Measuring Time and Comparing Clocks

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

Table of Contents

1. Abstract

I will discuss methods of comparing and synchronizing clocks and the procedure of characterizing their performance in terms of the two-sample Allan variance. I will describe methods that are used when the device under test and the reference device are in the same facility and when they are at different locations linked by a communications channel. The reference device can be a national time scale, and I will briefly describe how national time scales are defined. In each case I will use the Allan variance as a tool to understand the characteristics and limitations of the various methods. I will also introduce the concepts of traceability and of a cost-benefit analysis into the synchronization procedure.

2. Introduction

Time and time interval have played important roles in all societies since antiquity. The original definitions were based on astronomy: the solar day and year and the lunar month were widely used as measures of both time and time interval. As I will show below, the strong connection between astronomy and time persists even today, when both time and time interval are measured by means of clocks. I will begin by describing a generic clock and I will then discuss various means of comparing these devices and characterizing their performance using a combination of deterministic and stochastic parameters. I will conclude with a short description of calibrating them in terms of international standards of time and frequency.

3. A Generic Clock

All clocks consist of two components: a device that produces or observes a series of periodic events and a counter that counts the number of events and possibly also interpolates between consecutive events to improve the resolution of the measurement. The choice of which periodic event to use as the reference period for the clock plays a fundamental role in determining its performance, so that it is natural to characterize a particular clock design based on an evaluation of the reference period that it uses to drive its counter.

In addition to the two components discussed in the previous paragraph, real clocks and time scales have a time origin that is derived from some external consideration. As a practical matter, the time origin is generally chosen to be sufficiently far in the past so that most epochs of interest have positive times with respect to the origin.

In addition to a time origin, real time scales are used to construct a calendar – an algorithm that assigns names to clock readings. These considerations are very important, but are mostly outside of the scope of this discussion. Although I will not discuss the methods used to implement a calendar, I will discuss the methods that are currently used to define Coordinated Universal Time (UTC), and the discussions that are currently underway (as of 2013) about possibly modifying the definition of this time scale.

Two distinct parameters are important in characterizing the frequency reference of any clock: (1) The accuracy of the reference period -- how closely does the period conform to the definition of the second. (2) The stability of the reference period over both the short and long terms. (Stability is a necessary pre-requisite for an accurate device, but is not sufficient, and it is quite common for real oscillators to have a stability that is significantly better than the accuracy.) A number of methods have been developed for characterizing the stability of periodic events, and I will briefly describe the tools that implement these methods in the next section. I will discuss the question of accuracy in a subsequent section.

4. Characterizing the Stability of Clocks and Oscillators

The methods that are used for these purposes fall into two general classes: methods that characterize the worst-case performance of a clock and methods that characterize the average performance using statistical parameters derived from a root-mean-square calculation. In both cases, the analysis is based on a finite-length data set.

A worst-case analysis is usually sensitive both to the length of the data set that is analyzed and to the exact interval of the observation because large glitches usually occur sooner or later, and a data set either includes a large glitch or it doesn't. We might expect that the results of a worst-case analysis would show a large variation from sample to sample for this reason.

A statistical analysis, on the other hand, assumes implicitly that the data are stationary, so that neither the interval of observation nor the length of the data set is important in principle. A statistical analysis tends to attenuate the effect of a glitch, since even a large glitch may have only a small impact on an ensemble-average value. More generally, a statistical analysis is not suitable if the data are not stationary, since ensemble-average values will exist in a formal sense but will not be very useful in understanding the performance of the actual device.

In order to characterize the stability of a device under test, we can imagine that it is compared to a second clock that is perfect. That is, the perfect clock produces "ticks" that are exactly uniform in time. The interval between ticks of the perfect clock is τ, so that the ticks occur at times 0, τ, $k\tau$, $(k+1)\tau$, etc. where k is some integer. (As I discussed above, the times are relative to some origin that is defined outside of the measurement process, and the time of the first tick is 0 with respect to that origin. This is not a limitation in the current discussion, since the origin is simply an additive constant that is not important when discussing time stability.) The time of the clock under test is read each time the standard clock emits a tick, and the time differences are x_k, x_{k+1}, …, where x_k is a short-hand notation for the time-difference reading at time $k\tau$, etc. In general, the units of time are seconds and fractions of a second. The "frequency" of a clock is the fractional frequency difference between the device under test and a perfect clock operating at the same nominal frequency. For example, the frequency of the device under test, f, which generates a signal at a physical frequency of F is

$$f = \frac{F - F_o}{F_o} \quad , \tag{1}$$

where F_o is the output frequency of the perfect device used in the comparison. With this definition, the frequency of a clock is a dimensionless parameter and the time difference after a time τ has changed by $f\tau$.

4a. Worst-case Analysis

Analyzing the time difference data from the perspective of a worst-case analysis sounds easy. We simply look for the largest absolute value of *x* in the data set, assuming that the device under test and the standard reference device were set to the same time at the start of the measurement. A statistic that realizes this idea is MTIE[1], the maximum time-interval error, which is usually calculated as the difference between the largest and smallest time differences in the ensemble of measurements. (With this definition, a device that has an arbitrarily large and constant time difference has an MTIE value of 0, because MTIE is a measure of the evolution of the time difference, not the magnitude of the time difference itself. In this respect, the MTIE statistic is really a measure of the frequency offset between the device under test and the standard reference.) It is commonly used to characterize the oscillators in telecommunications networks.

The MTIE statistic depends both on frequency accuracy and frequency stability, since a clock with a frequency offset with respect to the reference device will eventually produce unbounded time differences even if the frequency offset is absolutely stable. As we mentioned above, the results of a worst-case analysis can show large variations from one set of data to another one, so that the MTIE statistic is normally defined as the largest MTIE value over all data sets of some fixed length within some larger time interval, T.[2] In an extreme case, the value of T is the life of the device, so that a device that exhibited one glitch after years of faithful service might not satisfy a specification based on MTIE in this case. This form of the definition is therefore unambiguous but not very useful.

An alternative view is to consider the MTIE value obtained from a single data set to be an estimate of the underlying "true" value of MTIE, which is characterized by the standard statistical parameters of a mean and a standard deviation. The standard statistical machinery can then be used to provide an estimate of the probability that the observed value is (or is not) consistent with a mean value for MTIE that is specified by some required level of performance. Since this analysis method characterizes MTIE as a statistical parameter, it usually requires some ancillary assumption about handling measurements that are not consistent with the mean and standard deviation of the distribution of the remainder of the observations. In other words, is a large outlier (1) treated as an error, which should be ignored, (2) accepted as a low-probability event that is consistent with the mean and standard deviation deduced from previous data, or (3) an indication that the mean or the standard deviation should be updated by including this new observation? One solution is to completely reject data that differ from the mean by more than four standard deviations and to provisionally accept data that differ by more than three but less than four standard deviations. The data in the provisional category are compared to subsequent values and all of these newer data may be used to provide an update to the mean or to the standard deviation as appropriate. The specific algorithm that is used is based more on administrative considerations and experience than on a rigorous statistical analysis, since errors are generally not statistical events by definition.

The evolution of the time difference data that are the input to an MTIE calculation are sensitive both to the frequency stability of the device under test and to the sampling interval, τ, and the data become increasingly insensitive to fluctuations in the frequency of the device under test with respect to the reference device that are much shorter than the sampling period. A fluctuation in the offset frequency whose period is an exact sub-multiple of the sampling period has not effect on MTIE. Therefore, the sampling period must be short enough so that these shorter-period fluctuations either are not important in the application supported by the device or are known to be small a-priori. Since the length of a data set is limited in practice, this consideration implies a trade-off between the shortest and longest frequency fluctuations that can be estimated from a set of measurements. However, a measurement that uses a longer sampling interval to detect longer-period fluctuations must guarantee (by means of digital or analog filtering) that the shorter-period fluctuations are not aliased by the longer sampling period.

Another way of addressing the aliasing problem is to use a very rapid sampling period (so as to minimize or eliminate the impact of aliasing) but to acquire these measurements in blocks separated by dead-time in which the clock is not observed. This method will also have a potential aliasing problem for frequency fluctuations that are synchronous with the sum of the sample period and the dead time. This problem can be addressed by varying the dead-time in a pseudo-random way, but this complicates the analysis somewhat, since the blocks are no longer equally spaced in time.

I will now describe the statistical estimates of time and frequency that are commonly used to characterize clocks and oscillators outside of the telecommunications domain. I will limit my discussion to the original Allan variance, since its significance can be explained intuitively. The more complicated versions of the Allan variance and the characterizations in the frequency domain using Fourier analysis are described in the literature. (See reference 3)

4b. Statistical Analysis and the Allan Variance

The statistical analysis starts with the same time differences that we discussed in the previous sections. The average frequency of the clock under test with respect to our perfect clock over the time interval τ between measurements is estimated as

$$y_k = \frac{x_k - x_{k-1}}{\tau}. \qquad (2)$$

The numerator and denominator on the right hand side of eq. 2 have the units of time, so that the frequency defined by this equation is a dimensionless quantity. If the device under test had a frequency that was constant with respect to the perfect clock, then equation 2 would give the same result for any value of k. (Note that the device under test need not have the *same* frequency as the perfect clock. We would get the same result for every value of k even if the frequency difference was *any* constant value.)

Real event generators are not perfect, and it is useful to characterize their performance by means of the estimator

$$y_{k+1} - y_k = \frac{x_{k+1} - 2x_k + x_{k-1}}{\tau} \quad . \tag{3}$$

Equation 3 gives the difference in the frequency of the device under test between two consecutive, equal measurement intervals with no intervening dead time. This estimator provides an estimate of frequency stability – not frequency accuracy. From the perspective of the measurement at time τ_k, this statistic is an estimate of the time difference that will be observed at the next measurement time with index *k+1* based on the evolution of the time difference in the time interval ending at index *k*. Its magnitude is not sensitive to a constant time difference *or* frequency difference between the device under test and the perfect reference device. (Compare this to MTIE, which was discussed in the previous section and which *is* sensitive to a constant frequency difference but not to a constant time difference.)

The final step in the definition of the estimator is to assert (or to hope) that the variations estimated by equation 3 are stationary. That is, the computation does not depend in a systematic way on the value of the index *k* – any choice of *k* would produce a value that is consistent (in a statistical sense) with the result for any other choice of *k*. Then, the root-mean-square (RMS) of equation 3 has a well-defined value, and that RMS value has an associated, well-defined, standard deviation. When various normalizing constants are added, the mean square value of equation 3 estimated over all possible values of *k* is the two-sample or Allan variance for an averaging time of τ, and the RMS value is the two-sample or Allan deviation for that averaging time. If there are *N* time difference data with indices *1, 2, …, N,* then the Allan variance at averaging time τ is defined as the average of the *N-2* calculations as

$$\sigma_y^2(\tau) = \frac{1}{2(N-2)\tau^2} \sum_{k=2}^{N-1} (x_{k+1} - 2x_k + x_{k-1})^2 \ , \tag{4}$$

and the Allan deviation is the square root of this value. Since the time-difference data are equally spaced in time, we can use the same data to compute the estimate of the Allan variance for any multiple of the sampling interval, τ:

$$\sigma_y^2(m\tau) = \frac{1}{2(N-2m)(m\tau)^2} \sum_{k=2}^{N-2m+1} (x_{k+2m-1} - 2x_{k+m-1} + x_{k-1})^2 \ . \tag{5}$$

The normalization is defined so that the Allan variance has the same value as the classical variance in the case of a random time-difference noise process, which we will discuss below. This type of process is often called "white phase noise." The number of terms that contribute to the sum in eq. 5 decreases as *m* is made larger, so that the estimates for large values of *m* are likely to exhibit more variation from one set of observations to another one. The maximum value of *m* that is used in eq. 5 is often limited to *N/3* for this reason.

It is important to emphasize that the two-sample Allan deviation is a measure of frequency *stability* – not frequency accuracy. Frequency stability is obviously a very desirable quality for a clock, and many real-world devices are characterized in this way. It is clearly not sufficient to characterize devices that are used to provide standards of time, time interval, or frequency.

A very powerful technique is to examine the dependence of the Allan variance on the averaging time, τ, because this dependence can provide insight into the noise processes that drive the magnitude of the Allan deviation. To take a simple example, suppose that the device under test has a true constant frequency offset with respect to the perfect device used for the calibration. If this constant fractional frequency offset is *f*, then the measured time differences at times $k\tau$ in the absence of any noise processes would be

$$x_k = fk\tau + x_0 , \tag{6}$$

where x_0 is the time difference between the device under test and the perfect clock when k=0. Then eq. 2 would estimate the average frequency as

$$y_k = \frac{x_k - x_{k-1}}{\tau} = f , \tag{7}$$

which is independent of *k*, so that the estimate of the Allan variance is 0 for all averaging times. As expected, the Allan variance is 0 for a clock with a constant offset frequency and it provides no information on the magnitude of this frequency.

A more interesting example is to suppose that the device under test had a constant frequency offset as in the previous example but that the time difference measurements are affected by a random noise process that might originate in the measurement hardware and not in the clock itself. The time difference measurements in this case would be given by

$$x_k = fk\tau + x_0 + \varepsilon_k \qquad , \tag{8}$$

where the noise contribution is characterized by a zero-mean signal with a well-defined variance:

$$\begin{aligned} \langle \varepsilon_k \rangle &= 0 \\ \langle \varepsilon_j \varepsilon_k \rangle &= \sigma^2 \delta(j-k) \end{aligned} \tag{9}$$

and δ is the Dirac delta function, which is 1 if its argument is 0 and 0 otherwise. The deterministic contributions to the summation in equations 4 or 5 cancel as in the previous example, and what is left is a sum that is proportional to the variance of the noise process but independent of the summation index, *k*. The Allan variance therefore decreases as the reciprocal of the square of the sampling interval, and this conclusion does not depend on the magnitudes of either the deterministic or stochastic contributions, provided only that the data satisfy equations 8 and 9. A plot of the logarithm of the Allan variance as a function of the logarithm of the measurement interval would have a slope of -2. If the frequency of the oscillator was not exactly constant but varied with the index *k*, then the summations in equations 4 or 5 will contain a contribution from the deterministic terms in the time differences that is some function of the interval between the measurements. The log-log plot will have a slope that is greater than -2, and the exact value of the slope will depend on the details of the noise process that is driving the frequency fluctuations. In general, the slope of the log-log plot of the Allan variance as a function of the measurement interval is an indicator of the type of noise process that is contributing to the measured time differences. The details of this relationship and the usefulness of other variances that are related to the simple Allan variance discussed here are described in the literature.[3]

The frequency dispersion estimated by the Allan deviation generates a corresponding time dispersion. The time dispersion is usually called $\sigma_x(\tau)$, and it is formally defined in terms of the Modified Allan Variance, which is described in reference 3.

We can provide a simple estimate of the time dispersion in the case of the simple white phase noise example that we considered in the previous paragraph. If we correct the measured time differences by the deterministic equation 6, the residual time dispersion for any measurement is simply ε_k, a random process defined by equations 8 and 9. If we substitute equations 8 and 9 into equation 4, then the summation is simply a sum of *N*-2 identical terms, and

$$\sigma_y^2(\tau) = 3\frac{<\varepsilon_k^2>}{\tau^2} \tag{10}$$

or

$$\sqrt{\left\langle \varepsilon_k^2 \right\rangle} = \frac{\sigma_y(\tau)\tau}{\sqrt{3}} \quad , \tag{11}$$

which relates the statistical RMS time dispersion to the Allan deviation for the case of white phase noise. Since the Allan deviation for white phase noise varies as the reciprocal of the measurement interval, the estimate of the RMS time dispersion does not depend on the interval between measurements. This is not a surprising result, since it follows directly from the assumption of the statistics of the noise process defined in eq. 9. A more conservative estimate is to take the RMS time dispersion as simply the product of the time interval between two measurements and the Allan deviation at that time interval, and I will generally use this more conservative estimate in the following discussion, since the constant in the denominator of eq. 11 is valid only for white phase noise.

4c. Limitations of the Statistics

Both MTIE and the Allan variance (including a number of variants that we have not discussed) are measures of stability, but they define stability in different ways. Neither statistic is sensitive to a constant time offset, but the value of MTIE is sensitive to a constant frequency offset whereas the value of the Allan variance is not. Therefore, the Allan variance is a measure of the predictability of the future time-difference of a clock based on its past performance and arbitrarily large *constant* frequency offsets do not degrade the prediction.

However, a clock with some other *deterministic (but not constant)* frequency offset produces time-difference values that are just as well-determined as a clock with only a constant frequency offset, but the Allan variance treats the two very differently. The frequencies of many types of oscillators (hydrogen masers, for example) can be approximated as varying linearly with time, and the Allan variance of the time difference measurements from such devices (which have a quadratic dependence on the time) does not give a realistic estimate of their stability if, by stability we mean how well can a future time difference be estimated based on previous performance. For this reason, it is common to estimate and remove a quadratic function of the time from these data before they are analyzed to compute the Allan variance. This process of “pre-whitening” the data is well known in the statistical literature and is often implemented using a Kalman filter[4] and in power-spectral analysis[5].

Many oscillators have frequencies whose fluctuations can be more easily (and intuitively) characterized in the Fourier frequency domain rather than as a stationary process in the time domain, which is basic assumption of the Allan variance. For example, an oscillator whose frequency was sensitive to ambient temperature could be expected to exhibit a diurnal frequency variation if it was operated in an environment without tight control of the ambient temperature. The Allan variance calculation will model these time differences as a stationary noise process, and it usually reports a large value at a time interval corresponding to roughly one-half of the period of the driving process. This is not too difficult to interpret correctly if there is only one such contribution, but it can be ambiguous if there are several “bright lines” in the power spectrum of the time differences, and pre-whitening the data is particularly important in this case.

Finally, it is important to keep in mind that the calculation of the Allan variance is based on a particular method for averaging the time differences. Therefore, while the *slope* of a log-log plot of the Allan deviation as a function of averaging time provides insight into the underlying noise processes that drive the time differences, the *value* of the Allan deviation at any averaging time is useful as an indicator of the performance to be expected from the device only if the clock is used in a manner that is consistent with the averaging procedure that is part of the definition. For example, the definition of the Allan variance that we have used is based on data that are equally spaced in time with no dead time between the measurements. There are other statistics that are related to the simple

Allan variance we have discussed (Mod Avar, the Modified Allan variance, as an example, and its close relative TVAR, the time variance), and these statistics have more complex averaging procedures, which are less likely to be used in a real application. Although the simple Allan variance has some limitations in identifying the underlying noise type in some circumstances, its definition is often closer to the way a device will actually be used, and it is often the preferred analysis tool for this reason.

5. Characteristics of Different Types of Oscillators

The frequency stability of an oscillator can be realized either actively, where the discriminator is actively oscillating at a resonant frequency derived from its characteristics, or passively, where the frequency of an oscillator is locked to the resonant response of a passive discriminator.

A quartz crystal oscillator is generally an active device, because the frequency is determined by the mechanical resonance in the quartz that is excited by an external power source. Atomic frequency standards fall into both categories. Atomic clocks that use a transition in cesium or rubidium as the frequency reference generally fall into the second category, where the discriminator is passive and is interrogated by a separate oscillator that is locked to the peak of the transition probability of the clock transition. Hydrogen masers can be either active or passive. In both types of devices, the actual output frequency is generally a function both of the resonant frequency of the discriminator and the method that is used to interrogate it. The parameters of the interrogation method may have a dependence on ambient temperature or may vary with time, so that even nominally identical oscillators generally have different output frequencies. (A primary frequency standard is designed so that these perturbing influences are minimized, and the residual perturbations are estimated by means of various ancillary measurements.)

The vast majority of oscillators currently in use are stabilized using a mechanical resonance in a quartz crystal. Newer devices are stabilized using a micro-electro-mechanical (MEMS) device.[6] The quartz crystals used as the frequency reference in inexpensive wrist watches can have a frequency accuracy of 1 ppm and a stability of order 10^{-7}. (A frequency accuracy of 1 ppm translates into a time dispersion of order 0.09 s/day.) These devices are generally sensitive to the ambient temperature, so that substantially better performance can be realized using active temperature control. The moderate stability of the frequency is exploited in many control applications and in the operation of Internet time servers, as we will discuss below. In spite of some very clever techniques that have been developed to improve the long-term stability of the frequency of these devices, none of the methods can totally eliminate the sensitivity of the frequency to environmental perturbations, to stochastic frequency fluctuations that are hard to model, and to a dependence on the details of manufacture that are hard to replicate.

Atomic frequency standards use an atomic or molecular transition as the frequency discriminator in an attempt to address the limitations of the frequency stability of mechanical devices that I discussed in the previous paragraph. The atoms can be used in a passive or active configuration. In the passive configuration, the atoms are prepared in the lower state of the clock transition and are illuminated by the output from a separate variable-frequency oscillator. The frequency of the oscillator is locked to the maximum in the rate of the clock transition. There are a number of different methods for preparing the atoms in the lower state and for detecting the transition to the upper state. The details are described in the literature.[7]

In the active configuration, the atoms are pumped into the upper state, and the radiation emitted when they decay to the lower state stimulates other atoms to decay by stimulated emission. The oscillating frequency is determined by the atomic transition frequency and by the properties of the cavity that is used to trap the radiation and provide the ambient field that induces stimulated emission. Most lasers and some hydrogen masers work this way. The cavity of an active hydrogen maser is generally tuned to improve the stability of the output frequency.[8]

In both active and passive hydrogen masers, the output signal is generated by an oscillator (typically a quartz-crystal device) whose frequency is locked to the transition frequency of the atoms. For sufficiently short averaging times (typically less than about 0.1 s), the stability of the output frequency is determined by the properties of the quartz oscillator and by the process used to generate the output ticks. The spectrum is generally white phase noise. At longer averaging times (greater than about 10 s), the stability of a maser is generally limited by the thermal noise in the oscillating field in the cavity and in the control loop that is used to lock the output oscillator. This is usually white frequency noise. Other types of atomic standards are generally operated in the passive configuration and have similar statistics.

6. Comparing Clocks and Oscillators

Comparing the times of two clocks is a simple process in principle, but the process becomes more complicated as the resolution of the measurement increases. I will discuss two classes of comparisons. In the first situation, the devices to be compared are in the same laboratory, so that the signals from both clocks are available locally and we don’t have to consider the characteristics of a transmission network. In the second configuration, one of the clocks is at a remote location, so that the statistics of the time comparison must include the effects of the transmission network.

The simplest time comparison is simply a one-time measurement. We read the time difference between the two clocks, using a time-interval counter, for example, and we use the measurement to adjust the time of one of them either by making an adjustment to its physical output or by noting its time offset for future administrative corrections. The implication of the process is that the reference clock is much more accurate than the device whose time difference we are measuring, and we need not consider the possibilities that the reference clock is broken or that the time comparison had a significant measurement error. Furthermore, this simple “set it and forget it” process does not provide any insight into the statistics of the device under test, so that we have no way of knowing how rapidly its time will diverge from the correct time after it has been set.

Thus, we have no way of knowing how often to repeat the measurement process in order to have the device being calibrated maintain some specified level of accuracy. These considerations lead to a more sophisticated measurement program.

To simplify matters, we will again assume that the clock under test is being compared to a second clock that is so much more accurate that it can be considered as perfect from the perspective of the measurement process. We will model the time differences of the clock under test by a combination of deterministic and stochastic parameters. The deterministic time differences are generally specified in terms of three parameters, the initial time difference, the frequency offset and the frequency aging. This formulation leads to a quadratic relationship with constant parameters whose independent variable is the elapsed time since the start of the measurement. This formulation is not adequate for most real devices and measurement processes.

In the first place, the frequency offset and frequency aging are not manifest constant parameters for any real device, and treating them as constants is neither adequate nor optimum. In addition, many applications depend on real-time estimates of the parameters, and it becomes increasingly cumbersome to re-evaluate a static quadratic form of the modeled time differences each time a new data point is measured. This type of analysis also requires that we save all of the measurements since the start of the experiment. Therefore, most analyses use an iterative form of the estimate, in which the current time difference is modeled based on the parameters estimated from the previous measurements. The previous measurements themselves are not needed.

In this method, we estimate the time difference at time t_k in terms of the previous data as

$$\hat{x}_k = x_{k-1} + y_{k-1}\Delta t + \frac{1}{2} d_{k-1}(\Delta t)^2 \qquad (12)$$

where x_k, y_k, *and* d_k are the time difference, frequency offset, and frequency aging at time t_k, and $\Delta t = t_k - t_{k-1}$. It is generally easiest to use measurements that are equally spaced in time, but the formulation of equation 12 is valid whether or not this is the case.

The measurement process measures the time difference at time t_k and returns the measured value X_k. The difference between the value we predicted and the value we observed is

$$\delta_k = X_k - \hat{x}_k . \qquad (13)$$

In the absence of any noise contributions, and assuming that the initial time difference, the frequency offset, and the frequency aging are perfectly constant values, eq. 12 has only three parameters, so that the time differences in equation 13 will converge to zero after three measurement cycles even if we are totally ignorant of the initial values of these parameters. But now we return to the real world, where the measurements have a stochastic component and the clock parameters are not absolutely constant. The goal of a real-world measurement process is to partition the time differences obtained in eq. 13 into a deterministic portion, which we use to update our estimates of the parameters in eq. 12 and a stochastic contribution, which we attenuate by averaging or ignore completely.

The measurement process in the previous paragraph cannot succeed in the most general case because there is only one observable (eq. 13) and multiple parameters that must be determined. The process can succeed in practice because the deterministic parameters in eq. 12 change slowly with time so that it is possible to treat them as substantially constant over short averaging times. Stated another way, the process of modeling the time differences by means of equation 12 will succeed if and only if the time interval between the measurements is short enough to validate this assumption. In the next section, I discuss this requirement quantitatively.

7. Noise Models

As I mentioned in the previous section, the limitations of the method come from our ignorance of the partition of the measured data into stochastic and deterministic components. To get insight into the characterization of the noise, it is useful to model an oscillator as a passive resonant system (such as an atomic transition) that is interrogated by a separate oscillator. The frequency of the oscillator is locked to the peak in the resonance response, and the output of the oscillator is used to generate the “ticks” that are used to drive the time display. Most oscillators that are stabilized using a transition in cesium or rubidium are configured this way. Lasers and most crystal oscillators cannot be modeled so easily because the oscillation frequency depends on a complicated combination of the gain of an amplifier and the phase shift in a resonant feed-back loop. Nevertheless, even these types of oscillators can be reasonably-well characterized using the machinery that I will discuss in the following sections.

7a. White Phase Noise

A common method for generating the ticks is to generate an output pulse each time the sine wave of the oscillator passes through zero with a positive slope. Real zero-crossing detectors have some uncertainty in the exact trigger point, and this uncertainty is often represented as an equivalent noise voltage at the input to the circuit. I will designate this noise voltage as V_n. This noise contribution is modeled as a zero-mean random process that is unrelated to the true input signal. If the output of the oscillator has amplitude A and an angular frequency ω, then the noise voltage, whose amplitude is much smaller than the amplitude of the signal, introduces a time jitter in the determination of the time of the zero-crossing whose amplitude is

$$\Delta t = \frac{V_n}{A\omega} \quad . \tag{14}$$

This fluctuation in the times of the output pulses is not associated with any frequency fluctuations in the oscillator itself. The noise voltage is inherent to the discriminator and has nothing to do with the input signal, so that the magnitude of the time fluctuation for any measurement is unrelated to the impact for any other measurement. As we showed in the previous section, the Allan deviation for this type of noise varies as the reciprocal of the interval between measurements. Since the fluctuation does not arise in the oscillator itself, correcting the apparent frequency jitter caused by this noise by steering the parameters of the oscillator (its frequency, for example) is not the optimum strategy. I will discuss this point in more detail later. For now we note that the time jitter defined by eq. 14 has a mean of zero. From the physical perspective, there is an underlying “true” time difference, which can be estimated by averaging multiple measurements. The standard deviation of the estimate decreases as more measurements are performed, and the mean value at any time is an unbiased estimated of the true time difference.

7b. White Frequency Noise

We next consider the control loop that locks the oscillator frequency to the peak of the resonance response. A common method of detecting the peak in the response is to lock the oscillator to the zero of the first derivative of the resonance response function. The first derivative signal is generated by applying a small modulation to the oscillator frequency and synchronously detecting the amplitude of the response of the resonant system at the frequency of the applied frequency dither. An alternate method locks the oscillator to a frequency that is between two frequencies above and below the resonance where the upper and lower frequencies are measured as the values where the response has fallen to some specified fraction of the peak. (This is often implemented by means of a zero-mean, bipolar, square-wave dither of the oscillator frequency.) The input to the discriminator is the difference between the response at the higher frequency and the response at the lower one. Both methods introduce a deterministic dither in the frequency output of the oscillator, which must be removed before the signal is used.

In either method, the discriminator locks the oscillator to a point where some voltage goes through zero. When the frequency is locked, the magnitude of the voltage specifies how much the frequency differs from the desired lock point and the sign of the voltage specifies whether the frequency output is higher or lower than the desired operating point.

The response of this discriminator is limited by the same noise problems that I discussed in the previous section, but now it is the frequency of the oscillator rather than the output of a pulse that is affected by the equivalent noise at the input of the discriminator. The same argument as in the previous section shows that the result is a random frequency modulation. As in the previous section, the noise contribution is assumed to be a zero-mean random process, so that frequency jitter is about a “true” value. This noise is identified by the slope of -0.5 in the log-log plot of the Allan deviation.

These random frequency fluctuations are integrated to produce time dispersion. Since the frequency fluctuations are characterized by a random process with a mean of zero, the impact on the time differences is a random walk with a variable step size, and so white frequency fluctuations are often described as a random walk in time (or phase, which is the same thing). Unlike the white phase noise case discussed in the previous section, the impact of white frequency noise on the measured time differences depends on the averaging time through eq. 12.

Since the impact of white frequency noise on the measured time differences is a function of the averaging time whereas white phase noise is independent of it, at least in principle it is possible to distinguish between the two by an appropriate choice of the averaging time; white frequency noise can be neglected at very short averaging times and white phase noise becomes negligible as the averaging time is increased.

7c. Long-period Effects – Frequency Aging

The effects I have discussed in the previous sections are modeled as being driven by stochastic noise processes that are assumed to have zero means. That is, there is an underlying "true" time difference in the white phase noise domain and there is an underlying "true" frequency offset in the white frequency noise domain. These models are useful because the behavior of many oscillators is well characterized in these terms.

Although there are oscillators that also exhibit a stochastic frequency aging that can also be modeled as a random zero-mean process just as we did above for time and for frequency, the frequency aging of many oscillators is often a combination of a random function that is only approximately characterized as a zero-mean process combined with an approximately constant aging value.

There are a number of processes that can produce frequency aging that varies only very slowly and is approximately constant over many measurement cycles. For example, the transition frequency that is used as the reference frequency in an atomic clock is generally sensitive to external electric and magnetic fields (the dc Stark and Zeeman effects, respectively), to collisions between the atoms, to frequency shifts that result from the interaction with the probing field (the ac Stark effect), and to many other effects. These perturbing influences may have long-period variations, which are translated into long-term frequency aging of the oscillator. The mechanical properties of quartz crystals, which determine the resonant frequency of a quartz-crystal oscillator, often have similar long-period changes. Stochastic frequency aging may be caused by more rapid variation in any of these parameters and in other effects whose quantitative driving term admittance is not known.

From equation 12, we can see that a constant frequency aging would produce time dispersion proportional to the square of the time interval between the measurements. Based on the relationship between time dispersion and Allan deviation, the Allan deviation for this type of aging would have a slope of +1 on a log-log plot of Allan deviation with respect to averaging time. The plot of the log-log plot of the Allan deviation with respect to averaging time has a slope of +0.5 for stochastic frequency aging.

Since the time dispersion due to frequency aging varies as the square of the interval between measurements, it is often possible to partition the measured variation in the time differences into three domains – a very short domain where white phase noise dominates the variance, an intermediate domain where frequency noise is the main contributor and a long-period domain where frequency aging dominates.

7d. Flicker Noise

In addition to the white time, frequency, and aging processes that I discussed in the previous sections there is another contribution to the variance of time-difference measurements that cannot easily characterized using a simple model analogous to the ones presented in the previous sections. I will characterize this type of noise process by contrasting it to the white phase noise and white frequency noise processes that I discussed above.

If a time-difference measurement process can be characterized as being limited by pure white phase noise, then there exists an underlying "true" 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 always 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.

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.

Suppose we measure the time differences between a perfect device and the device under test, where the device under test has the same nominal frequency as the perfect device, but its frequency stability is degraded by white frequency noise. If the time difference between the two devices at some epoch is X(t), then, since the deterministic frequency difference is zero, we would estimate the time difference a short time in the future as

$$X(t+\tau) = X(t) + y(t)\tau \quad (15)$$

where y(t) is the instantaneous value of the white frequency noise of the device under test, and

$$\langle y(t) \rangle = 0 \, , \quad (16)$$

$$\langle y^2(t) \rangle = \varepsilon^2 \quad . \quad (17)$$

Since $y(t)$ has a mean of 0 by assumption, eq. 15 predicts that the time difference at the next instant will be 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

$$X(t+\tau) = X(t), \tag{18}$$

but rather that

$$\left\langle X(t+\tau) - X(t) \right\rangle = 0 \quad , \tag{19}$$

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 simple random variable, and this smoothness is often mistaken for a pseudo-deterministic variation. Flicker phase noise is intermediate between white phase noise and white frequency noise, and we would therefore expect that the Allan deviation of the time-difference data would have a dependence on averaging time that is midway between white and random-walk processes. In fact, the simple Allan variance that we have discussed cannot distinguish between white phase noise and flicker phase noise. The more complicated Modified Allan Variance (ref. 3) is needed for this purpose, and the slope of this variance for flicker phase noise is indeed midway between the slopes for white and random-walk processes.

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.

The slopes of the log-log plot of both the Allan deviation and the Modified Allan deviation are zero for flicker frequency noise. In other words, the estimate of the frequency does not improve with longer averaging times. The Allan deviation for these averaging times is often called the “flicker floor” of the device for this reason.

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. For example, suppose that the optimum averaging time is T. Then, if Y_{k-1} is an estimate of some parameter at time t_{k-1}, and if y_k is the new data point received at time t_k, then we would estimate the update to Y at time t_k by means of

$$Y_k = \frac{wY_{k-1} + y_k}{w+1} \quad , \tag{20}$$

where w is a dimensionless parameter given by

$$w = \frac{t_k - t_{k-1}}{T} \quad . \tag{21}$$

The time interval between measurements is chosen so that $w \le 1$ – there is at least one measurement in the optimum averaging time. This method is often used in time scales, since these algorithms are commonly implemented recursively. Both the recursive and non-recursive 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.[9]

8. 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 – an exact power of 2, which simplifies 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 volts 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 waveform.

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.[10] 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 waveforms. 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 would have to be at least 1 GHz to realize this resolution without interpolation.

In addition to the limit on the resolution of time-difference measurements that results from the maximum rate of the oscillator that drives the time-interval counter, 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 the rise time by a significant amount. In addition to distorting the waveforms 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 and so have a mean that is not zero over short times.

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 that 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. As usual, we assume that any fluctuations in the reference oscillator are small enough to be ignored.

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 "dual-mixer" version of this idea developed at NIST[11], 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. Since the heterodyne process preserves the phase difference between the two signals, a phase measurement of 10^{-6} of a cycle is equivalent to a time interval measurement with a resolution of 0.2 ps at the 5 MHz input frequency. It would be very difficult to realize this resolution with a system that measured the time difference directly at the 5 MHz input frequency or by measurements of the 1 Hz pulses derived from this reference frequency by a process of digital division.

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 – 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 NIST system is only a few ps -- about a factor of 10 or 20 poorer than its resolution. Some of the factors that degrade the stability of a channel in a dual-mixer system are common to all of the channels in a single chassis, so that the differential stability of a pair of channels (which is what drives an estimate of the Alan variance) can be better than the stability of each channel alone.

Heterodyne methods are well suited to evaluating the frequency stability of an oscillator, but they often have problems in measuring absolute time differences because they usually have an integer ambiguity offset -- an unknown integer number of cycles of the input frequencies between the cycles that trigger the measurement system and the cycles that produce the 1 Hz output pulses that are the output "on-time" signals. For example, the time difference between two clocks measured using the NIST dual-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 generally changes 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 time 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.

9. Measurement Strategies

In the previous section, I discussed a number of methods that can be used to measure the time differences between two devices. All measurement techniques have some residual noise, which appears as jitter in the time difference measurements. I will assume for now that this jitter can be characterized as white phase noise. That is, it is a pure random process, and the impact on any measurement can be fully characterized by a distribution with a mean of zero and a standard deviation that is a property of the measurement system and does not depend on temperature, aging, or any of the other limitations that affect most real-world systems. Specifically, the noise of the measurement process can be characterized by relationships similar to eq. 9, above.

With this assumption, the optimum strategy for estimating the time difference between a device under test and a second device which we think of as perfect (or at least very much better than the device under test) is to make repeated measurement of the time difference and to average the results. This number of measurements that can be combined into a single average will be limited by the assumption that the measurements differ only by the white phase noise of the measurement process, perhaps combined with the white phase noise of the oscillator itself as described above.

I will model the time differences between the device under test and our perfect reference device in terms of an initial time difference, a frequency offset, y, and a frequency aging, *d*:

$$x = x_0 + yt + \frac{1}{2}dt^2 \quad , \tag{22}$$

and my initial goal is to determine the time difference, x_0 in the presence of the white phase noise of the measurement process, which has a standard deviation ε_m.

Equation 22 is not very useful as a tool to estimate the time difference directly. For example, it cannot be used to estimate the parameters x_0, y, and d by applying standard least squares to an ensemble of measurements, because the parameters y and d are not constants but change slowly with time and have both stochastic and deterministic contributions. The least-squares analysis will provide numerical estimates in a formal sense, but the estimates are neither physically significant nor statistically stationary, since eq. 22 is trying to fit a single quadratic relationship to an ensemble of data in which the parameters of the quadratic vary with the value of the independent variable. A simple least-squares analysis does not have the flexibility to provide a robust estimate of parameters that are themselves statistical variables. The only exception might be for a very short data set where the principal contribution to the time dispersion is the white phase noise of the measurement process. The frequency offset and frequency aging can be taken to be approximately constant in this situation. However, we will use the iterative form of this relationship, eq. 12, in the more general case.

Since the noise of the measurement process is a random function that depends only on the characteristics of the measurement device and not on time or on any external perturbations, I can average the measured time differences to attenuate the effect of the measurement noise. I can continue to do this as long as the assumption that the time differences are randomly distributed about a “true” value as a result of the measurement noise alone. The duration of my average is time T, where T is given by

$$yT + \frac{1}{2}dT^2 << \varepsilon_m . \tag{23}$$

That is, the average can continue as long as the deterministic evolution of the time differences is much less than the noise of the measurement process so that the measurements are extracted from an ensemble of measurements that have the same time difference within the measurement uncertainty. An averaging time that satisfies this constraint will usually satisfy the weaker constraint that is driven by the noise in the frequency of the oscillator:

$$\sigma_y(T)T << \varepsilon_m \ , \tag{24}$$

where $\sigma_y(T)$ is the two sample Allan deviation of the device under test for an averaging time of T. Equation 24 is a weaker condition because the frequency stability of most oscillators is generally better than the frequency accuracy.

These principles can be illustrated with a numerical example. Suppose that we wish to characterize a rubidium oscillator with a time-difference system that has a measurement noise of 1 ns (10^{-9} s). Based on the generic type of the device, we might estimate that the oscillator has a deterministic fractional frequency offset of 5×10^{-11}. (Recall that fractional frequency offsets are dimensionless.) The deterministic frequency aging is about 4×10^{-18} /s (about 1×10^{-11} per month). If we consider the limit imposed by equation 23, the first term requires that $T << 20$ s. The second term is negligible for that value of T, so that the requirement of equation 23 is primarily driven by a constant frequency offset with no deterministic aging. This is a common result, which we will discuss in greater detail below. The two-sample Allan deviation of a rubidium standard for an averaging time of about 20 s is of order 10^{-12}, so that eq. 24 is also easily satisfied with this averaging time.

Thus, if we know only generic values of the parameters that characterize the device, we can average the time differences for something less than 20 s. If we decide to use 16 measurements of the 1 Hz pulses from the device, we could average the 16 measurements and the standard deviation of the measurements would be improved by a factor of 4. This is about the best we can do without knowing more about the device. Note that we cannot make a robust estimate of the frequency yet because our measurements are dominated by the white phase noise of the measurement process.

If we average 16 measurements, the uncertainty in the time difference has been reduced to about 0.25 ns or 250 ps. If we continue to make time difference measurements, we can no longer average them directly, since the distribution becomes increasingly driven by the deterministic frequency offset of the device. We enter the domain where the deterministic frequency offset is making a contribution to the time differences, and the measurements are now limited by a combination of the white frequency noise of the device and white phase noise of the measurement. Thus, the measurement strategy changes from simply averaging the measured time differences to estimating the average frequency offset (the first derivative of the measurements) as well. This intermediate case is difficult to handle with this simple method, since we do not know how to partition the variance of the time difference data into a contribution of the phase noise of the measurement process and the frequency noise of the clock itself.

One way to handle this ambiguity is to reverse the inequality in eq. 24. For example, if we made a time-difference measurement every 10^4 s, the white phase noise of each of the measurements would still be only 1 ns, but the contribution of the frequency noise of the oscillator would now contribute about 10 ns. We have now moved into the complementary domain, where the variance of the time differences is dominated by the frequency noise of the oscillator and the contribution of the noise of the measurement process is small enough to be ignored. The contribution of the deterministic frequency aging to the time differences is of order 0.5×10^{-18} /s$\times$ 10^8 s^2 = 0.05 ns, so that we can make the reasonable assumption that *all* of the observed variance can be modeled as white frequency noise of the oscillator. (The ensemble algorithm that is used at NIST and other national laboratories often operates in this measurement domain where the data are can be modeled as pure white frequency noise to a good approximation.) The optimum strategy in this domain is to average the offset frequency estimates obtained from the first-differences of the time-difference data divided by the interval between these values.

This approach can continue as long as the measurements are in the white frequency noise domain, and the strategy will give an increasingly accurate estimate of the deterministic offset frequency of the device. We can derive the lower bound of this domain from the white phase noise of the measurement process as we have done above, but we don't have enough information to specify the upper end of this domain uniquely. We can make a rough estimate by comparing the time dispersion resulting from the frequency fluctuations to the time dispersion driven by the deterministic frequency aging, and defining the upper limit of the averaging time as the interval when these two contributions are equal. This upper limit to the averaging time is defined by:

$$\sigma_y(T)T = \frac{1}{2} dT^2 \ , \tag{25}$$

or

$$T = \frac{2\sigma_y(T)}{d} \ . \tag{26}$$

Eq. 26 highlights a fundamental problem with the model that we are using. The two sample Allan deviation of a typical rubidium standard is generally not less than 10^{-12} for intermediate averaging times, and tends to increase at longer averaging times because non-white frequency fluctuations usually become important there. On the other hand, the deterministic frequency aging, *d*, is of order 10^{-18} /s, so that the averaging time predicted by eq. 26 (where the contribution of deterministic frequency aging to the frequency fluctuations is equal to the stochastic contribution) is generally longer than 10^6 s. In other words, it is difficult to estimate the deterministic frequency aging of a device because the aging is masked by the stochastic frequency fluctuations for moderate averaging times. Furthermore, non-statistical considerations are often important at longer averaging times, which makes the numerator of eq. 26 larger and the problem of estimating the deterministic aging more difficult. On the other hand, the frequency aging contributes to the time differences as the square of the time interval, so that it will almost always dominate the time dispersion at sufficiently long times.

The conclusions of the preceding discussion are more general than the specific case that I discussed, and the result is that it is very difficult to characterize the deterministic component of the long-term performance of any oscillator. (A time scale, which is an ensemble of oscillators, is not immune to this problem.) The only difference among the different types of oscillators is where the "long-term" time domain begins. For example, the deterministic frequency fluctuation of a cesium standard is masked by the stochastic frequency fluctuations for almost any averaging time out to the life of the device (years), and cesium standards are generally modeled with no deterministic frequency aging for this reason. A hydrogen maser, on the other hand has a deterministic frequency aging of order 10^{-16} per day ($\sim 10^{-21}$/s), but its stochastic frequency fluctuations are small enough so that the frequency aging must be included in any model of the time differences. The frequency aging of a rubidium standard is often of order 10^{-11} per month ($\sim 3.9\times 10^{-18}$/s), and can be ignored only for short averaging times.

The basis of this discussion is that it is possible to find measurement domains where only one type of noise process dominates the variance of the measured time differences. This idea is widely used in modeling oscillators that are members of a time-scale ensemble. For example, the AT1 algorithm used to estimate the average time of an ensemble of clocks at the NIST laboratory in Boulder, Colorado is designed based on this principle.[12] The measurement system used at NIST has a measurement noise on the order of 1 ps and a time interval between measurements of 720 s, and the ensemble algorithm is based on the premise that the variance of the time differences can be modeled as white frequency noise. The hydrogen masers that are members of the ensemble have a constant frequency aging that is determined outside of the ensemble algorithm. This parameter is treated as a constant by the algorithm because it is difficult to compute a statistically robust estimate of the aging because of the problems discussed above.

The algorithm used by the International Bureau of Weights and Measures (the BIPM in French) to compute International Atomic Time (TAI) and Coordinated Universal Time (UTC) is also based on these principles. The measurement noise of the time differences of the clocks located at the various national timing laboratories is much larger than the local time-difference measurements at NIST, so that the interval between measurements has to be increased proportionally to guarantee that the contribution of the measurement noise to the time-differences is smaller than the contribution of the frequency variations of the clocks. However, this increase in the time interval between measurements increases the impact of the frequency aging of the masers that contribute to TAI and UTC, and the algorithm used by the BIPM has been modified to recognize this effect.[13] See also.[14] As we would expect from the previous discussion, including an explicit frequency aging term improves the long-term stability of the time scale by bringing the model closer to the actual behavior of the clocks that are members of the ensemble.

10. The Kalman Estimator

The preceding discussion depended on the assumption that the time interval between measurements was a free parameter that could be adjusted at will based on statistical considerations. In each case, it was chosen so that the variance of the time-difference measurements could be modeled as arising primarily from a single source. However, there can be systems where the time interval between measurements is constrained by other factors to values that do not support this simplifying assumption. The variance of the time difference measurements must be apportioned to more than one source in these configurations, but there is no way to do this using the machinery that we have developed so far. The Kalman estimator is one way of partitioning the variance in these situations, and I will discuss the general method in this section.

The Kalman estimator starts from the same recursive relationship for the time differences that we presented above in eq. 12, but it adds two additional recursive relationships describing the evolution of the offset frequency and the frequency aging. The “Kalman state” of the clock is characterized by the values of these three parameters at any instant. The three equations that characterize the evolution of the state as a function of time are:

$$
\begin{aligned}
x_k &= x_{k-1} + y_{k-1}\Delta t + \frac{1}{2} d_{k-1}(\Delta t)^2 + \xi \\
y_k &= y_{k-1} + d_{k-1}(\Delta t) + \eta \\
d_k &= d_{k-1} + \zeta
\end{aligned}
\qquad (27)
$$

The time and offset frequency components are defined recursively with a deterministic contribution and a stochastic contribution (ξ and η, respectively). The frequency aging might have an initial constant value but usually is assumed to start at zero with only a stochastic variation, ζ. In each case, the stochastic contribution to the corresponding parameter is assumed to be a noise process that has a mean of zero and a variance that is initially known from other considerations, at least to first order. The noise contributions are assumed to be uncorrelated both in time and with each other. In other words, all three noise parameters satisfy relationships of the form

$$
\begin{aligned}
&\langle \xi(t) \rangle = 0 \\
&\langle \xi(t)\xi(t') \rangle = \xi^2 \delta(t-t') \\
&\langle \xi(t)\eta(t) \rangle = 0
\end{aligned}
\qquad (28)
$$

for all combinations of the parameters and for all *t* and *t'*.

The deterministic terms in equations 27 describe how the state of the system evolves between measurements. The Kalman formalism can support measurements of any of the components of the state, but the most common arrangement is a measurement of the time difference of the clock with respect to a "perfect" reference as we discussed above. In general, the measurement of the state component (the time difference in our discussion) will not agree with the value predicted from the previous state values by equations 27, and the Kalman formalism provides a method of assigning the causes of this residual partially to updating the values of the deterministic parameters and partially to the noise parameters. The details of how to do this are in the literature.[15] Practical realizations of the Kalman algorithm applied to estimating the parameters of clocks often have the same difficulty in estimating the frequency aging term that we discussed above for basically the same reason – it is difficult to calculate a robust estimate of the frequency aging in the presence of stochastic frequency noise that is generally larger even at moderately large averaging times.

In addition, a Kalman algorithm can be no better than the accuracy of the model equations 27 and 28 that are used to define the evolution of the clock state. The assumption that the stochastic inputs to the frequency and frequency aging are random, uncorrelated, zero-mean processes is often not an accurate description of the true state of affairs. Even when the model equations accurately describe the steady-state evolution of the state parameters, Kalman algorithms often have start-up transients that can be troublesome in some real-time applications.

11. Transmitting time and frequency information

I now consider the problem of comparing the time difference between two clocks that are not at the same location, so that the time difference must be measured by means of a channel that connects the two devices. There is no difference between this configuration and the one we described above in principle, but there are a number of practical differences that make this type of measurement significantly more complicated.

The first issue that I will consider is the transmission delay that is introduced by the channel. The delay is at least 3 μs/km, so that it must be measured in all but the simplest measurement programs of time differences. (If the goal of the measurement program is a comparison of the frequency difference between the two devices, then the magnitude of the channel delay is not important, provided only that it remains constant to the level required by the measurement process. Although this sounds like an easier requirement to satisfy, designing a channel that satisfies this requirement and verifying that it does so is often not significantly easier than going the whole way and measuring the delay itself.)

There are many applications where the delay is small enough to be ignored. For example, there are a very large number of wall clocks, wrist watches, and some process-control devices that are calibrated and set on time by means of the radio signals transmitted by the NIST radio station WWVB in Fort Collins, Colorado. The transmission delay, which can be on the order of milliseconds, is simply ignored in these devices, since the required accuracy is generally only on the order of 1 s. Simple devices that are synchronized using the signals from the Global Positioning Satellites (GPS) often work the same way – the unmodeled portion of the transmission delay (due to the refractivity of the ionosphere and the troposphere, for example) is of order 65 ns in this case, but it is still much smaller than the required accuracy, which may be only to the nearest millisecond or even only to the nearest second. I will not consider these applications here, and I will focus on the applications where some measurement of the channel delay is needed to satisfy the demands of the application.

I will describe three methods that are currently used to estimate the channel delay: (1) modeling the delay by means of ancillary parameters and measurements; (2) the common-view method and its “melting pot” variant; and (3) two-way methods. I will also discuss the “two-color” method that is often used as an adjunct to one of the other methods when at least some of the path is through a medium that is dispersive. That is, the speed of the signal is a function of the frequency used to transmit the message. I will first discuss the general characteristics and assumptions of each of the methods, and I will then illustrate them with a more detailed discussion where the method is applied to a specific system.

The assumption that is implicit in this discussion is that the channel delay is not an absolute, unvarying parameter that can be measured once at the beginning by a process that may be complicated but need be done only once. There are some simple channels that satisfy this requirement – a coaxial cable between two parts of a building, for example.

A measurement of the delay of a long coaxial cable will often have some engineering complexity because coaxial cables are dispersive and have a frequency-dependent attenuation. In general, the attenuation, which is a result of the series inductance and shunt capacitance of the cable, increases with increasing frequency. Therefore, the rise time of a pulse transmitted on such a cable, which is a function of the high-frequency components of the signal, is increased as the signal propagates through the cable. The delay measurement is often further complicated by small impedance mismatches at the end of the cable, which cause reflections that interfere with the primary pulse used to measure the delay. These reflections can be exploited in the measurement process by leaving the far end of the cable un-terminated and measuring the round-trip travel time of a pulse sent from the near end and reflected back from the open remote end. (The measurement of the travel time is not sensitive to the details of the signal that is transmitted along the cable, and a measurement that transmits a pseudo-random code instead of a pulse can have some technical advantages because it can be less affected by the attenuation of the high-frequency components of the test signal.) Whichever method is used, the delay is normally considered to be a constant that is a characteristic of the cable, so that the measurement is normally a one-time effort. I will not consider this situation in detail and I will focus on measurements where the delay cannot be measured once as a calibration constant.

11a. Modeling the Delay

This method is based on the assumption that the channel delay can be estimated by means of some parameters that are known or measured from some ancillary measurement. For example, the geometrical path delay between a satellite and a receiver on the ground is estimated as a function of the position of the receiver, which has been determined by some means outside of the scope of the timing measurement, and the position of the satellite, which is transmitted by the satellite in real time.

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

11b. The Common View Method

This method depends on the fact that there are two (or more) receivers that are equally distant from a transmitter. Since the two path lengths are the same, any signal sent from the transmitter arrives at the same time at both receivers. Each receiver measures the time difference between its local clock and the received signal, and these two measurements are subtracted. The result is the time difference between the clocks at the two receivers. In this simple arrangement, the accuracy of the time difference does not depend on the characteristics of the transmitted signal or the path delay.

In the real world, it is difficult to configure the two receivers so that they are exactly equally distant from the transmitter, and some means must be used to estimate the portion that is not common to the two paths. This estimate is not as demanding as estimating the full path delay, so that the common view method attenuates any errors in the estimate of the path delay.

There are also a number of subtle effects that we must consider when the two path delays of a common-view measurement are not exactly equal. If a single signal is used to measure the time difference, then the signal does not arrive at the two receivers at the same time, since the path delays are somewhat different. Therefore, any fluctuation in the characteristics of the receiver clocks during this time difference must be evaluated. On the other hand, if the measurement is made at the same instant as measured by the receiver clocks, then the signals that are measured by the two receivers did not originate from the satellite at the same instant of time. Therefore, the fluctuations in the characteristics of the satellite clock during this time interval must be evaluated.

Finally, a common-view measurement algorithm does not support casual associations among the receivers. The stations that participate in the measurement process must agree on the source to be observed and the time of the observation. There must also be a channel between the two receivers to transmit the measured time differences. On the other hand, there need be no relationship between the receivers and the transmitter, and the transmitter need not even know that it is being used as part of a common-view measurement process. Signals from commercial analog television stations have been used as common-view transmitters, and the zero-crossings of the mains voltage can also be used to compare clocks with an uncertainty on the order of a fraction of a millisecond.

11c. The “Melting Pot” version of common view

The previous discussion of common view focused on a number of cooperating receivers, where each one measured the time difference between a physical signal and the local clock. However, there can be some situations where there is no single transmitter that can be observed by the receivers at the same epoch. For example, if the common-view method is implemented by means of signals from a Global Positioning System (GPS) navigation satellite, then receivers on the surface of the Earth that are sufficiently far apart cannot receive signals from any one satellite at the same time.

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

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

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

On the other hand, the melting pot method can usually use the observations from several satellites at the same time, so that the random phase noise of the measurement process can be attenuated by averaging the data from the multiple satellites. Therefore, a comparison between the simple two-way method and the melting-pot version depends on a comparison between the noise of the measurement process, which would favor a melting-pot measurement using multiple satellites, and the uncertainties and residual errors in the orbital parameters of the satellites and the offset between the clock in each satellite and GPS system time. The accuracy of the melting-pot method improves as more accurate solutions for the orbits and satellite clocks become available.[16]

11d. Two-way methods

There are a number of different implementations of the two-way method, but all of them estimate the one-way delay between a transmitter and a receiver as one-half of the round-trip delay, which is measured as part of the message exchange. The accuracy of the two-way method depends on the symmetry of the delays between the two end points. The accuracy does not depend on the magnitude of the delay itself; although the magnitude of the delay can be calculated from the data in the message exchange, the accuracy of the time difference does not require this computation.

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

The transmitter and receiver at the end points of the path are often sources of asymmetry. These hardware delays are often sensitive to the ambient temperature. However, the admittances to temperature fluctuations may be different at the two end stations and the temperatures at the two end points may be quite different. Finally, there are often components of the measurement process that are outside of the two-way measurement loop, and any delays originating in these components must be measured on every message exchange or measured once and stabilized.

11e. The two-color method

Suppose that there is a portion of the path that has an index of refraction that is significantly different from the vacuum value of one. This difference of the index of refraction relative to its vacuum value is the *refractivity* of the path. Suppose also that the refractivity is dispersive. That is, it depends on the frequency that is used to transmit the message. If the length of the path is measured using the transit time of an electromagnetic signal, the refractivity will increase the transit time, so that the effect of the refractivity will be to make the path length appear too long. If the length of the true geometric path is *D*, then the measured length will be *L*, where *L* is given by

$$L = nD = D + (n-1)D \ . \tag{29}$$

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

$$n - 1 = F(p)G(f) \ . \tag{30}$$

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

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

$$\begin{aligned} L_1 - L_2 &= F(p)\left(G(f_1) - G(f_2)\right)D \\ F(p) &= \frac{L_1 - L_2}{D}\frac{1}{\left(G(f_1) - G(f_2)\right)} \\ n_1 - 1 = F(p)G(f_1) &= \frac{L_1 - L_2}{D}\frac{G(f_1)}{G(f_1) - G(f_2)} \end{aligned} \quad . \tag{31}$$

If I substitute the last relationship of equation 31 into equation 29 evaluated for frequency f_1, I obtain

$$L_1 = D + (n_1 - 1)D = D + \left(L_1 - L_2\right)\frac{G(f_1)}{G(f_1) - G(f_2)} \quad , \tag{32}$$

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

$$\begin{aligned} H(f_1, f_2) &= \frac{G(f_1)}{G(f_1) - G(f_2)} \\ D &= (1 - H)L_1 + HL_2 \end{aligned} \quad . \tag{33}$$

The details of the function *G* are not important, provided only that it is known and that the medium is dispersive. (The denominator of the fraction on the right-hand side of eq. 31 or 32 is zero for a non-dispersive medium. The difference in the apparent lengths in eq. 31 will also be zero in this case.) Note that the second term on the right-hand side of eq. 32, which is the correction to the geometrical length *D* due to the dispersive medium, does not depend on *L*, the extent of that medium, but only on the apparent difference in this length for the measurements at the two frequencies. Thus this relationship is equally valid if only a portion of a geometric path is dispersive, and the correction term specifies the apparent change in the length of only that portion of the path. Note also that I do not have to know anything about the function *F(p)* – only that the separation into two terms expressed by eq. 30 represents the dispersion.

The measurements of both L_1 and L_2 will have some uncertainty in general, so that the two-color determination of the geometrical length, *D*, will have an uncertainty that is greater than it would have been if the medium were non-dispersive so that a measurement at one frequency would have been adequate. The magnitude of the degradation depends on the details of the function H, and I will discuss this point again when I describe measurements using navigation satellites such as those of the Global Positioning System (GPS).

12. Examples of the measurement strategies

In the following sections I will describe systems that use the various measurement strategies that I have outlined above. I will begin by describing the characteristics of the satellites of the Global Positioning System (GPS), since the data transmitted by these satellites are widely used for timing applications in the one-way, common-view and melting pot modes. The Russian GLONASS system, the European Galileo system, and the Chinese BeiDou system are different in detail, but the following discussion describes the general features of all of them. In general, the differences between the systems are hidden from the general user and are a concern only of the receiver designer.

12a. The Navigation Satellites of the Global Positioning System

The GPS system uses at least 24 satellites in nearly circular orbits whose radius is about 26 600 km (The number of satellites in the constellation that are active at any time is generally greater than 24.) 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 23h 56m).

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 from every satellite is modified by two effects that are common to all of them – a red shift due to the second-order Doppler effect and a blue-shift due to the difference in gravitational potential between the satellite and the observer. These two effects produce a net fractional blue-shift of about 4.4×10^{-10} (38 μs/day), and the proper frequencies of the oscillators on all of the satellites are adjusted downward 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 receiver computes and applies the corrections for these effects.

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. (The newer GPS satellites will transmit signals on additional frequencies, and the Galileo, GLONASS and BeiDou systems transmit signals at slightly different frequencies.) 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 “clear access” or 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. The codes are derived from the same 10.23 MHz primary oscillator. Under normal operating conditions, the C/A is present only on the L1 carrier. Many timing receivers process only the C/A code. Although the P code is normally encrypted with an encryption key that is not available to unclassified users, many receivers can operate in a “semi-codeless” mode where the P code data can be decoded with some increase in the noise of the process.

Each GPS satellite transmits at the same nominal frequencies but uses a unique pair of C/A and P codes. The codes are constructed to have very small cross-correlation at any lag and a very small auto-correlation at any non-zero 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, the repetition rate of the whole C/A code. The procedure locks the local clock to the satellite time with a time offset due to the transmission delay (about 65 ms), and allows the receiver to begin searching for the 50 bit/sec navigation message.

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 system time and UTC(USNO) is guaranteed to be less than 100 ns (modulo 1s), and the estimate of this offset, which is transmitted as part of the navigation message, is guaranteed to be accurate to 25 ns (also modulo 1 s). In practice, the performance of the system has almost always substantially exceeded its design requirements.

The UTC time scale includes leap seconds, which are added as needed to keep UTC with ±0.9 s of UT1, a time scale based on the position of the Earth in space. The GPS time scale 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 integer-second difference, GPS time – UTC, is currently (December, 2013) 16 s, and will increase as additional leap seconds are added to UTC. 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.

Most modern receivers can observe several satellites simultaneously, and can compute the time differences between the local clock and GPS system time using all of them at same time. In each case, the time of the local clock that maximizes the cross correlation with the signal from each satellite is the *pseudorange* – the raw time difference between the local and satellite clocks. (It is related to the geometrical time of flight with an additional time offset since the clock in the receiver is generally not synchronized to GPS system time.)

Using the contents of the navigation message, the receiver corrects the pseudorange for the travel time from the satellite to the receiver, 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 additional delay through the ionosphere due to its refractivity by applying the two-color method described above to the difference in the pseudoranges observed using the *L1* and *L2* frequencies. 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 system time.

In principle, the time difference between the local clock and GPS system time should not depend on the specific satellite whose data are used for the computation. In practice, the time differences computed using the data from different satellites will differ because of the noise in the measurement processes and because of errors in the broadcast ephemerides and the parameters of the satellite clocks. The group of time differences with respect to GPS system time, computed from the different satellites, forms a "Redundant Array of Independent Measurements" (RAIM), and some analysis methods compare these time differences in an attempt to detect a bad satellite. These "T-RAIM" algorithms succeed when the time difference computed using the data from one satellite differs from the mean of the differences computed from the other satellites by a statistically significant amount. The T-RAIM algorithm can be used in the one-way, common-view or melting pot algorithms that I discuss in the next sections. The same idea is used in the Network Time Protocol to identify a bad server. I will discuss this point in greater detail below.

12b. The one-way method of Time transfer: modeling the delay

This method is most often used with time transfer by means of signals from navigation satellites because they are the only systems that transmit enough information to support an accurate delay estimate. The estimate of the transit time of a message from a navigation satellite to a receiver on the ground can be divided into a number of components that are increasingly difficult to estimate.

The largest single estimate to the propagation delay is the delay resulting from the geometric path length. The magnitude of this delay depends somewhat on the position of the satellite in the sky, but is typically approximately 65 ms. The path length is computed from the position of the satellite, which is estimated from the orbital parameters transmitted as part of the navigation message, and the position of the ground receiver. In pure timing applications, the position of the ground receiver is assumed to be known from other data, and I will assume that this is the case in the current discussion. (If the position of the receiver is not known *a priori*, it can be estimated by computing the distances from the receiver to multiple satellites and solving for the 4 unknowns: the 3 Cartesian coordinates of the position of the receiver and the time offset of its clock with respect to satellite system time.) In a real-time application the accuracy of the estimate of the geometric path delay is limited by any uncertainty in the position of the receiver (the vertical coordinate usually has the largest uncertainty) and by errors in the broadcast ephemeris parameters. These combined uncertainties are generally on the order of a few meters, which is equivalent to an uncertainty in the time delay of about 10 ns or less.

Thus the uncertainty in the correction is much smaller than the magnitude of the correction itself.

The additional delay due to the passage of the signal through the ionosphere adds approximately 65 ns to the geometric delay. A receiver that can process both of the frequencies transmitted by a satellite can estimate the effect of the ionosphere using the two-color method that I have described above. Simpler, single-frequency receivers can use an estimate of the effect of the ionosphere that is broadcast by the satellite as part of the navigation message. This is a globally-averaged prediction and is therefore less likely to be accurate at any specific location.

The additional delay due to the passage of the signal through the lower atmosphere (the troposphere) is much smaller than the additional delay through the ionosphere, but there is no easy way of estimating it because the refractivity does not depend on the carrier frequency so that the two-color method cannot be used. Some more sophisticated analyses estimate this delay by means of local measurements of atmospheric pressure, temperature, and water-vapor content, but these data are not always available. Even when they are available at a site, these parameters often have significant azimuthal variation, which is generally not easily estimated. (Boulder, Colorado is potentially particularly bad in this respect, since the mountains to the West and the plains to the East would be expected to have quite different temperature profiles.) There are also models of the refractivity of the troposphere, which estimate this parameter as a function of the day of the year and, possibly, the coordinates of the receiving station.

The magnitude of this delay is typically on the order of 6 ns at the zenith, and it increases for satellites at lower elevation by a factor that is roughly proportional to the increase in the slant path through the troposphere relative to the zenith path length. The increase in the slant path delay relative to the zenith delay is usually estimated as proportional to the reciprocal of the sine of the elevation angle.

If an analysis assumes the slant-path model of the variation of the delay, it is possible to solve for the zenith delay by observing the apparent variation in the time difference estimates obtained from satellites at very different elevation angles. This estimate does not work as well as we might like because the slant-path model is only an approximation, because the tropospheric refractivity often has a significant azimuthal variation, and because the variation from one satellite to another is also affected by measurement noise and by errors in the broadcast ephemerides or any error in the coordinates of the receiver.

Many time-difference measurements ignore the effect of the troposphere altogether. This introduces a systematic error of order 10 ns in the time difference estimates; as I mentioned above, the magnitude of this error depends on the elevation of the satellites that are being observed.

The final contributions to the model of the delay are effects that are local to the receiver: the delay through the hardware and the motion of the station due to the Earth tides and other geophysical effects. The delay through the receiver hardware is normally assumed to be a constant that varies only very slowly over periods of years. The delay is often dominated by the delay through the antenna and the cable from the antenna to the receiver, and a value on the order of 100 ns is typical. (Delays through coaxial cables are of order 5 ns/m.)

It is possible to calibrate the delay through a receiver using a special signal generator that mimics the signals from the real satellite constellation. This type of equipment is not widely available, and most timing laboratories perform a differential calibration in which the delay through the receiver under test is compared to the delay of a “standard” receiver. This method is obviously not adequate for a one-way measurement, but is widely used because most timing laboratories use the satellites in common view, which I discuss in the following section.

The motion of the station and other geophysical effects contribute 1 – 2 ns to the overall delay. The magnitude of these effects can be calculated and included in more complicated post-processed analyses but are generally ignored for real-time applications.

12c. The common view method

I have already described most of the important features of the common-view method. It is most often used with signals from the navigation satellites, but it is more general than this and can be used with other sources as well. Signals from LORAN transmitters and even from television stations have been used in this way. There have even been some experiments to use the zero-crossings of the power line in common-view within a building or over a small area.

The method has two principal limitations:

(1) It is difficult in practice to configure the measurements so that the receivers are all equidistant from the source. Therefore, some correction is almost always necessary to model the differential delay. The differential delay is much smaller than the delay itself in most configurations, so that the required accuracy of the model of the delay is correspondingly easier to satisfy. However, ignoring the difference in the path delays is often not sufficiently accurate.

(2) The common-view method (and its melting-pot variant) cannot provide any help in mitigating the effects of delays that are local effects at a site. The differential effects of the ionosphere can be significant and are usually estimated using the two-color method. The differential effects of the troposphere cannot be estimated in this way and are often ignored. Ignoring the differential effects of the troposphere is often justified because the total contribution is relatively small and the differential contribution is correspondingly smaller.

Multi-path is a more serious local effect that is often too large to ignore. The effect is caused by copies of the signal that reach the antenna after they have been reflected from some nearby object. These signals always travel a longer distance than the primary one, and they arrive later than the primary signal as a result. A simple omni-directional antenna typically responds to these signals and the receiver computes a correlation that is a complicated sum of the direct and reflected signals.

The multi-path effect is a complicated function of the position of the satellite with respect to the antenna and the local reflectors, and it is therefore periodic with the orbital period of the satellite. From this perspective, the orbital periods of the GPS satellites are all very close to one sidereal day (23h 56m), so that the multi-path reflections have this periodicity. They can usually be estimated by comparing the time differences measured from any satellite at the same sidereal time on consecutive days.

The BIPM has exploited this sidereal-day periodicity in defining the common-view tracking schedules that are used by timing laboratories and National Metrology Institutes to compare time scales and to facilitate the computation of International Atomic Time (TAI) and Coordinated Universal Time (UTC). The observation time for each satellite is advanced by 4 minutes every day in the BIPM schedule, so that every satellite returns to the same point in the sky on every track each day relative to the antenna and to any multi-path reflectors. Thus, the multipath environment is a constant for each track, although it generally varies from track to track. This has the advantage of converting the varying effects of multipath to systematic offsets that are approximately constant for each track. The assumption of a sidereal-day periodicity is not exact, so that the offset due to the multipath contribution changes slowly with time. These long-period effects can be hard to distinguish from the contributions due to the random walk of frequency and frequency aging that I discussed previously.

Locating the antenna far away from reflecting surfaces can help minimize the impact of multipath reflections; adding choke rings and a ground plane to an antenna, which attenuate signals arriving from the side or from below can also help.

Another strategy to mitigate the impact of multipath is to exploit the sidereal-day periodicity and compute the average frequency of the local clock with respect to the GPS system time as an average over a sidereal day. The multipath contribution cancels in the sidereal-day time difference, so that the frequency estimated in this way is almost insensitive to multipath effects.

12d. Two-way Time Protocols

In the following sections, I will describe three two-way time protocols that are commonly used to compare clocks at remote locations. The list is intended to be descriptive rather than exhaustive. For example, I do not discuss time transmission using optical fibers because this method is generally too expensive to be used for long distances and because the underlying physics is basically the same as the other methods that I do describe. I also do not discuss the Precise Time Protocol (PTP, often called IEEE 1588) in any detail for much the same reasons. Its capabilities are very similar to a hardware-assisted version of NTP, and it is generally not well suited to long-distance time comparisons because it assumes that the delay is nearly constant so that it does not have to be measured on every message exchange.

12d-1. The Network Time Protocol

The Network Time Protocol (NTP) is widely used to transmit time and compare clocks that are linked together by a channel that is based on a packet-switched network such as the Internet. The NTP message format is based on the User Datagram Protocol[17] (UDP). The UDP message exchange is not sensitive to the details of the physical hardware that is used to transmit the packets. However, as with all two-way protocols, the accuracy of the NTP message exchange depends on the symmetry of the inbound and outbound delays, and this symmetry is often limited by the characteristics of the physical layer used to transmit the messages. In the following discussion I will focus on the time-difference accuracy of the message exchange; I will defer the question of how often a system should initiate such an exchange (the “polling interval”) to a later section and I will discuss only briefly the question of how a client system should discipline its local clock based on the exchange of messages with a server.

The protocol is initiated when station “A” sends a request for time information to station “B”. The two stations might have a client-server relationship, in which the client intends to adjust its clock based on the results of the exchange, or it could be a peer-to-peer exchange in which two systems exchange timing information with the goal of setting the times of both systems to agree with each other.

The message is sent at time T_{1a} as measured by the clock on system *A*. The transmission delay from station *A* to station *B* is δ_{ab} so that when the message arrives at station *B*, the time at station *A* is $T_{1a} + \delta_{ab}$. The time of arrival at station B, measured by the local clock at that station, is T_{2b}, and the time difference between stations *A* and *B* is:

$$(\Delta T)_{ab} = (T_{1a} + \delta_{ab}) - T_{2b} \ . \tag{34}$$

The *B* system responds by sending a message back to *A*. The message leaves the *B* system at time T_{3b}, arrives back at the *A* system at time $T_{3b}+\delta_{ba}$ and the time at the *A* system at that instant is T_{4a}.

The total round-trip transit time is measured at station *A* as the time that has elapsed during the message exchange as measured by the clock on station *A*, less the time between when station *B* received the request and when it replied, as measured by the clock on station *B*:

$$\Theta = \delta_{ab} + \delta_{ba} = \left(T_{4a} - T_{1a}\right) - \left(T_{3b} - T_{2b}\right) \ . \tag{35}$$

We now assert that the path delay is symmetric, so that the inbound and outbound delays are equal. Then the path delay from *A* to *B*, δ_{ab} is simply one-half of the expression on the right-hand side of eq. 35. If we substitute one half of the right hand side of eq. 35 into eq. 34, the time difference between stations *A* and *B* is

$$(\Delta T)^s_{ab} = \frac{T_{1a} + T_{4a}}{2} - \frac{T_{2b} + T_{3b}}{2} \quad . \tag{36}$$

The superscript *s* indicates that the time difference is computed using a symmetric path delay. If the path delay is not symmetric, then the inbound and outbound delays are not equal. We can parameterize this asymmetry as:

$$\delta_{ab} = (0.5 + \varepsilon)\Theta \quad . \tag{37}$$

The asymmetry parameter ε can take values from +0.5 to -0.5. The positive limit indicates that the path delay from *A* to *B* dominates the round-trip delay and the delay in the other direction is negligibly small, while the negative limit specifies the inverse: the delay from *A* to *B* is negligible compared to the reverse delay from *B* to *A*.

If we substitute eq. 37 into eq. 34, then the first term on the right hand side of eq. 34 reproduces the time difference expression of eq. 36, and the second term adds a correction to the time difference:

$$(\Delta T)^a_{ab} = (\Delta T)^s_{ab} + \varepsilon\Theta \quad . \tag{38}$$

Since we model the measurement based on the assumption of a symmetric delay (eq. 36), the time difference that we estimate is in error. The magnitude of the error is given by $\varepsilon\Theta$, the second term in eq. 38. This term is proportional both to the magnitude of the asymmetry and to the round-trip delay. Thus a smaller round-trip delay guarantees a smaller error due to any asymmetry. The lesson is that NTP servers should be widely located so that the round-trip delay to any user is minimized.

The round-trip delay is often of order 100 ms (0.1 s), and typical asymmetries are on the order of a few percent of the delay. Therefore, we might expect that a typical NTP message exchange would have an error on the order of 5 ms or 10 ms due to the asymmetry of the path delay, and errors of this order should be considered as routine for a server and a client on a wide-area network.

In addition to possible asymmetries in the network delay, there may also be additional asymmetries in the client system. For example, if the process that manages the NTP message exchange runs in a standard user environment then it must compete for processor cycles with all of the other processes that may be active on the system. In addition, it must issue a request to the system to retrieve the system time each time a message is sent or received in order to have the values to compute the time differences described above. All of these effects add to the network delay measurement; depending on the details of the system and the processes that are active, it may also add to the asymmetry.

In order to minimize these effects, the NTP process can be moved into the system space where it runs at much higher priority as a system service. The ultimate version of this idea would be to move the NTP process into the network driver that receives and transmits the network packets, and some version of NTP and its cousin, PTP (the Precise Time Protocol, also called IEEE 1588) operate in this mode.

Although moving the NTP process into the system space or into the network driver itself will make the NTP process appear more stable and more accurate, the overall timing accuracy of an application that uses the system time may be degraded. This application normally runs as a standard user process, and it therefore experiences the same jitter as a user-level NTP process would experience when it issues a request to the system for the current time. This jitter is inside of the measurement loop when the NTP process runs as a user process, and the time difference calculation therefore takes it into account (at least to some extent even if it also contributes to the asymmetry). However, this delay is completely outside of the measurement loop if the NTP process is pushed down to the system or driver levels, so that the application process experiences the full impact of the delay jitter in requesting system services. Thus, while the NTP statistics improve, the accuracy realized by a user process may be degraded, and this problem is not reflected in any of the NTP statistics.

12d-2. The ACTS protocol

The NIST (and a number of other timing laboratories) operate a time service that transmits the time in a digital format by means of standard dial-up telephone lines. The NIST system is called ACTS[18], the Automated Computer Time Service, and it corrects for the transmission delay using a variant of the two-way protocol I have described.

The ACTS servers transmit a text string each second with the time derived from the NIST clock ensemble and an on-time marker (OTM) character. This character is initially transmitted using a default advance. If the user echoes the OTM back to the server, the server measures the round-trip delay, estimates the one-way delay as one-half of this value, and adjusts the advance of the next OTM transmission so that it will arrive at the user's system on time. The server changes the OTM character from "*" to "#" to indicate that it has entered this delay-calibrated mode. In every case, the server includes the estimate of the one-way delay in the message, so that the client can determine the advance that was used. In the context of the previous discussion of the NTP protocol, the ACTS protocol assumes that $(T_{3b} - T_{2b})$ is essentially zero. That is, the client echoes the OTM character back to the server with only negligible delay.

This process continues on every transmission. The OTM character is advanced based on the one-way delays estimated from the average of the round-trip measurements of the previous seconds. The details of the averaging process are determined dynamically by the server based on the measured variation of the round-trip delay from second to second.

The original ACTS system was designed to minimize the complexity of the code in the receiving system. The receiver needs only to echo the OTM back to the server, and the next OTM will be transmitted so that it arrives on time. The receiver did not have to perform any calculations at all. The assumption of this design was that the delay variations were not accompanied by variations in the symmetry, and this assumption was largely confirmed by the original design, which could transmit time messages with an accuracy of order 0.5 ms RMS.

Placing the delay calculation in the server simplifies the design of the client system, but it has the unfortunate side-effect that the server cannot detect a change in the symmetry of the delay whether or not this change in symmetry is accompanied by a change in the total round-trip delay. However, the client system is in a better position to determine what is really going on.

Since the true time difference between the client and the server changes by less than 1 ms from second to second, any significant change in the time difference measured from one second to the next one indicates that the advance algorithm in the server has been fooled by a change in round-trip delay that was accompanied by a change in symmetry. For example, it is possible that the change in the measured round-trip delay was really largely confined to either the inbound or outbound paths. The client can detect this possibility by noting the change in the measured time difference between the ACTS time and its system clock and the change in the measurement of the round-trip delay that the server has inserted into the transmission. As a simple example, if the change in the delay is confined to the outbound path between the server and the client, the server will see this as a change in the total round-trip delay, and will advance the next OTM by one-half of this value. This is exactly one-half of the correct advance change, so that the next OTM will not arrive on time by one-half of the change in the advance. The client will detect that the advance parameter has changed and that the measured time difference has changed by one-half of that amount. This more sophisticated algorithm in the client system can almost completely compensate for the degraded stability of the dial-up telephone system, and the more sophisticated algorithm can transmit time over standard dial-up telephone lines with an uncertainty of order 0.5 ms - 0.8 ms RMS. This is about a factor of 10 better than the Internet Time servers because the delay through the telephone system is more stable and more symmetric than the delay through a wide-area packet network.

12d-3. Two-way Satellite Time Transfer

This method is used to compare the time scales of National Metrology Institutes and Timing Laboratories, and to transmit time and frequency information to the International Bureau of Weights and Measures (the BIPM, in French) for the purpose of computing International Atomic Time (TAI) and Coordinated Universal Time (UTC).

The method uses the same assumption as in the previous discussions: the one-way delay can be estimated as one-half of the measured round-trip value. The configuration of the message exchange is similar to the NTP exchange discussed previously.

Each station encodes the 1 Hz tick of its local time scale using a pseudo-random code and a sub-carrier whose frequency is about 70 MHz. The sub-carrier is transmitted up to a communications satellite, which is in a geostationary orbit. (The satellite is located above the equator. The radius of its orbit is approximately 40 000 km and its orbital period is 24 hours. It therefore appears to be stationary with respect to an observer on the Earth.) The satellite re-transmits the modulated signal down to the receiver, where the 1 Hz tick is recovered by a cross-correlation of the received pseudo-random code with a copy of the code generated in the receiver. The time difference between the recovered 1 Hz tick and the local clock is then stored. The message exchange is full duplex, and the time differences at each station are combined to estimate the time difference of the clocks at the two sites.

The up-link and down-link typically use different frequencies in the Ku band. The up-link frequency is nominally 14 GHz and the down-link is nominally 11 GHz. Both frequencies are used on a portion of the path in each direction, but the paths are not the same so that the delays in the two directions may not be exactly equal. The dispersion of the refractivity of the ionosphere and the troposphere is small at these frequencies so that this asymmetry is generally not an important limitation. Balancing the transmit and receive delays in the ground-station hardware is a more difficult problem, especially because these delays are often sensitive to fluctuations in the ambient temperature.

The transit time from one station up to the satellite and down to the other station is about 0.25 s, so that the rotation of the Earth during this time must be taken into account. This is called the Sagnac[19] Effect. The magnitude of this effect is $2\omega A/c^2$, where c is the speed of light, ω is the angular velocity of the Earth and A is the area defined by the triangle formed by the satellite and the two stations projected onto the equatorial plane. The effect is positive for a message traveling eastward and negative for messages in the opposite direction.

The design of this method treats both stations as equal partners in the exchange. The transmissions at each end are generated by the local time scale and are not a response to a message received from the other end as with NTP. In principle, this is an important difference between this method and the NTP and ACTS methods described above, but the underlying assumptions of all of the methods are the same and only the details of the analysis are somewhat different.

Based on the notation of the discussion of the NTP message exchange, a well-designed NTP system will have a very small time delay, T_{3b}-T_{2b}, between when a system receives a query and when it responds; the ACTS protocol assumes that this difference is negligible; the two-way system makes no assumptions about this difference except that it is accurately measured on each message exchange.

The time differences measured with the two-way satellite method are much more accurate than the measurements made with either of the systems discussed previously because the delays are much more stable and the assumption of the very small delay asymmetry is more accurate. The RMS uncertainty of the measurements is of order 0.1 ns. The spectrum of the noise is approximately white phase noise for short averaging times, but there are less favorable variations at longer periods; some links have a quasi-periodic approximately diurnal variation in the time-difference data. The source of this variation is not understood at present.

13. The Polling Interval: How often should I calibrate a clock?

There are three different considerations that should be used to determine the interval between time-difference measurements. The first consideration is based on a statistical estimate of the noise processes, the second is derived from concerns about non-statistical errors, and the third is based on a cost/benefit analysis. I will begin by considering choosing a polling interval based on a statistical analysis.

From the statistical perspective, the goal of the time-difference measurements is to improve the accuracy or the stability of the local clock once its deterministic properties have been determined and used to adjust the time, frequency, and aging (if appropriate) of the clock under test. The deterministic parameters can be applied directly to the physical device or they can be used to adjust the readings of the clock administratively. (In general, timing laboratories usually do not adjust the physical parameters of a clock but rather apply the deterministic offsets administratively.)

Once the deterministic parameters have been included in the readings of the device under test, I assume in a statistical analysis that its time dispersion can be determined solely from its statistical characterization. At least at the beginning, I will assume that both the deterministic and stochastic parameters are constants that do not depend on time or on perturbations such as fluctuations in the ambient temperature.

In this model, the design of the calibration procedure is driven by the requirement that the accuracy or stability of the remote clock as seen through the channel should be better than the corresponding parameters of the local clock for the same query interval. (The channel includes the physical medium used to transmit the time signal and any measurement hardware at the end stations.) As a practical matter, it often turns out that the statistical characteristics of the channel are much poorer than the statistics of the remote clock itself. In this situation, which is quite common, improving the accuracy or the stability of the remote clock will have almost no effect on the performance of the synchronization process, which will be dominated by the statistics of the channel. The following discussion depends only on being able to characterize the combination of the remote clock and the channel by means of the two-sample Allan deviation. The analysis is not sensitive to whether the source of the fluctuations is in the clock or the channel connecting it to the device under test.

Suppose that the statistical characteristics of the remote clock seen through the channel can be described as white phase noise for all averaging times. This is the best that we can hope for – the measurement process using this channel is degraded by a noise process that has a mean of zero and a standard deviation that does not depend on the time the measurement was performed or on any external parameter such as the ambient temperature. The magnitude of the two-sample Allan deviation (the square root of the variance) that describes the statistics of the remote clock seen through this channel varies as the reciprocal of the averaging time. The time dispersion in this configuration is independent of the averaging time. See eq. 10 and 11 above. This is not a surprising result. If the measurement noise is characterized as white phase noise, then the fluctuations of every measurement are derived from the same distribution with a mean of zero and a fixed standard deviation, and there is no relationship between one measurement and any other one, so that the time between measurements is irrelevant to the statistics of the time differences.

Now consider that the stability of the local clock is characterized as pure white frequency noise for moderate averaging times. Again, this is not a surprising result and is about the best we could ever hope to see; once the white phase noise of the measurement process has been accounted for, the next effect is the noise of the frequency control loop, which we also take to be a process with a mean of zero and a known standard deviation. (There are almost always longer period effects that modify the assumption of pure white frequency noise, but I will assume for now that the averaging times that will be used will not be large enough to make this consideration important.) The two-sample Allan deviation for white frequency noise varies as the reciprocal of the square root of the averaging time, so that time dispersion due to white frequency noise increases as the square-root of the averaging time. (I will use the conservative estimate that I discussed above for eq. 11.)

We can now combine these two results to define two measurement strategies. The noise of the time-difference measurement process is characterized by a standard deviation, *M*, which does not depend on averaging time. The time dispersion of the local clock due to its white frequency noise is characterized as a function of averaging time by $C\tau^{1/2}$. From the statistical perspective, the goal of the first synchronization procedure is to set the averaging time so that the remote clock seen through the channel is more stable than the local clock. In other words, the averaging time should be chosen so that the free-running time dispersion of the local clock due to its frequency noise is greater than the time dispersion of a time-difference measurement with respect to the remote clock seen through the channel:

$$\begin{aligned} C\tau^{1/2} &\geq M \\ \tau &\geq \left(\frac{M}{C}\right)^2 . \end{aligned} \tag{39}$$

The result may be surprising but is easily explained. As the remote clock seen through the network becomes less stable (increasing *M*), the crossover between its stability and the stability of the local device, which increases as the square root of the averaging time, moves to longer and longer averaging times. Thus we would expect that the optimum polling interval for a time transfer that used the wide-area Internet to communicate between the local and remote devices would be longer than the optimum polling interval for the same devices that exchanged messages on a local network connection because the stability of the transmission delay in a wide-area network would be poorer.

Since the channel connection back to the remote clock is characterized by white phase noise, a second strategy would be to make measurements of the time difference as rapidly as possible and average the data. For example, suppose we could make measurements every second for a time interval of T seconds. If we averaged these T measurements, the standard deviation of the mean would be reduced from M to $M/\sqrt{T}$. The time dispersion due to the white frequency noise of the local device would be the same as before, so that the comparison of eq. 39 becomes

$$\begin{aligned} CT^{1/2} &\geq \frac{M}{T^{1/2}} \\ T &\geq \frac{M}{C} \end{aligned} \quad , \tag{40}$$

where the value of T in eq. 40 specifies the point at which the noise of the local clock is greater than or equal to the standard deviation of the average, so that the averaging algorithm can improve the stability of the local clock starting at an averaging time of T. In this simple model, once this time is reached, additional averaging only makes things better because the standard deviation of the remote clock seen through the channel improves without bound while the stability of the local clock degrades without bound. In the limit of very large T, we are not using the data from the local clock at all.

This simple model will break down at some point for one of two reasons. The first possibility is that channel noise is not purely white phase noise starting at some averaging time – longer period fluctuations become important and they are not zero-mean random processes. These fluctuations can be incorporated by modeling the time dispersion of the remote clock, M, as constant at short times but increasing as some power of the averaging time starting at some averaging time, T_m. The second possibility is that the requirements of the application limit the averaging time – we cannot average forever because we need to use the time difference for some application. The assumptions that I used in this discussion are somewhat artificial in that they are often better than real-world devices and channels. Therefore, these calculations are more illustrative of the method than rigorous derivations with very general applicability.

14. Error Detection

Any measurement protocol that receives data from a remote device over a noisy channel should be prepared to consider the possibility that the received data are in error, either because the remote clock has failed or the channel characteristics have changed suddenly. A purely statistical analysis cannot be the whole story here, since there is generally no objective way of distinguishing between an error and a very low probability event that conforms to the statistical description. One method that is commonly used is to regard a measurement that differs from the mean (or from the predicted value) by more than three standard deviations as an error.

The machinery that I developed in the previous section can also be used to detect possible errors. For example, consider the averaging strategy presented in the discussion for eq. 40. Instead of waiting until all of the measurements have been completed to evaluate the average time difference, we could construct a running mean with an update each time a new time difference was acquired. The estimate of the mean at the k^{th} step, $\overline{X}_k$, after receiving the time difference x_k can be calculated iteratively based on the estimate of the mean at the previous step,

$$\overline{X}_k = \frac{(k-1)\overline{X}_{k-1} + x_k}{k} \quad , \tag{41}$$

where the estimate of the mean is initialized to zero.

One possibility is to ignore the k^{th} estimate as having a one-time error if it differs from the running mean by more than three times the running estimate of the standard deviation computed from the current average or from a previous one:

$$\left|(x_k - \overline{X}_{k-1})\right| > 3\sigma_{k-1} \tag{42}$$

The assumption that the difference is a one-time error would be confirmed if the next reading was consistent with the running mean value.

The situation becomes more complicated if the next measurement does not conform to the running mean either, and it may not be possible to distinguish between a problem with the remote clock, the local clock, or the channel. It is sometimes possible to decide this question if a second independent calibration source or an independent channel is available. This solution must be considered in the cost-benefit analysis that I discuss in the next section.

15. Cost-Benefit Analysis

In this section I will consider the situation where a time-difference measurement has a finite cost in terms of computer cycles, network bandwidth, or some other finite resource. The tradeoff between the accuracy of a time-difference measurement and the cost needed to realize it becomes important in this situation.

For example, consider again the simple case where the time difference measurements are characterized as white phase noise with a mean of zero. The standard deviation of the mean of N measurements decreases as $1/\sqrt{N}$, and this improvement can continue without bound in principle. On the other hand, the cost of the measurement process increases linearly with N, assuming that each measurement has the same cost. In this situation, the cost-benefit analysis is always unfavorable – the cost of the measurements always increases faster than the standard deviation improves, and the best cost-benefit strategy is to make the minimum number of measurements consistent with the standard deviation that is required to meet the needs of the application.

More generally, I assume that the total cost of a measurement procedure, *C*, is given by the cost of a single measurement, *c*, the interval between measurements, τ, and the total measurement time , *T*:

$$C = c\frac{T}{\tau} \quad . \tag{43}$$

I take the benefit of the procedure, *B*, as the time dispersion of the device for an averaging time, τ, where the time dispersion is calculated from the two-sample Allan deviation for that averaging time:

$$B = \sigma_y(\tau)\tau \ , \tag{44}$$

The goal is then to minimize the product *BC*, possibly with some additional constraint that the time dispersion must be less than some value required by the application.

Apart from the constants, the product *BC* is a function only of the two-sample Allan deviation, so that it will always improve with increasing averaging time as long as the slope of the Allan deviation is negative, and the best strategy will be the longest averaging time that satisfies the time-dispersion estimate in eq. 44. The slope of the two-sample Allan deviation is negative in white phase noise and white frequency noise domains, so that a pure cost-benefit analysis will always favor an averaging time that is at the onset of flicker processes where the slope of the Allan deviation approaches zero. The cost-benefit analysis product is a constant independent of averaging time in the flicker domain, but the time dispersion increases with averaging time (eq. 44), so that the dispersion may not satisfy the requirements of the application in this domain. The cost-benefit analysis becomes unfavorable in the random-walk of frequency domain where the slope of the two-sample Allan deviation is positive. The time dispersion of the local clock is increasing faster than the cost is decreasing in this region. This region might still be an acceptable choice if the accuracy requirement is very modest.

The method used for detecting errors also has a cost/benefit aspect. For example, if an Internet client queries N Internet servers on every measurement cycle in an attempt to detect an error, then the cost of the synchronization process has increased by a factor of N; the benefit will depend on how often this procedure detects a problem. Shortening the polling interval to detect a problem with the local clock more quickly is subject to the same considerations. That is, do problems happen often enough to justify the increased cost of the algorithm? In general, comparing the measured time difference with a prediction based on the statistics of the local clock (eq. 42. for example) and querying multiple servers only when that test fails is a better strategy because it exploits the statistics of the local clock as a method for detecting a possible error with the remote clock or with the channel.

A cost-benefit analysis is very important from the perspective of the operators of the network and the public time servers – increasing the polling interval and reducing the number of servers queried on each measurement cycle translates directly into the number of users that can be supported with available, scarce resources.

16. The National Time Scale

The official time in the United States (and in most other countries) is Coordinated Universal Time (UTC). The length of the UTC second is defined by the frequency of the hyperfine transition in the ground state of cesium. The frequency of this transition is defined to be 9 192 631 770 Hz, and counting this number of cycles defines the length of the second. The other time units (minutes, hours, …) are multiples of this base unit.

The length of the day, computed as 86 400 cesium seconds, is somewhat shorter than the length of the day in the UT1 time scale, which is a time scale based on the rotation of the Earth. The accumulated time difference is currently somewhat less than 1 s/year. In order to maintain a close connection between atomic time, defined by the cesium transition frequency, and the UT1 time scale, which characterizes the position of the Earth in space, additional seconds are added to UTC whenever the difference between UTC and UT1 approaches 0.9 s. The decision to add these "leap seconds" is made by the International Earth Rotation Service, and all national timing laboratories incorporate the leap second into their time services.

Leap seconds are normally added as the last UTC second of the last day of June or December. In the vicinity of a leap second, the time stamps are: 23:59:58, 23:59:59, 23:59:60, and then 00:00:00 of the next day. Digital time services and most clocks cannot represent the leap second time of 23:59:60 and stop the clock for one second at 23:59:59. The time services operated by the National Institute of Standards and Technology implement the leap second by transmitting a time value equivalent to 23:59:59 twice, and most other time services do the same thing. Assigning the same time stamp to two consecutive seconds is ambiguous and has obvious difficulties with respect to causality, and the question of continuing the leap second procedure is currently (as of 2013) being discussed.

The details of the leap second procedure are important for users who must synchronize a clock to the official time scale in the vicinity of a leap second. Unfortunately, some time services implement the leap second in different ways. One method adds the leap second by duplicating the time 00:00:00 of the next day. This eventually results in the same time as the NIST method, but it adds the leap second in the wrong day, and has time errors on the order of 1 s in the immediate vicinity of a leap second.

A more troubling method implements the leap second as a frequency adjustment during the last few minutes of the day. The clock is slowed down for some period of time until the additional second has been added. This method has both a time error and a frequency error during the time the leap second is being inserted. The clock is never stopped in this implementation, but both the time and the frequency are not correct with respect to national time standards during this interval. In addition, there is no generally accepted method for implementing the frequency adjustment, so that different implementations of this method will also have errors with respect to the national standards of time and frequency in this vicinity of a leap second.

In addition to the two proposals: (1) not to make any changes or (2) to stop adding future leap seconds to Coordinated Universal Time but to continue the number of leap seconds that have already been added, a number of other alternatives have been suggested. One proposal would be to switch to International Atomic Time (TAI) as the legal time scale, which would effectively reset the leap second count to zero in a single step. Another proposal would be to stop adding leap seconds to UTC, and to change the name of the time scale to reflect this change in its implementation.

17. Traceability

There are many applications that require time stamps that are traceable to a national time scale, and realizing this requirement requires clocks that are synchronized to a national or international standard of time. A clock is *traceable* to a national time scale if there is an unbroken chain of time-difference calibration measurements between the clock and the reference time scale by means of any of the methods that I have described above. Each one of these measurements must have an uncertainty estimate.

It can be difficult to establish the chain of measurements that is required for traceability. For example, the signal *in space* transmitted by a GPS satellite is traceable to Coordinated Universal Time (UTC), the national and international reference time scale, through the US Naval Observatory, which monitors the time signals broadcast by the GPS constellation and computes the offset between GPS system time and UTC as maintained by the Naval Observatory. This offset is uploaded into the satellites and is transmitted as part of the navigation message.

However, the traceability of the signal in space does not necessarily extend to the timing signals produced by a GPS receiver unless the receiver, the antenna, and the connecting cable have been calibrated. The traceability almost certainly does not automatically extend to the application that uses the timing signals to apply time stamps as part of some application. This discussion does not suggest that these links in the chain are known to be inadequate or in error, but rather that they do not satisfy the strict definition of traceability without some sort of calibration procedure.

There are some situations where the requirements of strict traceability can be satisfied without a complex calibration procedure. For example, if an application requires that time stamps be traceable to a national time standard with an uncertainty of less than 1 s, then simply certifying that the satellite timing equipment is working properly is likely to be good enough. The uncertainty of the time signals produced by a receiver synchronized using signals from a GPS satellite is several orders of magnitude smaller than the requirements of the application, so that the overall system is surely traceable at the level of 1 s if it is working at all. (Verifying that a GPS receiver is working properly may or may not be an easy job – it depends on the specific receiver that is used.)

A second aspect of traceability is *legal traceability*, by which I mean being able to establish in a legal proceeding that a time stamp was traceable to a national time scale. In this situation, "doing the right thing" might not be adequate if you can't prove it to a judge and jury.

Given that the technical aspects of traceability that I discussed above have been satisfied, establishing legal traceability is generally a matter of documentation – maintaining log files that show that the system was calibrated with an uncertainty consistent with the requirements of the application and that it was operating normally at the time in question. A log file that has entries only when there is a problem is unlikely to be adequate – it will have no entries when the system is working properly, and an empty log file is ambiguous and may not be of much help.

18. Summary

I have discussed a number of methods for synchronizing a clock using a reference device that can be located either in the same facility or remotely and linked to the device under test by a communications channel. I have discussed a number of methods of synchronizing a remote clock and the statistical considerations that characterize the accuracy of the procedure and how often to request a calibration. An important tool in these discussions is the two-sample Allan variance, and I have presented a simple introduction into how this estimator is calculated and used.

19 Bibliography

The following list contains a few of the very large number of publications that contain additional information on time and frequency standards and distribution methods.

1. The publications of the Time and Frequency Division of the National Institute of Standards and Technology, generally available on line at tf.nist.gov.
2. "Encyclopedia of Time," Edited by Samuel L. Macey, New York, 1994, Garland Publishing, Inc.
3. Special Issue on Time Scale Algorithms, Metrologia, Vol. 45, Number 6, December, 2008.
4. "Time and Frequency Measurement," Edited by Christine Hackman and Donald B. Sullivan, College Park, Maryland, 1996, American Association of Physics Teachers. This was also published as resource letter TFM-1 in the American Journal of Physics, Vol. 63, Number 4, pages 306-317, April, 1995.
5. David L. Mills, "Computer Network Time Synchronization," 2nd edition, New York, 2011, CRC Press.
6. Special Issue on Time and Frequency, Proceedings of the IEEE, Volume 79, Number 7, July, 1991.

7. "Understanding GPS: Principles and Applications," 2nd edition. Edited by Elliott D. Kaplan and Christopher J. Hegarty, Norwood, Mass., 2006, Artech House, www.artechhouse.com.
8. "Global Positioning System: Signals, Measurements and Performance," 2nd edition, Pratap Misra and Per Enge, Lincoln, Mass., 2006, Ganga-Jamuna Press, GPStextbook@G-JPress.com.

---

[1] Stefano Bregni, "Measurement of the Maximum Time Interval Error for Telecommunications Clock Stability Characterization", IEEE Trans. strumentation and Measurement, Vol. 45, pages 900-906, Oct. 1996.

[2] Synchronization Interface Standard, ANSI Standard T1.101-199X.

[3] Samuel R. Stein, "Frequency and Time – Their Measurement and Characterization," Precision Frequency Control, Vol. 2. ch. 12, published by Academic Press, New York, 1985. This paper and a number of others on the same topic are reprinted in NIST Technical Note 1337, edited by D. B. Sullivan, D. W. Allan, D. A. Howe and F. L. Walls, published by the US Department of Commerce, March 1990.

[4] Brian D. O. Anderson and John B. Moore, "Optimal Filtering," New York, Dover Publications, 1979, pages 54, ff.

[5] R. B. Blackman and J. W. Tukey, "The Measurement of Power Spectra," New York, Dover Publications,1959, pages 174, ff.

[6] Y. Lin, S. Lee, S. Li, Y. Xie, Z. Ren and C. Nguyen, "Series-resonant VHF Micromechanical Resonator Reference Oscillators," IEEE J. Solid-State Circuits, vol. 39, pp. 2477-2491, No. 12, Dec. 2004. See also Ville Kaajakari, "Practical MEMS," Small Gear Publishing, 2009. www.smallgearpublishing.com.

[7] Claude Audoin and Bernard Guinot, "The Measurement of Time: Time, Frequency and the Atomic Clock," Cambridge, UK, Cambridge University Press, 2001, chapter 6.

[8] H. G. Andresen and E. Pannaci, "Servo controlled hydrogen maser cavity tuning," Proc. 20th Annual Frequency Control Symposium, 1966, Atlantic City, New Jersey. See also, C. Audoin, "Fast Cavity auto-tuning system for hydrogen maser," Rev. Phys. Vol. A16, pages 125-130, 1981.

[9] George E. P. Box and Gwilym M. Jenkins, "Time Series Analysis: forecasting and control," San Francisco: Holden-Day, 1970.

[10] D. A. Howe, D. W. Allan and J. A. Barnes, "Properties of Signal Sources and Measurement Methods," Proc. 35th Annual Symposium on Frequency Control, A1-A47, 1981. This is document AD-A110870 available from National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

[11] S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe and L. Erb, Proc. 36th Annual Symposium on Frequency Control, 314-320, 1982. This is document AD-A130811 available from National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

[12] Judah Levine, "Introduction to time and frequency metrology," Rev. Sci. Instr. Vol. 70, No. 6, pages 2567-2596, 1999.

[13] G. Panfilo, A. Harmegnies and L. Tisserand, Proceedins of the 2011 Joint European Time and Frequency Forum and International Frequency Control Symposium, published by the IEEE, Psicataway, New Jersey, pages 850-855, 2011.

[14] Judah Levine, "The Statistical modeling of atomic clocks and the design of time scales," Rev. Sci. Instr. Vol. 83, 021101, 2012.

[15] R. H. Jones and P. V. Tryon, "Estimating Time From Atomic Clocks," J. Res. NBS, vol. 88, pages 17-28, 1983. See also A. Gelb, "Applied Optimal Estimation," Cambridge, Massachusetts, MIT Press, 1974, Chapter 4.

[16] Post-processed orbits and clock solutions are computed by the International GPS service, and are available on line at www.igs.org and www.igscb.jpl.nasa.gov.

[17] Many references. See, for example, Douglas E. Comer, "Internetworking with TCP/IP, Vol. 1, Ch. 11. Englewood Cliffs, New Jersey, Prentice-Hall, 1991.

[18] J. Levine, M. Weiss, D. D. Davis, D. W. Allan and D. B. Sullivan, "The NIST Automated Computer Time Service," J. Res. NIST, vol. 94, pages 311-322, 1989.

[19] N. Ashby and M. Weiss, “Global Positioning Receivers and Relativity,” National Institute of Standards and Technology Technical Note 1385, Boulder, Colorado, 1999.