value of Markowitz and Essen was used to define the length of the cesium second. In October of 1967, the 13th convocation of the General Conference on Weights and Measures [12] (which I will explain below) declared that, "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 cesium 133." This new definition replaced the previous one that was adopted in 1960: "The second is the fraction 1/31 556 925.9747 of the tropical year for 1900 January 0 at 12 h ephemeris time" [13]. In 2016, the cesium definition (with additional clarifications that I will discuss in the next section) is still the official definition of the length of the second. In order to provide continuity with the previous definition of time, the cesium time scale was set equal to the time of UT2 at 0 hours UT2 on 1 January 1958.

The consequence of this definition is that the fundamental parameter is really frequency (and, implicitly, time interval), and time is now a derived quantity. This is a fundamental change from the start of the discussion, where a clock was basically an interpolator and its driving frequency had to be adjusted to match an external time standard. The choice of an atomic transition frequency (rather than a frequency defined by a physical artifact, as with the kilogram) was a strong argument in favor of this definition at the time, and the principle of basing the standards of the fundamental quantities of metrology on the invariant properties of atoms continues to the present time. Although cesium clocks are available commercially, they are not suitable for everyday use, and the commercial devices have various systematic frequency offsets and do not accurately realize the definition of the second. A primary cesium standard that does realize the definition of the second is a laboratory-grade device and only a relatively small number such devices (generally fewer than 10) are currently operating in various standards laboratories or National Metrology Institutes.

7.1 Discussion

The evolution of the definition of time from apparent solar time to the length of the second defined by an atomic transition may seem like a natural progression, but at each step the need for the stability of time intervals (and, implicitly, the stability of the definition of frequency) played a central role in the decision. This was certainly an improvement from some perspective, but it had the consequence

of moving the technical definitions of time and time interval further and further away from the everyday notions of these quantities, which are implicitly related to apparent solar time, or, at worst, to mean solar time and UT1. The definitions of time and time interval in terms of ephemeris time and then atomic time removed the variability in astronomical observations from its impact on the technical definitions of these quantities, but it did nothing to remove the variability in these quantities themselves. The result was an inevitable and predictable tension between the fully physical cesium-based definitions of time and time interval, and the ordinary uses of these parameters, where I include astronomy and orbit calculations, which usually use UT1 as the measure of time, in the category of ordinary uses. This tension continues to the present day, and is behind the various proposals to modify the definition of Coordinated Universal Time (UTC), which is based on the cesium second, and which I will describe in the following sections.

The definition of the length of the cesium second is exact, and the uncertainty in the measurements of Markowitz and Essen plays no role in the definition of the length of the second or in practical metrology that determines a frequency or a time interval in units of the cesium transition frequency. The sole effect is to push the uncertainty in the length of the cesium second with respect to the ephemeris second into any measurement of an astronomical period in cesium seconds.

A more subtle point is that the numerical value in the definition is derived from the realization of the cesium transition frequency by the Essen atomic clock, with the implicit assumption that the transition frequency was a fundamental constant of nature and that *all* realization of the cesium frequency would produce the same numerical value within experimental uncertainties. This was (and remains) one of the primary arguments for realizing the definition of the frequency standard by the use of an atomic property rather than by an artifact, such as a precision resonant circuit or device, or by a complicated set of astronomical observation, such as was required for ephemeris time.

The assumption that all realizations of an atomic clock would yield the same numerical value within experimental error was a completely adequate assumption for this original determination, since its uncertainty was dominated by the difficulty of determining ephemeris time at the instant of observation of the occultation. However, this does not necessarily imply that different technical realization of the cesium transition frequency would agree within experimental uncertainty when compared against each other, and the definition of the length of the cesium second had to be qualified as different realization of the cesium frequency were developed and as the methods of comparing different cesium standards improved. I will discuss this point in more detail in the next section.

8 Accuracy of atomic clocks

The accuracy of the frequency of an atomic clock is generally described by presenting a list of known systematic perturbations to the frequency and the remaining contribution of each one to the uncertainty after the magnitude of each of the effects has been estimated and its impact removed from the result (the use of the word "accuracy" in this way is different from its use in other contexts, where it is often taken to indicate the difference between the result of a measurement and the accepted value). The most important contributions to the uncertainty are listed in Table 1 for the cesium fountain standard NIST-F2. The single largest correction is for the General Relativistic frequency shift with a magnitude of almost 1.8×10^{-13}. This frequency offset is a result of the fact that the standard is located in Boulder, Colorado, and is approximately 1800 m above the geoid, which is the reference gravitational potential

Table 1. The major systematic biases considered for the fountain standard NIST-F2. The units of both columns are 10^{-15} fractional frequency. The magnitude gives the size of the effect. The uncertainty is the residual contribution to the overall uncertainty after the effect has been measured or estimated and removed from the result.

Physical Effect	Magnitude	Uncertainty
Relativistic Frequency Shift	179.87	0.03
Second-order Zeeman	286.06	0.02
Blackbody radiation	–0.087	0.005
Spin exchange (low density)	–0.71	0.24
Total Standard Uncertainty		0.11
Data from Table 1 of [14]		

surface for the definition of the standard second. However, the height of the standard above the geoid is well known, so that the residual correction is very small. On the other hand, the spin exchange correction, which is small to begin with, makes the largest contribution to the overall uncertainty. The magnitudes of the most important effects are given in Table 1. For a complete list and an explanation of the effects, see [14].

The magnitude of the contribution of each effect can be estimated either theoretically or by an ancillary measurement, and the residual uncertainty is generally a measure of the estimated accuracy of the theoretical model or the uncertainty of the ancillary measurement. The accuracy estimated in this way is generally not statistical in nature and is usually not improved by acquiring more frequency data. It is often called a "type B" error, to distinguish it from "type A" errors, which are statistical in nature and which can typically be improved by acquiring more frequency data.

The effects that contribute to the "A" errors are statistical in nature, and the overall statistical uncertainty is generally computed as the square root of the sum of the squares of the contributions, which is the usual statistical method assuming that the contributing terms are not correlated with each other. The overall "B" error is more difficult to estimate since the uncertainties of the contributions are not statistical. The same difficulty is present in combining the "A" and "B" errors. Some experiments report the linear sum of the two contributions, while others calculate the square root of the sum of the squares.

8.1 The black body correction

The black body radiation emitted from the cavity surrounding the cesium atoms contributes to a broadening and a shift of the atomic resonance. The shift is a combination of the AC Stark effect (due to electric fields) and the AC Zeeman effect (due to magnetic fields). If the atoms have a random motion with respect to the cavity, the black body radiation results in a broadening of the resonance line due to the first-order Doppler shift. There is also a second-order Doppler effect caused by the time dilation predicted by Special Relativity.

The magnitude of the black body correction was estimated by Itano, Lewis, and Wineland in 1982 [15] based on the results of Gallagher and Cooke in 1979 [16]. The fractional frequency shift due to the AC Stark and Zeeman effects depends on the

temperature in Kelvin, T:

$$\frac{\delta\omega}{\omega} = -1.7 \times 10^{-14} \left(\frac{T}{300}\right)^2$$
$$\frac{\delta\omega}{\omega} = 1.3 \times 10^{-17} \left(\frac{T}{300}\right)^2 \quad (2)$$

where the first equation is frequency shift caused by the AC Stark effect and the second equation is the much smaller Zeeman shift. The uncertainty in the calculation was estimated as of order 1%, or about 2×10^{-16} in fractional frequency. This was a significant correction to the frequencies of the primary frequency standards that were available at that time. For example, the contemporary standard at the Physikalisch-Technische Bundesanstalt (PTB) in Germany had an uncertainty in the fractional frequency of about 6.5×10^{-15} [17] and the blackbody correction was not applied. Most of the primary frequency standards operating at that time operated at or near a temperature of 300 K, so that the correction was essentially the same for all of them, and the net effect was an offset in the definition of the cesium second relative to the spirit of the definition, which envisaged the frequency of an unperturbed cesium atom. To complicate matters further, it was difficult to estimate the uncertainty in the magnitude of the correction, since the cavity and vacuum chamber were not black bodies and the effective temperature was not well known.

The black body correction was extensively discussed at the BIPM Working Group on International Atomic Time (TAI) in 1995 and at the 13th meeting of the Consultative Committee for the Definition of the Second (CCDS) in 1996. After some discussion, the CCDS recommended that a correction for black-body radiation be applied to all primary frequency standards. The result was to insert a step in the realization of the cesium frequency of approximately 2×10^{-14}. The BIPM implemented this change in the TAI time scale by multiple steering corrections of amplitude 1×10^{-15} applied at 60 day intervals. The intent was to minimize the disruption that would result from applying the relatively large correction in a single step.

A strong motivation for applying the correction was the realization that cryogenic primary frequency standards were already being developed, and these standards would have a very different black body correction. Since the cooled-atom standards were likely to be the dominant type of standard in the future, a step in the realization of the SI second was almost inevitable. Furthermore, it would be difficult to combine data from the newer cryogenic and older warm-temperature primary frequency standards during the transition period when both types of devices were contributing to the definition of the cesium second.

8.2 The gravitational frequency shift

One of the consequences of General Relativity is that there would be an apparent difference in the frequency of an oscillator if that frequency was measured at a location where the gravitational potential was different from the potential at the source. If the difference in gravitational potential is $\Delta\phi$, the apparent change in the fractional frequency would be:

$$\frac{\delta\omega}{\omega} = \frac{\Delta\phi}{c^2} \quad (3)$$

where c is the speed of light. If we set the gravitational potential to be 0 at infinity, the potential at the surface of the Earth is

$$\phi = -\frac{GM}{R} \quad (4)$$

where G, M, and R are the gravitational constant, the mass of the Earth and the radius of the Earth, respectively. Near the surface of the Earth, the change in the potential for a vertical displacement ΔR and the fractional frequency change caused by the vertical displacement are given by:

$$\Delta\phi = \frac{GM}{R^2}\Delta R = g\Delta R \tag{5}$$

$$\frac{\delta\omega}{\omega} = \frac{g}{c^2}\Delta R = \frac{9.8}{9\times 10^{16}}\Delta R = 1.1\times 10^{-16}\Delta R \tag{6}$$

where g is the acceleration of gravity near the surface of the Earth. The parameter ΔR is in meters and is the vertical distance between the frequency standard and the observer, who is assumed to be located on the geoid. For example, a frequency standard in Boulder, Colorado is approximately 1800 m above the geoid. An observer with an identical clock on the geoid would see the standard in Boulder as having a fractional frequency that was higher by 1.98×10^{-13} relative to the local clock.

At its 81st meeting in 1992, the International Committee on Weights and Measures created a working group to study the question of how to extend the definition of the SI second to recognize the variation in the apparent frequency of a standard defined in the previous equations and to modify the definition of the cesium frequency to incorporate this effect. The report was published in 1997 [18].

The original definition of the SI second was taken to define proper time and proper frequency – the time and frequency that would be measured by an observer at rest with respect to the source and very close to it. The extended definition is that the duration of the second is to be defined as the frequency of the hyperfine transition in cesium as realized on the rotating geoid. This is a coordinate time scale; it lacks the theoretical purity of the initial definition in terms of the frequency of the unperturbed cesium atom (presumably in empty space and far away from any matter), but it has the significant advantage that it is an observable whereas the purist definition is not a practical observable. The practical realization of this definition requires a measurement of the distance between a primary frequency standard and the local geoid. It would be possible in principle to use this sensitivity to determine the position of the local geoid by comparing the frequency of a local primary standard with the frequency of a second device located at a reference location. This method is currently (in 2016) not competitive with more conventional methods for mapping the geoid.

9 The choice of cesium

The energy levels of atoms and the frequency of the transitions between them are affected by various external influences such as electric and magnetic fields and by collisions with other atoms. To minimize the impact of these perturbations, a frequency standard should isolate the atoms as much as possible from these effects.

The frequency shifts due to collisions of the clock atoms with the background gas could be minimized by placing the atoms in an evacuated chamber. The fundamental uncertainty in the measurement of the transition frequency would be inversely proportional to the time interval during which the atomic resonance could be probed, and this led naturally to the idea of a beam of atoms in a long evacuated apparatus. With the vacuum systems of the 1950s and 1960s, it was possible to reduce the pressure in a vacuum system to about 1.33×10^{-4} Pa (10^{-6} Torr), and the mean free path of a clock atom in the residual background gas was tens of meters at this pressure.

The alkali atoms in column 1 of the periodic table (Lithium, Sodium, Potassium, Rubidium, and Cesium) are particularly well suited to atomic-beam systems (Atomic

beams of hydrogen are neither easy to produce nor easy to detect, and Francium, the element in period 7 below cesium is radioactive and difficult to work with for that reason). It is relatively easy to produce a beam of these atoms by heating them in a small chamber with an exit slit. For various technical reasons, the slit was often rectangular, with a height of a few millimeters and a width of a few tenths of a millimeter. The atoms could be detected by surface ionization on a heated wire of tungsten or various platinum alloys. When the work function of the surface is greater than the ionization potential of the atom, an atomic electron can tunnel to the surface, leaving a positive ion behind. The ion can be collected on a nearby electrode and the resulting current measured using conventional electrical methods [19].

The atoms emerging from a thermal oven are in the electronic ground state, and the ground state of all of the alkali atoms is a single S electron (with spin 1/2) outside of a closed core. If the spin of the nucleus is not zero, there is a difference in energy between the spin of the nucleus and the spin of the electron parallel and anti-parallel. The frequency associated with this hyper-fine energy difference is the clock frequency. The spin of the nucleus, and therefore the magnitude of the hyper-fine splitting vary from atom to atom and within isotopes for the same atom. The nuclear spin in cesium-133 is 7/2, and the hyper-fine splitting between the parallel state with angular momentum $F = 7/2 + 1/2 = 4$ and anti-parallel state with angular momentum $F = 7/2 - -1/2 = 3$ is particularly large, making it a good candidate for a frequency reference.

The energy difference between the upper and lower hyper-fine states is about 6.5×10^{-24} J (about 4×10^{-5} eV), which is much smaller than the thermal energy at a nominal temperature of 300 K, which is about 4.1×10^{-21} J or 0.025 eV. Therefore, the atoms emerging from the oven are about equally divided between the upper and lower hyper-fine states. If the atoms enter a magnetic field, the field establishes an axis of quantization, and the hyperfine states are split into components based on the projection of the total angular momentum along the quantization axis. Each F level is split into $2F + 1$ components, with projections along the magnetic field, m_F, having integer values of F, $F - 1$, ..., $-F$. Thus the $F = 3$ level is split into 7 components and the $F = 4$ level has 9 components. The states have different magnetic moments, and a particular state can be selected by passing the beam through an inhomogeneous magnetic field, which is oriented perpendicular to the direction of motion of the beam. The field inhomogeneity exerts a force proportional to the product of the magnetic moment and the gradient (the magnetic moment is a function of the magnetic field in cesium, and this complicates the practical realization. This dependence implies that both the magnetic field and its gradient must be constant across the cross-section of the beam). If the magnetic field is small, the energy levels with $m_F = 0$ have energies that are almost independent of the value of the field, and making using of the frequency of this transition attenuates the contribution of the uncertainty of the magnetic field value at the position of the beam and its spatial variation. The frequency of this transition varies as the square of the magnetic field, and the measurement must be extrapolated to zero magnetic field to realize the definition of the length of the second. A physical slit is placed at the exit that allows only atoms with the one trajectory corresponding to the state with $m_F = 0$ to pass through. The inhomogeneous magnetic field was produced by the "A" magnet (see Fig. 5 and [20]). For small values of the magnetic field, the energy levels with m_F not equal to zero vary almost linearly with magnetic field, so that the transition frequency between these states is almost a constant independent of the applied field. These frequency offsets are used to determine the value of the magnetic field at the position of the beam.

If the atoms then pass through a region where they interact with an electromagnetic field whose frequency corresponds to the hyperfine transition frequency, they will

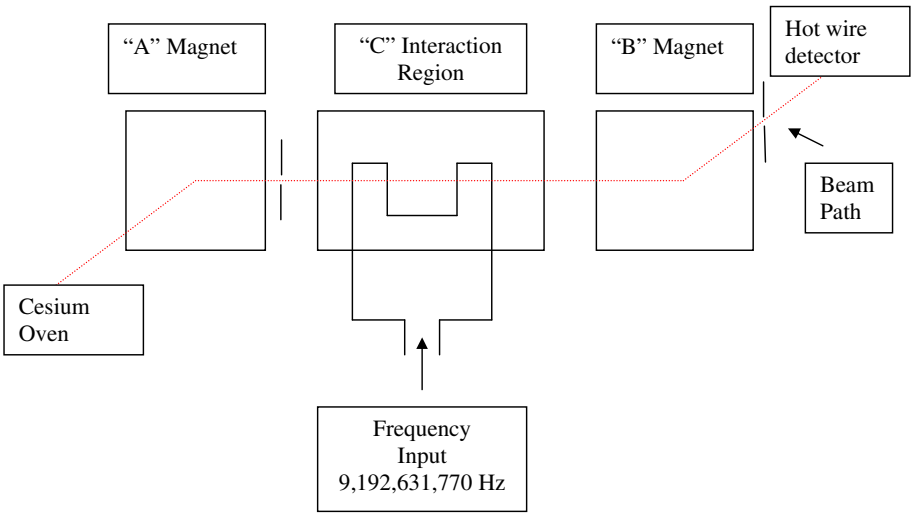

Fig. 5. The atomic beam portion of a cesium standard. A beam of cesium atoms in the electronic ground state emerge from the oven and travel to the A magnet. Those atoms that are in the lower hyperfine state are deflected by the inhomogeneous magnetic field of the A magnet and can pass into the C region. The atoms in other states are blocked by slit. The atoms in the C region pass through two interaction regions where they interact with the microwave field, and those that make a transition to the upper hyperfine state and therefore have the correct magnetic moment are focused onto the detector by the B magnet. As with the A magnet, atoms in other states are blocked by the slit. The selected atoms strike a heated wire of tungsten or an alloy of platinum and are surface ionized. The ions are then collected by a nearby plate and the resulting current is measured by conventional methods. The diagram shows the "flop in" method in which only atoms that make the transition to the upper state can reach the detector.

make a transition to the upper hyper-fine state, and the frequency of the electromagnetic field is locked to the maximum in this transition probability. If a small magnetic field is present in this region, the hyper-fine state is further split into sub-levels, whose energies depend on the projection of the atomic angular momentum on the quantization direction determined by the direction of the magnetic field. This magnetic field is called the "C" field in the literature. The clock frequency is chosen as a transition between particular sublevels whose energy difference is as independent as possible of the magnitude of the C field, and other transition frequencies, which do depend on the magnitude of the magnetic field, are used to calibrate the magnitude of the field.

The final step before the detector is the "B" magnet, which is similar to the "A" magnet, except that it now passes only those atoms that have made the transition to the upper state in the C region (the trajectory of the atoms is shown as a dotted line in Figure 5, and mechanical slits limit the atoms reaching the detector to those that have made a hyperfine transition in the C region. This configuration is referred to as a "flop in" configuration. The "flop out" configuration, which detects the decrease in the flux of atoms that do *not* make the hyperfine transition, generally has a poorer signal to noise ratio because of the larger shot noise in that beam).

The configuration that I have described was well-understood in the 1950s, and had been used for some time to study atomic properties, such as the quadrupole moment of the nucleus. It could have been made to work with any alkali atom in principle, but cesium had the additional advantage that it was the heaviest of alkali atoms and therefore had the smallest velocity at room temperature (a few hundred

meters per second). This simplified the design of the magnetic deflection "A" and "B" magnets and maximized the interaction time in the "C" region. The length of the C region could be as large as a few meters in a primary frequency standard, so that the interaction time was tens of milliseconds. Commercial devices had much shorter interaction regions with a correspondingly shorter interaction time on the order of 1 ms.

The combination of all of these advantages was definitive in the 1950s, and, until recently, cesium has remained the atom of choice for the definition of the length of the second for these reasons. The original "A" – "C" – "B" design has been improved over the years from the engineering perspective, but there was no fundamental change in the design for about 40 years. Even then, the main change was to replace the "A" and "B" magnets with state selection based on optical interactions rather than magnetic deflection.

9.1 Improvements to the cesium frequency standard

The conventional configuration that I described in the previous section mechanically selected the atoms in the correct hyperfine sub-level. The atoms in the desired sub-level passed through a series of narrow slits, while that were not in the correct sub-level were blocked by mechanical stops. The design did not make good use of the atoms emerging from the source because most of them were in the wrong hyperfine sub-level. This limitation could be overcome to some extent by increasing the flux of atoms emerging from the source, but the increase in the flux that could be realized in this way was limited by the increase in the velocity of the atoms passing through the interaction region and by atom-atom scattering in the beam, especially near the exit slit of the source.

The ability to optically pump atoms into the lower state of the clock transition by means of suitably designed diode lasers made it possible to replace the magnetic state selection of the "A" magnet with optical pumping. This was much more efficient than magnetic state selection, which did not use atoms that were not in the correct hyperfine state when they emerged from the oven.

The principle of optical pumping can be illustrated by a simple example. If cesium atoms in the $F = 4$ hyperfine level of the ground state are illuminated by light at 852 nm, the atoms can make a transition from this state to the $^2\mathrm{P}_{3/2}F' = 4$ excited level. The excited atoms decay back into both the $F = 4$ and $F = 3$ levels of the ground state with approximately equal probabilities. The atoms that decay back into the $F = 4$ level of the ground state are excited again; some fraction of them decay to the $F = 3$ state and some to the $F = 4$ state, where they are excited yet again. The net effect is to transfer the atoms into the $F = 3$ hyperfine state of the ground state. Approximately one-fifth of the atoms wind up in the $F = 3m_F = 0$ state, which is the lower state of the clock transition. More complicated optical pumping systems are possible if two lasers are used [21]. Without optical pumping, the population in the lower clock state would be about one sixteenth, since the various magnetic sublevels are approximately equally populated when the atoms leave the source. In addition, this technique is not sensitive to the velocity of the atomic beam, whereas the atoms that satisfy the arrangement of the mechanical stops must be traveling at the correct velocity in order that the deflection matches the mechanical configuration. A similar technique could be used to replace the "B" magnetic state selector that was used to detect the atoms that had made a transition to the upper state in the interaction region. These optical-pumping techniques were used in a number of primary frequency standards starting in the early 1980s [22]. This improved the signal to noise ratio of

the control loop, but the improvement was really not definitive from the accuracy perspective.

The more important change in the design of cesium standards was the ability to cool atoms by using the interaction with lasers and of trapping the cooled atoms in various electromagnetic configurations. If an atom is illuminated by a beam of photons whose frequency is somewhat lower than the transition frequency between two states, then the atom may absorb the photon and make a transition to the upper state. When it decays back to the ground state and re-emits a photon, the energy of the emitted photon will be greater than the absorbed photon on the average, so that the atom has lost energy and been cooled in the process [23].

Once an atom was trapped, the interaction time with an external electromagnetic field could be greatly increased, because it no longer was limited by the time of flight of a beam through an interaction region. The increased interaction time implied a correspondingly narrower resonance line-width. The ability to decrease the thermal velocities of atoms by laser cooling removed one of the original advantages of cesium, since the mass of the atom was no longer a determining factor in the length of the interaction time with the external field.

A second advantage of laser cooling was that it allowed a "fountain" geometry in which atoms where projected vertically with a speed of only a few meters per second, and then fell down again because of gravity (see Fig. 6). This idea had been proposed much earlier by Zacharias in 1954 but was not practical at that time because lasers and laser cooling had not been known at that time [24]. Since the average velocity of a beam at a temperature of 300 K is several hundred meters per second, Zacharias proposed using only the atoms whose velocities were at the lower end of the velocity distribution. The number of atoms that could be used was therefore quite small. Even so, the proposal required an evacuated flight path that was several meters tall. The interaction region with the electromagnetic field was near the bottom of the trajectory, so that the atoms passed through the region twice – once on the way up and again on the way down. In addition to an increased interaction time because of the slower speed, this design attenuated many of the systematic errors of the older thermal beam design and facilitated estimating the remaining systematic offsets. The fountain design was developed initially in 1995 [25], and has become pretty much the universal realization of the cesium frequency for these reasons. The most recent fountain design, reported in 2014, incorporates a cryogenically-cooled microwave cavity and flight region, which is expected to reduce the uncertainty due to the blackbody radiation shift [26]. It is quite likely that this design will be the ultimate frequency standard based on cesium, and the next generation of primary frequency standards will use a different atomic system.

10 A possible re-definition of the reference frequency

Improvements in vacuum systems have removed the effects of collisions with the background gas in the apparatus, but the slower velocity of the atoms in the fountain geometry and the increased flux of the atomic beam, which was used to improve the signal to noise ratio of the clock interaction, mean that atom-atom collisions within the beam itself are now more important – a problem that was not significant in the very low density thermal beams of old. The atom-atom collision cross-section for rubidium is smaller than for cesium, so that rubidium fountains are competitive with fountains based on cesium, and they can exhibit superior stability because they can operate at a higher beam flux without the limitation of frequency shifts due to atom-atom collisions. The community has recognized this advantage by designating a frequency standard based on rubidium as a "secondary representation" of the second [27, 28],

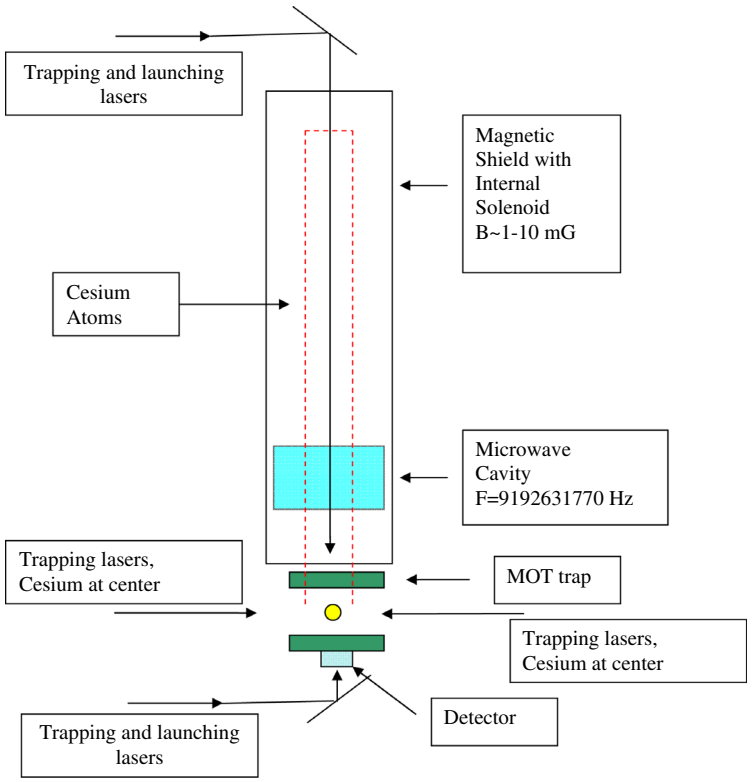

Fig. 6. A cesium "fountain" primary frequency standard. The cesium atoms are trapped and cooled in the magneto-optical trap (MOT) at the bottom of the device. After additional laser cooling, the magnetic field is switched off and the atoms are launched upward at a speed of a few m/s by detuning the upward laser beam to the blue and the downward laser beam to the red of the atomic transition frequency. The atoms pass through the interaction region twice – once on the way up and again on the way back down approximately 0.75 s later; this "Ramsey" configuration is essentially equivalent to a folded version of the separated interaction regions of the classical thermal-beam standard shown in the previous figure. However, the time between interactions is much longer in the fountain geometry, and the bi-directional flight path simplifies the evaluation of the systematic frequency offsets of the device.

even though the hyperfine frequency, 6.8 GHz, is somewhat lower than for cesium. Improvements in electronics have reduced the advantage of a higher microwave frequency to some extent.

A much more important development is the optical comb [29], which supports a direct link between an optical and a radio frequency. An optical frequency is about 5 orders of magnitude higher than the cesium frequency, and this higher frequency could support a much more stable standard of frequency in principle, assuming that the much shorter period of the optical frequency can be exploited to realize a corresponding increase in the resolution of the measurement process. The first caveat is that the systematic offsets of an optical-frequency device have to be understood and estimated or removed, and a number of promising solutions to these problems have been demonstrated in the last few years.

The second problem, and until recently the more fundamental one, was the link between any optical frequency and conventional time and frequency metrology, which

is based on measuring and counting electrical signals and pulses. The link between the optical frequency of the He-Ne laser (3.39 μm at first and then 632.8 nm) and the cesium frequency of 9 GHz had been established in the 1970s, but the technique that was used required a frequency synthesis chain involving several intermediate-frequency lasers and beat-frequency measurements that used whisker-diode detectors that were difficult to fabricate and were unreliable. It was a major one-time effort that was not at all suitable for normal operations [30].

The development of the optical "comb" has changed all of that, since it allows a link between almost any optical frequency and a conventional electrical frequency source to be established in one step. The basis of the comb is a laser that produces a chain of very short pulses that are equally spaced in time. In a simple configuration, a non-linear absorber in the laser cavity can produce these pulses by making the gain in the optical cavity vary very strongly with the amplitude of the signal. The interval between the pulses is determined by the round-trip optical transit time inside the laser cavity.

From the Fourier spectrum perspective, a series of delta functions in the time domain is equivalent to a series of delta functions in the frequency domain, where the spacing between the "teeth" in the two domains are reciprocals of each other. If the interval between the teeth is sufficiently short in the time domain, the Fourier spectrum can span an entire octave in frequency. The lowest frequency can be optically doubled and compared to the highest frequency, and the octave relationship can be detected and stabilized by measuring this beat frequency. The frequency between consecutive Fourier "teeth" is typically in the microwave frequency region, and can be measured by means of conventional time interval counters; the frequency of any laser can be determined by measuring the beat frequency between the laser output and the frequency of the nearest frequency "tooth" in the comb. The method actually involves a number of additional engineering details, but the comb technique has been demonstrated to transfer the stability of an optical signal to a conventional electronic oscillator without degradation [31].

The question of which optical frequency to use as a the next definition of the length of the second is still under study, but the comb technique pretty much guarantees that some optical frequency will be the next standard for the length of the second [31]. The usefulness of these devices in practical time and frequency metrology will depend on developing methods to distribute time and frequency information without degrading the characteristics of the primary standard. One possible transfer method is to use dedicated "dark" optical fibers [32], and high-accuracy time transfer across free space is also possible over relatively shorter distances [33]. Atmospheric turbulence is likely to result in scintillation that will make this method difficult to use over distances longer than 5 or 10 km.

The connection between the observed frequency of a device and the difference in the gravitational potential between the device and the observer (the gravitational "red shift") will become increasingly important as the potential accuracy of the clocks improves. As I discussed above, this difference introduces a fractional frequency difference of approximately 1×10^{-16} per meter of vertical height difference near the surface of the Earth, so that knowing the position of the transmitter and receiver with respect to the geoid at the level of a few cm will probably be needed in the foreseeable future. Alternatively, there have been some proposals to exploit this sensitivity to measure the position of the geoid. It is not clear that this technique would be competitive with more conventional means of determining the geoid in general, although it might be useful in measuring short-wavelength variations in the geoid height [34].

The published "accuracy" of the devices that are based on an optical reference frequency is estimated by a list of the residual contributions of known systematic offsets in the same way as I described above for cesium. The more conventional definition

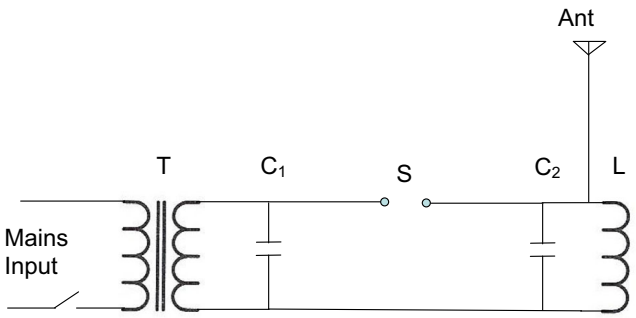

Fig. 7. A spark-gap transmitter. Transformer T steps up the input mains voltage and charges capacitor C_1. The spark-gap S is mechanically configured so that it fires at the peak of the mains voltage. It rapidly discharges into the parallel resonant circuit composed of capacitor C_2 and inductor L. The capacitor C_2 may be a physical capacitor or it is more often the distributed capacitance of the coil. The resonant circuit oscillates at its natural frequency and radiates a damped sine wave each time the spark gap fires. The radiated frequency is not controlled so that the radiated power has a very broad spectrum. The radiated power is proportional to the energy stored in capacitor C_1 and the peak voltage output of the transformer. The transmitter was keyed by switching the input voltage to the transformer.

of the word is not appropriate since the performance of these devices exceeds the accuracy of the definition of the second based on the cesium transition frequency.

11 The Bureau International de l'Heure (BIH)

The initiative for the worldwide unification of time-keeping came originally from the French Bureau des Longitudes in 1912 (there had always been a strong link between accurate time-keeping and determining longitude at sea dating back to the development of the sea-going chronometer by John Harrison in about 1750 [35]. The chronometer displayed Greenwich Mean Time, and the longitude of the ship could be estimated by comparing the reading of the clock at the instant of meridian transit of a star with the published time of meridian transit of the star at Greenwich for the same date).

To address this requirement, the Bureau International de l'Heure (BIH) was created in 1913, but its operation was not begun until after the end of the First World War. It was finally established in 1920 at the Paris Observatory under the auspices of the commission on time of the International Astronomical Union. Since time was purely an astronomically-determined quantity at that time, the initial function of the BIH was to determine and distribute astronomical time [36–38].

The BIH operated 4 pendulum clocks that were located in an underground chamber to minimize fluctuations in the ambient temperature. The periods of the pendulums were sensitive to atmospheric pressure, and the atmospheric pressure was recorded at the time of each observation. The pendulums were used to measure the times of meridian transit of various astronomical bodies, and the results were published in the *Bulletin Horaire du BIH* [39,40]. These observations calibrated the times of the pendulums in terms of local sidereal time. The time was transmitted by linking the pendulums to a remote transmitter and monitoring the transmissions. I will describe the link to the station at the Eiffel Tower, which was typical of the technique that was used.

The Eiffel Tower station ("Tour Eiffel" identified as station "FL") transmitted by means of a spark-gap transmitter (Fig. 7) [41]. The radiated frequency was on the

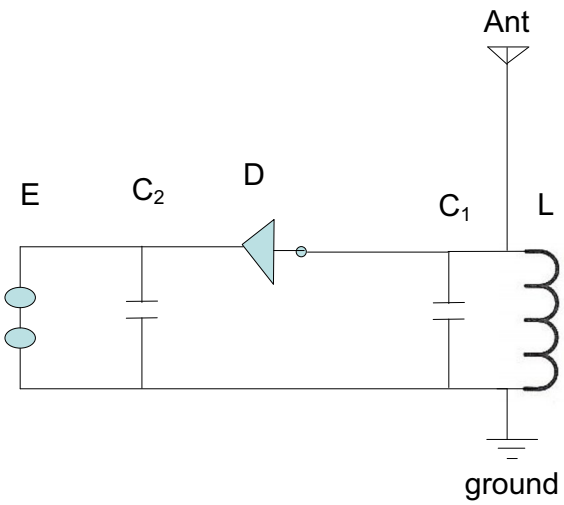

Fig. 8. A simple crystal receiver. Capacitor C_1 and coil L form a resonant circuit tuned to the transmitted frequency. The Q of this circuit was generally quite low, since the bandwidth transmitted by a spark-gap transmitter was quite large. Early crystal sets used a fine wire touching the surface of a crystal of galena or similar material for the diode D, which rectified the input radio frequency. The wire was moved around the surface of the crystal until a sensitive spot was found. Capacitor C_2 filtered the rectified output and its voltage was proportional to the amplitude modulation of the transmitted signal, which produced an audible signal in the headphones, E. Spark-gap transmitters generally produced pulses at the input frequency of the mains voltage, so that the signal heard in the headphones was a buzz at the mains frequency as long as the transmitter was on.

order of 150 kHz (a wavelength of 2000 m), but was not well controlled because the radiated signal was a damped sine wave whose decay time depended on the details of the circuit parameters. The signals had a repetition period determined by the mains frequency, which was generally 50 Hz. The step-up transformer in Figure 7 charged capacitor C_1, and a mechanical switch triggered the spark gap at the peak of the input cycle. This transferred all of the energy stored in the capacitor to the parallel LC circuit that was connected to the antenna. The circuit is not sensitive to the polarity of the input voltage, so that there could be two sparks on every cycle of the mains voltage, assuming that the capacitor C_1 transferred all of its stored energy to the LC circuit and the antenna in a time short compared to the period of the mains voltage. The radiated power was 50 kW, which is a significant amount of power even by contemporary standards. However, the power was spread over a large frequency range so that the power per unit bandwidth was quite small (the power per unit bandwidth is an important parameter in characterizing the signal to noise ratio at the receiver and therefore the range of the transmitted signal). The simple crystal diode receivers of that era (Fig. 8) had very poor frequency selectivity, which compensated to some extent for the very broad spectrum of the transmissions.

If we assume that the overall efficiency of the transmitter was 50%, then the input power was 100 kW. If there were 100 pulses per second (two on every cycle of the mains voltage), then the energy in each pulse, E, was 1000 J. If the input voltage was 10 000 V, then the capacitance needed to store 1000 J would be $C = 2E/V^2 = 20\ \mu$F. If the capacitor was constructed as two parallel plates separated by $d = 5$ mm of glass with electric susceptibility, $k = 5$, then the area of the plates would be approximately $Cd/k\varepsilon_0 = 2260\ \text{m}^2$.

The "Leyden Jar" was a typical capacitor of that era. It consisted of a glass jar and the two plates of the capacitor were made by attaching metal foil to the inside and outside of the glass. If I approximate the jar as about the size of a modern coffee cup, then it would be a cylinder 80 mm in diameter and 100 mm tall. The area between the foil plates would be about $2 \times \pi \times 0.040 \times 0.100 = 0.025$ m^2. If the thickness of the glass was 5 mm, the capacitance of a jar would be about $k\varepsilon_0 A/d = 221$ pf. About 90 000 jars of this size would have been needed! (A 20 μF capacitor that can operate at 10 000 V is not a trivial device even by contemporary standards).

The station transmitted two types of time signals – "ordinary" time signals, which were intended primarily to support navigation by transmitting the time at the prime meridian, and "scientific" time signals, which were intended to support applications requiring both time and frequency information at higher accuracy.

The "ordinary" signals were transmitted starting at 11:40 pm. The control of the transmitter at the Eiffel Tower was switched so that it could be keyed by an operator at the Paris Observatory. At approximately 11:40 pm, the operator transmitted the message, "*Observatoire de Paris, Signaux Horaires*". At 11:44 pm, the operator transmitted the time prefix consisting of dashes: "- - - - - - -". This transmission ended at about 11:44:55 pm. At 11:45 pm GMT, a pendulum at the observatory was connected to the transmitter at the Eiffel Tower and closed a circuit that turned on the transmitter for 1/4 s to send a "long dot". At 11:46 pm, the second prefix code, consisting of dashes separated by two dots, was transmitted: "-••-••-••-••-••" followed by another 1/4 s pulse at 11:47 controlled by a pendulum at the observatory. At 11:48 the third prefix code was transmitted, consisting of dashes separated by 4 dots: "-••••-••••-••••-" and the final 1/4 pulse was transmitted at 11:49 pm, again controlled by a pendulum at the observatory. The same signal sequence was also transmitted starting at 11:45 am. The accuracy of the time signal at the receiver would depend on how accurately the operator could identify the start of the 1/4 second time pulse. The quoted accuracy was about 0.1 s – somewhat better than one-half of the duration of the on-time marker pulse of the time signal. This accuracy would be perfectly adequate for navigation in the open sea, which would probably be limited by the accuracy of measuring the Greenwich Mean Time of the meridian transit of a star from a moving ship. The longitude of the ship could be estimated by comparing the time of local meridian transit with the published corresponding time at Greenwich.

A somewhat different format was instituted starting in July, 1913. The transmission started with a prefix code consisting of the letter X ("-••-") transmitted repeatedly for 50 s. The on-time marker signal was transmitted starting at 55 s. The signal consisted of 1 s pulse – 1 s space – 1 s pulse – 1 s space – 1 s pulse. The last pulse ended exactly at the start of the minute.

Several other time transmitting stations were established in the next few years; all stations transmitted a time signal in the same time format and each was assigned a specific transmission time to minimize interference. I will discuss the transmission formats used by the stations in the United States in the next section.

The "scientific" signals were controlled by a pendulum located at the Eiffel Tower transmitter. The nominal period of the pendulum was adjusted to be $(1 - 1/50)$ s $=$ 0.98 s, and the transmitter was keyed to emit a single pulse for each period of the pendulum. This arrangement was used to transmit 300 pulses, with the 60th, 120th, 180th and 240th pulses omitted to make it easier for a remote operator to count the pulses. The pulses were heard by an operator at the observatory, who recorded the relationship between the received pulses and the ticks produced by the local pendulum. The operator noted the times of coincidence – the times when the received signal from the Eiffel Tower transmitter was heard at the same instant as the tick from the local pendulum. The nominal interval between coincidences would be 49 s or 50 pulses, and the operator would expect to detect at least 5 such coincidences in the 300 pulses.

If the operator could determine the coincidence with an uncertainty of a single pulse, the average interval between coincidences would be $(250 \pm 1)/5 = 50 \pm 0.2$ pulses. This measurement calibrated the period of the pendulum at the transmitter in terms of the period of the standard pendulum at the observatory. The operator at the observatory also recorded the times of these coincidences as measured by the pendulum clocks at the observatory. The times of the pulses could be combined with the average interval between them to provide an estimate of the time of any pulse. The operator computed the times of the first and last pulse in the sequence and transmitted these times immediately after the conclusion of the pulse stream. The nominal elapsed time for the entire sequence would be 299×0.98 s $= 293$ s $=$ 4:53.02. The transmission format provided both standard time and standard frequency information. These data were published by the BIH in the *Bulletin Horaire* [39,40].

The same method of coincidences could be used by any user to calibrate a local clock to the time of the standard clock in the observatory. The user would measure the average time between coincidences to calibrate the frequency of the local clock and the time information transmitted by the observatory at the end of the sequence to calibrate the local time. If the time of coincidence could be determined to within a single pulse, the uncertainty would be of order 0.02 s. The frequency of a user's clock could be estimated by comparing the average interval between pulses measured by the user with the value published for this average by the Paris Observatory.

In addition to controlling the transmitter at the Eiffel Tower station, the BIH also monitored the reception times of signals from other timing transmitters. I will first describe the time services in the United States and I will then describe how the BIH averaged the times of the various services.

12 Early history of time-service stations in the United States: WWV, NAA and NSS

In contrast to the arrangement that I described previously where a single transmission from the Eiffel Tower provided both time and frequency calibrations, the transmission of standard time and standard frequency were initially realized by two different stations in the United States. The stations were operated by two independent agencies.

Radio station WWV was operated by the National Bureau of Standards (NBS) initially in Washington, DC, later in Beltsville (later named Greenbelt), Maryland (the transmitter was moved to Fort Collins, Colorado in the 1960s) [42]. The station started transmitting standard frequency signals in January, 1923. The initial format was a continuous wave signal with no modulation, and was intended to provide a standard reference frequency for other radio broadcast stations. Radio stations were assigned frequencies 10 kHz apart, and, by 1928, the stations were required to maintain the transmitted frequency within 500 Hz of the assigned value. The assigned frequencies were of order 1 MHz, so that the requirement corresponded to a fractional frequency accuracy of about 0.05%. Oscillators of that era, even when stabilized with quartz crystals, could marginally support this requirement. The service made a very important contribution to the operation of the radio stations of that era:

> "Probably no radio station has ever rendered the American radio world so great a service as that of WWV in transmitting the standard wave signals. Before these signals began both broadcast and amateur waves were uncertain and often wavemeters disagreed violently. Since the signals began those in the East have been able to make precision calibration on their own wavemeters and to pass the information on into the West" [43].

The transmission was a pure carrier and was not referenced to any time scale. The NBS maintained a number of frequency standards, which I will discuss below, and the frequency transmitted by WWV was controlled and monitored by the use of several different techniques. As I discussed above, frequency was defined in terms of an astronomical time interval at that time, so that any frequency standard was technically a secondary standard and had to be calibrated astronomically (the frequency calibrations based on the cesium atomic clock at the National Physical Laboratory that I discussed above were still several decades in the future).

The earliest WWV transmitter used the resonance frequency of a parallel inductor and capacitor to control the oscillator stage of the transmitter. The final stage of the transmitter operated in what we would now classify as Class C, and was implemented with several tubes in parallel. The grids of the final amplifiers were biased with a large negative voltage so that the tubes were cut off for most of the input signal. The oscillator drove the output tubes into saturation on the peaks of the input signal (often for about 1/4 of a cycle from a phase angle of 45° to 135°), which produced a series of output pulses at the period of the input signal. This series of pulses was used to drive a final resonant circuit that was coupled to the antenna. The Class C configuration provides relatively high efficiency because the final output tubes are driven from cut-off, where no current flows, to saturation, where the voltage drop across the tube is a minimum. The Class C configuration drives the tube between these two states as rapidly as possible. This minimizes the power dissipated in the tube itself, which is the product of the instantaneous current through the tube and the instantaneous voltage drop across it. The disadvantage of this mode of operation is that the output has appreciable harmonic distortion because the signal driving the output circuit is far from a single-frequency sinusoidal voltage. This harmonic distortion would become particularly troublesome when WWV transmitted signals simultaneously on multiple frequencies that were harmonically related (2.5 MHz, 5 MHz, 10 MHz, ...). The Class C configuration was also not suitable for the amplitude modulation that was added to the transmissions later.

The transmitted frequency was monitored by a number of different techniques. One of them used a precisely calibrated "NBS wavemeter" – the resonant frequency of a precision inductor and a variable capacitor in parallel [44]. The accuracy of the frequency was quoted as being better than 0.3% (this accuracy was adequate only for the early stages of radio broadcasting). A number of different coils were used and they ranged in inductance from 10 μH to 5000 μH. The variable capacitance was realized by a single device with a movable plate and a range of 0.0012 μF and four precision fixed capacitors (0.001 μF, 0.002 μF, 0.004 μF, and 0.008 μF) that could be connected in parallel with the variable device. The resonant frequency of the wavemeter could be adjusted from 3.5 kHz to 4.6 MHz.

A number of other methods for generating standard frequencies were based on a calibrated tuning fork [45]. The frequency of the tuning fork was determined by comparing it to the frequency of a standard pendulum. A simplified diagram of the method that was used is shown in Figure 9. By averaging the data for five seconds, the frequency of the tuning fork could be determined with a fractional uncertainty of about 4×10^{-6}.

Once the tuning fork was calibrated by this method, it could be used to calibrate higher frequency oscillators. One of the methods was based on Lissajous figures [46]. The vibrations of the standard tuning fork were converted to a varying electrical voltage and the voltage was displayed on the x axis of an oscilloscope. The signal from the oscillator to be calibrated was displayed on the y axis. The frequency of the oscillator to be calibrated was adjusted until a stable figure was displayed, at which time the ratio of the two frequencies was given by the ratio of the number of loops on the y axis to the number of loops on the x axis.

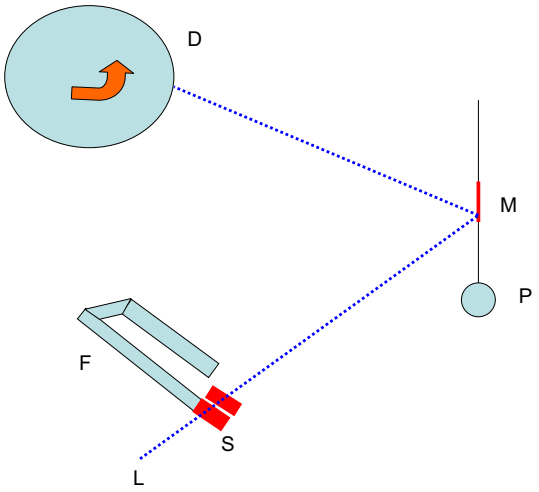

Fig. 9. A simplified method for comparing the frequency of a tuning fork to the frequency of a standard pendulum when the frequency of the tuning fork is approximately known to start with. A small slit, S, is mounted on the tuning fork to be calibrated, F. The slit allows light from source L to pass through once on each period of the vibrating fork. The light strikes mirror, M, on pendulum, P and is then reflected to the film that is mounted on a rotating drum, D. The rotational speed of the drum is fast enough so that consecutive images are displaced by a reasonable amount. Consecutive images will appear to lie on a straight line when the ratio of the period of the tuning fork and the period of the pendulum is an exact integer, and the images will trace out a periodic displacement when this condition is not exactly satisfied.

A second method was based on the non-linear class C amplifier that I described in the previous paragraphs. If the input to the amplifier is the standard frequency but the output circuit is tuned to a multiple of that frequency, then the non-linearity of the circuit will generate harmonics of the input frequency, and the output circuit will resonate strongly with one of them, suppressing all of the others. The resulting signal can be amplified and multiplied again if necessary.

The result of either process is a series of standard frequencies that are integer multiples of the original input. Other frequencies could be generated by amplitude modulation of one of these standard frequencies and tuning the output to resonance on only one of the sidebands. These "single-sideband" techniques were not practical until resonant circuits were developed that had sufficiently high Q to pass the sideband frequency while blocking (or at least significantly attenuating) the carrier frequency.

Starting in 1927, the transmitted frequency was calibrated by comparing it to a temperature controlled quartz-crystal oscillator [47]. The accuracy of the transmitted frequency improved to several parts per million by 1931 and to 2 parts in 10^{10} by 1958. The transmitted signals were used as a frequency reference with no timing information until 1945; time information was broadcast by the US Naval Observatory during that era. I will describe these transmissions in the next section.

12.1 Radio stations NAA and NSS

The US Naval Observatory had been transmitting a signal at 12 Noon every day starting in the summer of 1864. [48]The time was determined astronomically by observing the meridian transit of a number of "clock stars" as I have described above, and these

transit times were used to calibrate an ensemble of pendulums. The ensemble of standard pendulums was kept in a vault below ground to minimize the variation in the ambient temperature.

Starting in the summer of 1865, the transmissions were also sent to the fire stations in Washington, DC at 7 am, 12 Noon, and 6 pm. The signals were transmitted over the fire alarm signaling circuits. The service was gradually extended to more distant locations by the use of the telegraph circuits of the Western Union Company. The broadcast time service was started in January, 1905 from station NAA in Arlington, Virginia. The transmitted power was 100 kW, and was similar in design to the spark-gap configuration that was used at the Eiffel Tower as I discussed above.

The transmitter was controlled by a switch on the time pendulum in the observatory. This pendulum was calibrated by comparing its period to the periods of the standard pendulums. The comparison used the method of coincidences that I have described previously, except that the one-second pulses from both the standard and transmitting pendulums were recorded on a roll of paper, so that the times of coincidence could be easily observed.

The switch on the time pendulum was a toothed wheel that advanced by an angle corresponding to the width of a single tooth each second. The teeth corresponding to the 29th, 55th, 56th, 57th, 58th, and 59th seconds were missing, so that no pulse was transmitted at these times. The transmissions were switched on by each of the remaining teeth of the wheel starting at xx:55:00 – 5 minutes before the start of the next hour. Each pulse was about 0.3 s long. The transmission was controlled by this toothed wheel until time xx:59:49 – 11 seconds before the start of the next hour. No signal was sent from xx:59:50 to xx:59:59. A special tooth on the wheel closed the circuit at (xx+1):00:00 and the start of that pulse marked the beginning of the first second of the next hour. The duration of this special longer pulse was approximately 1.35 s. In addition to the pulse that signaled the start of the hour, a receiving station had 10 other opportunities to determine the transmitted time – by noting the 10 missing pulses at the minute and half-minute times.

The time delay between the swing of the pendulum and the transmission of the signal at Arlington was measured to be about 0.1 s, and this delay was removed by transmitting a signal back from Arlington to the Observatory and advancing the various pulses sent from the observatory so that the transmissions were on time. The residual error in the transmission times was estimated to be about 0.01 s [49].

The antenna towers at Arlington were removed in 1941 because they interfered with airplanes landing at Washington National (now Reagan) airport. The call letters NAA were re-assigned to a naval station at Cutler, Maine [50].

Radio station NSS was located at Annapolis, Maryland, and was also used to transmit time signals that were derived from the pendulums at the Naval Observatory [51].

13 International coordination of time and frequency

Although I have emphasized several times that time and frequency are really different aspects of the same phenomenon, the international coordination of these two quantities originally developed independently.

Oscillators that used quartz crystals as the frequency reference had frequencies that were very stable, but the accuracy of the frequency was much poorer than the stability. The most important customers for frequency measurements at that time were the radio broadcast stations. If the frequencies of all of the radio stations could be compared to the same physical oscillator, then its frequency stability and the reproducibility of the measurements would be much more important than the absolute accuracy of the frequency reference. The frequency accuracy of the oscillator would be

important only when the oscillator was used to realize the astronomical time interval that explicitly defined the length of the second and implicitly defined the standard frequency. Since radio signals did not stop at national boundaries, a national standard of frequency was not sufficient, and an international standard of frequency was needed to minimize the interference between radio broadcasts.

The first attempt at international comparison of frequency standards used the radio transmissions broadcast by the various national timing laboratories such as WWV. The accuracy realized by these comparisons was only of order 0.2%, and the stability and accuracy of the laboratory frequency standards was almost certainly better than this value. As I discussed above, by 1928 a frequency uncertainty of this magnitude was not adequate for ensuring the accuracy of the frequencies transmitted by radio stations.

Instead of comparing the frequencies of different laboratories by monitoring frequency broadcasts, work was started to define a portable standard oscillator that would be shipped to each station. Its frequency would be compared to the standard frequency of each laboratory, and this method could be used to compute a best estimate of the standard frequency derived from an average of all of these observations. The frequency of the traveling oscillator had to be very stable, but not necessarily very accurate. The method could compensate for the random variations in the frequencies of the stations but could not detect a systematic frequency offset in all of them. This was not a serious limitation from the radio broadcasting perspective.

The first test of this method was performed by Professor W.G. Cady of Wesleyan University. He patented the method of using quartz crystals as a frequency reference [52]. He carried quartz crystals to various laboratories and compared the frequency generated by these crystals with the standard frequency of the laboratory. This method was not completely satisfactory because the frequency of a quartz crystal depended on the details of the oscillator circuit, the supply voltage, and similar factors.

Starting in December, 1925, a complete oscillator [47], whose frequency was stabilized with a quartz crystal, was shipped from NBS to timing laboratories in England, France, Italy, and Germany. The frequency of the traveling oscillator was measured again when it was returned to NBS in 1927, and the mean value of the frequency before and after the trip was compared to the frequency of the standard at each laboratory. The frequencies of the various laboratories agreed to within a few parts in 10^4. These measured frequency differences were comparable to the measured change in the frequency of the traveling oscillator during the measurement campaign, so that the disagreements among the frequencies of the timing laboratories are only marginally statistically significant. The measurements were repeated later in 1927 with an oscillator whose frequency was stabilized by means of a temperature-stabilized quartz crystal, and the agreement among the laboratories improved by about a factor of 10 to a few parts in 10^5. All of the frequencies were compared by listening for the beat frequency between the oscillator under test and the traveling standard, and the frequency differences imply a beat frequency on the order of 1 Hz. This beat frequency was measured by combining the two frequencies and manually timing the period of the amplitude modulation of the combination (this is essentially the method of coincidences that was used with the signals from the Eiffel tower that I described earlier). Determining the beat frequency to 3+ significant digits would therefore have required several hours of listening to the beat frequency.

At the same time as these frequency comparisons were underway among the timing laboratories, an independent campaign of time comparisons was also in progress. I have already described the relationship between the BIH and the time transmissions from the station at the Eiffel Tower. The BIH also monitored the transmissions from other observatories, many of which used the same 300-pulse "scientific" format that

I described above. Each observatory had a unique transmission time, so that combining these data depended on reducing the observed astronomical time data based on the ephemeris of the stars that were observed and the longitude of the observing station. The uncertainties of this analysis gradually improved from about 0.010 s to 0.001 s.

The BIH published (*Bulletin Horaire*) the time differences between its best estimate of Greenwich Mean Time, called UT(BIH), and the time signals received from a number of other observatories. Initially (until 1929), the time reference for these calculations was derived from astronomical observations at the Paris Observatory. These data were called "*heure demi-definitiv*". The data did not include any corrections for the propagation delay between the various stations and the BIH, primarily because these values were poorly known at the time. This resulted in systematic time offsets that increased with the distance of the observatory from Paris. For example, the propagation delay from station NAA in Arlington, Virginia, was of the same order as the uncertainty that I quoted in the previous paragraph.

Starting in 1929, the BIH started publishing the "*heure definitiv*" based on the average of the data from a number of observatories. Initially these values were based on the data from 6 radio signals; the number slowly increased to 37 by 1963. There were large systematic offsets among these data, so that simply averaging all of the data would not necessarily eliminate these offsets. Each contributing observatory established a local version of UT called $UT(k)$, where k was the acronym of the observatory. The BIH computed a weighted averaged of the $UT(k)$ values from a fixed number of observatories to compute the time of a "*mean observatory*". The weights were proportional to the stability of the time of scale of each laboratory when compared to the ensemble average of all of them (Weighting schemes based on some form of this idea have formed the basis for all time scale algorithms up until now). This algorithm smoothed short-term irregularities in the computation of UT, but did not remove an overall systematic offset, which was the weighted sum of the offsets of each observatory whose data was used to compute UT. The propagation delays of radio signals have both diurnal and longer-term variations, and these variations would still have been present in the computed values for UT and in the published differences between UT and $UT(k)$. Even under favorable propagation conditions, this variation would probably have been on the order of several milliseconds for an averaging time of about one month. It would be difficult to separate the variation in the propagation delay from the fluctuations in the times of the clocks at the observatories, and this would have resulted in decreasing the weights of the more distant observatories.

I have already described the development of atomic clocks based on the hyperfine transition in the ground state of cesium. These devices became much more widely used in the 1950s and 1960s. The frequency generated by an atomic clock can be integrated to compute atomic time, and the BIH established a local atomic time scale to study the variation in $UT1$, which was still computed astronomically. The uncertainties in the values of the propagation delays were too large to compute an average atomic time scale based directly on time signals derived from cesium standards, and the atomic time scale was computed by integrating frequency data, since frequency transmissions were not sensitive to the systematic propagation delays but only to the (presumably) smaller fluctuations in these values. This would have improved the stability of the time scale, but not necessarily its accuracy, since an arbitrary constant of the integration process had to be determined by some other method.

Although higher frequency short-wave signals (frequencies on the order of 10 MHz and above) can be received at great distances from the transmitter, the propagation delay for a signal at these frequencies was quite variable. Very low frequency signals (100 kHz and below) had much better delay stability, and these signals were used to transmit time difference data starting in about 1960. Signals from the LORAN-C

navigation system [53] at 100 kHz were commonly used for this purpose, using the common-view method that I will describe below. The propagation delay of signals at these frequencies is dependent on the ground conductivity along the path, so that the delay over water is particularly stable. The short-wavelength spatial variation over land paths is often appreciable and difficult to model. The propagation delay has a diurnal variation, which is almost completely removed by 24-hour averages. There is also an annual variation in the propagation delay that depends on the details of the path conductivity. This annual variation, which was on the order microseconds, was especially important for the trans-Atlantic link that was used to compare clocks in the United States timing laboratories (The US Naval Observatory in Washington, D. C. and the National Bureau of Standards in Boulder, Colorado) with the BIH and the timing laboratories in Europe (these problems were much less serious for navigation applications, which determined a position by measuring the *difference* in the arrival times of signals from two LORAN-C transmitters, whose transmission times were synchronized. The absolute magnitude of the delay was not important for this reason, and the differences also significantly attenuated the correlated variation in the propagation delay).

In spite of these weaknesses, the LORAN-C system was widely used for time and frequency comparisons in the common-view mode because the peak transmitted power was very high, often approaching 1 MW, so that the signals could be detected even over continental baselines. In addition to the long-period variation in the propagation delay, the usefulness of a diurnal average is limited in practice by the stability of the clocks at both ends of the path. The long-period variation in the propagation delay between the United States and Europe was estimated by ancillary calibrations by means of portable clocks (generally rubidium standards, which could run on a reasonable number of batteries for the 24 h needed to make the trip). The clocks were generally carried from the NBS laboratory in Boulder, Colorado to the BIH at the Paris Observatory, with a stop at the airport in Washington, D.C., where the clock was compared with a portable clock carried to the airport from the Naval Observatory.

Although there was only a single atomic clock at NPL in 1955, a number of other laboratories had operating cesium standards by 1959. The BIH computed an atomic time scale by integrating the frequencies of the contributing laboratories. This time scale was initially called *AM* [54]. Starting in 1964, the BIH computed two atomic time scales: *AM*, computed as the mean of all of the contributing laboratories, and *A*3, computed from the three best contributions, where "*best*" was determined by the stability metric that I described above. The scale *A*3 was later re-defined to be the mean of the best atomic time data available, without regard to the number of clock that actually contributed to the computation. The *A*3 time scale was set to be equal to the $UT2$ time at 1958 January 1, 0 h $UT2$, and the arbitrary constant that resulted from the integration of the frequency of each contributing laboratory was adjusted to maintain the continuity of the average atomic time scale with this definition of the origin time. These (often large) integration constants have been carried forward to the contemporary values of the atomic time scales of the original contributing laboratories.

The *A*3 time scale was adopted by an increasing number of organizations, and was renamed *TAI* (International Atomic Time) in 1971. At about the same time, the computation of TAI was changed from a weighted average of the atomic time scales of the contributing laboratories to a weighted average of the clocks at each laboratory. The algorithm for computing this average was named *ALGOS* [55]. I will discuss time scale algorithms in detail below, but the general nature of the algorithm is to assign a weight to each contributing clock based on its stability with respect to the ensemble average time. The definition of the *"stability"* of a clock has evolved over time as I will discuss below.

The increasing importance of atomic clocks in the definition of time gradually transformed time from an astronomically-derived quantity to one that was derived from physical principles. As I mentioned above, the BIH originally considered universal time as the mean of a number of a time signals; this was changed in 1965 so that universal time was defined by a mathematical relationship to atomic time as maintained at the BIH. The result was Coordinated Universal Time (UTC). The term *"Coordinated"* originated from an earlier agreement in 1959 between the USA and the UK to disseminate a universal time scale derived from atomic time with an agreed-upon offset in frequency, and to adjust this offset as needed to maintain agreement with the astronomical definition of $UT1$. I will discuss the definition and realization of UTC below.

The practical realization of UTC from TAI effectively changed time from a fundamental quantity derived from astronomical observations to a derived quantity based on atomic physics and fundamental constants. This technical change in the practical realization of UTC had a corresponding change in how the realization would be managed; the result was the transfer of the realization of TAI from the BIH to the International Bureau of Weights and Measures (BIPM). This transition started in 1985, and was completed three years later, when the BIPM took over all aspects of the computation of TAI. The astronomical aspects of the program of the BIH (terrestrial reference frames, polar motion, etc.) were transferred to a new organization, initially named the International Earth Rotation Service (IERS). The name was changed to the International Earth Rotation and Reference System Service in 2003, but the acronym, IERS, remained the same.

13.1 The international bureau of weights and measures (BIPM)

The precise measurement of many physical parameters plays a central role in national and international trade, telecommunications, and scientific collaborations. With the start of the industrial revolution in the 19th century, it quickly became necessary to be able to compare measurements made in one laboratory or in one country with measurements made somewhere else. This requirement was the motivation for the *Convention du Mètre* (Treaty of the Meter), which was intended to rationalize standards and measurement practices internationally.

The treaty was signed by 17 initial member countries in 1875, and was ratified by the United States in 1878. The treaty established the Bureau International des Poids et Mesures (International Bureau of Weights and Measures, abbreviated in French as the BIPM, and universally known by that abbreviation) and defined its governance (see Fig. 10). The overall management is provided by the *Conférence Générale des Poids et Mesures* (CGPM), which is composed of a delegate from each of the member nations. The direct supervision of the BIPM is the responsibility of the *Comité International des Poids et Mesures* (CIPM). The members of the CIPM are typically chosen from the National Metrology Institutes (NMI) of the member states. The NMI of each country generally maintains the legal definitions of the various standards and provides calibration services.

The CIPM in turn has a number of *Comités Consultatifs* (Consultative Committees) who meet periodically and give technical advice to the CIPM with the assistance and collaboration of the technical staff of the BIPM. The consultative committee for time and frequency questions was initially named CCDS – Consultative Committee for the Definition of the Second, but is now named CCTF – *Comité Consultatif du temps et des fréquences* (Consultative Committee for Time and Frequency). The CCDS/CCTF in turn appoints various ad hoc and standing *Groupes de travail* (working groups) to discuss and provide advice on the definition and realization of the length

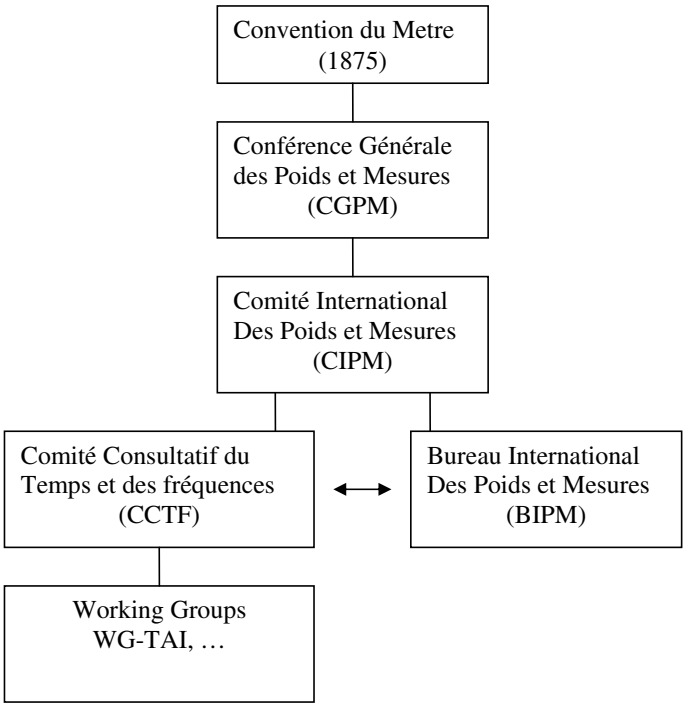

Fig. 10. The organization and governance of the International Bureau of Weights and Measures (BIPM). The figure shows the current configuration of the division of the BIPM that is concerned with questions of time and frequency. The CGPM consists of delegates from all member nations. The members of the CIPM are chosen by the CGPM. They are responsible for the supervision of the BIPM and the general affairs of the International system of units. The members of the Consultative committees are representatives of the National Metrology Institutes of the member states and experts in the definition and realization of time and frequency standards. The consultative committees appoint working groups that are charged with studying various matters that are relevant to time and frequency standards. One of the important working groups is the working group on International Atomic Time, WG-TAI, which considers matters relevant to the definition and realization of International Atomic Time and related questions.

of the second and international time. The most important working group is probably the working group on international atomic time (WG-TAI), which generally meets on the day before each meeting of the consultative committee.

This background material was not relevant in 1875 when the BIPM was founded, because time and time interval were the province of the astronomical community at that time. The initial role of the BIPM was focused on the international standards of mass and length, and it was the repository of the standard meter and kilogram. It also participated in comparisons of the primary standards of these quantities with copies that were distributed to the member states. However, the BIPM now plays a central role in the computation of both Atomic Time and Coordinated Universal Time.

14 Realization of coordinated universal time (UTC)

As I discussed above, the initial definition of the atomic time scale defined the length of the second based on the Markowitz and Essen value for the frequency of the hyperfine

transition in the ground state of cesium 133. The origin of the scale was set to be equal to $UT2$ on 1958 January 1. Coordinated Universal Time (UTC) would be implemented by means of an offset from atomic time both in time and in frequency as needed so that the resulting UTC scale would stay close to $UT1$.

From the very beginning, it was clear that the Markowitz and Essen value for the length of the atomic time second was too small relative to the then current length of the $UT1$ second [56]. The rate of the atomic time scale was therefore too large relative to the frequency calculated from the length of the $UT1$ day, so that the steering offsets that were required to realize UTC would be systematically negative. The initial fractional frequency offset, applied on 1 January 1961, was -1.5×10^{-8}, and this was followed by additional fractional frequency adjustments and fractional-second time steps. The resulting UTC time scale was very difficult to implement, since few clocks could accept either the frequency offsets or the small time steps. In many cases, the frequency offsets were applied administratively and selectively, depending on whether the physical signal was to be used to realize the standard of frequency or UTC. The standard frequency could therefore not be integrated to compute the standard of time without additional computations and ambiguities [57].

Whatever the merits of the process I described in the previous paragraph, it was very troublesome from the practical perspective because it was never clear whether a particular source of frequency data did or did not include the varying frequency offsets. In addition, it was almost impossible to study the long-term stability of a frequency source because of the steps in frequency. The realization of UTC was changed in 1972 as the result of these and other problems. On 1972 January 1, the rate of UTC was set exactly equal to the rate of the atomic time scale based on the Markowitz and Essen value for the frequency of the hyperfine transition in the ground state of cesium 133. The time difference, TA-UTC was set to exactly 10 seconds. The time difference based on the cumulative time steps and frequency adjustments that had been applied since 1961 was 9.999992 s [58].

The length of the UTC second was now systematically too small relative to $UT1$, which meant that the UTC frequency was systematically too large with respect to the corresponding astronomical value for $UT1$. The fractional frequency difference at that time was approximately 3×10^{-8}, which was the value of the last frequency steering adjustment that was applied in 1966. The time dispersion resulting from this rate difference would be approximately $3 \times 10^{-8} \times 3.16 \times 10^{7} = 0.95$ s/year. The magnitude of this time difference is based on astronomical observations and cannot be predicted exactly because of the variability in $UT1$. The prediction error is approximately 0.004 s for a prediction 30 days in advance. The measured and predicted values are published in Schedule A of the International Earth Rotation and Reference System Service (IERS) [59].

This time difference is accommodated by adding an exact integer second to UTC whenever the magnitude of the time difference reaches 0.9 s. This extra "leap" second can be added as the last UTC second of the last day of a month. The months of June and December are preferred, but any month can be used in principle, although only June and December have been used to date. The leap second is announced by the IERS in Bulletin C [60]. The name of the leap second is 23:59:60, and the name of the next second is 00:00:00 of the next day.

The initial interval between leap seconds was about a year as was predicted based on the frequency offset of 3×10^{-8} that was applied before the change in 1972, but the interval has slowly increased, which is contrary to expectations. Only 26 additional leap seconds have been added in the 43 years from 1972 to 2015. In other words, the rate of increase in the length of the $UT1$ day has slowed.

The leap second system has the advantage that UTC is close to $UT1$, and it is often possible to use the UTC time as an approximation for $UT1$ in applications

that treat time as effectively a proxy for the position of the Earth. However, the leap second system is awkward for applications that depend on frequency or time intervals measured across a leap second. The native time scales of navigation systems such as the Global Positioning System satellites do not use leap seconds for this reason. There is also no way of representing a leap second in most computer systems, which represent time as a number of seconds that have elapsed since some time origin. These systems typically stop the clock for an extra second just before the leap second at the time corresponding to 23:59:59 UTC, and this introduces an ambiguity in the internal time scale and in the computation of time intervals, which effectively do not include the extra second. This is particularly troublesome in Asia and Australia, where the leap second, which is defined relative to UTC, occurs in the morning of the next day in local time (A leap second at the end of December is also troublesome in the US Pacific time zone, where it occurs at 4 pm local time). If the leap second system were discontinued, the $UT1$ and UTC time scales would diverge; based on current (2016) data, the rate of divergence would be about one minute per century or less, but this rate is likely to be variable, so that any extrapolation would have a considerable uncertainty.

Although the initial definition of UTC was driven by the advantages of maintaining UTC close to $UT1$, the problems associated with the implementation of leap seconds have become more serious, and there are ongoing discussions to modify the method of coordinating UTC, and possibly abandoning the leap second system altogether [61]. The question was considered again at the World Radio Conference that was held in Geneva from 2 to 27 November, 2015, but no decision was reached at that meeting, and the question was postponed until the next meeting tentatively schedule for 2023. There has been some discussion of transferring the definition of UTC from the World Radio Conference of the International Telecommunications Union to the BIPM, since the BIPM already computes International Atomic Time, which is the basis for UTC. This question is also pending at this time.

15 National and international time scales

I have already mentioned the computation of the time scales $A3$ and later TAI by the BIH, and these computations were transferred to the BIPM in the 1980s. These time scales were initially derived from weighted averages of a number of time transmissions going back to the work of the BIH at the Paris Observatory in the 1920s.

When commercial atomic clocks based on the cesium transition first became generally available in the late 1950s and early 1960s, national laboratories started using these clocks to compute a realization of the official astronomical time based on local measurements of the cesium second. These data were often used to control time signal broadcasts, such as the transmissions from radio station WWV, which I have described earlier. In order to minimize the impact of a hardware failure, most timing laboratories operated multiple cesium clocks and devised an algorithm for averaging the times of the devices to generate a time signal that was more reliable and, at least in principle, more stable than any of the contributing member devices. The algorithm used at NIST is named $AT1$; it has been modified several times since it was developed in the early 1970s, but many of its essential features are unchanged from that time. The principles of time scales are discussed in [62].

The outputs of these local time scales were designated $TA(k)$, where k is the identifier of the laboratory. The initial operation of the BIH was then to compute a time scale based on a weighted average of the time scales of the contributing laboratories and to publish the differences $TA(k) - TAI$. UTC was then defined as an

offset in time and frequency from *TAI*, where the offset was computed from astronomical determinations of $UT1$. Each laboratory generally also computed a local version of *UTC*, called $UTC(k)$, which was derived from its $TA(k)$ by applying an offset in time and frequency as required to steer $UTC(k)$ to *UTC* as computed by the BIH. There was (and is) no standard method for realizing $UTC(k)$ or for applying the periodic steering corrections to steer it to *UTC*. In general, a steering method must be a compromise between time accuracy which would favor aggressive steering corrections, and frequency stability, which would favor smaller and less frequent adjustments. Time "accuracy" indicates how closely $UTC(k)$, the laboratory realization of *UTC*, agreed with the computation of the BIH, and later the BIPM. The method used at NBS/NIST has evolved from monthly small frequency adjustments, which favored frequency stability at the expense of time accuracy, to more frequent and aggressive frequency adjustments that emphasize time accuracy as the more important metric.

The *TAI* and *UTC* time scales were (and still are) computed after fact with a delay of several weeks. Therefore, while there was usually no long-term systematic difference between *UTC* and $UTC(k)$ for most laboratories, the short-term accuracy of $UTC(k)$-*UTC* was largely independent of the details of the steering algorithm and depended mostly on the free-running stability of the time scale of laboratory k. For example, the best commercial-grade cesium standards currently have a free-running fractional frequency stability of about $\delta f/f \sim 2\times10^{-14}$ for averaging times on the order of a few weeks, and the stability of an ensemble of such standards might improve by the square root of the number of contributing clocks. Hydrogen masers have significantly greater stability, and it is not unreasonable for the stability of an ensemble of commercial cesium clocks and hydrogen masers to have a frequency stability approaching (or even exceeding) 10^{-15} for averaging times of several weeks, which would result in a time dispersion of a few nanoseconds RMS in $UTC(k)$-*UTC* between successive computations of *TAI* and *UTC*.

In 1973 the ensemble algorithm at the BIH was changed from averaging the ensemble outputs of each laboratory to averaging the actual time data of the contributing clocks that made up the local ensembles. The intent was to design an algorithm that would be independent of the differences in the analysis methods at each timing laboratory and would be designed to maximize frequency stability at the expense of a real-time output. This algorithm was named *ALGOS* and its output was (and is) a paper clock that has no physical realization. It has been modified several times since 1973, but the essential features are unchanged.

In January, 2012 the BIPM established a pilot project to provide a rapid version of *UTC* that would be better adapted for near real-time applications. The rapid version could also improve the stability and the accuracy of the laboratory realizations of *UTC* by providing more timely steering information and thereby reducing the time dispersion resulting from flicker and random-walk frequency variations of the laboratory time scales. The rapid scale is called *UTCr*. It provides daily estimates of $UTCr$-$UTC(lab)$ computed every week [63]. The transmission of data from the timing laboratories to the BIPM and the general features of the time scale algorithm that are used by the BIPM to compute *UTCr* are the same as for the monthly computation of *UTC* (which did not change), except that the laboratories that participate in the *UTCr* computation are expected to transmit data on a daily basis rather than only every month. The *UTCr* time scale became an official product of the BIPM in July, 2013, and it has proven to be quite useful for the steering of the laboratory time scales, $UTC(k)$.

The $AT1$ and *ALGOS* time scale algorithms share many common features, and I will describe the general features of these (and most other) time scale algorithms in the next section.

16 Time scale algorithms

A time scale algorithm is a method for combining the measurements of the times of several clocks to generate an ensemble-average time. The procedure has two primary purposes: (1) to improve the reliability of the time signal by designing an algorithm that continues to function with only a minimal change in the output even if one of the member clocks fails. (2) To serve as the basis for a statistical procedure that can combine the data from different clocks in a statistically-optimum manner.

The statistical characterization of atomic clocks is derived from their internal design, and I will briefly describe the important aspects of the design in the following text. I will focus on cesium-based devices because the frequency of the hyperfine transition in cesium is the current definition of the length of the second, but similar considerations would apply to many other clocks whose reference frequency is derived from some other atomic transition frequency.

The design can be divided into two distinct parts: the "physics package", which provides the atomic transition that is the frequency reference and an "electronics package", which locks the frequency of a quartz-crystal oscillator to the atomic transition frequency. The electronics package also synthesizes the output signals, which are usually at standard frequencies such as 1 Hz and 5 MHz.

The frequency reference in a cesium atomic clock is the energy difference between two hyperfine levels in the ground state of the atom. The cesium atoms act as a passive discriminator – they do not generate the standard frequency, but are interrogated by an electronic frequency source (generally a quartz-crystal oscillator), which is locked to the frequency that stimulates the hyperfine transition in the atoms (this is a fundamental difference from a pendulum clock, for example, where the pendulum, which plays the role of the frequency reference, also is an active device and its oscillations drive the clock mechanism).

The control loop, which keeps the electronic oscillator locked to the cesium transition frequency, inevitably has some noise. This noise translates into fluctuations in the frequency of the electronic oscillator. In a well-designed system, this noise can be characterized as a random process with a mean of zero, so that the frequency of the oscillator fluctuates about the frequency defined by the atomic transition with no systematic offset. In addition, all control loops have a finite bandwidth. From the time-domain perspective, the finite bandwidth translates into a time delay between when the difference between the frequency of the oscillator and the frequency of the cesium-atom discriminator is detected and when the correction to the oscillator is applied (in engineering jargon, this delay would be described as a finite rise-time of the step response of the control loop). As a consequence of this delay, the statistical characteristics of an atomic clock are determined by the statistics of the internal quartz-crystal oscillator at sufficiently short averaging times. This "short" averaging time is typically on the order of a few seconds for most atomic clocks.

The counting circuit portion of the clock is triggered each time the output signal from the electronic oscillator completes a cycle. This is usually realized as a zero-crossing detector, which advances by one each time the output signal of the oscillator crosses zero in the positive direction. There is inevitably noise in the zero-crossing detector as well, and this noise translates into jitter in the output time signals, which are not related to the frequency of the device. In a well-designed system, this noise can also be characterized as a random, zero-mean process, so that the output signal fluctuates about the true output with no systematic offset.

The models of random, uncorrelated frequency and time fluctuations must be qualified in principle at very short averaging times. Since every real process can have only a finite first derivative, both frequency and time fluctuations must exhibit correlations

at sufficiently short averaging times. This is not a practical concern for almost all time-scale applications, which generally do not use averaging times shorter than 1 s.

Finally, the hyperfine transition frequency of the cesium atoms is affected by a number of influences such as electric and magnetic fields, collisions between the atoms or between the atoms and the residual gas in the chamber and similar effects. These perturbations often change slowly with time, and these changes produce corresponding drifts in the output frequency of the oscillator and the time signals that result from counting the frequency.

A measurement of the time difference between two nominally identical cesium devices will show a variation due to all of the effects that I have just enumerated. Since there is only a single measurement but multiple causes of its variation, it is impossible in principle to identify the source of the variation in the output time without some additional information or assumptions.

A common method for separating the different contributions to the variance in the time-difference measurements is to note that the contributions of the different effects depend differently on the interval between time-difference measurements. The fluctuations caused by the noise in the counting circuit, for example, are independent of the time between measurements; the time dispersion caused by frequency noise increases with the time between measurements, and the slow drift in the atomic-transition frequency results in fluctuations in the measured time differences that increase as the square of the time interval between measurements. From the Fourier frequency perspective, the noise in the counting circuit has a "white" Fourier spectrum that is constant at all Fourier frequencies, with the qualification with respect to very short averaging times, which correspond to very high Fourier frequencies that I discussed above. The time dispersion due to the white frequency noise is the integral of a white process and the time dispersion due to the frequency drift process is a double integral of a white process. Each of these integrations "reddens" the impact of the process on the time differences by a factor proportional to the reciprocal of the Fourier frequency squared. In addition, the time and frequency of all clocks exhibit "flicker" noise, whose power spectral density varies as the reciprocal of the Fourier frequency. There is no simple physical model that describes the cause of this contribution, although it can be modeled approximately by assuming a finite correlation time between consecutive time and frequency estimates.

All of the "red" contributions to the Fourier spectrum of the time differences imply a divergence at very low Fourier frequency that results in infinite total power. The power spectra of the time difference data of real clocks are bounded at low Fourier frequencies by a frequency on the order of the reciprocal of the total time duration of the measurements; this lower bound is in addition to the upper bound that I discussed above.

Time scale algorithms are usually implemented in a recursive form, where the computation at any time combines the values of the clock parameters computed on the previous measurement cycle with the current measurements. There are a number of reasons for using this method instead of a batch algorithm where all of the data are analyzed together by using an algorithm such as conventional least squares. In the first place, the deterministic clock parameters that are estimated by the algorithm are not strictly constant but change slowly with time as I discussed above in the model of an atomic clock. Conventional least squares analyses do not have the flexibility to support a variability of the estimated parameters. In the second place, adding an additional data point in a batch analysis potentially changes older parameter estimates, and this is not suitable for an algorithm whose output is used in real time. Finally, time scales never terminate, and a batch process becomes increasingly unwieldy as more and more measurements are acquired.

All time scale algorithms begin from a model of the behavior of each clock, which is driven by the physical considerations I outlined above. The clock model describes the state of clock j at time t_k based on the state at the time of the previous computation, t_{k-1} by

$$\begin{aligned} x_j(t_k) &= x_j(t_{k-1}) + y_j(t_{k-1})\tau + 0.5d_j(t_{k-1})\tau^2 + \xi_j \\ y_j(t_k) &= y_j(t_{k-1}) + d_j(t_{k-1})\tau + \eta_j \\ d_j(t_k) &= d_j(t_{k-1}) + \varsigma_j \end{aligned} \tag{7}$$

where x, y, and d are the deterministic time difference in units of seconds, the frequency in dimensionless units of seconds/second, and the frequency drift in units of 1/s (seconds/second^2), respectively for clock j. These parameters are defined with respect to the corresponding time, frequency, and frequency drift of the ensemble (the ensemble parameters are generally not realized by any physical device, so that these parameters are not directly observable at this stage). The parameter τ is the time difference between measurements, $\tau = t_k - t_{k-1}$. It is convenient to use a constant time difference between measurements, but this is not a requirement of the algorithm.

Each of the three deterministic parameters has an associated random noise process, identified by ξ, η, ζ for the white phase noise, the white frequency noise, and the white frequency drift noise, respectively. The noise processes are assumed to be uncorrelated with respect to each other, between the various clocks, and as a function of time. That is,

$$\begin{aligned} \langle \xi_j(t+\tau)\xi_j(t)\rangle &= \sigma_j^2(\xi)\delta(\tau) \\ \langle \xi_j\eta_j\rangle &= \langle \xi_j\eta_k\rangle = 0 \end{aligned} \tag{8}$$

for all combinations of the noise processes and for all j and k. The delta function, $\delta(\tau)$, is one for $\tau = 0$ and zero otherwise. Each of the three noise processes is characterized by a variance, $\sigma_j^2\ (\xi)$, for example, which is independent of time and may be different for each of the clocks that contribute to the ensemble.

All of this is an optimistic assumption that is generally only approximately realized by the real ensemble. The distinction between deterministic and stochastic parameters is one of degree: the deterministic parameters change at most very slowly with time, whereas the contributions of the stochastic parameters are assumed to be completely uncorrelated from one measurement to the next one. In addition, note that there is no term to model the flicker-noise contributions.

The Fourier transforms of the auto-correlation functions given in equation (8) give the power spectrum of each noise process. The power spectra of a δ function is a constant value for all Fourier frequencies [64] so that these noise terms are usually referred to as "white frequency noise", etc. As I mentioned above, a "white" power spectrum that was really constant for all Fourier frequencies would imply an unbounded, infinite total power, and real spectra, which can have only finite total power, must be band limited. A common assumption is that the high-frequency cut-off is of order $1/2\tau$ and the low-frequency cutoff is of order $T/2$, where T is the total time of the measurement process. A noise contribution that has a Fourier frequency that is much less than the low-frequency limit is modeled as a change in the corresponding deterministic parameter. Noise processes whose Fourier frequencies are above the high-frequency cut-off will be aliased by the sampling frequency, and will appear as lower-frequency contributions. This is an undesirable ambiguity, and additional limitations on the bandwidth may be realized by the measurement hardware to minimize this aliasing problem. Alternatively, we can view this requirement as setting an upper bound on the value of τ – if the time between measurements, τ, is made too large, stochastic fluctuations with significant amplitudes and with periods shorter than $1/2\tau$ will be

aliased by the measurement process, and this aliasing introduces an ambiguity that cannot be removed by subsequent analysis.

There are additional considerations that limit the range of values of τ, the time between measurements. For example, the first of the model equations treats the frequency offset of clock j, y_j, as a constant parameter over the time interval from t_{k-1} to t_k. This assumption is obviously not consistent with the second and third model equations, which predict a linear variation in this parameter during the interval between measurements. Similar considerations apply to the other parameters, and the implication of the model is that the time interval between measurements is small enough so that the deterministic component of the frequency offset and the deterministic frequency drift can be considered as substantially constant over this interval. On the other hand, if the time interval between measurements is made too small, the variation in the measured time difference is dominated by the white phase noise, and it will be difficult to determine anything about the other deterministic parameters. As I discussed above, at *very* short time intervals, the contribution of the white phase noise must reduce to a constant value. This can be the basis for a useful method for detecting outliers in the measured time-differences, since the time differences evolve completely deterministically at sufficiently short time intervals, where "sufficiently short" is defined by the magnitudes of the various contributions to the variance of the time difference data.

The white frequency drift noise is integrated by the model equations so that the spectrum of the frequency fluctuations for each clock has two components – a white noise component determined by η_j and a contribution from the integration of the frequency drift noise, ζ_j. This integration converts the white frequency drift noise to a contribution whose power spectrum is proportional to the reciprocal of the square of the Fourier frequency. The impact of the white frequency noise on the time dispersion has the same result – the integration of the spectrum of the frequency variations results in a time dispersion that has a strong long-period divergence. From the perspective of the time domain, the integration of drift into frequency and frequency into time introduces a random-walk with a variable step size into the statistical distribution of both targets. Note how the behavior of the model equations mirrors the physical description of the device.

The integrations implicit in the recursive form of equation (7) significantly complicate the analysis of time scale data. Without these integrations, the first equation giving the evolution of the time difference would be simple to analyze since the noise process would be a simple random variable whose variance could be improved by simple averaging, and the result would be an unbiased estimate of the time difference. In this simple situation, the standard deviation of the mean could be improved without limit by averaging more and more data. The integrations introduce a long-period divergence. From the physical perspective, the time difference of any real ensemble will increase without limit without some external data, and all time scales must be constrained in this way.

The effect of white frequency noise on the time differences is very different from the impact of white phase noise discussed in the previous paragraph. The evolution of the time difference due to zero-mean white frequency noise is uniformly distributed around the current value of the difference at all averaging times. Therefore, the optimum prediction of the next measured time difference is the *current value* with no averaging. This estimate will have a standard deviation proportional to the amplitude of the white frequency noise multiplied by the time interval between measurements. However, this estimate is optimal because it is unbiased, and no better estimate is available in this case when the only noise source is white frequency noise.

The model does not include any term that models the flicker contribution to the time dispersion. The frequency and frequency drift terms have the same deficiency.

The contribution of flicker noise to the time dispersion of real clocks also has a long-period divergence, and the flicker noise contributions adds to the walk off from the correct time.

From the discussion in the previous paragraphs, the contribution of white phase noise to the time differences can be reduced by averaging more and more data, and this improvement can continue without limit. The impact of white frequency noise is exactly opposite – the optimum strategy is to use the most recent point no averaging at all. Flicker noise can be considered as midway between these two extremes – the data exhibit a coherence and can be characterized by a mean and standard deviation, but only over some finite time interval. The data are not stationary over longer periods and these statistical parameters are valid only over short segments of the data.

Most time scale systems measure the time difference between one of the clocks that is designated as the reference device and all of the other clocks. This clock might be the "master clock" of the system, but this is not a requirement, and the algorithm can also support a configuration in which the reference clock plays no special role in the ensemble average. The measurement hardware periodically reports the time differences, $X_{rj}(t_k)$, between the reference clock and each of the other clocks in the ensemble. The details of the hardware that reports these time differences are not important to the running of the ensemble, except that the measurement process generally has an associated noise, ε, which is distinct from the noise of the clocks themselves.

Some algorithms can accept measurements of frequency differences, $Y_{rj}(t_k)$, between the reference clock and clock j, or even $Y_j(t_k)$, the frequency of clock j measured with respect to an external device (such as a primary frequency standard) that is not a member of the ensemble. An ensemble of this type might be useful for incorporating the data from a primary frequency standard into a clock ensemble, since many such devices do not run continuously as clocks and cannot be used as a source of time.

If there are N clocks in the ensemble, then the state equations have $6N$ parameters – three deterministic parameters and three noise terms for each clock. However, the usual time-difference measurement process provides only $N–1$ time differences, so that the general time scale problem has no unique solution. Every time scale algorithm must address this problem in some way, and the differences among the time scales are largely driven by how this problem is addressed.

All time scale algorithms implicitly assume that the values of the deterministic parameters, x, y, and d, are slowly varying, so that rapid variations in the measured time differences are largely assigned to the noise parameters, and the deterministic parameters are allowed to change only on a longer interval. This condition sets an upper bound on the time interval between measurements – it must be short enough that the deterministic parameters can be considered as constants between measurement cycles. There is also a lower bound on the time interval that is based on the measurement noise of the hardware process, ε, and ξ_j, the white noise of the clock itself, if that contribution is significant. If the interval between measurements is too short, then the noise of the measurement process itself masks the time dispersion of the clocks and makes it more difficult to estimate the deterministic parameters.

In addition to the equations that model the state of each clock with respect to the ensemble, the time scale algorithm must also include a method for combining the time-differences of the clocks into a single ensemble. The ensemble algorithm generally requires an additional assumption about the statistical behavior of the member clocks or about the deterministic and statistical behavior of the ensemble itself. Both the model equations and the measurements involve only differences, so that the calculation would be unchanged if a constant value were added to any of the model parameters for all of the clocks. Adding a constant time to every x_j, for example, has no effect on the measured or predicted time differences – the only effect is to change the time of the

ensemble average by the negative of the additional value. This change is not visible within the algorithm, since the ensemble average is not realized by a physical clock. The same effect is true for modifying any of the other model parameters. In other words, the ensemble average behaves like any other clock – its time, frequency, and frequency aging must be set by a method that is outside of the ensemble calculation itself. The statistical properties of the ensemble also are equivalent to a clock. The white noise of each clock parameter is attenuated by the ensemble averaging process in principle, but the parameters of an ensemble are likely to exhibit significant flicker noise because this contribution is not modeled by the ensemble equations. I will discuss the solution to these requirements that is implemented in the $AT1$ algorithm that is used at NIST. Other algorithms use variations of this general method.

16.1 The NIST AT1 time scale algorithm

The $AT1$ time scale algorithm [62] realizes a solution to the time scale problems by reducing the number of independent parameters in the model equations, and the parameters of the ensemble are determined from data received from the BIPM and from measurements of the frequency of the NIST primary frequency source.

The first assumption is that the frequency drift parameter varies so slowly that it can be treated as a simple noiseless constant that is determined outside of the time scale algorithm by comparing the frequency of the ensemble to the frequency of a primary frequency standard such as a cesium fountain. The contribution of the frequency drift to the time dispersion from one measurement to the next one is then a completely deterministic value. The second assumption is that there exists a measurement interval that is large enough so that the time dispersion due to the frequency noise dominates the noise of the measurement process, ε, and the white phase noise of the clock, ξ_j, for all j. On the other hand, the measurement interval must be short enough so that the contribution of the frequency drift parameter to the evolution of the time difference is small enough that the frequency has no deterministic variation over the interval between measurements and can be modeled as a constant value plus a noise term. Decreasing the interval between measurements also means that any measurement error or clock failure can be detected more quickly.

The result of all of these considerations is to favor the use of the shortest measurement interval that is only long enough so that the measurement noise does not make a significant contribution to the variance of the data. Improvements in the hardware that is used to measure the time differences should then be accompanied by a decrease in the interval between measurements, and this has been the case at NBS/NIST, where the interval between measurements has decreased from once per day, to once every two hours and finally to once every 12 minutes. The same considerations have been used at the BIPM to decrease the interval between computations of TAI from 10 days to 5 days. This interval used by the BIPM is determined as much by administrative issues as by statistics; the statistical considerations would support a shorter interval between computations. The BIPM has responded to this consideration by computing a "rapid" computation of TAI and UTC. This rapid computation, $UTCr$, is based on daily time difference data and is computed weekly.

When the conditions of the previous paragraphs have been satisfied, the stochastic contribution to all of the measured time differences is modeled as resulting totally from the white frequency noise of the clocks themselves. The optimum strategy under these circumstances is to compute the ensemble average time as the weighted average of the times of the contributing members with no averaging of older time differences, where the weights are determined from the average prediction error (the average difference of the measured time of the clock with respect to the ensemble and the value

predicted by Eq. (7)). The weight is therefore a measure of frequency stability, where "stability" is understood to include the contribution to the frequency estimate of any deterministic frequency drift. The accuracy, which is a function of the deterministic frequency and frequency drift, has no effect on the weight.

The frequency of each clock with respect to the ensemble is determined by examining the evolution of the time difference of the clock with respect to the ensemble average time. The mean value of this difference is the deterministic frequency offset and the variance determines the stochastic amplitude. The partition between deterministic and stochastic contributions is determined by the long-term stability of the clock measurements.

The $AT1$ algorithm does not have an explicit constraint on the frequency of the ensemble itself. This frequency is simply the evolution of the ensemble average time. Since the model of each clock does not include a term for flicker noise and stochastic long-period effects, we expect that the ensemble time and frequency will have these noise characteristics as well. The ALGOS algorithm used by the BIPM has the same potential problem. In both cases, the long-term accuracy is compromised by the difficulty of estimating the deterministic frequency drift of the member clocks (especially the hydrogen masers, which are very stable in short term but often have significant frequency drift that degrades the long-term stability). In both cases, the long-term stability of the respective time scale requires ancillary data. The *NIST* time scale is steered by ancillary data from the BIPM Circular T, and the BIPM computation is steered by data from the primary frequency standards. Both of these methods are discussed in the following text.

The final step in the $AT1$ time scale is to steer a physical clock to realize the ensemble average time. This idea is implemented by connecting the physical steered clock to the measurement system, and applying steering corrections that drive the time state parameter of the steered clock to zero. This is generally realized in hardware by steering the signal from one of the clocks in the ensemble by means of an external phase stepper (the physical clock itself is not steered). The output of the phase stepper is then added to the ensemble as a member clock with no weight in the ensemble average. The phase stepper is controlled after each measurement cycle so as to drive its output time to be zero with respect to the ensemble.

The output of the phase stepper is $UTC(NIST)$, and is used as the reference time for all of the *NIST* time services. In addition, it is transmitted to the BIPM, as I will discuss in the next section, and the BIPM returns the difference between UTC as computed by the BIPM and $UTC(NIST)$. This difference is returned in BIPM Circular T [65], and the time-difference data in this circular are used to adjust the phase stepper steering so that $UTC(NIST)$ will be driven towards UTC. The steering equation includes an offset in time and in frequency with respect to the ensemble average, $AT1$, and the time difference computed from the steering equation is added to the phase stepper control loop.

The BIPM Circular T is issued monthly based an analysis of the data of the previous month, so that the steering of $UTC(NIST)$ requires an extrapolation of about 30 days, and is limited by the inadequacies of the model parameters for this extrapolation time. The difference $UTC\text{-}UTC(NIST)$ is a slowly varying function whose amplitude is generally less than 10 ns. From the statistical perspective, this difference is driven by the flicker noise contributions to the time differences that are not included in the time scale model and inadequacies in the model partition of the measured time differences into deterministic and stochastic contributions. This inadequacy is particularly serious for the model of the frequency drift parameter of hydrogen masers. The model drives the frequency drift by a random zero-mean process, but the reality is generally more complicated, and usually cannot be described by any statistical estimator.

Starting in early 2016, *NIST* has started using data from the rapid *UTC* calculation, *UTCr* [63], to compute the steering commands used to generate *UTC*(*NIST*). This method will use weekly adjustments to the frequency of *UTC*(*NIST*) and should reduce the time dispersion resulting from the residual stochastic frequency fluctuations in the *NIST* clock ensemble. The time dispersion due to white frequency fluctuations is proportional to the square root of the time between measurements, so that decreasing the steering interval from one month to one week should improve the time dispersion by a factor of 2. The actual improvement is likely to be somewhat better than this value because a shorter interval between adjustments will result in a more rapid detection of non-statistical outliers.

16.2 The Algorithms of the BIPM

The BIPM computes *UTC* and *TAI* periodically by using the time-difference data from all of the national laboratories and timing centers. The computation proceeds in several steps. Each contributor sends two types of data to the BIPM. The first data set is a table of time differences between each physical clock at the timing facility and the local realization of *UTC*. For example, *NIST* sends the time differences *UTC* (*NIST*) – *clock j*, measured at 0 h *UTC* on all of the days in a month whose Modified Julian Day number ends in 4 or 9. This list also includes *UTC* (*NIST*) – *AT*1 for the same times. The monthly report typically includes six data points for each clock. The second data set is a measurement of the physical time difference between *UTC* (*NIST*) and *UTC* (*PTB*), the time scale of the Physikalische Technische Bundesanstalt, the German equivalent of *NIST*. This physical difference is currently measured both by means of two-way satellite time transfer and the melting-pot version of common-view observations of the US Global Positioning Satellites. I will describe these methods in the next section.

By combining these two data sets, the BIPM computes the difference between each physical clock at NIST and the time of the clock at PTB that is used to control the time-comparison hardware. This signal plays the role of the reference clock in the previous discussion. The BIPM receives the same type of data from every other timing facility, and uses these data to construct a time scale based on the *ALGOS* algorithm [66], which is very similar to *AT*1. Each contributing clock is weighted by its prediction error. The output of this time scale is a paper clock called Échelle Atomique Libre (*EAL*), which has no physical realization. The model for the instability of *EAL* is given by the BIPM as the quadratic sum of three components: a white frequency noise contribution of $1.7 \times 10\text{–}15/\sqrt{(\tau)}$, a flicker frequency noise contribution of 0.35×10^{-15} and a random walk frequency noise $0.55 \times 10\text{–}16 \times \sqrt{(\tau)}$ in 2012, and $0.4 \times 10\text{–}16 \times \sqrt{(\tau)}$ in 2013 and $0.2 \times 10\text{–}16 \times \sqrt{(\tau)}$ in 2014, with τ in days ([58], Tab. 7).

The frequency of *EAL* is compared to the frequency of any primary frequency standard data (such as a cesium fountain) that is available for the analysis period. For example, *NIST* transmits the frequency of its cesium fountain to the BIPM by reporting the offset of the fountain frequency with respect to one of the hydrogen masers that is part of the *NIST* clock ensemble. Since the frequency of this clock with respect to *EAL* has also been estimated by the *EAL* procedure, the BIPM can now compute the difference between the paper frequency of *EAL* and the frequency of the physical primary frequency standard at *NIST*. The BIPM repeats this analysis for every primary frequency standard that has submitted a data value for the current analysis period. From the statistical perspective, the data from the primary frequency standards compensates for the long-period divergence of the free-running *EAL* time scale. It is analogous to the use of the data from Circular *T* at *NIST*, which is used to adjust *UTC* (*NIST*) for the long-period divergence of *AT*1.

An important contribution to the divergence of both scales is the data from hydrogen masers. These devices have very good short-term stability, so that the time scale weighting algorithm, which is driven by relatively short-term frequency stability, assigns a large weight to these devices. However, these devices generally also have statistically significant frequency drift, which dominates the time difference data at sufficiently long times. Therefore, any inadequacy in modeling this frequency drift will introduce an inevitable frequency drift into the ensemble average time and frequency. Since the hydrogen maser is a member of the ensemble that is used to estimate its frequency drift, the maser data pulls the ensemble frequency to some extent determined by its weight. The effect of the correlation is to underestimate the frequency drift parameter.

If the paper frequency of *EAL* deviates significantly from the frequency of the primary frequency standards, the BIPM applies a steering correction to the frequency of *EAL* to drive it to realize the frequency of the primary frequency standards. The steered version of *EAL* is International Atomic Time, or *TAI*. The magnitude of a steering adjustment is a compromise between frequency accuracy, which would favor a relatively large steering adjustment to reduce the frequency difference as rapidly as possible, and the frequency stability of *TAI*, which would favor slower steering at the expense of frequency accuracy.

The steering applied to *EAL* to produce *TAI* is a result of the frequency drift of *EAL* that was not included in the clock models. To put this issue in perspective, the original implementation of *ALGOS* did not include a frequency drift term in the models for any of the contributing clocks. That is, the original ALGOS clock model set $d = 0$ in equation (7) for all clocks. The fractional frequency of *EAL* was adjusted by about 3.1×10^{-15} during calendar year 2008. The calculation of *EAL* at that time was based on data from about 400 clocks and about 100 of them were hydrogen masers, so that the masers might contribute a weight of order 25% to *EAL*. If each of the masers that contributed to EAL had a frequency drift comparable to the maser at *NIST* (10^{-21} s^{-1}) and if the aging parameters were uniformly distributed about 0, then the standard deviation of the aging of the 100 hydrogen masers in the *EAL* ensemble might be of order 10^{-22} s^{-1} or 3.2×10^{-15} yr^{-1}. If the weight of the masers was 25% of *EAL*, the masers would have been responsible for a frequency aging of *EAL* of about 8×10^{-16} yr^{-1}. This result is not very different from the observed frequency aging during 2008. Although this calculation has many approximations and assumptions, it illustrates the effect of frequency aging in masers and the need to model it accurately for time scales that are designed for maximum long-term stability. The BIPM recognized the impact of the drift parameter and modified the *ALGOS* computation to include a drift parameter starting with the computation of *EAL* for September, 2011 [67]. As expected, the addition of the frequency drift parameter to the model of hydrogen masers almost completely eliminated the need to apply additional steering corrections to *EAL* to produce *TAI*.

In the third step of the process, the time of *TAI* is compared to the astronomical time scale, *UT*1. The BIPM inserts a leap second into *TAI* whenever the magnitude of the difference *TAI*-*UT*1 reaches 0.9 s. The determination of when a leap second is needed is made by the IERS as described above. The result of this process is *UTC*, which is identical in both time and frequency to *TAI* except for the accumulated integer leap seconds in *UTC* that are never inserted into *TAI*. There is also a step in the frequency of *UTC* with respect to *TAI* when the estimate of the frequency difference includes a leap second.

When this analysis is complete, the BIPM publishes Circular *T*, which lists the differences between the computed values of *TAI* and *UTC* and $TAI(k)$ and $UTC(k)$, the corresponding time scales of laboratory k. This table is used at each laboratory to steer the local version of *UTC*, $UTC(k)$, to the *UTC* computed by the BIPM.

In general laboratories do not apply any steering to $TAI(k)$. Each laboratory has its own method for realizing this steering. The atomic time scales computed by the BIPM, $AT1$ and EAL, and the clocks at each timing laboratory that contribute to these time scales, are never steered.

The time scales computed by the BIPM are purely paper scales – there is no physical realization of TAI or UTC. Most timing laboratories also do not have a physical device that realizes $TA(k)$, and some laboratories do not compute an explicit TA scale at all (in principle, a timing laboratory can realize its $UTC(k)$ with a single clock that is steered to UTC or $UTCr$. The only requirement is that the time difference of this clock and the reference clock at PTB must be transmitted to the BIPM by some means). There are no clocks or primary frequency standards at the BIPM that contribute to any of the time scales.

The BIPM also computes a post-processed scale that incorporates the data from all of the primary frequency standards available since 2000. The monthly estimates of EAL are smoothed and integrated to obtain TTBIPMxx, where xx is the two-digit year of the computation. The goal is to realize a reference time scale with the best possible long-term stability. The results are computed according to the method described by Azoubib et al. [68] and are published in the annual report of the BIPM, available online at [58].

17 Transmitting time and frequency information

One of the advantages of the time standards of antiquity that were derived from apparent solar time (or any other observed astronomical event) was that there was no need for the infrastructure to transmit time and frequency information. An event that is linked to a specific astronomical observable (local sunrise, for example), will occur at different times at different locations, but can be widely observed with no sophisticated instrumentation. The fact that the event would occur at different times at different locations was not a problem in antiquity, and, even today, religious authorities prepare tables giving the local civil time corresponding to the apparent solar time for religious observances that are defined in this way. Although apparent solar time still plays an important role in everyday life, time and frequency have been defined by more complicated and less obvious methods, and some means for distributing the information is required.

I have already mentioned a number of radio stations that were used to transmit time signals. In general, there was little or no correction for the propagation delay of these signals. The refractivity of the atmosphere (the difference between the atmospheric index of refraction and the vacuum value of 1) at radio frequencies is a function of the pressure, temperature, and water-vapor density along the path, and is about 3×10^{-4} [69] for average conditions. The variation in these parameters along the path can be significant, so that a measurement of these parameters only at the end points is not adequate in many situations. This value does not depend strongly on frequency, so that the two-wavelength method that I will describe below cannot be used. Ignoring this correction was adequate in the early days of national and international time coordination, but the fluctuations in the atmospheric refractivity and the variation in the actual physical path taken by a radio signal eventually limited the accuracy of time transmissions to an unacceptable degree, and some means of correcting for the path delay became necessary.

The propagation delay through the atmosphere of low frequency signals (generally less than 100 kHz) is much more stable than the higher frequency signals that were used initially, especially when the path is over water, and low frequency transmitters such as WWVB operated by NIST proved very useful when only moderate accuracy

(millisecond-level) time transfer was required [70]. The stability of the path delay could also support transmissions of frequency information, a feature that was important in the work of Essen and Markowitz that I discussed above. The variation in the path delay for low-frequency signals has a strong diurnal component, and the uncertainty in the comparison of the fractional frequency differences of remote clocks can be as small as 10^{-11} for an averaging time of one day, assuming that the frequency of the clocks is stable enough to support this averaging time interval [70] (note that this is a result of the *stability* of the path delay averaged over 24 h, and not its magnitude. A large part of the variability in the path delay occurs near local sunrise and sunset because the characteristics of the ionosphere change significantly at these times. The variations at sunrise and sunset are of opposite sign and cancel each other to a significant degree. The variability at other times is significantly smaller). However, low-frequency signals could not support the accuracy required for international time coordination when the required accuracy approached 1 μs, and more sophisticated methods had to be developed. I will discuss each of these methods in the following sections.

17.1 Modeling the delay

This method is based on the assumption that the channel delay can be estimated by means of some parameters that are known or measured from some ancillary measurement. For example, the geometrical path delay between a navigation satellite and a receiver on the ground is estimated as a function of the position of the receiver, which has been determined by some means outside of the scope of the timing measurement, and the position of the satellite, which is transmitted by the satellite in real time in the "navigation message" [71]. For example, a signal transmitted from a global navigation system satellite (such as GPS) takes about 0.65 s to reach a receiver located on the Earth. This delay is very large, but it can be modeled based on the known position of the receiver and the transmitted orbital parameters of the satellite. The residual uncertainty in the propagation delay is on the order of nanoseconds, many orders of magnitude smaller than the delay itself. This residual uncertainty can be reduced even further if the position of the satellite is determined from the precise ephemerides that are available with some delay [72].

There are only a few situations where the channel delay estimated from a real-time model is sufficiently accurate to satisfy the requirements of an application. In most cases, the delay estimate has significant uncertainties, even if the model of the delay is well known. For example, the delay of a signal from a navigation satellite through the troposphere to a station on the Earth is a known function of the atmospheric pressure, temperature and the partial pressure of water vapor, but these parameters are likely to vary along the path so that end-point estimates may not be good enough. This limitation is bad enough for a nearly vertical path from a ground station to a satellite, but the uncertainties become much larger for a lower-elevation path between two ground stations or from signals from a satellite that is near the horizon.

When the satellite is at the zenith with respect to an observing station, the approximate refractivity of the troposphere can be estimated from the atmospheric parameters on the ground at the receiver. If we assume for simplicity that the atmosphere on the ground is an ideal gas at standard temperature and pressure, then the density of air is approximately 2.7×10^{25} molecules/m^3 (or about 44 moles/m^3) at the position of the receiver. The average atomic weight of air is about 30 g/mole, so that the weight of air is about $44 \times (30 \times 10^{-3}) \times 9.8 = 13$ N/m^3. The atmospheric pressure at the surface of the Earth is approximately 101×10^3 N/m^2, and this pressure implies an effective column height of air of $101 \times 10^3/13 = 7700$ m. The refractivity of the atmosphere at standard temperature and pressure is approximately

3×10^{-4}, so that the refractivity adds approximately $7700 \times 3 \times 10^{-4} = 2.3$ m to the geometrical path length. The equivalent increase in the travel time through the troposphere is approximately 2.3 m $\times$ 3.3 ns/m = 7.6 ns. A more careful calculation gives a somewhat smaller value of approximately 6 ns for the effect of the refractivity on a signal from a satellite at the zenith at a typical mid-latitude location [73].

There are a number of different estimates for the increase in the path delay when the satellite is at lower elevations with respect to the receiver. The simplest model assumes that the troposphere is homogeneous and isotropic, so that increase in the path delay is due solely to the geometrical increase in the length of the slant path. With this assumption, the contribution increases as the reciprocal of the sine of the angle between a vector to the satellite and the zenith direction. This dependence only on elevation angle assumes that the properties of the troposphere are the same at all elevations and at all azimuths. These assumptions are almost always too optimistic, especially for a satellite at a low elevation angle or at a station that has very different topography along different azimuths (as is true in Boulder, Colorado, for example). More sophisticated mapping functions, which give the refractivity as a function of elevation and the time of the year are in [74, 75].

17.2 The common view method

This method depends on the fact that there are two (or more) receivers that are approximately equally distant from a transmitter. Since the two path lengths are the same, any signal sent from the transmitter arrives at the same time at both receivers. Each receiver uses its local clock to measure the arrival time of the received signal, and these two measurements are subtracted. The result is the time difference between the clocks at the two receivers. In this simple arrangement, the accuracy of the time difference does not depend on the characteristics of the transmitted signal or the path delay. The detailed structure of the signal and any timing or positioning information that it carries is not used, and the format of the transmission need not be known, provided only that both receivers process the received signal in the same way.

In the real world, it is difficult to configure the two receivers so that they are exactly equally distant from the transmitter. Either this difference must be considered sufficiently small that it can be ignored, or else some means must be used to estimate the contribution of the delay that is not common to the two paths. This estimate is not as demanding as estimating the full path delay, so that the common view method can attenuate any errors in the estimate of the path delay.

There are a number of subtle effects that we must consider when the two path delays of a common-view measurement are not exactly equal. If a single signal is used to measure the time difference, then the signal does not arrive at the two receivers at the same time, since the path delays are somewhat different. Therefore, any fluctuation in the characteristics of the receiver clocks during this time difference must be evaluated. On the other hand, if the measurement is made at the same instant as measured by the receiver clocks, then the signals that are measured by the two receivers did not originate from the source at the same instant of time. Therefore, the fluctuations in the characteristics of the transmitter clock, and the motion of the transmitter during this time interval must be evaluated. Finally, the common-view method does not attenuate the contributions of local effects – delays in the antenna cable or the receiver, for example. Any variation in these delays, due to a dependence of the delay on ambient temperature, for example, is difficult to estimate and is not removed by a one-time calibration.

Another important local effect is caused by multipath reflections – the signal reaching the antenna is a combination of the signal transmitted directly from the satellite

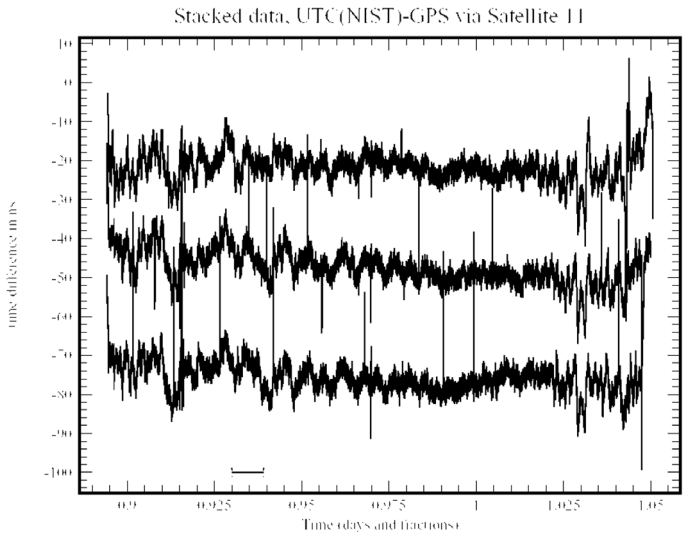

Fig. 11. The time difference in ns between UTC(NIST) and the time transmitted by GPS satellite SV11. The satellite is tracked from horizon to horizon. The data from successive days is displaced vertically for clarity so that the absolute values on the y axis are not significant. Each plot is also advanced (data shifted left) by 4 minutes in time. The significant correlation in the time differences from day to day is an example of the effect of multipath reflections. Note that the magnitude of the fluctuations increases as the elevation of the satellite decreases. The data set ends when the satellite is no longer visible.

and a second signal that reaches the receiver after reflection from a nearby surface (although I discuss multipath problems in this section, they may also be important in other transmissions methods – even when atmospheric propagation is not involved. An impedance mis-match in the termination of a signal cable, for example, will result in a reflection back along the cable, and this reflection behaves very much like the multipath effect that I discuss here). For atmospheric multi-path, the contribution of a reflected signal depends on the geometrical relationship between the satellite, the antenna, and the reflector, and can change relatively rapidly as the satellite position changes. The magnitude of multi-path reflections, which can be of order 10 ns peak-to-peak, can often be estimated by observing this variation, which is generally much too rapid to be caused by the frequency noise of the local clock.

An example of the impact of multipath reflections is shown in Figure 11. Each trace shows the measured time difference every second between UTC(NIST) and the time transmitted by GPS satellite 11. The three traces show the data acquired by tracking the satellite from horizon to horizon on three consecutive days. Each trace is displaced vertically for clarity so that the y-axis scale is preserved but the absolute values on the y axis are not significant. The geometrical relationship between the position of the satellite and any ground reflector is periodic with a sidereal-day period (in first order). Therefore, the multipath contribution has a period of approximately 23:56 and each trace in the figure is advanced (shifted to the left) by 4 minutes to compensate for this effect (the advance is somewhat different for each satellite [76]; this variation is not significant for the current discussion and is not included in the figure). The correlation in the data from one sidereal day to the next is quite apparent, especially near the end of the plot when the satellite is low in the sky so that the direct signal is weaker. The magnitude of the variation depends on the elevation of the satellite, and it is usually largest when the satellite elevation is a minimum near the end of the track as can be seen in the figure.

The common-view method depends on a collaboration between the receiving stations – they must agree on the source of the signal and on the time of observation. In the early days of the use of the GPS system for international time and frequency comparisons (from about 1980 to 2014), the BIPM facilitated this collaboration by publishing tracking schedules that were used by the timing laboratories to schedule the satellites to be used and the times to observe them. The start time of each track was advanced by 4 minutes every day to compensate for the approximate sidereal period of the multipath contribution. These tracking schedules were needed because the early GPS receivers could observe only a single satellite at any one time. The start times of the tracks were originally calculated to use data from satellites that were as close to the zenith as practical for all of the stations in a region. The start times of the tracks were converted to a fixed 16-minute grid in October, 1997, so that the start time of any track could be computed from the origin time of the grid and the 4 minute advance, and a formal schedule was no longer needed. With the development of GPS receivers that could observe all of the satellites in view at any time the publication of these schedules was discontinued in May, 2014, and the common-view between any two stations is calculated after the fact (in general, the BIPM uses the melting-pot method, which I will describe in the next section).

The 4-minute advance in the common-view method used by the BIPM corrected for the day-to-day variation caused by the multipath reflections. Since the reflected signals always travel a greater distance than the direct signal, there is a systematic offset due to multipath. The 4-minute advance therefore converted the multipath effect into a very slowly varying systematic offset that is very difficult to estimate because the long-period variation is hard to separate from the long-period variations in the frequencies of the receiver clocks themselves.

The multipath effect is a systematic offset and cannot be removed by averaging the time differences. However, it is possible to attenuate the effect of multipath by computing the average sidereal-day *frequency difference* between the local clock and the GPS signal and then integrating the result to obtain the time difference. This method exploits the correlation between the multi-path effects on the time differences measured at the same time on consecutive sidereal days (a time interval of approximately 23:56).

A common-view measurement algorithm does not support casual associations among the receivers. The stations that participate in the measurement process must agree on the source to be observed and the time of the observation. There must also be a channel between the two receivers to transmit the measured time differences. On the other hand, there need be no relationship between the receivers and the transmitter, and the transmitter need not even "know" that it is being used as part of a common-view measurement process. Signals from the LORAN-C navigation system [77] were used to compare the clocks of national laboratories in the 1960s and 1970s before GPS satellites were available. The transmissions from commercial analog television stations have also been used as source of common-view signals [78], although multipath reflections can be a serious problem. The zero-crossings of the mains voltage have also be used to compare clocks in the same building with an uncertainty on the order of a fraction of a millisecond [79].

17.3 The "melting pot" version of common view

The previous discussion of common view focused on a number of cooperating receivers, where each one used a local clock to measure the arrival time of a physical signal. However, there can be some situations where there is no single transmitter that can be observed by the receivers at the same epoch. For example, if the common-view

method is implemented by means of signals from a Global Positioning System (GPS) navigation satellite, then receivers on the surface of the Earth that are sufficiently far apart cannot receive signals from any one satellite at the same time.

However, determining the position of a receiver by means of signals from the GPS satellites depends on the fact that the clock in each satellite has a known offset in time and in frequency from a system-average time that is computed on the ground and transmitted up to the satellites. Each satellite broadcasts an estimate of the offset between its internal clock and this GPS system time scale ([73], Chap. 2). By means of this information, two stations that observe two different satellites can nevertheless compute the time difference between the local clock and GPS system time, rather than computing the difference in the time of arrival of the physical signal transmitted by a single satellite. In other words, the common-view time difference in this case is not with respect to a physical transmitter but rather with respect to the computed paper time scale, GPS system time.

In general, a receiver may be able to compute the time difference between its clock and GPS system time by means of the signals from several satellites, which explains the origin of the term "melting pot" method. All of these measurements should yield the same time difference in principle, but this is not the case in practice for a number of reasons.

In the first place, the path delays between the receivers and the various satellites are not even approximately equal, so that any error in computing the path delays is not attenuated as it is in the simpler common-view method described above. In addition, the method depends on the accuracy of the offset between each of the satellite clocks and the system time. As a practical matter, the full advantage of the melting pot method is realized only when the orbits of the satellites and the characteristics of the on-board clocks have been determined using post-processing – the values broadcast by the satellites in real time are usually not sufficiently accurate to be useful for this method.

On the other hand, the melting pot method can usually make use of the observations from several satellites at the same time, so that the random phase noise of the measurement process can be attenuated by averaging the data from the multiple satellites. Therefore, a comparison between the simple two-way method and the melting-pot version depends on a comparison between the noise of the measurement process, which would favor a melting-pot measurement using multiple satellites, and the uncertainties and residual errors in the orbital parameters of the satellites and the offset between the clock in each satellite and GPS system time. The accuracy of the melting-pot method improves as more accurate solutions for the orbits and satellite clocks become available [80].

For baselines not longer than a few hundred km, both stations can generally receive signals from the same satellites, so that the simple common-view and the melting-pot methods are essentially equivalent. The choice is not so clear for longer baselines, and depends on the considerations I mentioned in the previous paragraph. The melting-pot method may be the only choice for very long baselines because no single physical signal source can be observed simultaneously by the receiving stations. For example, this is the situation between stations in Australia and most places in the United States.

17.4 Two-way methods

There are a number of different implementations of the two-way method, but all of them estimate the one-way delay between a transmitter and a receiver as one-half of the round-trip delay, which is measured as part of the message exchange. The accuracy

of the two-way method therefore depends on the symmetry of the delays between the two end points. The accuracy does not depend on the magnitude of the delay itself; although the magnitude of the delay can be calculated from the data in the message exchange, the accuracy of the time difference does not require this computation.

There are generally two aspects of the delay asymmetry that must be considered. The first is a static asymmetry – a constant time difference in the propagation delays between the end points for signals traveling in the opposite directions. In general, this type of asymmetry cannot be detected from the data exchange, and it places a limit on the accuracy of the time differences that can be realized with any two-way implementation (frequency differences are not degraded by a static asymmetry in the delay). The second type of asymmetry is a fluctuation in the symmetry that has a mean of zero. In other words, the channel delay is symmetric on the average, but this does not guarantee the symmetry of any single exchange of data. The impact of this type of fluctuating asymmetry can be estimated with enough data. As I will show in more detail below, a smaller measured round-trip delay is generally associated with a smaller time offset due to any possible asymmetry.

In addition to the delay in the path itself, the transmitter and receiver hardware at the end points of the path are often sources of asymmetry. These hardware delays are often sensitive to the ambient temperature. These effects may not cancel because the admittances to temperature fluctuations may be different at the two end stations and the temperatures at the two end points may be quite different. Finally, there are often components of the measurement process that are outside of the two-way measurement loop, and any delays originating in these components must be measured on every message exchange or measured once and stabilized.

The two-way method can be implemented either in a query-response mode, where one station requests the time and a second station replies, or asynchronously where both stations transmit time messages continuously. I will discuss the query-response mode in detail, because it is the method that is commonly used in network-based digital time services. However, the description also applies to the asynchronous message exchanges, as I will show below.

The message exchange is initiated when station 1 sends a request for the time to station 2 at time T_{1s}, which is measured by the clock at station 1. The message is received at station 2 at time T_{2r}, which is measured by the clock at station 2. At some other time, station 2 sends a message to station 1. This message is transmitted at time T_{2s} (measured by the clock at station 2) and is received at time T_{1r} (measured by the clock at station 1).

The protocol does not depend on any particular relationship among these times, but different implementations may impose a relationship between T_{2r} and T_{2s}. For example, the Network Time Protocol (NTP) [81], which is widely used in digital network time services, generally operates in a query-response configuration, so that station 2 will reply to a message from station 1 but not initiate an exchange. Therefore, $T_{2s} > T_{2r}$. The delay between the time that the query was received by station 2 and the time that the response was transmitted is not important provided that it is measured accurately and that both clocks are well behaved during this interval. There is no specific relationship between the times in two-way satellite time transfer – both stations generally transmit signals continuously. It is simplest to analyze this situation and the query-response configuration in the same way. For the analysis of the two-way satellite time transfer, this means that a transmission from station 2 is paired with an earlier received message from station 1. Since the message exchange is symmetrical, there is no loss of generality in limiting the discussion to this case where station 1 initiated the exchange and $T_{2s} \geqslant T_{2r}$.

If d_{12} is the time it takes a message to travel from station 1 to station 2, then the outbound message provides an estimate of the time difference between the clocks at

stations 1 and 2, Δ_{12}, given by

$$\Delta_{12} = (T_{1s} + d_{12}) - T_{2r}. \tag{9}$$

The first term is the time of clock 1 when its request reaches station 2, and the second term is the reading of clock 2 at that instant. The clock at station 1 is fast with respect to station 2 if this time difference is positive. In a similar manner, we can compute a second estimate of this time difference using the second message, which travels in the opposite direction. As in the previous equation, the first term is the time when the reply is received at station 1 and the second term is the reading of the clock at station 2 at that instant.

$$\Delta'_{12} = T_{1r} - (T_{2s} + d_{21}). \tag{10}$$

These two estimates refer to the same time difference. We estimate the time difference as the average of these two equations

$$\Delta = \frac{\Delta_{12} + \Delta'_{12}}{2} = \frac{(T_{1s} + T_{1r}) - (T_{2s} + T_{2r}) + (d_{12} - d_{21})}{2}, \tag{11}$$

$$D = d_{12} + d_{21} = (T_{1r} - T_{1s}) - (T_{2s} - T_{2r}), \tag{12}$$

where D is the measured round-trip travel time. The first term in equation (12) gives the total elapsed time for the message exchange as measured by the clock in station 1, and the second term is the time interval between when station 2 received the request and when it replied as measured by its clock. All two way methods assume that the path delay is symmetric, so that the third term in the numerator of equation (11) $(d_{12} - d_{21}) = 0$, and equation (12) can be used to find the one-way path delay in either direction. Thus,

$$d_{12} = d_{21} = \frac{D}{2}. \tag{13}$$

If the path delays are not the same in both directions, then equation (12) still gives the correct round trip path delay, D, but equation (13) is no longer correct and the third term in the numerator of equation (11) is no longer 0. We can characterize the symmetry of the path delay using a parameter k, which is a fraction between 0 and 1:

$$\begin{aligned} d_{12} &= kD \\ d_{21} &= (1-k)D, \end{aligned} \tag{14}$$

where the assumption of symmetry in the delay implies that $k = 0.5$. For other values of k, the third term in the numerator of equation (11) introduces a time offset whose magnitude is

$$\varepsilon = (k - 0.5)D, \tag{15}$$

so that the effect of the asymmetry becomes more important as the round-trip path delay itself increases. In other words, the effect of any asymmetry can be minimized by keeping the path delay itself small.

In addition to the obvious requirement that the path delay must be symmetric, there is a more subtle requirement that the clocks be well behaved during the interval in which the messages are exchanged. When the path delay is perfectly symmetric and the third term in the numerator of equation (11) is 0, the time difference given by this equation can be thought of as comparing the time of the clock at station 1 at the epoch corresponding to the midpoint between the time when the initial request was sent and when the final reply was received with the epoch corresponding to the midpoint of the time between when the request was received at the second clock

and when the reply was transmitted. The actual message exchanged spanned a time from $D/2$ before this epoch to $D/2$ after it, and the protocol depends on the fact that the time difference was well behaved over this interval. In practice, this requirement will be satisfied if the time difference between the two clocks varies no faster than linearly with epoch and has no discontinuities. The algorithm will produce a value even if this requirement is not satisfied, but the computed time difference will not be correct.

The two-way method is advantageous even when the path delay is asymmetric. As a simple example, suppose that the round-trip delay is measured as 1 s, but the actual outbound delay is 0.6 s and the inbound delay is 0.4 s, so that there is a static asymmetry of 0.2 s. Equation (14) gives $k = 0.6$, and equation (15) shows that the time error is $(0.6 - 0.5) \times 1 = 0.1$ s. The application of the two-way method has attenuated the impact of the asymmetry by a factor of 2. The same attenuation would also be true for any variation in the asymmetry or if the asymmetry was caused by a single component that contributed to the delay in one of the directions but not in the other one.

For example, suppose that the delay in the outbound direction from station 1 to station 2 is $d + \delta$, while the delay in the opposite direction is d, where d is a constant value and δ is the asymmetry in the delay. The round-trip delay is $2d + \delta$, and, from equation (14), the asymmetry parameter, k, is given by

$$k = \frac{d_{12}}{D} = \frac{d + \delta}{2d + \delta}. \tag{16}$$

From equation (15), the time error is given by:

$$\varepsilon = (k - 0.5), \quad \Delta = \left(\frac{d + \delta}{2d + \delta} - 0.5 \right) (2d + \delta) = \frac{\delta}{2}. \tag{17}$$

The two-way method attenuates *any* asymmetry (whether it is constant or varying) by a factor of 2. Thus there is an advantage to keeping as much of the path delay as possible inside of the two-way measurement loop because delays that are outside of the loop are not attenuated in this way. Although the two-way method will attenuate asymmetric delays that are inside of the measurement loop, it cannot totally compensate for their impact, and the accuracy of all two-way message exchanges is always limited by one-half of whatever residual asymmetry is present.

In many general-purpose computer systems there is a delay between when a message is received by the hardware driver and when the application applies a time stamp to the message. These delays are inside of the measurement loop in principle, but the inbound and outbound contributions are often not equal. A consequence of this asymmetry is that the magnitudes of outbound and inbound time differences, estimated by equations (9) and (10), respectively, are very different because of the corresponding difference in the inbound and outbound hardware latency. It is sometimes possible to detect this problem by noting that either the time difference or the round-trip delay estimated from these data is significantly different from the value expected based on previous data. From equation (15), a smaller round-trip delay guarantees a smaller maximum asymmetry error, and the "huff-'n-puff" filter is based on this principle [81].

17.5 The two-color method

Suppose that there is a portion of the path that has an index of refraction that is significantly different from the vacuum value of one. This difference of the index of refraction relative to its vacuum value is the *refractivity* of this portion of the path.

Suppose also that the refractivity is dispersive. That is, it depends on the frequency that is used to transmit the message. If the length of the path is measured using the transit time of an electromagnetic signal, the refractivity will increase the transit time, so that the effect of the refractivity will be to make the path length appear too long. If the length of the true geometric path is D, and if the index of refraction is n, then the measured length will be L, where L is given by

$$L = nD = D + (n-1)D, \tag{18}$$

where the refractivity is $(n-1)$. I will now consider the special case where the refractivity can be expressed as a product of two functions: $F(p)G(f)$. That is,

$$n - 1 = F(p)G(f). \tag{19}$$

The first function, F, includes parameters that characterize the transmission medium, including any dependence of these parameters on the environment such as the ambient temperature, relative humidity, etc. The second function, G, describes the dispersive characteristics of the path. Both of these function can be arbitrarily complex and non-linear – the only requirement is that the separation be complete. The function F cannot depend on the frequency that is used to transmit the signal and the function G cannot depend on the parameters that describe the characteristics of the path.

If I measure the apparent length of the path using two frequencies, f_1 and f_2, I will obtain two different values for the apparent length because the index is dispersive. These two measured values are L_1 and L_2, respectively. Since the geometrical path length is the same for the measurements at the two frequencies, the difference between the two measurements can be used to solve for the value of the function $F(p)$, and this value of $F(p)$ can be used to calculate the refractivity at frequency f_1:

$$\begin{aligned} L_1 - L_2 &= F(p)\left(G(f_1) - G(f_2)\right)D \\ F(p) &= \frac{L_1 - L_2}{D}\frac{1}{(G(f_1) - G(f_2))} \\ n_1 - 1 = F(p)G(f_1) &= \frac{L_1 - L_2}{D}\frac{G(f_1)}{G(f_1) - G(f_2)}. \end{aligned} \tag{20}$$

If I substitute the last relationship of equation (20) into equation (18) evaluated for frequency f_1, I obtain

$$L_1 = D + (n_1 - 1)D = D + (L_1 - L_2)\,\frac{G(f_1)}{G(f_1) - G(f_2)}, \tag{21}$$

which allows me to compute the geometrical path length, D, in terms of L_1, the length measured using frequency f_1, and the difference in the lengths measured at the two frequencies f_1 and f_2 multiplied by a known function of the two frequencies. If I call this function of the two frequencies H, then

$$\begin{aligned} H(f_1, f_2) &= \frac{G(f_1)}{G(f_1) - G(f_2)} \\ D &= (1 - H)L_1 + HL_2. \end{aligned} \tag{22}$$

The complexity and linearity of the function G are not important, provided only that it is known and that the medium is dispersive (the denominator of the fraction on the right-hand side of equations (21) or (22) is zero for a non-dispersive medium. The difference in the apparent lengths in the first equation (20) will also be zero in

this case). Note that the second term on the right-hand side of equation (21), which is the correction to the geometrical length D due to the dispersive medium, does not depend on D, the geometrical extent of the dispersive medium, but only on the apparent difference in this length for the measurements at the two frequencies. Thus this relationship is equally valid if only a portion of a geometric path is dispersive, and the correction term specifies the apparent change in the length of only that portion of the path. Note also that I do not have to know anything about the function $F(p)$– only that the separation into the product of two terms as expressed by equation (19) represents the dispersion.

The measurements of both L_1 and L_2 will have some uncertainty in general, so that the two-color determination of the geometrical length, D, will have an uncertainty that is greater than it would have been if the medium were non-dispersive so that a measurement at one frequency would have been adequate. The magnitude of the degradation depends on the details of the function H.

For example, the original GPS satellites transmit information on two frequencies to facilitate computing an estimate of the refractivity of the ionosphere. The two frequencies are $f_1 = 1575.42$ MHz and $f_2 = 1227.6$ MHz. The refractivity of the ionosphere is proportional to $1/f^2$. With these values, $H = -1.55$ from the first equation (22). The second equation (22) becomes

$$D = 2.55L_1 - 1.55L_2. \tag{23}$$

The distance defined by equation (23) is often called the "ionosphere free" combination, and is commonly used in both geodetic and timing applications. The noise in this combination is approximately $\sqrt{(2.55^2 + 1.55^2)} \approx 3$ times the noise of either component by itself, so that the two-frequency estimate of the ionosphere degrades the signal to noise ratio by about a factor of 3, assuming that the inherent signal to noise ratios of the L_1 and L_2 signals are equal. Therefore, the two-frequency estimate of the ionosphere is not appropriate for a common-view measurement where the distance between the two stations is short enough so that they see the same ionosphere. The refractivity of the ionosphere will cancel in the common-view subtraction in this case, and the two-wavelength method is not needed and unnecessarily degrades the signal to noise ratio of the measurement.

The two-color method depends on the assumption that the two wavelengths travel along the same path and are affected by the same refractivity. However, any spatial variation in the refractivity will tend to separate the paths taken by the two different frequencies unless the signals travel along (or close to) the direction of the gradient in the refractivity. This is a reasonable approximately for signals traveling from a satellite near the zenith, but is less valid for satellites at lower elevations, and it is a common practice to ignore data from a satellite whose elevation is less than about 20 degrees with respect to the horizon of the observer. This problem also limits the usefulness of the two color method to estimate the refractivity along a horizontal path between two ground stations.

18 Practical time and frequency distribution

Systems for time and frequency distribution use one or a combination of the methods that I have described. The earliest comparisons of the times and frequencies of atomic clocks were realized by means of various forms of the common-view method. I mentioned the use of signals from radio station WWV in common-view by Essen and Perry in the UK and Markowitz at the US Naval Observatory in the 1950s. These experiments were based on the transfer of frequency, and were not sensitive to the path delay provided only that it was stable over short averaging times. Signals from

LORAN transmitters were used in common-view time comparisons for many years starting in the 1950s and 1960s until the satellites of the GPS constellation became available in the 1980s. These LORAN common-view comparisons attenuated any instability in the transmitter clock, but the path delays between the transmitter and the receiving stations were not equal and the difference in delays was not stable, and these problems limited the accuracy of the comparisons.

The signals transmitted by GPS satellites made possible a significant increase in the accuracy of time and frequency distribution. The common-view method based on signals from GPS satellites is usually augmented by the two-color method to estimate the differential delay of the ionosphere and a model of the difference in the geometrical path lengths derived from the ephemeris of the satellite and the known positions of the receivers. This combination has been used to compare the time scales of National Metrology Institutes and other timing laboratories starting in about 1980 when the GPS satellites were first launched and continuing for more than 30 years. The simple common-view method based on the observation of a physical signal was replaced by the melting-pot technique starting in about 2014. This transition was facilitated by the general availability of precise post-processed ephemerides of the satellites in the GPS constellation.

Although the GPS system was the first global navigation satellite system, a number of other systems can be used for distributing time and frequency by using any of the methods that I have described. The Russian GLONASS system, the European GALILEO system and the Chinese BeiDou system are either operational now or will become operational in the next few years. In principle, all of the systems can provide comparable accuracies, although the designs of the systems differ in a number of important technical details, and these differences may impact the accuracy of the time comparisons [73].

The common-view method is generally limited by the local effects that I have discussed, because these effects are not attenuated by the common-view subtraction. The effect of multipath reflections is particularly difficult to estimate. When the common-view method is based on the signals from navigation satellites such as the GPS system, these local effects can introduce long-period systematic offsets on the order of a few ns and larger short-period variations that can be of order 10 ns.

The accuracy of a two-way time transfer depends very strongly on the medium used to transmit the information. Two-way message exchanges that use transponders on communications satellites as part of the transmission path routinely support accuracies on the order of 0.2 ns, and are the basis for the current clock comparisons between many timing laboratories and the BIPM that are used to compute *EAL*, *TAI*, and *UTC* [82, 83]. On the other hand, a two-way message exchange that transmits messages over the standard public Internet supports only millisecond-level accuracies because the network delay is neither stable nor perfectly symmetric [84]. A fluctuating asymmetry on the order of a few percent of the network delay is quite common. Internet-based time services are very widely used when only moderate time accuracy, on the order of milliseconds, is required because the infrastructure that is required to support these messages is already present to support electronic mail, web services and similar applications. There are techniques for improving both the symmetry and the stability of the network delay in digital circuits, but they are typically confined to local area networks or to special circuits that have ancillary hardware to improve the stability of the delay. The routers and switches in digital networks often make a significant contribution to the fluctuations in the delay, and the ancillary hardware ("boundary clocks") is designed to estimate and remove these variations [85]. Communications circuits based on dedicated "dark" optical fibers can support very accurate distribution of time and frequency information by using the standard two-way method. The high accuracy is made possible by the stability and symmetry of the path delay [86].

An important aspect of any practical time and frequency distribution method is the detection of outliers. An important technique for identifying these problems is to compare the time data that are received with an expectation of what values we were expecting. This is usually done by constructing a model of our local clock based on the techniques that I discussed in the time scale text with respect to equation (7). The most difficult part of this job is to estimate the values of the various noise parameters, since these parameters provide an estimate of what is an "outlier" and what is a data point that conforms to our model in a statistical sense. The Allan variance plays a central role in this estimation process. The variance was first developed in 1964 and 1965 by James A. Barnes and David Allan [87]. The variance proved to be particularly useful to characterize the data from clocks and oscillators, which could not be characterized by the usual statistical techniques that use the mean and the standard deviation because the data are generally not stationary over long periods. Therefore, the values of the usual statistical parameters depend on the length of the data set that is used to compute them. The original definition of the variance has been improved over the years, and is now a universal standard for characterizing clocks, oscillators, and the communications channels that are used to transmit the data from these devices [88,89].

19 Summary

The definitions of time and frequency have evolved from a definition of time and time interval based on astronomical observations such as apparent solar time, to a definition of frequency based on the properties of atoms that defined time and time interval as quantities derived from the primary standard of frequency. The relationship between the two definitions has been preserved in the current definition of UTC, which combines frequency-standard data with observations of the astronomical time-scale $UT1$. The details of this connection have both advantages and problems as I have discussed, and it is possible that the connection between atomic time and astronomical time will be modified in the future. This will complete the transition away from the everyday definitions of time and time interval to a more uniform but more abstract realization in which applications that depend on stable frequencies and time intervals play a more fundamental role than time itself. If the link between UTC and $UT1$ is completely eliminated, the times of the two scales will slowly diverge. At present (2016), the rate of divergence is somewhat less than one minute per century.

20 Glossary

Accuracy. The accuracy of a device or measurement normally refers to how closely the measurement or the device produces a value that agrees with one that is considered to be correct. This definition is not always useful in time and frequency metrology for two reasons. In the first place, there is often no device or measurement that is considered to be absolutely correct. In the second place, the term is often used to characterize a new device or method whose performance exceeds the resolution of the current definition. This latter situation arises routinely in the development of new frequency standards.

The accuracy of the newly developed standard is then evaluated by means of two methods. The first method makes use of a comparison between nominally identical, independently operating copies of the device. The agreement between the devices is taken as an indication of the accuracy. The second method makes use of a list of all of the known systematic offsets that affect the device and an estimate of the contribution each offset makes to the result. The contribution is a measure both of the

magnitude of the offset and the residual uncertainty in a subsidiary measurement or calculation intended to estimate and correct for its impact. Both of these techniques have possible short-comings. The first one will be compromised by offsets or problems that are common to both devices and the second one will be compromised by offsets or corrections that are not included in the list of effects or when the effect is included but its impact is not estimated correctly, and both methods are typically used to mitigate the impact of these possibilities.

Apparent lunar month. The time interval between the observations of consecutive lunar first crescents. The observations were usually made just after sunset.

Apparent Solar Day. The time interval between consecutive observations of the position of the Sun at the same point in the sky relative to the horizon. For example, sunrise to sunrise, noon to noon, etc. The observations were often made by noting the position of the shadow of a vertical post.

Apparent Solar Year. The time interval between consecutive observations of a given position of the Sun with respect to the fixed stars. For example, from one summer solstice to the next one or one Vernal equinox to the next one (see Fig. 1).

BIPM, bureau international des poids et mesures. A treaty organization established in 1875 and located near Paris, France. The BIPM receives data from clocks located at various National Metrology Institutes and timing laboratories and uses the data to compute various time scales. The organization of the BIPM is shown in Figure 10.

Day. A day consists of 86 400 seconds exactly, but the length of a second must be specified as described below.

Declination. The position of an astronomical object measured in degrees North (positive) and South (negative) with respect to the equatorial plane (see Fig. 1).

Échelle Atomique Libre (*EAL*). A time scale computed by the BIPM based on data from approximately 400 clocks at various National Metrology Institutes and timing laboratories. The time scale is a weighted average of the contributing clocks, where the weights are derived from the stability of each contributor.

Ephemeris Time. A time scale that is the independent argument of the ephemerides of the sun, the moon, or the planets. It was often determined by observing the position of the moon with respect to the fixed stars.

Ecliptic. The plane containing the apparent path of the Sun around the Earth as shown in Figure 1.

Equation of time. The difference between the apparent position of the sun (apparent solar time) and the position of the mean sun. The difference is usually expressed in minutes (see Fig. 2).

Equatorial plane. The projection of the equator of the Earth onto the sky (see Fig. 1).

Equinox. The two times per year when the path of the Sun along the ecliptic crosses the equatorial plane, which is the projection of the equator on the sky (see Fig. 1).

Frequency. The frequency is the number of events per second, where the second must be specified as described below. The unit of frequency is the Hz, which is equivalent to cycle/second. However, the source need not be cyclical or sinusoidal, and the same unit is used for a source that generates pulses/second.

The term frequency is also used to characterize the dimensionless fractional frequency difference between two devices, where one of the devices is often assumed to be correct and noiseless. The evolution of the "time" of a device (as defined below) is the product of this fractional frequency and the time interval of the observation. For example, a clock whose frequency is 10^{-6} will have a time that will diverge by 1 μs/s relative to a second, nominally perfect device.

GPS Satellites. An ensemble of about 30 satellites that orbit the Earth with a period of about 12 h. The satellites have on-board atomic clocks that provide the reference frequency for the transmitted signals. The signals include information about the orbit of the satellite and the time of the satellite clock with respect to the ensemble average time. The message also includes an estimate of the time offset to the *UTC* time scale as maintained by the US Naval Observatory.

Greenwich Mean Time. A time scale defined by the observation of the position of the mean sun at the Greenwich meridian. Equivalent to Mean Solar Time on the Greenwich Meridian. Although defined with respect to the motion of the mean sun, it is actually computed from observations of stars.

Heliacal Rising. The time of the first appearance of a star above the horizon just before sunrise.

Hour. The hour consists of 3600 seconds exactly, but the length of a second must be specified as described below.

Mean Solar Time. A time scale defined by the motion of a fictitious sun in orbit along the equator. The position of the fictitious sun on the equator follows the position of the real sun in its motion along the ecliptic.

Leap Second. A second added to *UTC* so that the difference between *UTC* and *UT*1 will be less than ± 0.9 s. Leap seconds are announced by the International Earth Rotation and Reference Service about 6 months in advance.

Minute. The minute consists of 60 seconds exactly, but the length of a second must be specified as described below.

NBS. The National Bureau of Standards. The NBS maintains the standards of the International System of Units and conducts research on the best methods for realizing and distributing calibration information. The NBS laboratories were originally located in downtown Washington, DC. The Time and Frequency activities moved to Boulder, Colorado in the 1950s, and the main NBS laboratories moved to Gaithersburg, Maryland shortly after that.

NIST. The National Institute of Standards and Technology. The NBS was renamed *NIST* in 1988. There was no change in its mission or operation from the perspective of time and frequency standards.

Photographic Zenith Tube. A device for determining the meridian transit of a star. The image of the star is reflected from a pool of mercury and is then imaged on a screen or photographic plate.

Right Ascension. The angle measured Eastward from the Vernal equinox along the equatorial plane. The value may be given in degrees or in time units, where 1 h corresponds to 15 degrees (see Fig. 1).

Time. Time is the reading of a clock when an event occurs. The clock and the event are at rest with respect to each other in the same inertial frame, so that the effects of relativity need not be considered. The time is given in units of seconds, and the length of the second must be specified as described below.

When used to discuss the statistics or accuracy of a clock, the time is the difference in the times of a device under test and a second nominally identical device. The second device is often assumed to be absolutely correct, although this is not always the case, and time is also used to specify the time difference between a clock and an ensemble of other devices. When used in this context, the units of time are second and fractions of a second.

TAI, International Atomic Time. A time scale that is based on *EAL* and incorporates frequency information from primary frequency standards located at various laboratories.

Time interval. The difference in the TIMEs of two events. The time interval is given in units of seconds, and the length of the second must be specified as described below.

Tropical Year. The period between consecutive passages of the Sun through the same point in the ecliptic. For example, the time between two consecutive Vernal Equinoxes.

Second. The second is the basic unit of all time data. The definition of the length of a second depends on the era as I have discussed in the text. The early definitions were based on the length of the apparent solar day, the mean solar day, the length of the year 1900, and the independent variable in the ephemeris of the moon. The current (in 2016) technical definition is that the length of a second is exactly 9 192 631 770 cycles of the frequency of the hyperfine transition in the ground state of cesium 133. The frequencies of a number of transitions in other elements are also recognized as "secondary representations" of the second.

Transit Circle. A telescope that is used for determining the meridian transit of stars so as to determine universal time. It is constructed so that it can move only North-South along the local meridian.

Universal Time. With no modifiers, this is a synonym for Greenwich Mean Time.

***UT*0.** An astronomical time scale based on the time of meridian transit of a star at the local meridian. The stars were observed at a number of observatories at approximately the same latitude. These data are sensitive to polar motion.

***UT*1.** The *UT*0 observations from a number of observatories corrected for polar motion.

***UT*2.** The *UT*1 time scale corrected for the annual variation caused by the variation in the moment of inertial of the Earth.

UTC, Coordinated Universal Time. A time scale that adds leap seconds as needed to TAI so that the difference between UTC and UT1 is less than ± 0.9 s.

Very Long Baseline Interferometry (VLBI). A method for determining the position of the Earth with respect to distant radio sources. The signals from the radio sources are received at widely-spaced antennas, and the received signals are correlated after the fact. The time evolution of the peak in the correlation can be used to compute the motion of the Earth.

References

1. J. Jespersen, J. Fitz-Randolph, *From Sundials to Atomic Clocks: Understanding Time and Frequency* (Dover Publications, Inc., Mineola, New York, 1999), Chap. 3.
2. T. Jones, *Splitting the Second: The Story of Atomic Time* (Philadephia, PA, Institute of Physics, 2000), Chap. 2.
3. G.S. Hawkins, J.B. White, *Stonehenge Decoded* (Hippocrene Books, New York, 1988).
4. E.G. Richards, *Mapping Time: The Calendar and its History* (Oxford University Press, Oxford, 1998), Chap. 2.
5. Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac, Her Majesty's Stationary Office, London, 1961, Chap. 14.
6. F. Cabrol, *The Catholic Encyclopedia* (Robert Appleton Company, New York, 1912), Chap. 13.
7. Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac, Her Majesty's Stationary Office, London, 1961, Chap. 3.
8. L. Essen, J.V.L. Parry, An Atomic Standard of Frequency and Time Interval: a Cesium Resonator, Nature **176**, 280-282 (1955.) See also The Cesium Resonator as a Standard of Frequency and Time, Phil. Trans. Roy. Soc. London A **250**, 45-69 (1957) by the same authors.
9. W. Markowitz, R. Glenn Hall, L. Essen and J.V. L. Parry, Frequency of Cesium in Terms of Ephemeris Time, Phys. Rev. Lett. **1**, 105-107 (1958).
10. D.D. McCarthy, P.K. Seidelmann, *Time: From Earth Rotation to Atomic Physics* (Weinheim, Germany, Wiley-VCH GmbH), Chap. 12.

11. S. Leschiutta, The Definition of the Atomic Second, Metrologia **42**, S10-S19 (2005).
12. Resolution 1 of the 13th Conférence Générale des Poids et Mesures (CGPM), available at: www.bipm.org/en/CGPM/db/13/1 (1967).
13. Resolution 9 of the 13th Conférence Générale des Poids et Mesures (CGPM), available at:www.bipm.org/en/CGPM/db/11/9 (1960).
14. T.P. Heavner, E.A. Donley, F. Levi, G. Costanzo, T.E. Parker, J.H. Shirley, N. Ashby, S. Barlow and S.R. Jefferts, First accuracy evaluation of NIST-F2, Metrologia **51**, 174-182 (2014).
15. W.M. Itano, L.L. Lewis and D.J. Wineland, Shift of $^2S_{1/2}$ Hyperfine Splittings due to Blackbody Radiation, Phys. Rev. A **52**, 1233-1235 (1982).
16. T.F. Gallagher, W.E. Cooke, Interactions of Blackbody Radiation with Atoms, Phys. Rev. Lett. **42**, 835-839 (1979).
17. G. Becker, Uncertainty of Cesium-Beam Time Standards due to Beam Asymmetry, IEEE Trans. Inst. Meas. **29**, 297-300 (1980).
18. B. Guinot, Application of General Relativity to Metrology, Metrologia **34**, 261-290 (1997).
19. N.F. Ramsey, *Molecular Beams* (The Clarendon Press, Oxford, 1956), Chap. XIV.
20. N.F. Ramsey, *Molecular Beams* (The Clarendon Press, Oxford, 1956), Chap. X.
21. H.J. Gerritsen, G. Niehuis, Multidirectional Doppler pumping: A new method to prepare an atomic beam having a large fraction of excited atoms, Appl. Phys. Lett. **26**: 347-349 (1975).
22. M. Arditi, J.L. Picque, A cesium beam atomic clock using laser optical pumping, Preliminary tests, J. Phys. Lett. **41**: L379-L381 (1980).
23. C. Salomon, J. Dalibard, W. Philips, A. Clairon and S. Guellati, Laser cooling of cesium atoms below 3 μK, Europhys. Lett. **12**: 683-688 (1990).
24. J.R. Zacharias, Precision Measurements with Molecular Beams, Minutes of the 1954 Annual Meeting of the American Physical Society, 28-30 January, 1954, Phys. Rev. **94**: 751 (1954).
25. A. Clairon, P. Laurent, G. Santarelli, S. Ghezali, S.N. Lea and M. Bouhara, A cesium fountain frequency standard: preliminary measurements, IEEE Trans. Instr. Meas. **44**: 128-131 (1995).
26. T.P. Heavner, E.A. Donley, F. Levi, G. Costanzo, T.E. Parker, J.H. Shirley, N. Ashby, S. Barlow and S.R. Jefferts, First accuracy evaluation of NIST-F2, Metrologia **51**: 174-182 (2014).
27. J. Guéna, P. Rosenbusch, Ph. Laurent, M. Abgrall, D. Rovera, G. Santarelli, M.E. Tobar, S. Bize and A. Clairon, Demonstration of a Dual Alkali Rb/Cs Fountain Clock, IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control **57**, 647 (2010).
28. The International System of Units, Bureau International des Poids et Mesures, Appendix 2, "Practical Realization of the Unit of Time", Published online as SIApp2_s_en.pdf at: www.bipm.org.
29. J.L. Hall, Nobel Lecture: Defining and measuring optical frequencies, Rev. Mod. Phys. **78**: 1279-1295 (2006).
30. L. Hollberg, S. Diddams, A. Bartels, T. Fortier and K. Kim, The Measurement of Optical Frequencies, Metrologia **42**, S105-S124 (2005).
31. A.D. Ludlow, M.M. Boyd, J. Ye, E. Peik and P.O. Schmidt, Optical Atomic Clocks, Rev. Mod. Phys. **87**, 637-701 (2015).
32. S. Droste, F. Ozimek, Th. Udem, K. Predehl, T.W. Hasch, H. Schnatz, G. Grosche and R. Holzwarth, Optical Frequency Transfer over a Single-Span 1840 km Fiber Link, Phys. Rev. Lett. **111**, 110801 1-5 (2013).
33. L.C. Sinclair, F.R. Giorgetta, W.C. Swann, E. Baumann, I.R. Coddington and N. Reynolds Newbury, The impact of turbulence on high accuracy time-frequency transfer across free space, Optical Society of America, Conference on Imaging and Applied Optics, Arlington, Virginia, June 23-27, 2013, available at: http://dx.doi.org/10.1364/PCDVT.2013.PTu2F.2
34. J. Flury, Relativistic Geodesy with Clocks, Proc. 2015 Joint Conference of the IEEE International Frequency Control Symposium and the European Frequency and Time Forum, Denver, Colorado, 12–16 April 2015, in press.

35. Web page of The Royal Museum at Greenwich, available at: http://www.rmg.co.uk.
36. B. Guinot, *History of the Bureau International de l'Heure, Astronomical Society of the Pacific, Conference Series*, edited by S. Dick, D. McCarthy and B. Luzum (2000), Vol. 208, pp. 175-184.
37. C. Audoin, B. Guinot, *Les Fondements de la mesure du temps* (Masson, Paris, 1998).
38. A. Lambert, Le Bureau International de l'heure, son rôle, son fonctionnement, Annuaire du Bureau des Longitudes, Paris, Gauthier-Villars, 1940.
39. G. Bigourdan, A. Lambert, N. Stoyko and B. Guinot, Bulletin Horaire, Paris, Observatoire de Paris, 1922-1967 (19 volumes).
40. G. Bigourdan, Corrections des signaux horaires déterminées par le Bureau International de l'heure, Paris, Gaithier-Villars, 1920-1924.
41. Wireless Time Signals: Radio-Telegraphic Time and Weather Signals Transmitted from the Eiffel Tower and Their Reception, published by Paris Bureau of Longitudes, E & F. N. Spon Ltd., New York, 1915.
42. M.A. Lombardi, G.K. Nelson, WWVB: A Half Century of Delivering Accurate Frequency and Time by Radio, J. Res. of NIST **119**, 25-54 (2014). See also the references in this article.
43. H.J. Walls, QST Magazine, October, 1924, p. 9.
44. R.T. Cox, Standard Radio Wavemeter, Bureau of Standards Type R 70B, J. Opt. Soc. Am. **6**, 162-168 (1922).
45. C. Moon, A Precision Method of Calibrating a Tuning Fork by Comparison with a Pendulum, available on the web at: `dx.doi.org/10.6028/jres.004.016` (1929).
46. Lissajou Figures, Encyclopedia Britannica. See also, J.D. Lawrence, *A Catalog of Special Plane Curves* (Dover, New York, 1972), pp. 178-183.
47. W.G. Cady, United States Patent 1,472,583, October 30, 1923.
48. Captain J.L. Jayne, The Naval Observatory Time Service and How to Use its Radio Signals, The Keystone: Annual Convention of the American National Retail Jewelers' Association, 1913, pp. 129-135.
49. A.H. Orme, Regulating 10 000 Clocks by Wireless, Technical World Magazine, October 1913, pp. 232-233.
50. www.navy-radio-com/commsta/cutler.htm. See also Wikipedia article "VLF Transmitter Cutler".
51. T.H. White, D.C. Washington, AM Station History, 2006. Web page at EarlyRadioHistory.us.
52. W.G. Cady, United States Patent 1,450,246, April, 1923.
53. G. Hefley, The Development of Loran-C Navigation and Timing, National Bureau of Standards Monograph 129, Washington, D. C., Government Printing Office, 1972.
54. B. Guinot, E. Felicitas Arias, Atomic Time-Keeping from 1955 to the present, Metrologia **42**, S20-S30 (2005).
55. B. Guinot, Some Properties of Algorithms for Atomic Time Scales, Metrologia **24**, 195-198 (1987).
56. R.A. Nelson, D.D. McCarthy, S. Malys, J. Levine, B. Guinot, H.F. Fliegel, R.L. Beard and T.R. Bartholomew, The Leap Second: Its History and Possible Future, Metrologia **38**, 509-529 (2001).
57. D.D. Davis, B.E. Blair, J.F. Barnaba, Long-term Continental U. S. Timing System via Television Networks, IEEE Spectrum **8**: 41-52 (1971).
58. Annual Report on Time Activities of the BIPM, Sèvres, France, BIPM, 2008, Vol. 3. Table 1. The more recent reports are available on line at: www.bipm.org/en/bipm/tai/annual-report.html.
59. IERS Bulleting A is available at: http://datacenter.iers.org/eop/-/somos/5Rgv/latest/6 and can also be received by e-mail with a request to http://maia.usno.navy.mil/docrequest.html.
60. IERS Bulletin C is available at: http://datacenter.iers.org/eop/-/somos/5Rgv/latest/16 and can also be received by e-mail.
61. http://www.itu.int/net/pressoffice/press_releases/2012/03.asp

62. J. Levine, The Statistical Modeling of Atomic Clocks and the Design of Time Scales, Rev. Sci. Instr. **83**, 012201-28 (2012).
63. G. Petit, F. Arias, A. harmegnies, G. Panfilo and L. Tisserand, UTCr: A rapid realization of UTC, Metrologia **51**, 33-39 (2014).
64. R.B. Blackman, J.W. Tukey, *The Measurement of Power Spectra* (Dover Publications, New York, 1958).
65. Circular T is published monthly and is available on the BIPM web site as: ftp://ftp2.bipm.org/pub/tai//publication/cirt/
66. G. Panfilo, E.F. Arias, Algorithms for TAI, IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control **57**, 140-150 (2010).
67. G. Panfilo, A. Harmegnies and L. Tisserand, A New Prediction Algorithm for the Generation of International Atomic Time, Metrologia **49**, 49-56 (2012).
68. J. Azoubib, M. Granveaud and B. Guinot, Estimation of the Scale Unit Duration of Time Scales, Metrologia **13**:87-93 (1977).
69. E.K. Smith, S. Weintraub, The constants in the equation for atmospheric refractive index at radio frequencies, Proc. IRE **41**, 1035-1037 (1953).
70. B.E. Blair, Time and Frequency Dissemination: An overview of Principles and Techniques, National Bureau of Standards Monograph 140, Chapter 10, Annex A, Washington, DC, US Government Printing Office, 1974.
71. P. Misra, P. Enge, *Global Positioning System: Signals, Measurements, and Performance* (Massachusetts, Ganga-Jamuna Press, Lincoln, 2006). Chap. 2.
72. International GNSS Service, available on the web at: http://www.igs.org. The service provides a number of precise ephemerides and clock products with delays ranging from a few hours for the "Ultra-Rapid" ephemerides derived from observations to 12–18 days for the "Final" products. The accuracies of the final products are approximately 2.5 cm for the satellite orbits and 75 ps RMS for the satellite and stations clocks.
73. E.D. Kaplan, C.J. Hegarty, *Understanding GPS: Principles and Applications*, 2nd edn., edited by Boston M.A. (Artech House, 2006), Chap. 7, p. 311ff.
74. J. Guo, R.B. Langley, A New tropospheric propagation delay mapping function for elevation angles down to 2°, *Proc. Of the Institute of Navigation ION/GPS Conference*, Portland, Oregon, Sept. 9–12, 2003.
75. A.E. Niell, Global mapping functions for the atmosphere delay at radio wavelengths, J. Geophys. Res. B **2**, 3227-3246 (1972).
76. P. Axelrad, K. Larson and B. Jones, Use of the correct satellite repeat period to characterize and reduce site-specific multipath errors, Proceedings of the 18th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2005), Long Beach California, Sept. 13-16, 2005, pp. 2638-2648.
77. G. Hefley, The Development of Loran-C Navigation and Timing, NBS Monograph 129, Boulder, Colorado, National Bureau of Standards, October, 1972.
78. D.D. Davis, J.L. Jespersen and G. Kamas, The use of television signals for time and frequency dissemination, Proc. IEEE **58**, 931-933 (1970).
79. D.W. Allan, Time transfer using nearly simultaneous reception times of a common transmission, Proc. IEEE **60**, 625-627 (1972).
80. G. Petit, Z. Jiang, GPS all in view time transfer for TAI computation, Metrologia **45**, 35-45 (2008).
81. D.L. Mills, *Computer Network Time Synchronization*, 2nd edn. (CRC Press, Boca Raton, Florida), p. 64ff.
82. Z. Jiang, H. Konaté and W. Lewandowski, Review and Preview of Two-way Time Transfer for UTC generation – from TWSTFT to TWOTFT, Proc. Joint Conference of the Frequency Control Symposium and the European Time and Frequency Forum, 2013, pp. 501–504. Available on the web at: http://www.eftf.org/proceedings/proceedingsEFTF2013.pdf.
83. Z. Jiang, G. Petit, Combination of TWSTFT and GNSS for accurate UTC time transfer, Metrologia **46**: 305-314 (2009).
84. P. Misra, P. Enge, *Global Positioning System: Signals, Measurements, and Performance* (Massachusetts, Ganga-Jamuna Press, Lincoln, 2006). Chap. 2, p. 48ff.

85. J.C. Eidson, *Measurement, Control, and Communication using IEEE 1588* (Springer-Verlag, London, 2006), Chap. 5.
86. M. Rost, D. Piester, W. Yang, T. Feldmann, T. Wübbena and A. Bauch, Time transfer through optical fibers over a distance of 73 km with an uncertainty below 100 ps, Metrologia **49**: 772-778 (2012).
87. J.A. Barnes, D.W. Allan, Two papers on the statistics of precision frequency generators, National Bureau of Standards Technical Report 8878, 1965. Available from the publications list at: http://tf.nist.gov, publication 224.
88. D.W. Allan, J.A. Barnes, A modified Allan variance with increased oscillator characterization ability, *Proc. 35th Annual Frequency Control Symposium, 1981*, pp. 470-475. Available from the publications list at: http://tf.nist.gov, publication 560.
89. D.A. Howe, A total estimator of the Hadamard function used for GPS operations, *Proc. 32nd Precise Time and Time Interval Planning and Applications Meeting, Nov. 29, 2000*, pp. 255-268. Available from the publications list at: http://tf.nist.gov, publication 1431.