

REVIEW ARTICLE

Introduction to time and frequency metrology

Judah Levine^{a)}

JILA and Time and Frequency Division, NIST and the University of Colorado, Boulder, Colorado 80309

(Received 24 July 1998; accepted for publication 22 February 1999)

In this article, I will review the definition of time and time interval, and I will describe some of the devices that are used to realize these definitions. I will then introduce the principles of time and frequency metrology, including a discussion of some of the types of measurement hardware in common use and the statistical machinery that is used to analyze these data. I will also introduce various techniques of distributing time and frequency information, with special emphasis on the global positioning system satellites. I will then discuss the advantages of clock ensembles and a prototype time-scale algorithm. I will conclude with a discussion of how clocks are synchronized to remote servers using noisy and poorly characterized transmission channels.

[S0034-6748(99)00106-9]

I. INTRODUCTION

Clocks and oscillators play important roles in many areas of science and technology, ranging from tests of general relativity to the synchronization of communication systems and electric power grids. Hundreds of different types of devices exist, ranging from the cheap (but remarkably accurate) crystal oscillators used in wrist watches to one-of-a-kind primary frequency standards that are both seven or more orders of magnitude more accurate and almost the same number of powers of ten more expensive than a wrist watch. In spite of this great diversity in cost and performance, almost all of the devices can be described using basically the same statistical machinery. The same is true for methods of distributing time and frequency signals—the same principles are used to transmit time over dial-up telephone lines with an uncertainty of about 1 ms and to transmit time using satellites with an uncertainty of about 1 ns.

In this article I will begin by describing the general characteristics of clocks and oscillators. I will then discuss how time and frequency are defined, how these quantities are measured, and how practical standards for these quantities are realized and evaluated. This characterization, and the analysis of time and frequency data in general, is complicated because the spectrum of the variation is almost never white and the conventional tools of statistical analysis are therefore usually not appropriate or effective.

I will then discuss the basic principles that govern methods for distributing time and frequency information. The channels used for this purpose are often noisy, and this noise is usually neither white nor very well behaved statistically. It is often possible to separate the degradations due to the noise in the channel from the fluctuations in original data. Although the details of how this can be done vary among the

different applications, we will see that all of them can be understood in terms of a few basic principles. I will discuss the details of specific applications only insofar as is needed to illustrate these basic ideas.

I will not discuss the very large number of optical frequency standards that have been developed. Such devices would certainly qualify as standards from the stability point of view, but their outputs must be handled by optical rather than by electronic means at the current time, and the resulting technology is very different as a result. In addition, it is extremely difficult to use these devices as clocks in the usual sense of that word—counting an optical frequency has been done using various heterodyne methods, but these measurements are far from routine.

II. CHARACTERISTICS OF OSCILLATORS AND CLOCKS

An oscillator is comprised of two components: a generator that produces periodic signals and a discriminator that controls the output frequency. In many configurations, the discriminator is actively oscillating and the output frequency of the device is set by a resonance in its response. Pendulum clocks and quartz-crystal oscillators are of this type—in both cases the frequency is determined by a mechanical resonance which must be driven by an external power source. In other configurations, the discriminator is passive—its frequency-selective property is interrogated by a variable-frequency oscillator whose frequency is then locked to the peak of the discriminator response function using some form of feedback loop.

Atomic frequency standards fall into both categories. Lasers are usually (but not always) in the first category, while cesium and rubidium devices are almost always in the second. Hydrogen masers can be either active or passive—both designs have advantages, as we will see below.

^{a)}Electronic mail: jlevine@boulder.nist.gov

Either type of oscillator can be used to construct a clock. The periodic output is converted to a series of pulses (using a mechanical escapement or an electrical zero-crossing detector, for example), and the resulting “ticks” are counted in some way. While the frequency of an oscillator is a function of its mechanical or physical properties, the initial reading of a clock is totally arbitrary—it has to be set based on some external definition. This definition cannot be specified from the technical point of view—for historical and practical reasons the definition that is used is based on the motion of the earth.¹

The vast majority of oscillators currently in use are stabilized using quartz crystals. Even the quartz crystals used as the frequency reference in inexpensive wrist watches can have a frequency accuracy of 1 ppm and a stability a factor of 10 or more better than this. Substantially better performance can be achieved using active temperature control. They are usually the devices of choice when lowest cost, robust design, and long life are the most important considerations. They are not well suited for applications (such as primary frequency standards) where frequency accuracy or long-term frequency stability are important. There are two reasons for this. The mechanical resonance frequency depends on the details of the construction of the artifact and is therefore hard to replicate. In addition, the resonance frequency is usually affected by fluctuations in temperature and other environmental parameters so that its long-term stability is relatively poor. A number of very clever schemes have been developed to mitigate these problems, but none of them can totally eliminate the sensitivity to environmental perturbations and the stochastic frequency aging that are the characteristics of an artifact standard.²

Atomic frequency standards use an atomic or molecular transition as the discriminator. In the passive configuration, the atoms are prepared in one of the states of the clock transition and are illuminated by the output from a separate oscillator. The frequency of the oscillator is locked to the maximum in the rate of the clock transition. An atomic frequency standard therefore requires a “physics package” that prepares the atoms in the appropriate state and detects the fact that they have made the clock transition following the interaction with the oscillator. It also requires an “electronics package” that consists of the oscillator, the feedback circuitry to control its frequency, and various synthesizers or frequency dividers to convert the clock frequency to standard output frequencies such as 5 MHz or 1 Hz. Since the transition frequencies are determined solely by the atomic structure in principle, these standards would seem to be free of many of the problems that limit the accuracy of artifact standards. This is true to a great extent, but the details of the interaction between the atoms and the probing frequency and the interaction between the physics and electronics packages affect the output frequency to some degree, and blur the distinction between a pure atomic frequency standard and one whose frequency is determined by a mechanical artifact.

III. DEFINITION OF TIME INTERVAL OR FREQUENCY

Time interval is one of the base units of the International System of Units (the SI system). The current definition of the second was adopted by the 13th General Conference on Weights and Measures (CGPM) in 1967:

The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the cesium 133 atom.³

This definition of the base unit can be used to derive the unit of frequency, the hertz. Multiples of the second, such as the minute, hour, and day are also recognized for use in the SI system. Although these common units are defined in terms of the base unit just as precisely as the hertz is, they do not enjoy the same status in the formal definition of the international system of units.⁴

The definition of the SI second above implies that it is realized using an unperturbed atom in free space and that the observer is at rest with respect to the atom. Such an observer measures the “proper” time of the clock, that is the result of a direct observation of the device independent of any conventions with respect to coordinates or reference frames.

Practical timekeeping involves comparing clocks at different locations using time signals, and these processes introduce the need for coordinate frames. International atomic time (TAI) is defined in terms of the SI second as realized on the rotating geoid,⁵ and TAI is therefore a coordinate (rather than a proper) time scale.

This distinction has important practical consequences. A perfect cesium clock in orbit around the earth will appear to observers on the earth as having a frequency offset with respect to TAI. This difference must be accounted for using the usual corrections for the gravitational redshift, the first- and second-order Doppler shifts, etc. These frequency corrections are important in the use of the global positioning system (GPS) satellites, as we will discuss below.

A. Realization of TAI

TAI is computed retrospectively by the International Bureau of Weights and Measures (BIPM), using data supplied by a world-wide network of timing laboratories. Each participating laboratory reports the time differences, measured every five days (at 0 h on those modified Julian day numbers which end in 4 or 9), between each of its clocks and its laboratory time scale, designated UTC(lab). To make it possible to compare data from different laboratories, each laboratory also provides the time differences between UTC(lab) and GPS time, measured using an algorithm defined by the BIPM and at times specified in the BIPM tracking schedules. (The GPS satellites are used only as transfer standards, and the satellite clocks drop out of the data.) Approximately 50 laboratories transmit data from a total of about 250 clocks to the BIPM using these methods.

Using an algorithm called ALGOS, the BIPM computes the weighted average of these time-difference data to produce an intermediate time scale called *echelle atomique libre* (EAL). The length of the second computed in this way is compared with data from primary frequency standards and a correction is applied if needed so as to keep the frequency as

close as possible to the SI second as realized by these primary standards. The resulting steered scale is TAI. As of December, 1998, the fractional frequency difference, $f(\text{EAL}) - f(\text{TAL}) = 7.13 \times 10^{-13}$. At that same epoch, the fractional frequency offset between TAI and the SI second as realized on the rotating geoid was estimated by the BIPM to be

$$d = (u\text{TAI} - u0)/u0 = (-4 \pm 10) \times 10^{-15}.$$

The determination of the uncertainty for the estimate of d is complex. It includes the uncertainties (usually the ‘‘type B’’ uncertainty, which we define later) associated with the data from the different primary frequency standards; extrapolating these data to a common interval also requires a model for the stability of EAL. The details are presented in reports published by the BIPM.⁶

B. Coordinated universal time and leap seconds

It is simple in principle to construct a timekeeping system and a calendar using TAI. When this scale was defined, its time was set to agree with UT2 (a time scale derived from the meridian transit times of stars corrected for seasonal variations) on 1 January 1958. Unfortunately, the length of the day defined using 86 400 TAI seconds was shorter than the length of the day defined astronomically (called UT1) by about 0.03 ppm, and the times of the two scales began to diverge immediately. If left uncorrected, this divergence would have continued to increase over time, and its rate would have slowly increased as well as the rotational rate of the earth continued to decrease.

Initially, this divergence was removed by introducing frequency offsets between the frequency used to run atomic clocks and the definition of the duration of the SI second. Small time steps of 0.05 or 0.1 s were also introduced as needed.⁷ This steered time scale was called coordinated universal time (UTC)—it was derived from the definition of the SI second, but was steered (that is, ‘‘coordinated’’) so that it tracked the astronomical time scale UT1.

This method of coordinating UTC turned out to be awkward in practice. The frequency offsets required for coordination were on the order of 0.03 ppm, and it was difficult to insert these frequency offsets into real-world clocks. On 1 January 1972, this system was replaced by the current system which uses integral leap seconds to realize the steering needed so that atomic time will track UT1. The rate of UTC is set to be exactly the same as the rate of TAI, and the divergence between UTC and astronomical time is removed by inserting ‘‘leap’’ seconds into UTC as needed. The leap seconds are inserted so as to keep the absolute magnitude of the difference between UTC and UT1 less than 0.9 s. These leap seconds are usually added on the last day of June or December; the most recent one was inserted at the end of 31 December 1998 and the time difference TAI-UTC became exactly 32 s when that happened. The leap seconds are always positive, so that UTC is behind atomic time, and this trend will continue for the foreseeable future.

Although this system of leap seconds is simple in concept, it has a number of practical difficulties. A leap second

is inserted as the last second of the day (usually on the last day of either June or December), after 23:59:59, which would normally have been the last second of the day. Its name is 23:59:60, and the next second is 00:00:00 of the next day. The first problem is that it is difficult to define a time tag for an event that happens during the leap second. This time does not even exist on conventional clocks, for example, and a time tag of 23:59:60 might be rejected as a format error by a digital system. Furthermore, there is no natural way of naming the leap second in systems that do not use the hh:mm:ss nomenclature. Many computer systems fall into this category because they keep time in units of seconds (and fractions of a second) since some base date.

Calculating the length of a time interval that crosses a leap second presents additional ambiguities. An interval based on atomic time or on some other periodic process will differ from a simple application of a calculation based on the difference in the UTC times, since the latter calculation effectively ‘‘forgets’’ that a leap second has happened once it has passed.

There is no simple way of realizing a time scale that is both smooth and continuous and simultaneously tracks UT1. This is because the intervals between leap seconds are not exactly constant and predictable, and there is no simple algorithm that could predict them automatically very far into the future. Each leap second is announced several months in advance by the International Earth Rotation Service in Paris.⁸ Thus, changing the SI definition of the second (which would produce difficulties in the definitions of other fundamental constants anyway) could reduce the frequency of leap seconds, but would not entirely remove the need for them. The problem would return in the future anyway as the length of the astronomical day increased.

It is possible to avoid the difficulties associated with leap seconds by using a time scale that does not have them, such as TAI or GPS time. There are a number of practical difficulties with this solution, such as the facts that most timing laboratories disseminate UTC rather than TAI, and that GPS time may not satisfy legal traceability requirements in some situations. Practical realizations of this idea often require ancillary tables of when leap seconds were inserted into UTC.

IV. CESIUM STANDARDS

Cesium was chosen to define the second for a number of practical reasons. The perturbations on atomic energy levels are quite small in the low-density environment of a beam, and atomic beams of cesium are particularly easy to produce and detect. The frequency of the hyperfine transition that defines the second is relatively high and the linewidth can be made quite narrow by careful design. At the same time the transition frequency is still low enough that it can be manipulated using standard microwave techniques and circuits.

In spite of these advantages, constructing a device that realizes the SI definition of the second is a challenging and expensive undertaking. Although the frequency offsets produced by various perturbations (such as Stark shifts, Zeeman shifts, Doppler shifts,...) are small in absolute terms, they are large compared to the frequency stability that can be

achieved in a well-designed device. A good commercial cesium frequency standard, for example, might exhibit fractional frequency fluctuations of 2×10^{-14} for averaging times of about one day. The frequency of the same device might differ from the SI definition by 1×10^{-13} or more (sometimes *much* more), and this frequency offset may change slowly with time as the device ages. (This frequency offset is what remains after corrections for the perturbations mentioned above have been applied. If no corrections are applied, the fractional frequency offset is usually dominated by Zeeman effects, which can be as large as 1×10^{-10} .)

Constructing a device whose accuracy is comparable to its stability is a difficult and expensive business, and only a few such primary frequency standards exist. It is usually impossible to reduce the systematic offsets to sufficiently small values, and the residual offset must be measured and removed by means of a complicated evaluation process that often takes several days to complete.⁹ The operation of the standard may be interrupted during the evaluation, so that many primary frequency standards cannot operate continuously as clocks. In addition, the systematic offsets may change slowly with time even in the best primary standard, so that evaluating their magnitudes is a continuing (rather than a one-time) effort.

A simple oven can be used to produce a beam of cesium atoms. The oven consists of a small chamber and a narrow slit through which the atoms diffuse. The operating temperature varies somewhat, but is generally less than 100 °C, so that the thermal mean velocity of the cesium atoms is somewhat less than 300 m/s. Operating the oven at a higher temperature increases the flux of atoms in the beam and therefore the signal to noise ratio of the servo loop to control the local oscillator, but the velocity of the atoms also increases as the square root of the temperature. This decreases the interaction time and increases the linewidth by the same factor, but the more serious problem is usually that the collimation of the beam is degraded. The decrease in interaction time resulting from raising the temperature can be offset by using only the low-frequency “tail” of the velocity distribution.

Conventional frequency standards use an inhomogeneous magnetic field (the “A” magnet) as the state selector. The atoms enter the A magnet off axis and are deflected by the interaction between the inhomogeneous magnetic field and the magnetic dipole moment of the atom. The geometry is arranged so that only atoms in the $F=3$ state emerge from the magnet—atoms in the other state are blocked by a mechanical stop. It is not practical to make the magnetic field of the A magnet large enough to separate the m_F sub levels of the $F=3$ state.

The atoms that emerge from the A magnet enter an interaction region (the “C” region), where they are illuminated by microwave radiation in the presence of a constant magnetic field, which splits the hyperfine levels because of the Zeeman effect. Although the transition between the $F=3$ $m_F=0$ to $F=4$ $m_F=0$ is offset somewhat as a result of this field, the sensitivity to magnetic field inhomogeneities is reduced, and the clock transition is independent of the magnetic field value in first order. The actual value of the field is

usually measured by observing the frequencies of field-dependent transitions between other m_F states.

The microwave radiation that induces the clock transition is applied in two separate regions separated by a drift space (Ramsey configuration¹⁰). The atoms experience two kinds of transitions: “Rabi” transitions due to a single interaction with the microwave field, and “Ramsey” transitions due to a coherent interaction in both regions. The Rabi transitions have a shorter interaction time and therefore a larger linewidth. The magnitude of the magnetic field in the C region is usually chosen so that the Rabi transitions between the different sub levels will be well resolved. A typical value for a conventional thermal-beam standard would be about 5×10^{-6} T. It is possible to use a lower value of magnetic field which does not completely resolve the Rabi transitions. While this reduces the Zeeman correction, it complicates the analysis of the Ramsey line shapes.

The length of this drift space and the velocity of the atoms determine the total interaction time, which in turn sets the fundamental linewidth of the device. The microwave frequency is adjusted until the transition rate in the atoms is a maximum; the transitions are detected using a second inhomogeneous magnetic field, which is created by the “B” magnet. The gradient of the B field is usually configured so as to pass only those atoms that are in the final state of the clock transition (a “flop-in” configuration). This configuration eliminates the shot noise in the detected beam due to atoms that do not make the clock transition in the C region. However, it is more difficult to test and align, since there is no signal at all until everything is working properly.

The state-selection process in a conventional cesium device is passive—it simply throws away those atoms that are not in the proper state. There are 16 magnetic sub levels in the ground state: 9 for $F=4$ and 7 for $F=3$ (that is, $2F+1$ in both cases). All of them have essentially the same energy and are therefore roughly equally populated in a thermal beam. As a consequence, only 1/16 of the flux emerging from the oven is in a single F , m_F state, and only this small fraction of the atoms is actually used to stabilize the oscillator. This unfavorable ratio can be improved using active state selection—that is, by optically pumping the input beam so as to put nearly all of the atoms into the initial state of the clock transition. A similar technique can be used to detect the atoms that have undergone the clock transition in the interaction region. This design increases the number of atoms that are used in the synchronization process and therefore the signal to noise ratio of the error voltage in the servo loop that controls the oscillator. In addition, it removes some of the offsets and problems associated with the fields due to the A and B magnets. Although the optical pumping techniques that are needed to realize this design have been known for many years,¹¹ realizing a standard using these principles was not practical until recently when lasers of the appropriate wavelengths became available.

Another technique for improving the performance of the device is to increase the interaction time in the C region by increasing its length or by decreasing the velocity of the atoms. The ultimate version of this idea may be an “atomic fountain” in which atoms are shot upward through the inter-

action region, reverse direction, and fall back down through the same region a second time. This idea was proposed more than 40 years ago,¹² but has only recently become feasible as a result of developments in laser cooling and trapping. The fountain idea has a number of advantages over one-way designs. The atoms are moving more slowly and the interaction time can be longer as a result. In addition, the fact that atoms traverse the interaction region in both directions can be used to cancel a number of systematic offsets that would otherwise have to be estimated during the evaluation process.

V. OTHER ATOMIC STANDARDS

Cesium is not the only atom that can be used as a reference for a frequency standard; transitions in rubidium and hydrogen are also commonly used. In order to relate their frequencies to the SI definition, devices based on rubidium or hydrogen must be calibrated with respect to primary cesium-based devices if accuracy is important to their use. As we have discussed above, this is true for commercial cesium devices as well if the accuracy of the output frequency is to approach the frequency stability of the device. In spite of this similarity, there are real differences among the devices.

Rubidium devices usually use low-pressure rubidium vapor in a cell filled with a buffer gas such as neon or helium rather than the atomic-beam configuration used for cesium. The clock transition is the $F=1$, $m_F=0$ to $F=2$, $m_F=0$ transition in the $^2S_{1/2}$ ground state of ^{87}Rb , its frequency is about 6.83 GHz. The two states of the clock transition are essentially equally populated in thermal equilibrium.

The atoms are optically pumped into the $F=2$ state. There are a number of ways of doing this. One common method uses a discharge lamp containing ^{87}Rb whose light is passed through an absorption cell containing ^{85}Rb . The light emerging from the lamp could induce optical frequency transitions from both the $F=1$ and $F=2$ states to higher excited levels. However, there is a coincidence between transitions in ^{85}Rb and ^{87}Rb . The effect of this coincidence is that the light which would have induced transitions originating from the $F=2$ state is preferentially absorbed in passing through the ^{85}Rb filter, whereas the component that interacts with the $F=1$ lower state is not. When the filtered light enters the cell, it preferentially excites the $F=1$ state of the clock transition because the light that would interact with the $F=2$ state has been preferentially absorbed in the ^{85}Rb cell. The excited atoms decay back to the two ground sub levels. Those atoms which decay back to the $F=1$ state are preferentially excited again, while those atoms that decay back to the $F=2$ state tend to remain there. The result is to build up a nonequilibrium distribution in these two sub levels. As in cesium, a magnetic field is applied to split the hyperfine sub levels and to reduce the sensitivity to any residual inhomogeneities and fluctuations in the ambient field. When the microwave clock frequency is applied, the $F=1$, $m_F=0$ state is partially repopulated by the induced transition between the two clock states, the absorption of the incident light from the lamp increases again, and the transmitted intensity drops. This decrease in transmitted intensity (typically on the order

of 1%) is used to lock the incident microwave frequency to the clock transition.

The beauty of this design is that the microwave transition modulates an optical intensity, and changes in the visible photon flux are much easier to detect than the same flux of microwave photons would be. On the other hand, the decrease in the transmitted light through the cell must be detected against the shot noise in the full background light from the lamp. Using a discriminator based on the visible flux from the lamp makes the device simpler and cheaper, but it also makes for poorer stability and larger and more variable frequency offsets than would be found in a good cesium-based device. This is because of pressure shifts in the cells, light shifts due to the lamp, a large dependence of the clock frequency on the ambient magnetic field, and other effects. The values of many of these parameters must be chosen as a compromise between accuracy and stability. A higher-intensity lamp, for example, increases the signal to noise ratio in the resonance-detection circuit, but also increases the light shift (the ac Stark effect on the clock transition). A typical rubidium-stabilized device might have a stability and an accuracy 100 times (or more) poorer than a device using cesium in the standard beam configuration. In terms of the Allan deviation (to be defined later), the performance is generally about $3 \times 10^{-11}/\tau^{1/2}$ for averaging times (that is, values of τ) between about 1 and 10^4 s; the long-term frequency aging is typically on the order of 5×10^{-11} /month. On the other hand, they are usually substantially cheaper, lighter, and smaller than standard cesium devices.

Both the advantages and the limitations of rubidium devices are due to the implementation in an optically pumped cell with a buffer gas rather than to any inherent difference between cesium and rubidium. A cesium-based device that used a cell rather than a beam, for example, might have many of the advantages of both systems, and such a device has been under development for some time. Furthermore, a fountain based on rubidium would have a number of advantages over a cesium-based device. Although the clock transition has a lower frequency in rubidium, the atom-atom collision cross section in the cold beam of a fountain is substantially smaller than in cesium¹³ so that the beam flux can be greater. The short-term stability will be improved as a result; a number of other systematic offsets would also be smaller. (Improving the short-term stability is a very valuable advantage, since it eases the requirements on the stability of the local oscillator which interrogates the atoms and which acts as a flywheel between cycles of the fountain.)

As with cesium and rubidium, the clock transition in a hydrogen maser is a hyperfine transition in the ground state of the atom. The $F=0$ and $F=1$ hyperfine states are separated using an inhomogeneous field as in a cesium device, but instead of using the beam geometry of a cesium-based device, the atoms in the initial state of the clock transition enter a bulb that is inside of a microwave cavity.¹⁴ In a passive maser, the atoms in the cavity are probed with a microwave signal at 1.42 GHz, and the interaction with the atoms produces a phase shift in the reflected signal that is dispersive about line center. In an active maser, the losses of

the cavity are small enough so that the device oscillates at the hyperfine transition frequency.

The frequency of oscillation of an active maser is a function of both the cavity resonance frequency and the atomic transition frequency, so that some means of stabilizing the cavity is usually required. Collisions with the walls of the bulb also affect the frequency of the device. Various strategies are used to minimize these effects, including applying special "nonstick" coatings to the inside of the bulb and actively tuning the resonance frequency of the cavity so that it is exactly on atomic line center.

In spite of these techniques, hydrogen masers usually have frequency offsets that are large compared to the stability of the device. (The fractional frequency offset of a hydrogen maser is often of order 5×10^{-11} , while the fractional frequency fluctuations at one day can be as small as 2×10^{-16} .) In addition to being significant, these offsets usually change slowly with time in ways that are difficult to predict. Nevertheless, the frequency stability over short periods (usually out to at least several days) can be much better than the best cesium device. Hydrogen masers are therefore the devices of choice when the highest possible frequency stability is required and cost is not an issue.

Finally, a number of other transitions have been used to stabilize oscillators. Some of the most stable devices are those using a hyperfine transition in the ground state of an alkali-like ion,¹⁵ such as $^{199}\text{Hg}^+$.¹⁶ This ion has a structure similar to cesium, with a single $6s$ electron outside of a closed shell. The clock transition is the hyperfine transition in the ground state analogous to the transition in cesium; it has a frequency of 40.5 GHz. Using a linear ion trap to confine the ions, frequency standards with a stability of $2 \times 10^{-14}/\tau^{1/2}$ (where τ is the averaging time in seconds) have been reported¹⁵ and devices based on other ions are being developed.¹⁷ These devices have the advantage that the interaction time can be much longer than in a conventional beam configuration.

VI. QUARTZ CRYSTAL OSCILLATORS

Each of the devices we have discussed above is implemented using a quartz-crystal oscillator whose frequency is stabilized to the atomic transition using some form of feedback loop. The details of the loop design vary from one device to another, but the loop will always have a finite bandwidth and therefore a finite attack time. The internal oscillator is essentially free running for times much shorter than this attack time, and the stability of all atomic-stabilized oscillators is therefore limited at short periods to the free-running stability of the internal quartz oscillator. (The boundary between "short" and "long" periods varies from one device to another, of course. A typical value for many commercial devices is probably on the order of seconds.)

An obvious simplification would be to use the bare quartz-crystal oscillator and to dispense altogether with the stabilizer based on the atomic transition. The stability of such a device is clearly the same as an atomic-based system at sufficiently short times. The stability at longer times will depend on how well the quartz crystal frequency reference

can be isolated from environmental perturbations, or how well the effects of these perturbations can be estimated and removed.² Quartz-crystal oscillators can be surprisingly good in this respect. Even cheap wrist watches have oscillators that may have a frequency accuracy of about 1 ppm and a stability 10 or 50 times better than this. The residual frequency fluctuations are largely due to fluctuations in the ambient temperature, and stabilizing the temperature can improve the performance.

VII. MEASURING TOOLS AND METHODS

The oscillators that we have discussed above generally produce a sine-wave output at some convenient frequency such as 5 MHz. (This frequency may also be divided down internally to produce output pulses at a rate of 1 Hz. Crystals designed for wrist watches and some computer clocks often operate at 32 768 Hz to simplify the design of these 1 Hz dividers.) A simple quartz-crystal oscillator might operate directly at the desired output frequency; atomic standards relate these output signals to the frequency appropriate to the reference transition by standard techniques of frequency multiplication and division. The measurement system thus operates at a single frequency independent of the type of oscillator that is being evaluated. The choice of this frequency involves the usual trade-off between resolution, which tends to increase as the frequency is made higher, and the problems caused by delays and offsets within the measurement hardware, which tend to be less serious as the frequency is made lower.

Measuring instruments generally have some kind of discriminator at the front end—a circuit that defines an event as occurring when the input signal crosses a specified reference voltage in a specified direction. Examples are 1 V with a positive slope in the case of a pulse, or 0 V with a positive slope in the case of a sine-wave signal. The trigger point is chosen at (or near) a point of maximum slope, so as to minimize the variation in the trigger point due to the finite rise time of the wave form.

The simplest method of measuring the time difference between two clocks is to open a gate when an event is triggered by the first device and to close it on a subsequent event from the second one.¹⁸ The gate could be closed on the very next event in the simplest case, or the N th following one could be used, which would measure the average time interval over N events. The gate connects a known high-frequency oscillator to a counter, and the time interval between the two events is thus measured in units of the period of this oscillator. The resolution of this method depends on the frequency of this oscillator and the speed of the counter, while the accuracy depends on a number of parameters including the latency in the gate hardware and any variations in the rise time of the input wave forms. The resolution can be improved by adding an analog interpolator to the digital counter, and a number of commercial devices use this method to achieve sub-nanosecond resolution without the need for a reference oscillator whose frequency is 1 GHz or greater.

In addition to these obvious limitations, time-difference

measurements using fast pulses have additional problems. Reflections from imperfectly terminated cables may distort the edge of a sharp pulse, and long cables may have enough shunt capacitance to round it by a significant amount. In addition to distorting the wave forms and affecting the trigger point of the discriminators, these reflections can alter the effective load impedance seen by the oscillator and pull it off frequency. Isolation and driver amplifiers are usually required to minimize the mutual interactions and complicated reflections that can occur when several devices must be connected to the same oscillator, and the delays through these amplifiers must be measured. These problems can be addressed by careful design, but it is quite difficult to construct a direct time-difference measurement system whose measurement noise does not degrade the time stability of a top-quality oscillator, and other methods have been developed for this reason. Averaging a number of closely spaced time-difference measurements is usually not of much help because these effects tend to be slowly varying systematic offsets, which change only slowly in time.

Many measurement techniques are based on some form of heterodyne system. The sine-wave output of the oscillator under test can be mixed with a second reference oscillator which has the same nominal frequency, and the much lower difference frequency can then be analyzed in a number of different ways. If the reference oscillator is loosely locked to the device under test, for example, then the variations in the phase of the beat frequency can be used to study the fast fluctuations in the frequency of the device under test. The error signal in the lock loop provides information on the longer-period fluctuations. The distinction between “fast” and “slow” would be set by the time constant of the lock loop.

This technique can be used to compare two oscillators by mixing a third reference oscillator with each of them and then analyzing the two difference frequencies using the time-interval counter discussed above. In the version of this idea developed at NIST,¹⁹ this third frequency is not an independent oscillator, but is derived from one of the input signals using a frequency synthesizer. The difference frequency has a nominal value of 10 Hz in this case. The time interval counter runs with an input frequency of 10 MHz and can therefore resolve a time interval of 10^{-6} of a cycle. This is equivalent to a time-interval measurement with a resolution of 0.2 ps at the 5 MHz input frequency.

All of these heterodyne methods share a common advantage: the effects of the inevitable time delays in the measurement system are made less significant by performing the measurement at a lower frequency where they make a much smaller fractional contribution to the periods of the signals under test. Furthermore, the resolution of the final time-interval counter is increased by the ratio of the input frequencies to the output difference frequencies (5 MHz to 10 Hz in the NIST system). This method does not obviate the need for careful design of the front-end electronics—if anything the increased resolution of the back-end measurement system places a heavier burden on the high-frequency portions of the circuits and the transmission systems. As an example, the stability of the National Institute of Standards and Technol-

ogy (NIST) system is only a few ps—about a factor of 10 or 20 poorer than its resolution.

Heterodyne methods are well suited to evaluating the frequency stability of an oscillator, but they often have problems in measuring time differences because they usually have an offset, which is an unknown number of cycles of the input frequencies. Thus the time difference between two clocks measured using the NIST mixer system is offset with respect to measurements made using a system based on the 1 Hz pulse hardware by an arbitrary number of periods of the 5 MHz input frequency (i.e., some multiple of 200 ns). To further complicate the problem, this offset can change if the power is interrupted or if the system stops for any other reason.

The offset between the two measurement systems must be measured initially, but it is not too difficult to recover it after a power failure, since the step must be an exact multiple of 200 ns. Using the last known time difference and frequency offset, the current time can be predicted using a simple linear extrapolation. This prediction is then compared with the current measurement, and the integer number of cycles is set (in the software of the measurement system) so that the prediction and the measurement agree. This constant is then used to correct all subsequent measurements. The lack of closure in this method is proportional to the frequency dispersion of the clock multiplied by the time interval since the last measurement cycle, and the procedure will unambiguously determine the proper integer if this time dispersion is significantly less than 200 ns. This criterion is easily satisfied for a rubidium standard if the time interval is less than a few hours; the corresponding time interval for cesium devices is generally at least a day.

Given the difficulties of making measurements using 1 Hz pulses, it is natural to consider abandoning them in favor of phase measurements of sine waves. The principal reason for not doing this now is that almost all of the methods that are used to distribute time are designed around 1 Hz ticks. As we will see below, extracting 1 Hz ticks from a GPS signal is a straightforward business in principle, while phase comparisons between a local 5 MHz oscillator and the same GPS signal have a number of awkward aspects and ambiguities. Very similar problems also make measurements on 1 Hz pulses the method of choice in two-way satellite time transfer.

VIII. NOISE AND MEASURES OF STABILITY

A. Introduction to the problem

We now turn to looking at ways of characterizing the performance of clocks and oscillators. The simple notions of Gaussian random variables will turn out to be inadequate for this job, because many of the noise processes in clocks are not even close to Gaussian. In particular, the common notions of mean and standard deviation (which completely characterize a Gaussian distribution) must be applied with great care—while their values usually exist in a formal sense, they do not have the nice properties that we usually think of when these parameters are used to characterize a true Gaussian variable. In particular, calculating either the mean or the

standard deviation of one of these data set results in estimates that are not stationary and that depend on the length of the time series used for the analysis.

When we speak of the time or frequency of a clock in the following discussion, we almost always mean *relative* values. Time is measured as the difference (in seconds) between the device under discussion and a second identical noiseless one, and frequency is the *fractional* frequency difference between the same two devices. If the nominal frequency of the devices is ν_0 , and if the phase difference between them at some instant is ϕ , then the fractional frequency difference, y , is given in terms of the time derivative of the phase by

$$y = \frac{\dot{\phi}}{2\pi\nu_0}. \quad (1)$$

Using this convention, frequencies are dimensionless quantities. We will also use the term frequency aging to denote slow changes in the fractional frequency (where slow means long with respect to $1/y$ and certainly with respect to the much smaller $1/\nu_0$). The unit of frequency aging is s^{-1} .

B. Separation of variance

Noiseless measuring systems and perfect clocks are scarce commodities, and we must use real devices in a test system and extract from these data the performance parameters of each of the devices. One of the most important aspects of this job is the concept of separation of variance—that is identifying which part of the system is responsible for the observed fluctuations in the data. This idea most often arises in two contexts:

- (a) We are trying to evaluate the performance of a device or system and we would like to be able to separate its performance from the noise due to the system that we are using as a reference. Both systems can be physical clocks or pseudo clocks, such as time scales. This job is easy if the reference is much less noisy than the device under test; if this is not the case then the “three-cornered hat” method can be used.²⁰

Suppose we have three devices whose individual variances are σ_i^2 , σ_j^2 , and σ_k^2 . We wish to estimate these parameters, but we can only compare the devices to each other using pair-wise comparisons. The measured variances of these pair-wise comparisons are σ_{ij}^2 , σ_{jk}^2 , and σ_{ki}^2 . If the variances of the three devices are roughly equal, then we can estimate the three individual variances by

$$\begin{aligned} \sigma_i^2 &= \frac{\sigma_{ij}^2 + \sigma_{ki}^2 - \sigma_{jk}^2}{2}, \\ \sigma_j^2 &= \frac{\sigma_{jk}^2 + \sigma_{ij}^2 - \sigma_{ki}^2}{2}, \\ \sigma_k^2 &= \frac{\sigma_{jk}^2 + \sigma_{ki}^2 - \sigma_{ij}^2}{2}. \end{aligned} \quad (2)$$

These estimates can be misleading if the variances are

correlated—the variances estimated in this way can even be negative in this case despite the fact that this is impossible on physical grounds.

- (b) We are trying to evaluate (or synchronize) a device using data that we receive from a distant calibration source over a noisy channel. Even though the reference device may be much more stable than the device under test, its data are degraded by the channel noise.

It would be very helpful to be able to separate the contributions to the variance due to the channel and the local clock in this configuration. This separation is important for two reasons. In the first place, it tells us which part of the overall system is limiting its overall performance, and where we should spend time and effort to improve things. In the second place, separation of variance plays a central role in designing methods for synchronizing clocks when the calibration data are degraded in this way.

This situation arises in a number of difference contexts. We discuss later two common situations: synchronizing local clocks to GPS signals which are degraded by selective availability and other effects and synchronizing clocks using time signals transmitted over packet networks such as the Internet.

C. Measurement noise

All of the instruments that are used for time and frequency measurements have some kind of discriminator at the input—something that is triggered when the input wave form crosses some preset threshold. Any noise on the input signal combines with the internal noise of the circuit to produce a jitter in the trigger point. These effects will produce a fluctuation in the measured time difference between two clocks that has nothing to do with the clocks themselves—it arises purely in the measurement process and there is no corresponding jitter in the underlying frequency difference between the two devices. It is especially important that we develop techniques that can identify the source of these fluctuations in the data in applications where the goal is to synchronize the local clock to a second standard. Adjusting the frequency of the local clock to compensate for this measured time jitter is almost always the wrong thing to do, as we will see below.

If the input data are a sine wave with amplitude A and frequency ω , then additive noise at the input to the discriminator of amplitude V_n will result in a time jitter whose amplitude is approximately

$$\Delta t \cong \frac{V_n}{A\omega}. \quad (3)$$

We usually assume that in each discriminator the equivalent noise voltage at the input is not correlated with the corresponding input signal, so that the time jitter in each channel has no mean offset relative to the corresponding noise-free trigger point. Consecutive measurements of the time difference are then randomly distributed about a mean value that represents the actual time difference between the two clocks. If the instantaneous noise-induced offset in the trigger point

at some epoch t is $\epsilon(t)$, then the independence between the signal and the noise in each channel implies that

$$\langle \epsilon(t)\epsilon(t+\tau) \rangle = \langle \epsilon^2(t) \rangle \delta(\tau). \quad (4)$$

Since the autocorrelation of $\epsilon(t)$ is a delta function, its power spectrum is constant for all frequencies, and measurement noise of this type is often called white phase noise for this reason. (Although white phase noise is often thought of as a Gaussian variable, this is not necessarily true. While the properties described above are *necessary* for the statistical distribution of the phase noise to be Gaussian, they are not *sufficient* to guarantee this. Nevertheless, although the individual noise-induced offsets may or may not be Gaussian, consecutive estimates of their mean will exhibit a Gaussian distribution as a result of the central limit theorem of statistics.)

Note that Eq. (4) is only an approximation—we cannot make τ either very small or very large and so the autocorrelation is in fact bounded at small times by the finite bandwidth of the measurement system and at long times by the length of the experiment. A real-world spectrum will therefore be not quite “white” for very short or very long measurement intervals but this is usually not important. The finite response time of the overall system limits the importance of very high-frequency noise, and other sources of noise always become important at longer averaging times.

If the white phase noise is superimposed on a constant frequency offset between the two devices, then consecutive measured time differences will scatter uniformly about a line whose slope is given by this offset frequency. Because of this uniform scatter, the frequency difference between the two oscillators can be estimated using the usual least-squares machinery. Although the white phase noise will contribute to the uncertainty of this estimate, it will not introduce a bias—frequency estimates computed from different blocks of data will scatter uniformly about the true frequency offset between the two devices. The statistical distribution of these frequency estimates can be calculated using the standard methods of propagation of error.

D. Frequency noise processes

Although white phase noise arises in the measurement process and is not due to the oscillator itself, there are also noise sources within the oscillator that we must consider. As we discussed above, an atomic clock (or any other oscillator for that matter), is comprised of two basic components: a frequency reference (such as an atomic transition or a vibrational mode in a piece of quartz) and an oscillator that interrogates this reference frequency and is locked to it by means of some kind of feedback loop. As with all circuits, there will inevitably be some noise in this feedback loop (due to shot noise in the number of atoms that are interrogated or Johnson noise in the circuit resistors), and this noise will produce jitter in the lock point of the oscillator. In the best of all possible cases, this noise will not be correlated with the signals in the feedback loop. Using exactly the same argument as above, this noise will produce jitter in the output fre-

quency of the device whose autocorrelation function is a delta function. By analogy with the argument above, we would call this white frequency noise.

The frequency fluctuations in the oscillator result in time fluctuations as well. If we started the device at some time $-T$ in the past, and if the instantaneous frequency at any time is $y(\tau)$, then the time difference at some epoch t between this oscillator and a second noiseless but otherwise identical device would be given by

$$X(t) = \int_{-T}^t y(\tau) d\tau. \quad (5)$$

The integration of the instantaneous frequency offset to produce a time difference has an important consequence for the power spectrum of the resulting time fluctuations. If the fluctuations in $y(\tau)$ are dominated by white frequency noise, then its spectral density is a constant at all frequencies (keeping in mind the comment above that the infinities implied by this formal description are not a problem because of the finite limits both in time and in frequency of a real-world situation). That is, the Fourier expansion of $y(\tau)$, the frequency offset of the oscillator as a function of time, is given in terms of the Fourier frequency Ω by

$$y(\tau) = \int Y(\Omega) e^{i\Omega\tau} d\Omega, \quad (6)$$

where

$$|Y(\Omega)| = C \quad (7)$$

and C is a constant. Keep in mind that the variable y represents the real fractional frequency difference between two nominally identical physical oscillators. Unfortunately, this quantity is not a constant. The expression $y(\tau)$ expresses the variation in y in the time domain as a function of epoch τ , while $Y(\Omega)$ expresses this variation in the Fourier frequency domain as a function of Fourier frequency Ω . Also note that Eq. (7) cannot be literally correct for a true noise process. We will discuss this point in more detail below.

Using Eqs. (5), (6), and (7), we can see that the power spectral density of $X(t)$ is not white but rather varies as $1/\Omega^2$ with an amplitude determined by C in Eq. (7). Thus there is a fundamental difference between white phase noise and white frequency noise. The former results in time-difference fluctuations that still have a white spectrum, while the integration implied by the relationship between time and frequency causes white frequency noise to have a “red” Fourier-frequency dependence when its effect on time-difference measurements is considered. This is an important conclusion because it means that we can distinguish between frequency noise and phase noise by examining the shape of the Fourier spectrum of the time-difference data. (The form of the dependence on Fourier frequency distinguishes the two cases—not the magnitude of the spectral density itself.)

Unlike the white phase noise case, if there is a deterministic frequency offset between two devices whose noise spectra are characterized by white frequency noise, it is no longer true that the time differences between the devices will scatter uniformly about a simple straight line determined from this offset. A least-squares fit of a straight line to these time

differences will not result in a statistically robust estimate of this frequency offset for this reason. (We mean by this that the mean difference between these estimates and the underlying deterministic frequency offset is zero, *and* that the mean and variance are characteristics of the data set and are independent of the sample size or of which portion of the data is used for computing the estimate.) However, consecutive frequency estimates (i.e., the first difference of the time differences normalized by the time interval between them) are distributed uniformly about the mean frequency offset, and the average of these first differences will produce a statistically robust estimate of the deterministic frequency offset with a well-defined variance.

These results illustrate a more general conclusion, namely that there is no single optimum method for estimating clock parameters. The best strategy depends on the underlying noise type. Unfortunately, there is no simple optimum strategy for some types of noise (flicker processes, for example), and the best we can do is an approximate analysis whose predictions are not statistically robust.

It is clear that these ideas could be extended to higher-order derivatives. If we had an oscillator in which the frequency *aging*, $d(\tau)$, had uncorrelated fluctuations about a (possibly nonzero) mean value, then, by analogy with Eq. (5), the frequency at any time would be given by

$$y(t) = y(0) + \int d(\tau) d\tau. \quad (8)$$

Using the same argument as above, this white frequency aging would result in fluctuations in the frequency difference between this clock and an identical noiseless one that varied as $1/\Omega^2$ relative to the spectral density of the aging itself. The time differences between the two devices would have a spectrum that varied as $1/\Omega^4$ relative to the aging, since the aging must be integrated once to compute the frequency and a second time to compute the time difference. Again, we could distinguish the different contributions by examining the shape of the Fourier spectrum of the time differences.

Using the same sort of argument as above, when the noise spectrum of the device is dominated by white frequency aging, we cannot make a statistically robust estimate of the frequency of a device by averaging the consecutive frequency estimates, since those estimates are no longer distributed uniformly about a mean value. The best we can do now is to estimate a mean value for the frequency aging.

The previous discussion used the power spectrum of the time differences as the benchmark, but it is also common in the literature to find a discussion in which the relative dependencies on Fourier frequency are described with respect to the power spectrum of the frequency fluctuations. The relationship between frequency and phase specified in Eq. (5) is unchanged, of course—only the benchmark spectrum is different. White phase noise would then be characterized as having a relative dependence proportional to the square of the Fourier frequency in this case. A white frequency-aging process would be described as having a relative dependence on Fourier frequency of $1/\Omega^2$, where both of these spectra would be evaluated relative to the power spectrum of the frequency fluctuations.

Since the aging and the frequency specify the time evolution of the frequency and the time, respectively, white frequency aging is sometimes called random-walk frequency noise in the literature and white frequency noise is sometimes called random-walk phase noise.

There is another way of thinking of the differences among the types of noise processes. If a time-difference measurement process can be characterized as being limited by pure white phase noise, then there is an underlying time difference between the two devices. The measurements scatter about the true time difference, but the distribution of the measurements (or at worst the mean of a group of them) can be characterized by a simple Gaussian distribution with only two parameters: a mean and a standard deviation. We can improve our estimate of the mean time difference by averaging more and more observations, and this improvement can continue forever in principle. There is no optimum averaging time in this simple situation—the more data we are prepared to average the better our estimate of the mean time difference will be. As we discussed above, a least-squares fit of a straight line to these time differences can be used to compute a statistically robust estimate of the frequency difference between the two devices.

The situation is fundamentally different for a measurement in which one of the contributing clocks is dominated by zero-mean white frequency noise. Now it is the frequency that can be characterized (at least approximately) by a single parameter—the standard deviation. If the time difference between two identical devices at some epoch is $X(t)$, we would estimate the time difference a short time in the future using

$$X(t + \tau) = X(t) + y(t)\tau, \quad (9)$$

where $y(t)$ is the instantaneous value of the white frequency noise, and

$$\langle y(t) \rangle = 0, \quad (10)$$

$$\langle y^2(t) \rangle = \epsilon^2. \quad (11)$$

Since $y(t)$ has a mean of 0 by assumption, the time difference at the next instant is distributed uniformly about the current value of X , and the mean value of $X(t + \tau)$ is clearly $X(t)$. In other words, for a clock whose performance is dominated by zero-mean white frequency noise, the optimum prediction of the next measurement is exactly the current measurement with no averaging. Note that this does not mean that our prediction is that

$$X(t + \tau) = X(t), \quad (12)$$

but rather that

$$\langle X(t + \tau) - X(t) \rangle = 0, \quad (13)$$

which is a much weaker statement because it does not mean that our prediction will be correct, but only that it will be unbiased on the average. This is, of course, the best that we can do under the circumstances. The frequencies in consecutive time intervals are uncorrelated with each other by definition, and no amount of past history will help us to predict what will happen next. The point is that this is the opposite

extreme from the discussion above for white phase noise, where the optimum estimate of the time difference was obtained with infinite averaging of older data.

Clearly, there must be an intermediate case between white phase noise and white frequency noise, where some amount averaging would be the optimum strategy, and this domain is called the “flicker” domain. Physically speaking, the oscillator frequency has a finite memory in this domain. Although the frequency of the oscillator is still distributed uniformly about a mean value of 0, consecutive values of the frequency are not independent of each other and time differences over sufficiently short times are correlated. Both the frequency and the time differences have a short-term “smoothness” that is not characteristic of a true random variable. Flicker phase noise is intermediate between white phase noise and white frequency noise, and we would therefore expect that the spectral density of the time difference data would vary as the inverse of the Fourier frequency relative to the spectral density of the phase fluctuations themselves.

The same kind of discussion can be used to define a flicker frequency noise that is midway between white frequency noise and white aging (or random walk of frequency). The underlying physical effect is the same, except that now it is the frequency aging that has short-period correlations. We could think of a flicker process as resulting from a very large number of very small jumps—not much happens in the short term because the individual jumps are very small, but the integral of them eventually produces a significant effect. The memory of the process is then related to this integration time.

Data that are dominated by flicker-type processes are difficult to analyze. They appear deterministic over short periods of time, and there is a temptation to try to treat them as white noise combined with a deterministic signal—a strategy that fails once the coherence time is reached. On the other hand, they are not quite noise either—the correlation between consecutive measurements provides useful information over relatively short time intervals, and short data sets can be well characterized using standard statistical measures. However, the variance at longer periods is much larger than the magnitude expected based on the short-period standard deviation.

The finite-length averages that we have discussed can be realized using a sliding window on the input data set. This is simple in principle but requires that previous input data be stored; an alternative way of realizing essentially the same transfer function is to use a recursive filter on the output values. This method is often used in time scales, since these algorithms are commonly cast recursively. Both the recursive and nonrecursive methods could be used to realize averages with more complicated transfer functions. These methods are often used in the analysis of data in the time domain.²¹

E. Two-sample Allan variance

The ability to predict the future performance of a clock based on past measurements can be quantified by consider-

ing the two-sample or Allan variance. Suppose we have measured the time difference between two clocks at times t_1 and $t_2 = t_1 + \tau$. If we want to predict the time difference at a time $t_3 = t_2 + \tau$, then we would use the two measurements to estimate the frequency difference between the two clocks during the first interval

$$y_{12} = \frac{X(t_2) - X(t_1)}{\tau} \tag{14}$$

and we would combine this estimate with the measurement at t_2 to estimate the value at time t_3 :

$$\hat{X}(t_3) = X(t_2) + y_{12}\tau = 2X(t_2) - X(t_1), \tag{15}$$

where we have assumed (for lack of any better choice) that the frequency in the interval between t_2 and t_3 is the same as the value in the previous interval. If this is not the case, the prediction error will be proportional to the difference between the actual frequency in the interval between t_2 and t_3 , which is y_{23} , and our estimate of it, which is that it was the same as the average frequency over the previous interval

$$\epsilon = X(t_3) - \hat{X}(t_3) \approx y_{23} - y_{12}. \tag{16}$$

Expressed in terms of time difference measurements, the prediction error would be proportional to

$$\frac{X(t_3) - 2X(t_2) + X(t_1)}{\tau} \tag{17}$$

and the mean-square value of this quantity, usually designated as $\sigma_y^2(\tau)$, is called the two-sample (or Allan) variance for an averaging time of τ . (The variance is actually defined as one half of this mean-square value so that it will be consistent with other estimators when the input data are characterized by white phase noise.)

If the frequency of the oscillator is absolutely constant over these time intervals, then the expressions in Eqs. (16) and (17) are nonzero only because of the white phase noise that contributes to the measurements of X [and indirectly through Eq. (14) to our estimates of y]. The numerators of Eqs. (14) or (17) are independent of τ in this case so that the Allan variance for measurements dominated by white phase noise will vary as $1/\tau^2$.

If the frequency of the oscillator varies with time, then the numerator in Eq. (14) will have a dependence on τ . At least when τ is sufficiently small, this dependence must contain only non-negative powers of τ , because it must be that

$$X(t + \tau) - X(t) \rightarrow 0 \quad \text{as} \quad \tau \rightarrow 0. \tag{18}$$

The Allan variance must therefore decrease more slowly than $1/\tau^2$ in this case and may actually increase with τ if the expansion of the numerator in Eq. (17) contains a sufficiently large power of τ .

The Allan variance measures predictability in the rather narrow sense of Eq. (15). However, it does not make a clear distinction between oscillators that have stochastic frequency noise and those that may have deterministic frequency variation (a constant frequency aging, for example), even though such a deterministic parameter might be included in a modified version of Eq. (15). It also does not give any information

TABLE I. Summary of noise types.

Name	Slope of spectral density		Slope of two-sample variance		
	Phase	Frequency	Simple	Modified	Time
White phase noise	0	Ω^2	-2	-3	-1
Flicker phase noise	$1/\Omega$	Ω	-2	-2	0
White frequency noise (Random-walk phase)	$1/\Omega^2$	0	-1	-1	1
Flicker frequency noise	$1/\Omega^3$	$1/\Omega$	0	0	2
White frequency aging (Random-walk frequency)	$1/\Omega^4$	$1/\Omega^2$	1	1	3

about frequency accuracy—arbitrarily large frequency offsets will make no contribution to Eqs. (16) or (17).

A constant frequency offset will result in time differences that increase linearly with time, and a constant frequency aging results in a quadratic dependence of the time differences. It is tempting to remove either (or both) of these effects using standard least-squares methods before computing the Allan variance. As we have discussed above, least-squares estimators may produce biased estimates if they are used on data containing appreciable variance due to random-walk or flicker processes; the resulting Allan variance estimates are likely to be too small as a result.

F. Relationship between the Fourier decomposition and the Allan variance

The Allan variance and Fourier decomposition views of the noise process are always valid in principle, but they are particularly useful if the measurements are dominated by a single noise process, since both parameters have particularly simple forms in these special cases. Specifically, if the slope of the power spectral density as a function of frequency on a log-log plot has a slope of α , that is if the spectral density of the frequency fluctuations is such that

$$S_y(\Omega) = h_\alpha \Omega^\alpha, \quad (19)$$

where h_α is a constant scale factor and α is an integer such that

$$-2 \leq \alpha \leq 1, \quad (20)$$

then the two-sample ordinary Allan variance we have defined above [as one half of the mean square of Eq. (17)], has a slope of μ when plotted as a function of averaging time on a log-log plot. That is,

$$\sigma_y^2(\tau) = b_\mu \tau^\mu \quad (21)$$

with

$$\mu = -\alpha - 1. \quad (22)$$

Measuring the slope of the Allan variance as a function of averaging time therefore determines the slope of the spectral density and the noise type. This is a substantial advantage, because computing the Allan variance is usually a much simpler job than computing the full power spectrum. This procedure is most useful when the spectrum is dominated by a single noise type in each range of Fourier frequency and averaging time, which is quite often the case for many clocks and oscillators. Both the Allan variance and the power spec-

tral density can be approximated by a series of straight-line segments on a log-log plot in this case, and the approximation that only a single type of noise dominates the spectrum in each one of the corresponding Fourier frequency/averaging time domains is well justified in practice. We have already mentioned the five most common noise processes in clock data, and their properties are summarized in Table I.

This relationship between Allan variance, spectral density, and noise type has three important limitations. The first is that the slope of the simple Allan variance that we have defined above cannot distinguish between flicker phase modulation and white phase modulation—in both cases the slope is $1/\tau^2$. This defect can be remedied by defining a modified Allan variance²² (MOD AVAR) which replaces the time differences in Eq. (17) [or the frequencies of Eq. (16)] with averages over a number of adjacent intervals. The principal advantage of this modification is that it provides a way of distinguishing between white phase noise and flicker phase noise, as is shown in Table I. A relative of the modified Allan variance is the time variance, usually called “TVAR” and designated $\sigma_x^2(\tau)$.²³ It is an estimate of the time dispersion resulting from the corresponding frequency variance, and it is normalized so as to be consistent with the conventional definition of the variance when the data are dominated by white phase noise. It is defined by

$$\sigma_x^2 = \frac{\tau^2}{3} \text{mod } \sigma_y^2(\tau). \quad (23)$$

It has the same statistical properties as the modified Allan variance, but it is somewhat easier to interpret in applications where time dispersion rather than frequency dispersion is of primary concern. Using TVAR, it is also somewhat easier to identify the onset of the domain in which white frequency noise dominates the spectrum by simply looking at the plot of the variance as a function of averaging time. This is because it is easier to see the point at which the slope of TVAR changes from 0 to +1 than it is to identify the point at which the slope of MOD AVAR changes from -2 to -1 (see Table I).

The second limitation is that the relationship between the Allan variance and the slope of the power spectral density depends on the fact that the input data are really noise and do not have any coherent variation. The most common exception to this assumption is the case of deterministic frequency aging—the relatively constant drift in the frequency of hydrogen masers, for example, or a nearly diurnal frequency

fluctuation in quartz-crystal oscillators driven by changes in the ambient temperature. Both the Allan variance and the spectral density function are well defined in these cases, but the simple relationships we have described are no longer valid. The spectral density function of a periodic signal is, of course, not a smoothly varying function but a peak at the corresponding frequency, while the Allan variance oscillates at the period of the perturbation and no longer has a simple monotonic slope either.

The third limitation is that the variance is a statistical parameter, and it does not give any information on worst-case performance. This problem is not unique to the Allan variance, of course—all statistical estimators share this failing. It is possible, however, to define an uncertainty in an Allan variance estimate,^{24,25} while this does not speak to worst-case performance, it does provide some information on the width of the distribution function—that is on the spread that might be expected from measurements on an ensemble of identical oscillators. It is also important to remember that both mod avar and tvar are computed using an average over adjacent frequency estimates, and the resulting magnitudes are not the same as what would be obtained from the same device measured without that averaging.

G. Uncertainty of the Allan variance

The uncertainty in the Allan variance estimate for a given averaging time is proportional to the number of differences that contribute to it. If the entire data set spans a time T , for example, then an estimate using a time difference of τ will be comprised of about $T/2\tau$ estimates. This number clearly decreases with averaging time, so that the uncertainty in each estimate must increase as the averaging time increases. In the limit we will have only one estimate for an averaging time of $\tau=T/2$, and we will have almost no information at all about its uncertainty as a result. To make matters worse, consecutive estimates of the frequency have a correlation that depends on the noise type, so that simple-minded estimates of the number of degrees of freedom can be misleading even for small averaging times.

The summation of Eq. (17) that is required to compute the Allan variance can be considered to be a convolution of the input data set $X(t_j)$ with the series 1, -2 , 1, 0, If the series of time differences are dominated by noise processes (rather than deterministic effects) then it is likely that the series will be at least approximately stationary. If this is the case, then we can replace the ordinary convolution with a circular one,²⁶ and this change removes the difficulty in the calculation for large time differences. If the data are not stationary initially, then it may be possible to pre-whiten them using digital filter techniques or by removing a low-order polynomial first.²⁷

H. Uncertainty using Fourier decomposition

The uncertainties in Fourier frequency estimates of noise processes are discussed in the literature.²⁸ Although Eq. (7) (and other Fourier estimates of noisy series) accurately characterizes the power spectral density of white noise in the mean, there are significant uncertainties in the estimates. In

other words, the Fourier decomposition of a white noise process is not a smooth straight line but is itself a random variable whose statistical properties can be derived from the original time series. (Although more sophisticated analyses are usually used, the spectral densities are linear combinations of the time series data, and the standard methods of propagation of error can be applied to estimate the variance of the spectral density.)

It is common practice to reduce the variance in the Fourier coefficients by averaging a number of adjacent values. The simplest method is a straight average which effectively weights the adjacent points equally, but more complicated “triangular” functions (such as 1/6, 2/3, 1/6 or 0.23, 0.54, 0.23) are also used. These averages provide a trade-off of stability against frequency resolution. The practice of averaging adjacent estimates using a sliding window is mathematically equivalent to convolving the Fourier coefficients with the square or triangular-shaped function derived from the weights. This convolution in turn is equivalent to multiplying the original time series by a “windowing” function that varies smoothly from a maximum at the midpoint of the data set to zero at the end points.

The choice of windowing function is particularly important when analyzing time series that have significant power at discrete frequencies, since the windowing process can affect both the magnitude and shape of these “bright lines.” More generally, the windowing operation can cause ambiguities when the width of the windowing function (in frequency space) contains statistically significant variation in the power spectral density of the data, since quantitative estimates of the spectral density will depend on the shape of the window in this case.²⁹ Analyzing a data set that contains a number of discrete spectral components requires considerable care in the choice of a windowing function because a poor choice may result in frequency aliases—spurious peaks in the spectral density corresponding to frequencies that are not really present in the time series. It is common practice to remove such bright lines (using simple least squares, for example) before estimating the spectral density. This process is called “pre-whitening.”

I. Other noise estimators

A number of applications require an estimator that provides some measure of the worst-case (rather than the average) performance of an oscillator. The most common application is the use of oscillators to synchronize the multiplexors in communications systems. This hardware is used to combine data from several circuits into one higher-speed data stream using time-division multiplexing (in which each input circuit is given a periodic time slice). This combined stream is de-multiplexed at the other end of the line using receiving hardware that routes the data to the appropriate output line based on the same time-slice logic. A certain amount of frequency difference between the hardware at the end points can be accommodated by “slip” buffers, but these have finite capacity, and the *maximum* time dispersion and frequency difference between the two end stations must be controlled to prevent overflowing these buffers.³⁰

This requirement leads naturally to the concept of MTIE—the maximum peak-to-peak time-interval error in some observation time, T .³¹ A number of different versions of MTIE exist—in some cases the reference for the measurement is the input frequency to the network element and in other cases it is an external standard reference frequency.

In addition to specifying T , measurements of MTIE usually also specify some observation period for the measurements of the time dispersion. This is often implemented as a sliding window of width $\tau < T$. The value of MTIE for the complete data set is the maximum of the values for each of the windows.

The variation of MTIE with τ can provide some insight into the character of the frequency offset between the two clocks. A static frequency offset, for example, produces a value of MTIE that increases strictly linearly with τ . The converse is not necessarily true, however. Random-walk processes can produce almost the same effect.

Another estimator that is often used in telecommunications work is ZTIE. This estimator uses averages of the first difference of the time data [Eq. (14)] instead of the second differences [Eq. (17)] as in the Allan variance. As with MTIE, the value of the estimator is the maximum value of the statistic over the averaging window.

Both ZTIE and MTIE are sensitive to static frequency offsets, whereas AVAR (in all its variations) is not. A clock with an arbitrarily large and constant frequency offset is a “good” clock by AVAR standards—while its time dispersion will eventually become very large, its performance is *predictable*. The same clock is not nearly so useful by MTIE or ZTIE standards because its static frequency offset will produce repeated overflows of the slip buffers when it is used as the reference for a communications switch. The fact that these overflows are regular and predictable is beside the point. In addition, MTIE is more sensitive to transients—it will give a better estimate of the worst-case performance because of this, but it can also become saturated by a single event that will mask its response to the remainder of the data.

IX. REPORTING THE UNCERTAINTY OF A MEASUREMENT

In 1978, the International Committee on Weights and Measures (CIPM) requested that the BIPM address the problem of the lack of an international consensus on the expression of uncertainties in measurements. The BIPM convened a working group, which produces a number of recommendations. These were approved by the CIPM in 1981³² and were reaffirmed in 1986.³³ These recommendations have been published as the “Guide to the Expression of Uncertainty in Measurement.”³⁴

The guide divides uncertainties into two categories: type A uncertainties, in which the method of evaluation uses statistical analysis of a series of observations, and type B uncertainties, in which the method of evaluation is by means other than statistical analysis. Note that there is no direct correspondence with the terms “random” and “systematic” errors.

Type A uncertainties are usually specified using a vari-

ance (or its square root, the standard deviation) combined with the number of degrees of freedom in the estimate. If needed, covariances are also specified. These terms have their conventional statistical definitions.

Type B uncertainties are also characterized by a variance (or its square root, the standard deviation). The variances are approximations to the corresponding statistical parameters, the existence of which is assumed. If needed, covariances are included in the same way.

The combined uncertainty is computed in the usual way as the simple sum of the variances, and the combined standard deviation is the square of the sum of the squares of the contributing standard deviations. (The guide recognizes that in some applications it may be appropriate to multiply the combined uncertainty by an additional “coverage” factor; this factor must be specified when it is used.)

The procedure of simply summing the variances to compute the combined uncertainty is justified if the contributions are statistically independent of each other—an assumption that is easy to make but hard to prove. In principle, it is possible to deal with data where the observations are not statistically independent of each other, but this is difficult in practice because the correlation coefficients are usually not well known. Furthermore (as we discussed above), summing variances (or averaging data) that are dominated by flicker or random-walk noise processes usually does not produce estimates that are statistically robust. As a specific case in point, while adding more and better clocks to EAL certainly helps matters, the accuracy of TAI is still evaluated by using data from primary frequency standards.

If the combined uncertainty is computed as the simple sum of variances of all types then identifying a particular contributor as being of type “A” or “B” does not affect the final result, since only the magnitude of each contribution is significant. However, it is useful in that it makes it easier to understand how the result was obtained. Many time and frequency applications depend more on the stability (or predictability) of a device or on its frequency accuracy. In these applications the question of type A versus type B is less important than the magnitude of the Allan variance (for stability/predictability) or MTIE (for accuracy).

X. TRANSMITTING TIME AND FREQUENCY INFORMATION

We now turn to a discussion of how time and frequency information is transmitted. Distribution systems based on relatively simple radio broadcasts were adequate in the early days, because the clocks themselves were not very good. The transit time of an electromagnetic signal is about $3 \mu\text{s}/\text{km}$; in the “old days” errors in estimating this delay were usually not significant compared to the time dispersion of the clocks themselves.

Short-wave signals from radio stations such as WWV and low frequency (60 kHz) signals from stations such as WWVB continue to be widely used for transmitting time and frequency information. (A complete list of stations that transmit time and frequency information is contained in the Annual Report of the BIPM.³⁵) However, the uncertainties in

the one-way propagation delay of either signal can be significant—perhaps as much as tens of milliseconds for WWV and a few milliseconds for WWVB. Short-wave signals tend to travel as both a ground wave and a sky wave, which is reflected from the ionosphere (possibly multiple times). The interference between these two signals and the variations in the height of the ionospheric reflecting layer result in significant variation in the path delay. These problems are less serious for lower-frequency signals, but even these signals show a strong diurnal variation in the path delay, which tends to vary most rapidly near sunrise and sunset. This diurnal variation can be almost completely removed by averaging the observations for 24 h (leaving a much smaller uncertainty on the order of microseconds) but many oscillators have too much flicker or random-walk frequency noise to support averaging for this long a time. Even with this level of averaging, the uncertainties are too large for more demanding applications, and other techniques have been developed to meet these needs. Roughly speaking, the design goal for any distribution system is to transmit the data to a user without degrading the accuracy or stability of the source. This goal is not easy to meet now, and it will grow even more difficult in the future as clocks continue to improve.

Although time and frequency are closely related quantities, there are important differences between them, and these differences affect how they are distributed and what uncertainties are associated with these processes. Time distribution is, of course, directly affected by the delay in the transmission channel between the clock and the user, and uncertainties in this delay enter directly into the error budget. A frequency, on the other hand, is a time interval (rather than an absolute time) so that the uncertainty in its distribution is limited more by temporal fluctuations in the transmission delay rather than by its absolute magnitude. In other words, the time delay of a channel must be accurately known if that channel is to carry time information, while the delay need only be stable if frequency information is to be transmitted. This distinction is not of purely academic interest; many common channels have large delays that are difficult to measure but which change slowly enough with time that they may be thought of as constant for appreciable periods; such channels are clearly better suited to frequency-based applications.

The simplest way of dealing with the transmission delay is to ignore it, and many users with modest timing requirements do just that. Many users of the time signals broadcast by short-wave stations (NIST radio stations WWV and WWVH, for example), fall into this group. A somewhat more sophisticated strategy is to approximate the actual delay using some average value—perhaps derived from some simple parameter characterizing the distance between the user and the transmitter.³⁶ Users of the signal from NIST radio station WWVB, for example, often do not measure the transmission delay, but rather model it based on average atmospheric parameters and the physical distance between themselves and the transmitter. (Many users of WWVB are interested in frequency calibrations and do not model the delay at all. Instead they use the fact that the variation in the

delay is dominated by a diurnal effect that can be canceled almost perfectly by averaging the data for 24 h.³⁷ This is an important example of the difference in the requirements for time and frequency distribution. Although the uncertainty in the transit time of a WWVB signal might be as large as 1 ms, the variation in the delay averaged over 24 h is probably almost 1000 times smaller than this value.)

The relatively simple models used to correct transmissions from high-frequency radio stations (like WWV) are not very accurate, and they are only adequate for users with very modest timing accuracy requirements on the order of 1 ms. This is a result of the inadequacy of the models, however, and is not a fundamental limitation of the technique. Transmissions from the global positioning system (GPS) satellites, for example, can be corrected to much higher accuracy using much more complex models of the satellite orbit and the ionospheric and atmospheric delays.³⁸ In addition, multiple-wavelength methods (to be described below), can also be used to estimate the delay through dispersive media.

The fundamental limitation to modeled corrections is often due to the fact that the adequacy of the model cannot be evaluated independently of the transmission itself. The GPS system is a notable exception because of the redundancy provided by being able to observe several (almost always at least 4 and sometimes as many as 12) satellites simultaneously, and this technique is especially powerful when they are located at different azimuths with respect to the antenna.

When modeling is not sufficiently accurate to estimate the path delay, users must measure the delay with an uncertainty small enough for the task at hand. There are a number of methods that can be used for this task, and we will describe them in the following sections. Many applications use more than one of them to estimate different components of the delay. The full accuracy of time transfer between a global positioning system satellite and an earth station, for example, can only be realized by combining a model of the tropospheric delay with a measurement of the delay through the ionosphere using multiple wavelength dispersion. Using GPS signals to transfer time between two earth stations often benefits from common-view subtraction as well.

A. Delay measurements using portable clocks

The simplest method for measuring the time delay along a path is to synchronize a clock to the transmitter and then to carry it to the far end of the path where its time is compared to the signals sent from the transmitter over the channel to be calibrated. Many timing laboratories used this technique in the past. This method is limited in practice by the finite frequency stability of the portable clock while it is being transported along the path and by uncertainties in a number of relativistic corrections that may have to be applied. These corrections will probably be needed if the transport velocity is high enough that time dilation must be considered, if the path is long enough that the Sagnac correction is important, or if points along the path are at different heights above the geoid so that gravitational frequency shifts must be included. While this method can be used to calibrate a static transmission delay, it is too slow and cumbersome to evaluate the temporal fluctuations in the delay through real transmission

media (due to changes in temperature or other environmental parameters, for example).

Even though various electrical or electromagnetic distribution methods are usually faster and more convenient, the portable clock method remains quite useful because it tends to be affected by a very different set of systematic uncertainties than the other methods to be described below, and it may provide estimates of these uncertainties that are difficult to achieve in any other way.³⁹

B. Two-way time transfer

A second method for measuring delay is two-way time-transfer, in which signals are sent in both directions along the path (the traditional “Einstein” clock synchronization method). If the path is reciprocal, the one-way delay is estimated as one-half of the round-trip transit time. Real paths are unlikely to be perfectly reciprocal—there is likely to be a small lack of reciprocity even in the atmosphere⁴⁰ and the receiver and transmitter at each of the end points are unlikely to be perfectly balanced.⁴¹ Even if this balance between incoming and outgoing delays at each station could be achieved, the delays through the different components are unlikely to have identical temperature coefficients so that the balance cannot be maintained in normal operations without repeated re-calibrations.⁴²

Two-way time transfer can be further characterized by how transmissions in the two directions are related. A half-duplex system uses a single transmission path that is “turned around” to transmit in the reverse direction. A full-duplex system, on the other hand, transmits in both directions simultaneously using either two nominally identical unidirectional channels or some form of multiplexing on a single bidirectional path. All of these arrangements have advantages and drawbacks. The half-duplex system is more likely to be reciprocal since the same path is used in both directions, but that advantage is partially offset by the fact that fluctuations in the delay may degrade the time delay measurement—especially if these fluctuations have periods on the order of the turn-around time. A full-duplex system is not degraded as much in this way since the measurements in the two directions are more nearly simultaneous, but incoherent fluctuations in the delays of the two unidirectional channels (or the cross talk between the two counter-propagating signals in the case of a single, bidirectional channel) may be an equally serious limitation.

Two-way methods can be used in many different hardware environments. Kihara and Imaoka⁴³ realized frequency synchronization of better than 10^{-13} and time synchronization of better than 100 ns using the two-way method in a 2000 km synchronous digital hierarchy based on optical fibers. Even higher accuracy has been realized using communications satellites to transfer time between timing laboratories⁴⁴ and using dedicated optical-fiber circuits.⁴⁵

The same two-way method is used in the NIST automated computer time service (ACTS)⁴⁶ to transmit time information using standard dial-up telephone lines with an uncertainty of a few milliseconds. Although the method is the same as that used in the previous cases, the timing uncer-

tainty is much greater because of variations in the reciprocity of the path—that is, in the lack of equality of the delays over the inbound and outbound circuits. Note that jitter in the overall delay itself does not produce a problem if it does not affect the reciprocity, but it may be difficult to exploit this fact in a real-time system, which often uses the measured delay of a previous round-trip measurement to correct the current one. While this strategy is simple in principle, it results in a system that is sensitive both to jitter in the overall delay itself and to its asymmetry.

C. Common-view time transfer

The third commonly used method is called common view.⁴⁷ Two receivers, which are approximately equidistant from a timing source, each measure the arrival time of the same time signal using their local clocks. Since the signal arrives at the two stations at very nearly the same time, the difference between the two time tags is approximately the time difference between the two clocks. The two stations can be synchronized with an uncertainty that is proportional to the difference between the two paths back to the transmitter. Common-mode fluctuations in the two paths and fluctuations in the timing source itself cancel out (at least in first order). There need not be any explicit correction for the path delay if the two paths are very nearly the same, but it is hard to arrange this in practice, and some correction for the residual path difference is usually required. The receivers must know their locations relative to the transmitter to compute this correction—a requirement that has no analogue in the two-way scheme. If the transmitter and the receivers are moving (as in the global positioning system) then computing these distances may involve complicated transformations.

Common-view techniques lend themselves naturally to broadcast services, since many receivers can be synchronized simultaneously using signals from a single transmitter. They have been an important technique in time transfer for many years. Early systems used common-view observations of television signals, Loran stations, and similar transmissions. There have even been proposals to use local power distribution systems for common-view time transfer within a single distribution region. In many of these cases the common-view transmitter is designed for some other purpose, and it does not even need to “know” that it is being used for time distribution. All of the receivers must know about each other, however, and must agree on a common observation schedule and processing algorithm. As we will show below, this knowledge can be acquired after the fact in some cases.

Two-way systems and common-view systems are sensitive to fluctuations in the path delay in different ways. Two-way systems depend on the reciprocity of a single path—on the fact that it has the same delay in both directions. Its limitation arises from fluctuations in the symmetry of the path and not on its overall delay. Common view, on the other hand, depends on the equality of the delays in two separate paths transmitting signals in the same direction, and it is limited by effects that are not common to both paths. When the paths are mostly through the atmosphere, both methods

are often limited in practice more by fluctuations in the delays through the hardware than by problems in the path itself.

XI. GLOBAL POSITIONING SYSTEM

The signals from the global positioning system satellites are widely used for high-accuracy time transfer using both one-way and common-view methods. A similar system, operated by Russia, is called GLONASS. We will first describe the satellites and the characteristics of the signals they broadcast, and we will then discuss how these messages can be used for time and frequency distribution.

The GPS system uses 24 satellites in nearly circular orbits whose radius is about 26 600 km. The orbital period of these satellites is very close to 12 h, and the entire constellation returns to the same point in the sky (relative to an observer on the earth) every sidereal day (very nearly 23 h 56 m).

The satellite transmissions are derived from a single oscillator operating at a nominal frequency of 10.23 MHz as measured by an observer on the earth. In traveling from the satellite to an earth-based observer, the signal frequency is modified by two effects—a redshift due to the second-order Doppler effect and a blueshift due to the difference in gravitational potential between the satellite and the observer. These two effects produce a net fractional blueshift of about 4.4×10^{-10} (38 $\mu\text{s/day}$), and the proper frequencies of the oscillators on all of the satellites are adjusted to compensate for this effect, which is a property of the orbit and is therefore common to all of them. In addition to these common offsets, there are two other effects—the first-order Doppler shift and a frequency offset due to the eccentricity of the orbit, which vary with time and from satellite to satellite.

The primary oscillator is multiplied by 154 to generate the $L1$ carrier at 1575.42 MHz and by 120 to generate the $L2$ carrier at 1227.6 MHz. These two carriers are modulated by three signals: the precision “ P ” code—a pseudo-random code with a chipping rate of 10.23 MHz and a repetition period of 1 week, the coarse acquisition (“ C/A ”) code with a chipping rate of 1.023 MHz and a repetition rate of 1 ms, and a navigation message transmitted at 50 bits/s. Under normal operating conditions, the C/A is present only on the $L1$ carrier.

The satellites currently in orbit also support “anti-spoofing.” (AS) When the AS mode is activated (which is almost always at the present time), the P code is encrypted and the result is called the Y code. The P and Y codes have the same chipping rate, and this code is therefore often referred to as $P(Y)$.

Each satellite transmits at the same nominal frequencies but uses a unique pair of codes which are constructed so as to have very small cross correlation at any lag and a very small auto-correlation at any nonzero lag (CDMA—code division multiple access). The receiver identifies the source of the signal and the time of its transmission by constructing local copies of the codes and by looking for peaks in the cross correlation between the local codes and the received versions. Since there are only 1023 C/A code chips, it is feasible to find the peak in the cross correlation between the local and

received copies using an exhaustive brute-force method. When this procedure succeeds, it locks the local clock to the time broadcast by the satellite modulo 1 ms (with an offset due to the transmission delay), and allows the receiver to begin searching for the 50 b/s navigation data. Once this data stream has been found, the receiver can read the satellite time from the navigation message, use it to construct a local copy of the P code, and begin looking for a cross-correlation peak using this faster code. This cross-correlation process fails when a nonauthorized receiver processes the Y code, and these receivers usually drop back automatically to using the C/A code only in this case.

The pseudo-random codes are Gold codes,⁴⁸ which are implemented using shift registers that are driven at the chipping frequency of the code. The details of the design of the code generators, the correlation properties of the codes, and the relationship between the codes and the satellite time are discussed in the GPS literature. A particularly clear exposition is in Ref. 49.

The navigation message contains an estimate of the time and frequency offsets of each satellite clock with respect to the composite GPS time, which is computed using a weighted average of the clocks in the satellites and in the tracking stations. This composite clock is in turn steered to UTC(USNO), which is in turn steered to UTC as computed by the BIPM. The time difference between GPS time and UTC(USNO) is guaranteed to be less than 100 ns, and the estimate of this offset, which is transmitted as part of the navigation message, is guaranteed to be accurate to 25 ns. In practice, the performance of the system has almost always substantially exceeded its design requirements.

GPS time does not incorporate additional leap seconds beyond the 19 that were defined at its inception; the time differs from UTC by an integral number of additional leap seconds as a result. This difference is currently 13 s, and will increase as additional leap seconds are added to UTC. The accuracy statements in the previous paragraph should therefore be understood as being modulo 1 s. The number of leap seconds between GPS time and UTC is transmitted as part of the navigation message, but is not used in the definition of GPS time itself. Advance notice of a future leap second in UTC is also transmitted in the navigation message.

The GLONASS system is similar to GPS in general concept. In its fully operational mode, there will be 24 satellites orbiting in 3 planes 120° apart. At the present time (October, 1998) there are 13 operational satellites. The GLONASS P code has a chipping frequency of 5.11 MHz and a repetition period of 1 s; the C/A code chipping frequency is 511 kHz and repeats every millisecond. All satellites transmit the same codes, but the carrier frequencies are different (FDMA—frequency division multiple access). These frequencies are defined by

$$f_1 = 1602 + 0.5625k \text{ MHz} \quad (24)$$

and

$$f_2 = 1246 + 0.4375k \text{ MHz}, \quad (25)$$

where k is an integer that identifies the satellite. These rather widely spaced frequency allocations complicate the design of

the front end of the receiver—especially the temperature stability of the filters that are usually needed to reject strong out-of-band interference. On the other hand, cross talk between signals from the different satellites is much less of a problem; interference from a single-frequency source is also less of a problem with GLONASS, since the interfering signal is less likely to affect the signals from the entire constellation.

In the original design, k was a unique positive integer for each satellite. The resulting frequencies interfered with radio astronomy measurements, and a set of lower frequencies with negative values of k will be phased in over the next few years. In addition, satellites on opposite sides of the earth may be assigned the same value for k . This would not present a problem for ground-based observers, since they could not see both of them simultaneously, but satellite-based systems may require additional processing (similar to the tracking loops in GPS receivers) to distinguish between the two signals.

As with GPS, the satellites transmit the offset between the satellite clock and GLONASS system time, but the relationship to UTC is not as well defined. Signals from the GLONASS satellites are therefore usually used in common view, since the time of the satellite clock cancels in this procedure.

The signals from the GPS constellation are usually affected by anti-spoofing (AS), which encrypts the P code, and by selective availability (SA), which dithers the primary oscillator and degrades the frequency stability of the signal. (The implementation of SA may also dither parameters in the ephemeris message which produce a dither in the computed position of the satellite. This aspect of SA is usually switched off.) From the point of view of a nonauthorized user, the primary effect of AS is to deny access to the P code, since it can only be processed by a keyed receiver when it is transmitted in its encrypted Y -code form. An authorized user would see this encryption as authenticating the transmission, since only a real satellite can produce valid encrypted messages. Hence the name anti-spoofing. The description of the GLONASS P code has not been published and is officially reserved for military use, but the GLONASS C/A code is not degraded.

The operation of the receiver is basically the same in either system. The receiver generates a local copy of the pseudo-random code transmitted by each satellite in view, and finds the peak in the cross correlation between each of these local copies and the received signal. The local code generator is driven from a local reference oscillator, and the time offset of the local clock needed to maximize the cross correlation is the *pseudo range*—the raw time difference between the local and satellite clocks. Using the contents of the navigation message, the receiver corrects the pseudo range for the travel time through the atmosphere, for the offset of the satellite clock from satellite system time, etc. (If the receiver can process both the $L1$ and $L2$ frequencies, then the receiver can also estimate the delay through the ionosphere by measuring the difference in the pseudo ranges observed using the $L1$ and $L2$ frequencies. The details of this calculation are presented below. If the receiver can process only

the $L1$ signal then it usually corrects for the ionospheric delay using a parameter transmitted in the navigation message.) The result is an estimate of the time difference between the local clock and GPS or GLONASS system time.

The time difference between the local clock and satellite system time can be used directly to discipline a local oscillator or for other purposes. However, these signals have deterministic fluctuations due to imperfections in the broadcast ephemeris which is used by the receiver to correct the pseudorange. In addition, signals from the GPS constellation are usually also degraded by the frequency dithering of selective availability. These effects are attenuated or removed in common view; if the stations are not too far apart, the common-view method also attenuates fluctuations in the delay due to the troposphere and the ionosphere since the paths to the two stations often have similar delays and the fluctuations in these delays often have significant correlation as well.

To facilitate time transfer among the various national laboratories and timing centers, schedules for observing the signals from the GPS and GLONASS satellites are published by the International Bureau of Weights and Measures (BIPM),⁵⁰ and all timing laboratories adhere to them as a minimum. (Not all timing laboratories can receive GLONASS signals at the present time.) Other users often observe this schedule as well so as to facilitate the evaluation of their local clocks in terms of national time scales. The comparison with national time scales may be required for legal traceability purposes, and using GPS time directly may not satisfy these legal requirements.

The need for common-view schedules arose initially because the first GPS receivers could only observe one satellite at a time. Newer receivers can observe many satellites simultaneously (as many as 12 for the newest units), and this makes it possible to eliminate formal schedules in principle. Each station records the difference between its local clock and GPS or GLONASS system time using all of the satellites in view, and the time difference between any two stations is computed after the fact by searching the data sets from both stations for all possible common-view measurements. In addition to eliminating the need for formal tracking schedules, this method can be used to construct common views between two stations that are so far apart that they never have a satellite in common view. A receiver in Boulder, Colorado, for example, might have a common view link to Europe using one satellite and a simultaneous second link to Asia using a second one. These data could be combined to produce a common-view link between Europe and Asia that was independent of the intermediate clocks in Boulder and on the satellites.

XII. MULTIPLE FREQUENCY DISPERSION

In practice, it is difficult to arrange things so that the two receivers in a common-view arrangement are exactly equally distant from the transmitter, and common-view data must be corrected for this path difference. If the transmission path is through the atmosphere or some other refractive medium, then knowing the difference in the physical path lengths is not enough—it is also necessary to know the refractivity of

the medium (the difference between the actual index of refraction and unity). If the refractivity is dispersive (that is, if it depends on the frequency of the carrier used to transmit the information), and if the dispersion is separable into a part that depends on the carrier frequency and a part that is a function of the physical parameters of the path, then it is often possible to compute the refractivity by measuring the dispersion, even when both of these parameters vary from point to point along the path. If the length of the physical path is L then the transit times measured using two different carrier frequencies f_1 and f_2 will be

$$t_1 = \int \frac{[1+n_1(r)]}{c} dr = \frac{L}{c} + \frac{F(f_1)}{c} \int G(\dots) dr \quad (26)$$

and

$$t_2 = \int \frac{[1+n_2(r)]}{c} dr = \frac{L}{c} + \frac{F(f_2)}{c} \int G(\dots) dr, \quad (27)$$

where the quantities $n_1(r)$ and $n_2(r)$ are the refractivities at the two frequencies at a coordinate point r along the path. These refractivities are decomposed into a known function F that depends only on the frequency of the carrier and a second known function G that depends only on the parameters of the medium such as its density, pressure, temperature, The integrals must be evaluated along the transmission path. This would be difficult or impossible for a real-world path through the atmosphere, since the required parameters are not known and cannot be easily measured. Since the integral of G is independent of the carrier frequency, we can eliminate G in terms of the difference in the two measured transit times

$$\int G(\dots) dr = \frac{c}{L} \frac{(t_1 - t_2)}{[F(f_1) - F(f_2)]}. \quad (28)$$

Substituting Eq. (28) into Eq. (26)

$$t_1 = \frac{L}{c} + (t_1 - t_2) \frac{F(f_1)}{F(f_1) - F(f_2)} = \frac{L}{c} + (t_1 - t_2) \frac{n_1}{\Delta n}. \quad (29)$$

The first term on the right-hand side is the time delay due to the transit time of the signal along the path of physical length L , and the second term is the additional delay due to the refractivity of the medium. Note that this second term depends only on the measured time difference between the signals at the two frequencies and on the known frequency dependence of the refractivity. It does not depend on the physical length L itself or on the details of the spatial dependence of G , and we would have obtained exactly the same result if only a small piece of the path $L' < L$ was actually dispersive. This is an important conclusion, because it means that a receiver can estimate the delay through the ionosphere by measuring the difference in the arrival times of identical messages sent using the $L1$ and $L2$ frequencies. Note that this method fails if the medium is refractive without being dispersive—that is, if the index of refraction differs from unity but is independent of the carrier frequency. This is the situation with the effect of the troposphere on microwave frequencies near 1 GHz, and other means must be found to model the refractive component of its delay. (There is no

easy way to estimate the correction due to the troposphere, and its effect is usually ignored at the present time as a result. This is not adequate when the phase of the carrier is used as in geodetic measurements, and more sophisticated estimates of the troposphere are required in those applications.⁵¹⁾

Although the time difference arising from the dispersion may be small, it is multiplied by the refractivity divided by the dispersion, a ratio that is larger than unity. (The value is about 1.5 for the $L1/L2$ frequencies used by GPS.) The result is that the ionospheric refractivity makes a significant contribution to the time of flight of a signal from a satellite; ignoring this contribution is only adequate for common-view observations over short base lines where the effect cancels in the subtraction. Furthermore, the noise in the dispersion measurement ($t_1 - t_2$) is amplified by this same ratio, so that the price for removing the refractivity of the path is an increase in the noise in the measurements by roughly a factor of 3 compared to a single-frequency measurement using only $L1$.

XIII. EXAMPLES OF TIME TRANSFER USING GPS

In this section we illustrate the principles of time transfer using signals from the GPS satellites. (The same principles would apply to time transfer using GLONASS signals.⁵²⁾ Since the entire GPS constellation usually has AS enabled, the P code is not available to most time-transfer users. (Even when AS is turned on, the P code can still be used for dispersion measurements. A cross correlation of the P -code signals on $L1$ and $L2$ can be performed without knowing the actual P code—the fact that it is the same on both carriers is enough.) All of the time-difference results in this section are based on measurements using the C/A code or the carrier itself. In addition, the code-based data were acquired with single-frequency receivers, which cannot measure the $L1-L2$ dispersion and must estimate or measure the contribution of the ionosphere using some other method. This is generally done using the parameters in the navigation message transmitted by the satellite, although other techniques exist.⁵³

The C/A code cross-correlation process results in a series of ticks with a period of 1 ms, and decoding the ephemeris message allows the receiver to identify the ticks in this stream that are associated with integer GPS seconds.⁵⁴ When the arrival times of these ticks are corrected for the offset of the satellite clock from GPS time, for the path delay and for the other effects described above, the result is a string of 1 Hz pulses that are synchronized to GPS time. (In order to apply these corrections, the receiver must parse the data stream transmitted by the satellite.) The time differences between these pulses and the corresponding ticks from a local clock are then measured using a standard time interval counter. Single-channel receivers can only perform this measurement on one satellite at a time, but newer “all in view” receivers can measure the time difference between the tick of the local clock and GPS time using up to 8 or 12 satellites simultaneously. In the simplest version of this scheme, the receiver measures the time difference between the local clock and a composite 1 Hz tick. This tick is positioned

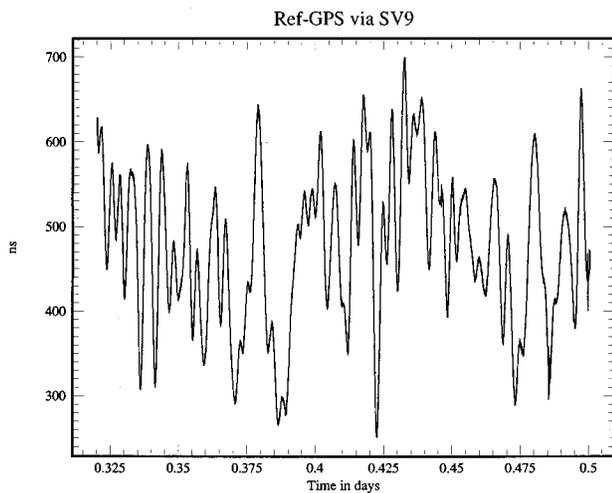


FIG. 1. The time difference between a local cesium clock and GPS time measured using data from satellite 9 only. The observed variation is due to the effects of SA on the GPS signal; the contributions of the clock and receiver to the variance are negligible on this scale.

using data from all of the satellites that are being tracked, but the better receivers can provide independent time differences between the local clock and GPS time using each satellite. This is done by providing a single 1 Hz tick to the time interval counter together with a data stream that specifies the offset between the time of the physical tick and GPS time as computed using data from each satellite being tracked. Unless the two stations are very close together, this latter capability is necessary for common-view measurements, as we will show below.

Figure 1 shows an example of these time-difference measurements. The plot shows the time difference between a local cesium standard and GPS time measured using data from a single satellite—SV 9. The time of the local clock is offset from GPS time by about 500 ns, and this explains the mean value of the data that are displayed. The time dispersion in the figure is totally due to the effects of SA on the GPS signal—the contributions due to receiver noise and to the frequency dispersion of the clock are negligible on this scale. These data are *not* white phase noise at periods shorter than a few minutes, and the mean time difference over a few minutes is not stationary. Averages over longer periods begin to approximate white phase noise, but the contribution of SA to the variance does not become comparable to the noise in the clock itself until the data are averaged for several days. (Unfortunately, averaging white phase noise only decreases the standard deviation of the mean at a rate proportional to the square root of the averaging time, and this improvement does not even begin until we enter the domain where the variance is dominated by white phase noise.)

As we pointed out above, the common-view procedure cancels the contribution of the satellite clock to the variance of the time differences, and we would therefore expect that this procedure would cancel SA. This expectation is confirmed by the data in Fig. 2. This figure shows the time differences between two identical receivers whose antennas were located about 1 m apart. The two receivers used the same local clock as a reference; the mean time difference

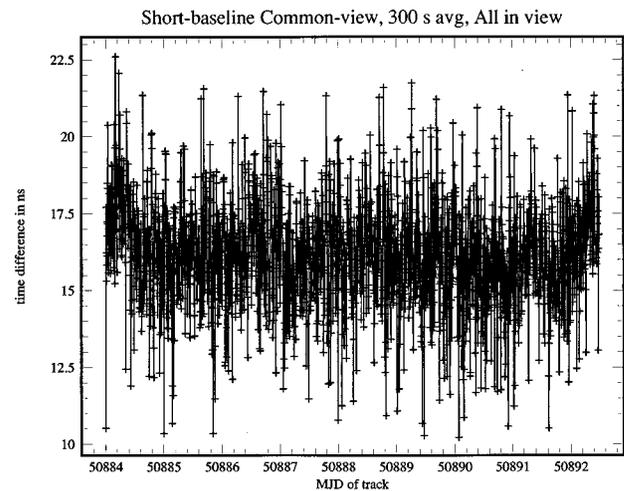


FIG. 2. The time differences measured in a short-baseline common-view experiment between two multichannel GPS receivers. The two receivers were connected to the same clock and the antennas were about 1 m apart. The average time difference between the two receivers was computed using all satellites in view.

shown in the figure is due to the difference in the lengths of the two cables between the receivers and the common clock. These data have an amplitude of 1.7 ns rms, and they are much closer to white phase noise—at least at short periods.

The data shown in Fig. 2 represent a best-case result: the reference clock is the same for both units and the antennas are very close together so that almost all propagation effects are the same for both of them and therefore cancel in the common-view difference. Likewise, the fluctuations in the clock itself are also common to both channels. However, the advantages of common view are still significant even over longer base lines and different clocks. Figure 3 shows data from a common-view experiment between NIST in Boulder, Colorado, and the United States Naval Observatory (USNO) in Washington, DC. Even though these stations are about

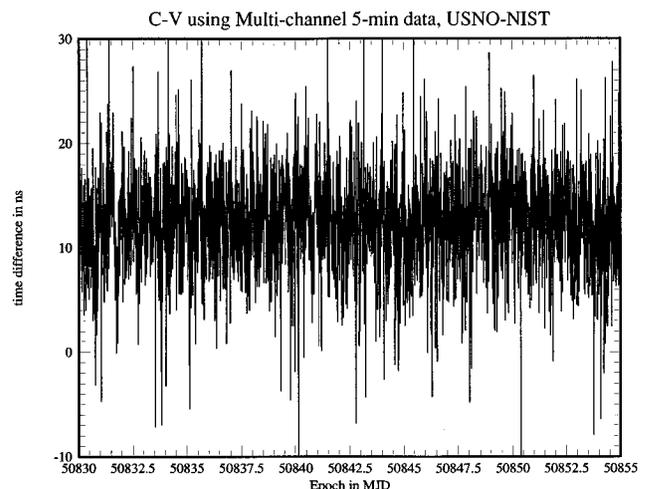


FIG. 3. The time differences between a receiver at NIST in Boulder, Colorado, and a second receiver at USNO in Washington, DC, about 2400 km away. As in the previous figure, the average time difference between the two stations was computed using all satellites in common view at each time. The members of the common-view ensemble change slowly during the observation period.

2400 km apart, the common-view time difference data show a short-term noise of about 3 ns rms—not as good as the short base line results, but still much better than the one-way data shown in the first figure.

The common-view differences shown in both of these figures use data from all satellites that are in common view at any time (at least four and usually five or six). It is important to compute the common-view differences on a satellite by satellite basis and then average these differences (rather than subtract the average time difference computed using all satellites in view at each station). The reason for this is that stations that are 500 km or more apart are likely to be tracking a different set of satellites with only some of them in common. The fluctuations due to SA are different for each satellite, so that the two stations will have a different average SA degradation that will not be cancelled in the common view procedure. It is important that multichannel receivers provide the time-difference data by satellite for this reason. Single-channel receivers can do almost as well at canceling the SA, although the noise is somewhat higher because there are fewer independent measurements to average.⁵⁵

Since the GPS carrier is derived from the same oscillator that produces the code, several groups,^{56–59} have begun experimenting with time transfer (or frequency calibration⁶⁰) using the phase of the carrier rather than the code.⁶¹ There are a number of conceptual difficulties with this approach. The most serious are probably the difficulty in calibrating the effective delay through the receiver and the ambiguities in estimating the integer number of cycles in the time of flight between the satellite and the receiver. These experiments are almost always done using common-view methods, since it is difficult to estimate the various systematic corrections with sufficient accuracy in a one-way experiment. In addition, the increased resolution that is available using the phase of the carrier requires a corresponding increase in the accuracy of the satellite ephemerides, and most carrier-phase analyses used postprocessed ephemerides for this reason.

A number of newer code-based receivers use the carrier tracking loop to aid the code-based correlator. These techniques include delta pseudo range and integrated Doppler.⁶² The ionospheric effect on the code and the carrier have opposite signs, so that it is also possible in principle to use the code-carrier dispersion to estimate the ionospheric refractivity.

One of the primary problems with GPS data is the effect of multipath on the propagation of the signals through the atmosphere. This effect arises from the fact that the signal may travel along many different paths in going from the satellite to the receiver. Reflectors that are near the receiving antenna are most troublesome, but even objects that are quite far away can cause trouble. If the local clock is sufficiently stable, it is possible to average the multipath effects by tracking each satellite for as long as possible and averaging the results. This procedure is most effective in averaging the contributions due to reflectors that are many wavelengths away from the antenna, since the multipath contribution tends to fluctuate rapidly in this case.⁶³

The magnitude of the multipath effect varies from satellite to satellite and from second to second as the satellites

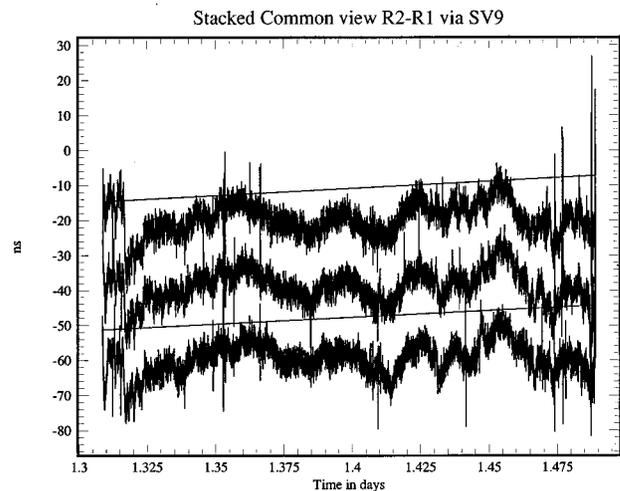


FIG. 4. The results of a short-base line common-view experiment using data from SV 9 only. The figure shows 1 s measurements with no averaging. Time differences from successive days are offset by -240 s and are displaced vertically by an arbitrary amount for clarity.

move through the sky and the geometry of the path and the reflectors changes. However, as viewed by an observer on the earth, the satellites return to the same point in the sky every sidereal day (approximately every 23 h 56 min), and the contribution should repeat with this period. Figure 4 demonstrates this. These data show the common-view time differences (computed using 1 s measurements with no averaging) between two receivers whose antennas were about 1 m apart (the same setup as was used for the results shown in Fig. 2). Time differences for consecutive days have been “stacked”—that is, they have been offset by -240 s in time and displaced vertically on the plot for clarity. The points with the same X coordinate on successive plots were therefore recorded at the same sidereal time on consecutive days. Note that although the pattern of the time differences is very complex, it shows very high correlation from one trace to the next. This is exactly what we would expect from multipath. The geometry and the details of the reflection change with time in a complicated way, but the entire pattern repeats on the next sidereal day when the satellite comes back to the same position relative to the ground-based observer and the reflectors that are near the antenna. Also note that this figure shows the *differential* multipath effects on two antennas that are only 1 m apart. We do not know the full multipath effect on either signal, although we would expect that it would be not smaller than the magnitude shown in this figure.

It is tempting to try to correct for the multipath effect by measuring it once as in Fig. 4 and by then correcting the measurements on subsequent days using these data as a template. This does not work for three reasons. In the first place, the plot shows the differential multipath effect, and it is not clear what the actual effect on each receiver is. In the second place, the pattern is not absolutely stationary in time because the satellites do not return to *precisely* the same position every sidereal day. This drift is small over the few days shown in the figure, but it becomes significant after a week or so, so that the template would not be useful for very long. Finally, it is often difficult to separate multipath effects from

a sensitivity of the antenna or its connecting cable to the diurnal fluctuation in the ambient temperature. The frequency signatures of the two effects differ by only 1 c/y, and are therefore difficult to separate. Diurnal temperature fluctuations are much less repeatable from day to day, so that these effects cannot be modeled so effectively in this way.

The tracking schedules published by the BIPM include an advance of 4 min every day so that tracks on consecutive days are observed in approximately the same multipath environment. This method does not remove or correct for the effect of multipath—it simply converts it from a fluctuating signal to a nearly static offset which is different for each satellite. This is not necessarily a desirable result, since there are already satellite-to-satellite offsets due to errors in the ephemerides.

XIV. DELAY MEASUREMENTS IN VIRTUAL CIRCUITS

We now turn to a discussion of time distribution using communications networks. Everything that we have said so far about measuring the transmission delay assumed implicitly that the path between the transmitter and receiver had the well-defined characteristics that we intuitively associate with a physical circuit or a well-defined path through the atmosphere. These ideas are still useful in characterizing most paths through the atmosphere, but transmission circuits on cable and fiber networks are increasingly packet switched rather than circuit switched. This means that there is no end-to-end physical connection between the transmitter and the receiver. Packets are sent from the transmitter to the receiver over a common high-bandwidth physical channel, which is shared among many simultaneous connections using time-domain multiplexing (or some other equivalent strategy). Each packet passes through a series of routers which read the destination address and direct the packet accordingly. Data appear to enter at one end and emerge at the other, but there need not be a physical connection between the two. In an extreme case, consecutive packets may travel over very different paths because of congestion or a hardware failure. This variability complicates any attempt to measure the effective delay of the circuit.

These problems are likely to become more important as multiplexed digital circuits become more widely used. A number of strategies have been proposed for dealing with these problems. A shorter path is likely to have smaller variability as well, since it would probably have fewer routers and gateways, and this suggests that the absolute delay might be used as an estimate of its variability. It is tempting to assert that the absolute magnitudes of these problems will be smaller on higher-speed channels, but this is by no means a sure thing.

XV. ENSEMBLES OF CLOCKS AND TIME SCALES

You will probably want to have more than one clock if you are serious about time keeping. Although a second clock can provide a backup when the first one definitively fails, three clocks provide the additional ability to detect partial failures or occasional bad data points using a simple majority-voting scheme. Unfortunately, such schemes are

not as simple as they first appear to be for reasons to be outlined below, and the full advantage of using several clocks can be realized only when they are combined in a more complex time-scale procedure.

The problem with simple majority voting (or with comparing the time of a “master” clock with a number of subsidiary ones) is that the stability of the frequency of almost any commercial clock (even a cesium-based device) is better than its accuracy by three orders of magnitude or more. Comparisons among clocks are almost always realized using time-difference measurements, and the large and stable frequency differences between nominally identical devices result in linearly increasing time differences as well. (Many clocks have deterministic frequency aging as well, and this further complicates the issue by adding a quadratic dependence to the time-difference data.) It is no longer a simple matter to detect a hardware failure in this situation. It would be nice to think that the deterministic frequency offsets of an ensemble of clocks would scatter uniformly about the true SI frequency, but there is no reason to believe that this is true in general. In any case, it is much less likely to be true for a small group of devices. Thus neither a simple average of the times nor of the frequencies of the ensembles of devices has good statistical properties. The machinery of a time scale provides a natural way of dealing with this problem.

A second advantage of a time scale is in providing a framework for combining a heterogeneous group of devices so as to achieve a time scale that has the best properties of each of them. For example, a time scale might be used to combine the excellent short-term stability of a hydrogen maser with the better long-term performance of a cesium-based device. Finally, a time scale can provide a flywheel for evaluating the performance of primary frequency standards or other devices such as prototype standards. Such evaluations can be done using only one clock, of course, but there is no way of identifying the source of a glitch in that configuration.

As we will show below, the ensemble-average time and frequency have many of the properties of a physical clock. In particular, the ensemble time and frequency must be set initially using some external calibration. Conceptually, this is done by measuring the time difference and frequency offset between each member clock and the external standard, although other methods in which a single member clock is calibrated and the other clocks are then compared to it are often used in practice. These initial time and frequency offsets can be used to define the corresponding parameters of the ensemble or an additional administrative offset can be added to them. This offset might be used to steer the scale to a primary standard, for example. As we discussed above, TAI has been steered from time to time in this way using data from primary frequency standards.

A. Clock models

As we discussed above, one of the problems with not using a formal time-scale algorithm but simply comparing clocks using time differences, is the deterministic frequency differences between nominally identical devices. A natural way to deal with these deterministic effects is to include

them in the definition of a model of how we expect the time differences to evolve, and to compare our observations with the predictions of the model in order to detect problems. Since the physical devices that make up an ensemble change as old devices fail and new ones are added as replacements, it is customary to define the model parameters of each member clock with respect to the ensemble average rather than with respect to any particular master device. The goal is to make the characterization of each of the clocks in the ensemble independent of the parameters of the others.

The deterministic performance of most clocks can be very well characterized using three parameters: x , the time difference, y , the frequency difference, and d , the frequency aging, where each of these parameters characterizes measurements between the clock and the ensemble. These definitions imply that the “ensemble” is a clock too, although we have not yet specified how it can be read, set or characterized. In fact, different algorithms address this problem in different ways.

Using these three parameters, we can construct a model of the deterministic performance of the clock. Assuming that we have the values of the three model parameters at some time t , the values at some later time $t + \tau$ are given by

$$\begin{aligned}x(t + \tau) &= x(t) + y(t)\tau + \frac{1}{2}d(t)\tau^2, \\y(t + \tau) &= y(t) + d\tau, \\d(t + \tau) &= d(t).\end{aligned}\quad (30)$$

It is possible to extend these equations to include higher-order terms, but this is not necessary (or even useful) for most devices. These model equations assume that there is no *deterministic* variation in the frequency aging—an assumption that is justified more by experience than by rigorous analysis.

In addition to these deterministic parameters, our previous analysis of oscillator performance would lead us to expect that our data would also have stochastic contributions to the variation of each of these parameters. For simplicity, we could add terms for white phase noise, white frequency noise, and white frequency aging. These names are chosen to parallel the deterministic description in Eqs. (30)—we can also use the alternative names we have defined above. (As we discussed above, the variance in each parameter has contributions both from its own white noise term and from all of the noise terms below it. The combined phase or frequency noises are very definitely not white.)

The next step would be to compare the predictions of our model equations with measurements and to test if the differences are consistent with the assumed model for the noise. Our measurements usually consist of pairwise time differences between our physical clocks, and there is no obvious, universal prescription for using these data to evaluate a model of the type specified in Eqs. (30). There are two difficulties: the first is that the time difference data are ambiguous to within an overall additive constant and the second is that we have not specified a model for how the ensemble clock itself is measured or how it evolves.

B. AT1 algorithm

Several versions of the AT1 algorithm have been used at NIST for many years. Although it works quite well with real data, I discuss it in detail not because it is theoretically optimum but because it is relatively simple to implement. In addition, it illustrates the general ideas that underlie a number of other algorithms including ALGOS, which is used by the BIPM to compute EAL and TAI.

The starting point for a discussion of AT1 is the observation that it is difficult to make robust estimates of the deterministic frequency aging for most clocks. Even at the largest values of τ (where the effects of frequency aging are greatest), these deterministic effects are usually swamped by the flicker and random-walk frequency noises in the device. Almost the only exceptions are the very best hydrogen masers, whose frequency noise is so small that robust estimates of the deterministic frequency aging are possible, and rather poor quartz oscillators (which are not normally used as members of ensembles anyway) whose frequency aging is so large that it is readily observable even for small averaging times. Thus the assumption that $d(t) = 0$ for all t is consistent with the data for almost all clocks, except for the best hydrogen masers. For those clocks we would use a constant for the aging that is determined outside of the ensemble algorithm. A typical value for such a maser would be of order 10^{-21} s^{-1} .

If we model the frequency aging as a noiseless constant value then this is equivalent to assuming that the frequency fluctuations are dominated by white noise processes. This is a reasonable assumption provided that τ is not too large. The maximum value of τ that still satisfies this requirement varies with the type of device; a typical value for a cesium oscillator would be less than a day or two. On the other hand, it must not be made too small either, because otherwise the time difference data will be dominated by the white phase noise of the measurement process. (This is not necessarily a bad thing, but taking full advantage of the data in this case requires a change in strategy. This point is discussed below in connection with techniques used for transmitting time using the Internet.) As specific examples, the NIST time scale, which uses measurement hardware with a noise floor of several ps uses $\tau = 2 \text{ h}$, while the BIPM computation of ALGOS, which uses the much noisier GPS system to measure the time differences between laboratories, uses $\tau = 5d$.

As we pointed out above, the initial values of the time and frequency of the ensemble must be known *a priori*. The epoch associated with these initial values is t_0 . The algorithm then continues with periodic measurements of the time differences between each clock and one of them, which is designated as the reference clock or working standard. These measurements are made at epochs t_1, t_2, \dots, t_n . These measurements are usually equally spaced in time, but this is more for convenience than for any formal requirement of the algorithm.

If at some epoch t_k the time difference and frequency offset between the j th clock and the ensemble were $x_j(t_k)$ and $y_j(t_k)$, respectively, then AT1 models the time difference at some later epoch t_{k+1} as being given by

$$\hat{x}_j(t_{k+1}) = x_j(t_k) + y_j(t_k)\tau, \quad (31)$$

where

$$\tau = t_{k+1} - t_k. \quad (32)$$

Note that the right side of Eq. (31) is an estimate based on the two model parameters of the clock using a simple linear model. As we discussed above, this is appropriate for all but the best hydrogen masers. We would use the first of Eqs. 30 with a constant value for d for these devices. This constant frequency aging does not really have any effect at this point. Since the algorithm normally operates with a fixed interval between computations, the effect of adding frequency aging is to add a constant offset to each of the predictions. Such an offset can be absorbed into the starting values without loss of generality.

The m th clock in the ensemble is the working standard. It is characterized using the same model as any other clock [i.e., Eq. (31) with $j = m$]. The working standard is not necessarily the clock with the best statistical properties—since it may be awkward to change this designation in a running ensemble, it is often the clock that is likely to be the most reliable or to live the longest.

At each epoch we measure the time differences between the working standard and each of the other clocks using a time interval counter or any of the other methods we described above. These measured time differences are $X_{jm}(t)$, where j takes on all possible values except for m . Combining the predicted time of each clock with respect to the ensemble using Eq. (31) with the corresponding measured time difference between that clock and the working standard, we can estimate the time difference between clock m and the ensemble via clock j using

$$\hat{x}_m^j(t_{k+1}) = \hat{x}_j(t_{k+1}) - X_{jm}(t_{k+1}), \quad (33)$$

where the superscript j on the left side indicates that the estimate has been made using data from clock j . Note that Eq. (33) could equally well have been thought of as a definition of the time of the ensemble with respect to physical clock m estimated using data from clock j . Also note that, as expected, Eq. (33) reduces to Eq. (31) when $j = m$ since $X_{mm}(t) \equiv 0$. Thus Eq. (33) applies to the working standard just as it does to any other clock in the group.

If Eq. (31) is an accurate model of the time of clock j , then Eq. (33) is an unbiased estimate of the time of the working standard with respect to the ensemble since the deterministic time and rate offset of clock j have been properly estimated. Note that clock m may have a deterministic time or frequency offset with respect to the ensemble—there is no requirement that its parameters be the same as those of the ensemble. Its statistical parameters are also not important for this procedure to work.

If there are N clocks in the ensemble, there are $N - 1$ independent time-difference measurements X_{jm} , and the same number of equations identical to Eq. (33) with all possible values of j except $j = m$. The previous state of clock m is used to predict its time with respect to the ensemble using Eq. (31). (Alternatively, we could use Eq. (33) for this purpose, setting $X_{mm} = 0$. We adopt this latter strategy in what

follows, since it simplifies the notation and emphasizes the fact that clock m is not a special clock with respect to the ensemble—only with respect to the time-difference hardware.) If we assume that the fluctuations in the time-difference measurements are not correlated with each other, the N estimates of the time of clock m with respect to the ensemble are characterized by a well-defined mean, which we can compute using

$$x_m = \sum_j w_j \hat{x}_m^j. \quad (34)$$

The weights w_j are arbitrary at this point, subject only to the usual normalization requirement that

$$\sum_j w_j = 1. \quad (35)$$

If the clocks are all identical then a simple assumption would be to set all of the weights equal to $1/N$, so that Eq. (34) reduces to a simple average, but this is usually not the optimum choice in a real ensemble.

Equation (34) can be equally well thought of as defining the time of the ensemble with respect to the working standard, clock m , computed by averaging the independent estimates obtained from the time difference between clock m and each of the member clocks. The time of clock j with respect to the ensemble can then be calculated by combining Eq. (34) with the measured time difference X_{jm} :

$$x_j = X_{jm} + x_m. \quad (36)$$

These values are then used as the starting point for the next measurement cycle.

The difference between Eq. (31), which predicted the time of clock j with respect to the ensemble based on its previous state, and Eq. (36), which gives the value of clock j based on the data from the current measurement cycle, is the prediction error, ϵ_j . The average of this prediction error is a measure of the quality of the clock in predicting the time of the ensemble. The average of this value can be used to obtain an estimate for the weight to be used for clock j in Eq. (34)

$$w_j = \frac{C}{\langle \epsilon_j^2 \rangle}, \quad (37)$$

where C is chosen so that the weights satisfy the normalization condition, Eq. (35). Defined in this way, the weights are related to the TVAR statistic, which we defined above. As in our previous discussions, a good clock is a predictable clock—not necessarily one that has a small deterministic frequency offset, because we can always remove that offset using Eq. (31).

The dispersion in the prediction error can also be used to test for a clock failure. That is, on each cycle we could examine the statistic

$$\frac{\epsilon_j^2 - \langle \epsilon_j^2 \rangle}{\langle \epsilon_j^2 \rangle} \quad (38)$$

and decide that clock j has a problem if this statistic is greater than some threshold σ_{\max} . We might drop clock j from the ensemble calculation in this case and repeat the

computation [Eq. (34)] with its weight set to 0. If the statistic is close to σ_{\max} , we could choose to decrease the weight of the clock gradually so that there is no abrupt step in performance exactly at σ_{\max} . The AT1 time scale currently used at NIST sets $\sigma_{\max}=4$, and linearly decreases the weight of a clock starting when its average prediction error reaches 75% of this value.

This procedure is a more formalized version of a simple majority-voting scheme. Its advantage is that the systematic differences between the clocks have been removed by the model equations and therefore do not affect the error-detection process. Furthermore, the averaging procedure is applied to quantities that have no first-order deterministic variation. The working standard enters into these calculations exactly as any other member of the ensemble, and its data can still be rejected by the prediction-error test just discussed even though it continues to play a central role in the time-difference measurements.

If the clock frequencies were noiseless and constant then that would be the end of the story. The time-difference measurements would be composed of deterministic frequency and time offsets which would be accounted for by the model equations, and the residual time difference data would be characterized by pure white phase noise, which would be optimally attenuated by the averaging procedure. Unfortunately, the frequency of the clock is characterized by white frequency noise at intermediate times and by flicker and random-walk fluctuations at longer times. As we showed above, these frequency fluctuations do not produce time differences whose variation is characterized as pure white phase noise, and the average in Eq. (34) is no longer completely justified.

These frequency fluctuations can be accommodated by replacing the static frequency parameter y_j in Eq. (31) with a parameter that is allowed to evolve with time. In order to compute this dynamic frequency, the short-term frequency estimates

$$f_j(t_{k+1}) = \frac{x_j(t_{k+1}) - x_j(t_k)}{\tau} \tag{39}$$

are averaged using a simple exponential low-pass filter to produce the frequency estimate used in the model equation

$$y_j(t_{k+1}) = \frac{y_j(t_k) + \alpha f_j(t_{k+1})}{1 + \alpha}, \tag{40}$$

where the time constant α is set to the transition at the upper end of the domain in which white frequency fluctuations dominate the spectrum of the clock noise. The time constant parameter is actually a dimensionless constant less than 1; the effective time constant is τ/α . The value of the time constant is set to attenuate white frequency noise at intermediate periods; the gain of the filter approaches unity at very long periods. As we discussed above, the primary reason for using an exponential filter of this type is that it can be realized using the same kind of iteration as is used for the other parameters of the model. At each step in the process the

model parameters computed on the previous step are combined with the current measurements—older data need not be saved for this computation.

Several versions of this algorithm have been used at NIST for many years.^{64,65} The AT1 time scale provides optimum performance when the noise in the measurement system and the clocks is no more divergent than white phase noise. Unfortunately, clocks exhibit flicker and random-walk frequency noise at longer term, and the scale itself tends to become dominated by this type of noise at longer periods (on the order of months). The ALGOS algorithm used to compute EAL (and from that TAI and UTC) tends to have the same problem; as we discussed above, the solution in that case is to steer TAI using data from the primary frequency standards. This limitation has been addressed in the design of AT2.⁶⁶

The AT1 algorithm has other limitations. It cannot deal with a device (such as a primary frequency standard) that operates only occasionally. We would like to be able to characterize such a device using only a well-defined frequency—its time with respect to the ensemble is not interesting or important. In fact, since it operates only intermittently, its time with respect to the ensemble may not be well defined. As we mentioned above, the data from a primary frequency standard can be incorporated into the ensemble frequency by administrative adjustment as is done with TAI.

The AF1⁶⁷ procedure presents another solution to this problem. Using an algorithm analogous to AT1, it estimates the *frequency* (rather than the time) of the working standard using the measured frequency differences between it and the other members of the ensemble. Data from a primary frequency standard can be incorporated into this scale in a natural way, and the fact that the standard operates only intermittently is not a problem. Since it is quite similar to AT1 in its basic design, it will also become dominated by flicker and random-walk processes at longer periods—probably on the order of 1–2 yr.

In addition to ALGOS, there are a number of other algorithms that are similar to AT1. These algorithms have been widely discussed in the literature.⁶⁸ Other algorithms are based on Kalman filters.^{69,70} This formalism has some advantages in dealing with clocks that operate irregularly; it can also provide more robust definitions of the ensemble time and frequency than is possible using AT1 or its relatives. On the other hand, Kalman filter algorithms often have startup transients before the filter enters the steady-state regime. These transients can result in instabilities when the membership in the ensemble changes frequently.

A fundamental problem with many time scales is the way the weights of the member clocks are calculated. The AT1 algorithm and its relatives assign a weight to each clock based on its prediction error—a good clock is one that is predictable over the time interval between measurement cycles. (We have already seen this principle in the discussion of the Allan variance.) However, a clock whose weight is increased because of this will tend to pull the ensemble through Eq. (34). A small increase in weight thus tends to produce positive feedback, and a clock that starts out a little bit better than the other members of the ensemble can

quickly take over the scale. The problem, of course, is that Eqs. (31) and (36) are correlated, so that the prediction error is always systematically too small. The correlation clearly gets worse as the weight of a clock increases.

This correlation can be addressed by administratively limiting the maximum weight any clock can have, either to an absolute maximum value (as in ALGOS) or to a maximum percentage of the scale (as in the NIST implementation of AT1). It is also possible to estimate the magnitude of the correlation and correct for it.⁷¹

These methods limit the maximum weight that any clock can have in the ensemble, even if that large weight would have been deserved on statistical grounds. The reason is that one of the motivations for using an ensemble is safety and robustness in case of a failure. An ensemble whose time is effectively computed using only one clock will have a serious problem when that high-weight clock fails or has a data glitch. These problems can be more easily addressed in scales that are not computed in real time (such as TAI, which is computed about one month after the fact). It is much easier to recognize a problem retrospectively and to administratively remove a clock from the ensemble when later data shows that it was starting to behave erratically.

Since the weights are normalized using Eq. (35), any method that limits the maximum weight that a good clock can have implicitly gives poorer clocks more weight than they deserve. The price for keeping good clocks from taking over the scale is therefore a scale that is not as good as it could be because poorer clocks have more weight than they deserve based on statistical considerations.

A similar effect happens when a member clock has a large prediction error and is dropped from the ensemble as a result. The contribution that this clock would have made to the ensemble is divided among the remaining members, with the result that the ensemble time (and, indirectly, its frequency) may take a step when this happens. This can be especially serious in a small ensemble with one clock that is significantly poorer than the others. If one of the good clocks is dropped because it fails a test based on its prediction error, its weight will be distributed among the remaining members. If the weights of the other clocks are already limited by the administrative constraints that we discussed above, the result is to transfer the weight of this good clock to the poorest clock in the ensemble.

The methods used to calculate the frequency and the weight of each clock are effectively trying to separate the observed variance in the time difference data into deterministic and stochastic components. This separation, which is implicitly based on a Fourier decomposition of the variance, can never be perfect. These methods tend to have the most difficulty distinguishing between deterministic frequency aging and random-walk frequency noise because the two have similar “red” Fourier decompositions. This is the reason AT1 and similar time scales often do not include a term for deterministic frequency aging into the model equations for the time differences of the clocks—it is very difficult to estimate these parameters in the presence of the random-walk noise in the data. This can be a serious problem in an ensemble containing hydrogen masers, which tend to have sig-

nificant long-term frequency aging combined with very small short-term prediction errors. The very good short-term performance means that these clocks will be given a very high weight in an AT1-type procedure, and this high weight will then adversely affect the long-term stability of the scale due to the frequency aging of these devices. One solution is to incorporate a constant frequency-aging term that is determined administratively using procedures outside of the scale computations (using data from primary frequency standards, for example). Another solution is to modify the algorithm to include explicitly different weights for short-term and long-term predictions.

The time and frequency of the scale itself can be made almost totally insensitive to a change in the membership of the ensemble, provided that the initial parameters of a new clock are accurately determined. These initial parameters are usually determined by running a new clock in the scale with its weight administratively kept at 0.

One of the practical difficulties in running an ensemble is that there is no physical clock that realizes its time or frequency. This deficiency must be remedied if the ensemble time is used to control a physical device. One solution is to steer a physical clock so that its time and frequency track the corresponding ensemble parameters. There are a number of different ways of realizing this steering: the algorithm can emphasize time accuracy or frequency smoothness, but not both. An algorithm that emphasized time accuracy would use large frequency adjustments and time steps to keep the steered clock as close as possible to ensemble time. An algorithm that emphasized smoothness would limit the magnitude of frequency adjustments and never use time steps. The choice between the two extremes (or the relative importance of accuracy and smoothness in a mixed solution) cannot be specified *a priori*—the choice depends on how the data are to be used. Historically, steering algorithms at NIST have tended to place a higher value on frequency smoothness than on accuracy, with the inevitable result that the time dispersion of UTC(NIST) is larger than it would be if time accuracy were emphasized more strongly.

XVI. SYNCHRONIZING A CLOCK USING TIME SIGNALS

Finally, we turn to the question of how best to use a time signal that we receive from a distant source of uncertain accuracy over a channel whose delay may be variable. In the simplest distribution system, the receiver is a simple, “dumb” device. Its internal state (whether time or frequency) is simply reset whenever a signal is received. The signal may be corrected for transmission delay or some other systematic correction might be used, but the important point is that the previous state of the receiver makes little or no contribution to the process. This may be the best strategy if the uncertainty in the receiver state is so high as to make it effectively worthless, but it is important to recognize that this is not universally true, and that the state of the receiver before the new information is received also contains information that may be useful.⁷²

The performance of the receiver clock can be character-

ized in terms of deterministic and stochastic components. The deterministic aspects of its performance can be used to predict the data that will be received from the transmitter at some future time. This prediction is never perfect, of course, because the deterministic components may not be known precisely, because even if they are known at one instant they are likely to evolve in time, and because the stochastic components of the model are always present and add noise to all measurements. Although this prediction might not be as accurate as we would like, it nevertheless can often provide an important constraint on the received data.

The most useful situation arises when the spectrum of the noise in the receiver clock has little or no overlap with the spectrum of the noise in the transmission channel.⁷³ This happy situation supports the concept of “separation of variance”—the ability to distinguish between a “true” signal, which should be used to adjust the state of the receiver clock, and a “false” signal which is due to an un-modeled fluctuation in the channel delay and which should therefore be ignored. In an extreme situation, this strategy could be used to detect a failure in the transmitter clock itself.

Decisions based on separation of variance are probabilistic (rather than deterministic) in nature, and will only be correct in an average sense at best. Whether or not they improve matters in a particular situation depends on the validity of the assumption of separation of variance and on the accuracy of the various variance estimators. Since the estimators must be based on the data themselves, a certain amount of ambiguity is unavoidable.

This method obviously fails in those spectral regions where separation is impossible. There is no basis for deciding whether the observed signal is “real” and should be used to adjust the local clock or represents transmission noise and should be ignored. If the resulting ambiguity results in unacceptably large fluctuations, then either the receiver or the channel must be improved. The crucial point is that neither the receiver clock nor the transmission channel need be good everywhere in the frequency domain—it is perfectly adequate if they have complementary performance characteristics, and this may in fact be the “optimum,” least-cost alternative for achieving a given accuracy in the distribution of time or frequency.

These considerations explain why it may not be optimum to use the signal received over a noisy channel “as is.” It may be possible to remove the channel noise if the received data can be used to discipline an oscillator whose inherent noise performance can be characterized by a variance that separates cleanly from the noise added by the channel in the sense we have discussed. These ideas can be used to filter the intentional degradations imposed on the GPS satellites because of selective availability⁷⁴ and to improve the distribution of time on widearea networks such as the Internet.⁷⁵

These strategies can become quite complicated in detail, but all of them tend to share a number of general characteristics. Individual measurements of the time difference are usually dominated either by white phase noise arising from the measurement process itself or by something like selective

availability, which is a pseudo-random process that looks like noise in the short term. In the usual state of affairs, the frequency of the local clock is more stable than the time difference measurements would indicate—in other words these fluctuations in the measured time difference do not imply a corresponding fluctuation in the frequency of the local clock. The optimum strategy is clearly to leave the local clock alone in the short term, since its free-running stability is better than the time-difference data would indicate. This implies some form of averaging of the measurements, with the details depending on the application.

When the variance in the time-difference data can be modeled as white phase noise, then the optimum strategy is clearly to make a large number of rapid-fire time-difference measurements and to average these data. This process can continue for as long as the variance continues to look like white phase noise. For synchronization over packet networks such as the Internet, it is quite common to find periods of up to 15–30 s when this assumption is valid, and it is a simple matter to make 50 (or more) time-difference measurements during this period. (The same idea can be used to improve the measurement noise in a time-scale system. In that case as well, the short-term variance is well modeled as white phase noise.) These data would be averaged to form a single point, which was then used to discipline the local clock using a model such as Eqs. (30) or (31). Although the average cost of the synchronization process depends only on the number of calibration cycles, clustering the measurements in this way may improve the performance of the synchronization algorithm by optimally averaging the white phase noise of the measurement process. Although the average cost of the synchronization process would be the same if the measurements were spread out uniformly in time, the performance would not necessarily be as good because the increased interval would put the measurements outside of the time domain where the fluctuations are dominated by white phase noise.⁷⁶ While the cost of realizing this synchronization process increases linearly with the number of calibration cycles that are used in each group, the variance only improves as the square root of this number, so that incremental improvements in the performance of the algorithm become increasingly expensive to realize.

At some longer time, the variance of the averaged time differences drops below the variance of the free-running clock, and that is the point at which some form of correction to the time of the local clock is appropriate. (Note that the crossover is not determined by the frequency offset of the local clock but by the stochastic noise in this parameter. Deterministic offsets may be a nuisance in practice, but they can always be removed in principle and should not play a role in a variance analysis. Furthermore, if the variance of the local clock can be characterized as white frequency noise, then averaging consecutive, closely spaced frequencies will yield a statistically robust frequency estimate. The domain in which white frequency noise dominates the variance is often quite small, so that frequency averaging often does not produce an appreciable improvement in the performance of the synchronization process.) This crossover point

is never absolute, however. The local clock will always contribute some information to the measurement process—and even if this information is only used to detect a gross failure of the distant clock or of the calibration channel it should probably never be completely ignored.⁷⁷

In the usual situation, the noise in the data from the distant clock continues to decrease (relative to the local clock) at longer averaging times, so that the weight attached to the time-difference data in steering the local clock increases by a corresponding factor. This is certainly true for the degradation imposed on the GPS satellites by selective availability, and it is also a common feature of noise in a communications channel. Since both of these fluctuations are bounded in time, they make contributions to the Allan variance that must decrease at least as fast as the averaging time for sufficiently long intervals.

These considerations tend to imply relatively slow correction loops—the time constant for steering a good commercial cesium standard to a GPS signal is usually a day or longer. The corresponding time constant for a computer clock synchronized using messages transmitted over the Internet is usually a few hours. Although such loops maximize the separation of variance and uses both the local hardware and the calibration channel in a statistically optimum way, they have problems nevertheless. The very slow response of the correction loop means that it takes a long time to settle down when it is first turned on and also after any transient. A step of the local clock either in frequency or in time may persist for quite some time before it is finally removed. The loop may never reach steady-state operation in the extreme case. More frequent calibrations and shorter time constants are usually used when worst-case (rather than long-term average) performance is the principal design metric.

Another situation which requires more frequent calibrations and shorter time constants occurs when the local oscillator has a relatively rapid variation in frequency that cannot be characterized as a noise process. An example would be a large sensitivity to ambient temperature. Such an oscillator will tend to have large nearly diurnal frequency variations, and much of the variation will often be concentrated near the start and end of the working day. Although the time dispersion averaged over 24 h will tend to be small, the maximum dispersion during the day may be unacceptably large, and relatively frequent calibrations may be required to keep the time dispersion within specified limits.

The opposite problem happens when it is the channel that has a large diurnal fluctuation in the delay (as with WWVB, for example). The optimum strategy is clearly to increase the interval between calibrations so as to reduce the sensitivity of the measurements to these spurious fluctuations. This is not too hard to do in the case of WWVB, since it is usually used with local oscillators that are stable enough to support the required increase in averaging time. The analogous problem in computer networks is more difficult to deal with, since the clock circuits in many computers are not stable enough to support the increase in averaging time that

would be required to remove diurnal and similar load-related fluctuations in the network delay.

XVII. DISCUSSION

I have described some of the principles that are used to design clocks and oscillators, and I have introduced the statistical machinery that is used to describe time and frequency data. This machinery can be used to characterize the oscillators themselves and also to understand the degradation that results from transmitting these data over noisy channels with varying delays.

Time and frequency measurements will continue to play an important role in science, engineering, and life in general. Several new requirements are likely to be most important in the near term:

- (a) The need for higher-accuracy frequency standards for applications ranging from research in atomic physics and astrophysics to synchronizing high-bandwidth communications channels. Devices such as cesium or rubidium fountains and trapped-ion devices are already in the prototype stage, and devices that can potentially realize the SI second with an uncertainty of significantly less than 10^{-15} are already being designed. There is no reason to believe that the trend towards improving the accuracy and stability of frequency standards will stop, or that applications for these improved standards (both in science and in engineering) will not follow their realization.
- (b) Making use of these new devices will require improvements in distribution methods. Methods based on the phase of the GPS carrier may be adequate—at least initially, but other methods will probably be needed as well. Digital transmission methods—especially methods using optical fibers—are likely to become more important for many of these applications.
- (c) If these devices are to make a contribution to TA1, then incorporating them will also require a better understanding of the transformation from the proper time realized by the device to coordinate time—the exact shape of the geoid, for example, especially at real-world laboratories, which are often located at places (such as Boulder, Colorado) which have inhomogeneous and inadequately documented crustal structure.
- (d) Applications such as electronic commerce, digital notaries, and updating distributed databases need methods for simple, reliable, authenticated, and legally traceable distribution of time information with moderate accuracy on the order of milliseconds. The demand for these services is growing exponentially, and the servers must be protected against attacks by a significant number of malicious users. These requirements for security and reliability are not unique to digital time services, of course, but addressing them while maintaining the accuracy of the time messages presents some unique problems, which may not be completely solved by simply scaling up the current designs.

ACKNOWLEDGMENT

The work on the algorithms and statistics used for network synchronization was supported in part by Grant Nos. NCR-941663 and NCR-911555 from the NSF to the University of Colorado.

General References

The following references provide more detailed information about many of the subjects discussed in this article. They are not specifically referenced in the text.

- Applied Optimal Estimation*, edited by Arthur Gelb (MIT Press, Cambridge, MA, 1992). See especially Chaps. 4 and 9 on Kalman filters and real-time control.
- C. Hackman and D. B. Sullivan, *Am. J. Phys.* **63**, 306 (1995). This resource letter, together with a number of other papers on time and frequency metrology, was reprinted in 1996 by the American Association of Physics Teachers, College Park, MD. The title of the reprint volume is, "Time and Frequency Measurement."
- A. Leick, *GPS Satellite Surveying* (Wiley, New York, 1990).
- F. G. Major, *The Quantum Beat: The Physical Principles of Atomic Clocks* (Springer, New York, 1998).
- D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, *Characterization of Clocks and Oscillators*, NIST Technical Note No. 1337. (U.S. Government Printing Office, Washington, DC, 1990). This is a book of reprints that contains some of the papers cited in the specific reference section. J. Vanier and C. Audoin, *The Quantum Physics of Atomic Frequency Standards*, in 2 volumes. Bristol: (Hilger and Institute of Physics, London, 1989).

Specific References

- ¹D. D. McCarthy, *Proc. IEEE* **79**, 915 (1991).
- ²F. L. Walls, and J. J. Gagnepain, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **39**, 241 (1992).
- ³*Le Systeme International d'Unites*, 7th ed. (Bureau International des Poids et Mesures, Sevres, France, 1998).
- ⁴A. O. McCoubrey, *Guide for the Use of the International System of Units*, NIST Special Publication No. 811 (NIST, Gaithersburg, MD, 1991).
- ⁵T. J. Quinn, *Proc. IEEE* **79**, 894 (1991).
- ⁶C. Thomas, BIPM Report No. 98/2, Sevres, France, 1998.
- ⁷*Annual Report of the BIPM Time Section* (Bureau International des Poids et Mesures, Sevres, France, 1995), Vol. 8, pp. 17–19.
- ⁸Central Bureau of the IERS, Observatoire de Paris, Avenue de l'Observatoire, 61, 75014 Paris, France. See also Ref. 6, p. 5.
- ⁹R. E. Drullinger, W. D. Lee, J. H. Shirley, and J. P. Lowe, *IEEE Trans. Instrum. Meas.* **44**, 120 (1995); R. E. Drullinger, J. P. Glaze, J. P. Lowe, and J. H. Shirley, *ibid.* **40**, 162 (1991).
- ¹⁰N. F. Ramsey, *Phys. Today* **33**, 25 (1980). See also the references to Ramsey's original papers at the end of this article.
- ¹¹A. Kastler, *J. Opt. Soc. Am.* **47**, 460 (1957).
- ¹²J. R. Zacharias, *Phys. Rev.* **94**, 751 (1954).
- ¹³K. Gibble, and S. Chu, *Phys. Rev. Lett.* **70**, 1771 (1993).
- ¹⁴D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. Vessot, H. E. Peters, and J. Vanier, *Phys. Rev.* **138**, 972 (1965).
- ¹⁵W. M. Itano, *Proc. IEEE* **79**, 936 (1991).
- ¹⁶J. D. Prestage, G. J. Dick, and L. Maleki, *IEEE Trans. Instrum. Meas.* **40**, 132 (1991).
- ¹⁷J. J. Bollinger, D. H. Heinzen, W. M. Itano, S. L. Gilbert, and D. J. Wineland, *IEEE Trans. Instrum. Meas.* **40**, 126 (1991).
- ¹⁸D. A. Howe, D. W. Allan, and J. A. Barnes, Proceedings of the 35th Annual Symposium on Frequency Control, 1981, pp. A1–A47. This is document AD-A110870, available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.
- ¹⁹S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe, and L. Erb, Proceedings of the 36th Annual Symposium on Frequency Control, 1982, pp. 314–320. This is document AD-A130811, available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

- ²⁰J. A. Barnes, *Proc. IEEE* **54**, 207 (1966).
- ²¹G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day, San Francisco, 1970).
- ²²S. R. Stein, *Frequency and Time, Their Measurement and Characterization, Precision Frequency Controls* (Academic, New York, 1985), Vol. 2, Chap. 12.
- ²³D. W. Allan, M. A. Weiss, and J. L. Jespersion, Proceedings of the 45th Annual Symposium of Frequency Control, 1991, pp. 667–678. IEEE Catalog No. 91CH2965-2.
- ²⁴P. Lesage, and C. Audoin, *IEEE Trans. Instrum. Meas.* **IM-22**, 157 (1973).
- ²⁵P. Lesage and T. Ayi, *IEEE Trans. Instrum. Meas.* **IM-33**, 332 (1984).
- ²⁶D. A. Howe, Proceedings of the 49th Symposium on Frequency Control, 1995, pp. 321–329. IEEE Catalog No. 95CH3446-2.
- ²⁷D. A. Howe and K. J. Lainson, Proceedings of the 50th Symposium on Frequency Control, 1996, pp. 883–887. IEEE Catalog No. 96CH35935.
- ²⁸R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra* (Dover, New York, 1959), pp. 15–22.
- ²⁹F. L. Walls, D. B. Percival, and W. R. Ireland, Proceedings of the 43rd Annual Symposium on Frequency Control, 1989, pp. 336–341. IEEE publication 89CH2690-6.
- ³⁰P. Kartaschoff, *Proc. IEEE* **79**, 1019 (1991).
- ³¹W. C. Lindsey, F. Ghazvinian, W. C. Hagmann, and K. Dessouky, *Proc. IEEE* **73**, 1445 (1985).
- ³²P. Giacomo, *Metrologia* **18**, 43 (1982).
- ³³P. Giacomo, *Metrologia* **24**, 49 (1987).
- ³⁴Published by the International Organization for Standardization (ISO), Geneva, Switzerland, 1993.
- ³⁵Published annually by the International Bureau of Weights and Measures (BIPM), Sevres, France.
- ³⁶B. E. Blair, *Time and Frequency Dissemination, An Overview of Principles and Techniques*, NBS Monograph No. 140 (National Bureau of Standards, Boulder, CO, 1974).
- ³⁷G. Kamas and M. A. Lombardi, *Time and Frequency Users Manual*, NIST Special Publication 559 (National Institute of Standards and Technology, Boulder, CO, 1990).
- ³⁸W. Lewandowski, and C. Thomas, *Proc. IEEE* **79**, 991 (1991).
- ³⁹P. Wheeler, D. Clamers, A. Davis, P. Koppang, A. Kubig, and W. Powell, Proceedings of the 26th Precise Time and Time Interval Planning and Applications Meeting, 1995, pp. 51–62. This is NASA Conference Publication No. 3302, and is available from NASA Goddard Space Flight Center, Greenbelt, MD 20771.
- ⁴⁰J. Jespersion, in Ref. 29, pp. 186–192.
- ⁴¹D. Kirchner, *Proc. IEEE* **79**, 983 (1991).
- ⁴²G. de Jong and M. C. Polderman, in Ref. 39, pp. 305–318.
- ⁴³M. Kihara and A. Imaoka, Proceedings of the 7th European Time and Frequency Forum, 1993, pp. 83–88, Available from EFTF Secretariat, Rue de l'Orangerie 8, CH-2000, Neuchatel, Switzerland.
- ⁴⁴J. A. DeYoung, W. J. Klepczynski, J. A. Davis, W. Powell, P. R. Pearce, C. Hackmann, D. Kirchner, G. de Jong, P. Hetzel, A. Bauch, A. Soering, P. Grudler, F. Baumont, J. Ressler, and L. Veenstra, in Ref. 39, pp. 39–50.
- ⁴⁵L. Primas, R. Logan, Jr., and G. Lutes, in Ref. 29, pp. 202–212.
- ⁴⁶J. Levine, M. Weiss, D. D. Davis, D. W. Allan, and D. B. Sullivan, *J. Res. Natl. Inst. Stand. Technol.* **94**, 311 (1989).
- ⁴⁷W. Lewandowski, G. Petit, and C. Thomas, *IEEE Trans. Instrum. Meas.* **IM-42**, 474 (1993).
- ⁴⁸R. Gold, *IEEE Trans. Inf. Theory* **33**, 619 (1967).
- ⁴⁹*Understanding GPS, Principles and Applications*, edited by Elliott D. Kaplan (Artech, Boston, 1996), Chap. 4.
- ⁵⁰*Annual Report of the BIPM Time Section* (Bureau International des Poids et Mesures, Sevres, France 1995), Vol. 8, pp. 55–61 and 81–103.
- ⁵¹Y. Bar-Sever, P. Kroger, and J. Borjesson, *J. Geophys. Res.* **103**, 5019 (1998); J. Davis, G. Elgered, A. Niell, and C. Kuehn, *Radio Sci.* **28**, 1003 (1993).
- ⁵²W. Lewandowski *et al.*, Proceedings of the 11th European Frequency and Time Forum, 1997 pp. 187–193.
- ⁵³D. Davis, M. A. Weiss, K. Davies, and G. Petit, Proceedings of the 4th Ion GPS Intl. Technical Mtg., Albuquerque, NM, 1991, pp. 89–95.
- ⁵⁴Interface Control Document, Navstar GPS Space Segment/Navigation User Interfaces, ICD-GPS-200, Seal Beach, CA, Rockwell International Corporation, 1987.
- ⁵⁵J. Levine, *IEEE Ultrason. Ferroelectr. Freq. Control* (submitted).
- ⁵⁶L. E. Young, D. C. Jefferson, S. M. Lichten, R. L. Tjoelker, and L.

- Maleki, Proceedings of the 1996 International Frequency Control Symposium, 1996, pp. 1159–1162.
- ⁵⁷D. C. Jefferson, S. M. Lichten, and L. E. Young, in Ref. 56, pp. 1206–1210.
- ⁵⁸G. Petit and C. Thomas, in Ref. 56, pp. 1151–1158.
- ⁵⁹P. Baeriswyl, Th. Schildknecht, T. Springer, and G. Beutler, Proceedings of the 1996 European Frequency and Time Forum, 1996, 430–435.
- ⁶⁰G. Kramer and W. Klische, in Ref. 52, pp. 331–340.
- ⁶¹K. Larson and J. Levine, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 539 (1998). See also the reference cited in this article.
- ⁶²*Understanding GPS, Principles and Applications*, edited by Elliott D. Kaplan (Artech, Boston, 1996), pp. 190–192.
- ⁶³C. Michael Volk and J. Levine, NIST Technical Note No. 1366, Boulder, National Institute of Standards and Technology, 1994.
- ⁶⁴D. W. Allan, J. E. Gray, and H. Machlan, *The National Bureau of Standards Atomic Time Scale: Generation, Stability, Accuracy and Accessibility*, NBS Monograph No. 140 (National Bureau of Standards, Boulder, CO, 1972), pp. 205–230.
- ⁶⁵M. A. Weiss, D. W. Allan, and T. K. Peppler, IEEE Trans. Instrum. Meas. **38**, 631 (1989).
- ⁶⁶M. Weiss and T. Weissert, Metrologia **28**, 65 (1991).
- ⁶⁷J. Levine, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 629 (1997).
- ⁶⁸P. Tavella and C. Thomas, Metrologia **28**, 57 (1991).
- ⁶⁹S. R. Stein, U.S. Patent Nos. 5, 155, 695 (13 October 1992) and 5, 315, 566 (24 May 1994).
- ⁷⁰R. H. Jones and P. V. Tryon, J. Res. Natl. Bur. Stand. **83**, 17 (1983).
- ⁷¹P. Tavella, J. Azoubib, and C. Thomas, Proceedings of the 5th European Time and Frequency Forum, Neuchatel, Switzerland, 1991, pp. 435–441.
- ⁷²J. Levine, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 450 (1998).
- ⁷³J. Levine, IEEE/ACM Trans. Networking **3**, 42 (1995).
- ⁷⁴J. A. Kusters, R. P. Giffard, L. S. Cutler, and D. W. Allan, in Ref. 39, pp. 89–94.
- ⁷⁵D. L. Mills, “Measured Performance of the Network Time Protocol in the Internet System,” Request for Comments 1128. Available on the Internet via ftp to nic.ddn.mil
- ⁷⁶J. Levine, Proc. IEEE International Frequency Control Symposium, 1998, pp. 241–249. This document is IEEE Catalog No. 98CH36165.
- ⁷⁷J. Levine, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (in press).