Incorporating Data from a Primary Frequency Standard Into a Time Scale

Judah Levine

Abstract—**It is difficult to combine data from a primary frequency standard with the time-difference measurements that are usually used as input to most time scales because a primary standard usually operates only occasionally on an irregular schedule and because the fundamental output of a primary frequency standard is a *frequency* rather than a time, and there is often no natural way of inserting this kind of datum into the scale in a manner that is statistically robust. We will present a new time-scale algorithm that seeks to address these problems. We call this frequency-based algorithm AF1, by analogy with the time-based algorithm AT1 that has been used at NIST for many years. Unlike AT1 in which frequency is simply a parameter that specifies how the time of a clock evolves between measurements, however, the frequency of each clock is a fundamental parameter in AF1. This change in focus provides a natural way for incorporating data from a primary frequency standard into the ensemble. We will present the details of the algorithm and results using data from our primary frequency standard, NIST-7.**

I. Introduction

A primary frequency standard differs from the commercial cesium devices that are used for timekeeping in one important respect: the device is constructed so that the systematic difference between its output frequency and the International System of Units (SI) definition of the second can be evaluated each time a measurement of the standard frequency is performed. This evaluation process is quite time-consuming and involves several diagnostic experiments which affect—and may even interrupt—the output signal. As a result, the output of such a device is only occasionally "on frequency," and there are often significant periods of time when it is not producing any output at all. If the data are to be generally useful, there must be a procedure for realizing this standard frequency in real-time by transferring these aperiodic observations to a system that operates continuously.

Many laboratories (including NIST) use a hydrogen maser as the transfer oscillator between the primary standard and other systems. The evaluation procedure for the primary standard uses the maser as a local oscillator, and effectively calibrates the frequency of the maser in absolute SI units. The output of the maser (corrected either electrically using a phase stepper or administratively via software) then serves as the real-time realization of the SI frequency. The frequency stability of the maser (or, more precisely, the stability of its frequency aging) determines how well this procedure will work. This stability enters explicitly into the uncertainty with which the calibration can be "remembered" between evaluations. It may also enter into the error budget of the evaluation process itself because the maser is often used as the flywheel oscillator to monitor the effects of reversing the direction of the beam, changing the microwave power level, and the other experiments that are part of the evaluation protocol. Its stability (rather than its frequency accuracy) is all that is required here, so that it is important to be able to detect a glitch in its frequency that occurs during the evaluation procedures.

These considerations suggest that an algorithm that could provide real-time estimates of the frequency and frequency-aging of the maser would be very useful. It could be used to detect glitches in the operation of the maser during an evaluation and would provide an estimate of the performance of the maser between evaluations. Ideally, this algorithm would not depend on the data from the primary frequency standard itself (at least in the short term) so that its data would provide a statistically independent measure of the performance of the maser and of the procedure for realizing the standard frequency using it.

Although a conventional time-scale (such as AT1) might be used for this purpose, there are a number of reasons why this is not an optimum solution. The AT1 scale as currently implemented at NIST [1] contains no algorithms for modeling frequency aging. This is appropriate for a scale that is based primarily on cesium devices, since such oscillators tend to have very small aging parameters which are difficult to determine in the presence of the flicker and random-walk frequency modulations that tend to characterize these devices at periods longer than a few days. A scale that makes heavy use of hydrogen masers, on the other hand, must have aging parameters, since they are usually statistically significant for most masers. A second concern is dealing with the white frequency noise that is important at shorter times. The need to detect frequency glitches during an evaluation process implies that the frequency stability of the scale at short times (on the order of a few hours) is an important parameter. Since the underlying noise processes tend to be white at these averaging times, a natural solution is to measure the clocks much more frequently than is usually needed for time-scale operations.

Another fundamental consideration is the need to design an algorithm that makes optimum use (in a statistical sense) of the calibration data from the primary standard.

Manuscript received May 16, 1996; accepted November 14, 1996.

The author is with Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80303 (e-mail: jlevine@boulder.nist.gov).

If the frequency and frequency aging of a maser are carefully modeled, its residual fractional frequency noise can be considerably smaller than the fractional uncertainty in the evaluation of the primary frequency standard, which is currently on the order of 10^{-14} for NIST-7. If we think of the primary standard as having white frequency noise at about this level, then simply steering the maser (electrically or administratively) to follow the evaluation data is not optimum because the frequency stability of the maser (as distinct from its accuracy) is probably at least a factor of 10 better than this value. In other words, it is important to evaluate the primary frequency standard relatively frequently so as to average its white frequency noise using some type of flywheel oscillator, and the link between it and its transfer oscillator must be designed to facilitate this averaging process so as to make the best use of both devices. Loosely speaking, the algorithm should "average" a number of consecutive evaluations and steer the maser only when the uncertainty in this average has dropped below the frequency noise of the maser itself. An important design criterion for the algorithm is to provide an objective measure of determining this cross-over point using data from devices whose behavior is not correlated with the noise in the maser which is used as the transfer oscillator.

II. The Measurement Protocol

It would be nice if there was an easy way to measure frequency directly, but there is no easy way to do this with the requisite resolution. Therefore, our primary data are time differences. Our measurement hardware reports the time *difference* between each clock in the local ensemble and the hardware reference clock. The time difference between the reference clock and the j-th clock, measured at epoch t_k, is $x_j(t_k)$. The subscript value $j = m$ will be used to designate the device (usually a maser) which is used in the evaluation process for the primary frequency standard and whose output realizes the standard frequency between evaluations. By analogy with the terminology used to describe AT1, we will call this device the working frequency standard. It is connected to the measurement system just like any other clock. In particular, it is not necessarily the hardware reference clock of the measurement system.

The parameter measured by the j-th hardware channel is the time of the reference clock minus the time of clock j. The average frequency difference between the j-th clock and the working frequency standard during the time interval Δt between t_{k-1} and t_k is:

$$f_{jm}(t_k) = f_j(t_k) - f_m(t_k), \tag{1}$$

where

$$f_j(t_k) = \frac{x_j(t_{k-1}) - x_j(t_k)}{\Delta t} \tag{2}$$

and

$$f_m(t_k) = \frac{x_m(t_{k-1}) - x_m(t_k)}{\Delta t}. \tag{3}$$

Note that the performance of the reference clock cancels in (1), but that measurement noise associated with this procedure for determining f_{jm} is approximately $\sqrt{2}$ larger than a procedure that measured the time differences between clocks j and m directly because two independent time differences are measured at each epoch in this method.

We now define a pseudo-clock called "the ensemble," and we use the subscript "e" to denote a parameter measured or estimated with respect to it. We will define its frequency to be our best estimate of the frequency of International Atomic Time (TAI) [2], [3] at any epoch. The initial value for the average frequency of each clock with respect to the ensemble, y_{je}, is then the value published by the Time Service of the International Bureau of Weights and Measures (BIPM) for the frequency of the clock in question (with respect to TAI) for the epoch which we designate as the origin time for the algorithm. The frequency aging for each clock with respect to the ensemble, d_{je}, is analogously defined using the first difference of the frequency estimates published by the BIPM for the two consecutive epochs ending with the origin time of the algorithm. These two parameters can be used to predict the frequency of the clock with respect to the ensemble at any future time. Expressed in an iterative form, the frequency prediction becomes:

$$\hat{y}_{je}(t_k) = y_{je}(t_{k-1}) + d_{je}(t_{k-1})\Delta t. \tag{4}$$

We assume initially that the frequency aging terms do not evolve with time but are fixed at the initial values determined from the BIPM data. This assumption will be relaxed in the manner to be described below.

In the AF1 algorithm, each clock is characterized by only these two parameters—a frequency term which is estimated by the procedure after each measurement cycle and a frequency aging term which is adjusted much more infrequently using data from the primary frequency standard in a subsidiary procedure to be discussed below. The algorithm does *not* incorporate time states either for the member clocks or for the ensemble itself, nor does it estimate the frequency aging using the same data as is used to estimate the frequencies (2) and (3). These are significant differences from AT1 [1], TAI [3], or the Kalman formulation of the algorithm discussed by Stein [4], [5].

These equations have been presented as defining the parameters of the physical clocks with respect to a pseudo-clock called "the ensemble," but they could equally well be thought of as definitions (or predictions) of the ensemble pseudo-clock with respect to the physically observable hardware. It is useful to keep both of these points of view in mind in the following discussion.

It is also important to distinguish between the frequency of the ensemble, which is initialized using data from TAI but is then free-running using only data from our local clocks and TAI itself, which is computed by the BIPM using a much larger number of clocks located in many laboratories.

III. Estimating the Ensemble Parameters

We can combine the equations above with the time-difference data obtained on each measurement cycle to update our estimate of the frequency of the working standard with respect to the ensemble. Specifically, the average frequency of the maser with respect to the ensemble during the time interval of length Δt ending at time t_k can be estimated using the data from clock j. To do this, we combine the predicted frequency difference between clock j and the ensemble from (4) with the measured frequency difference between clocks j and m over the most recent time interval from (1):

$$f^j_{me}(t_k) = y_{je}(t_{k-1}) + d_{je}(t_{k-1})\Delta t - f_{jm}(t_k), \quad (5)$$

where the superscript j on the left-hand side indicates that the estimate is made using the time-difference data of clock j. The working standard requires no special handling. Equation 5 reduces to (4) in this case, since $f_{mm} \equiv 0$.

There is one of these equations for each clock in the local ensemble, and each one of them provides an estimate of the frequency of the working standard with respect to it. Since *any one* of them uniquely defines the relationship between the working standard and the ensemble pseudo-clock, all of them together over-determine this parameter. The various definitions may be mutually inconsistent because of noise in the measurement processes and in the clocks themselves. If we have done our job properly, this noise will have a white spectrum about an underlying mean value, and we might suppose that a suitably weighted average of these definitions would provide an unbiased estimate of the ensemble frequency. In other words, if this algorithm worked properly, the weighted average (the ensemble frequency, in other words) would provide an unbiased prediction of the frequency of TAI. The frequency (with respect to TAI) of any member clock in general, or the working standard in particular, could be predicted using the equations above.

Making the assumption that the estimates in (5) are characterized by white frequency noise about an underlying mean value, the average ensemble frequency over the time interval Δt ending at time t_k is computed using:

$$f_{me}(t_k) = \sum_j w_j f^j_{me}(t_k), \quad (6)$$

where the weights are proportional to the reciprocal of the prediction variance (as defined below). They are normalized by:

$$\sum_j w_j = 1. \quad (7)$$

The operation defined by (6) will always exist in a formal sense, but the result is meaningful if, and only if, the estimates defined by (5) are dominated by white noise and are, therefore, normally distributed about an underlying mean value.

There are two types of noise processes that contribute to the estimates in (5): measurement noise from the time difference hardware and frequency noise from the clocks. It is reasonable (although perhaps a bit optimistic) to assume that both contributions are dominated by white processes at short-time intervals, so that the weighted sum defined by (6) is still limited by white noise processes. Averaging this frequency would then yield an estimate that converged to the mean frequency difference between the working standard and the ensemble if the averaging interval is not too long. Decreasing the time interval between measurements should improve matters both because the clock noise is more likely to be white and because the standard deviation of the mean computed over a given time interval decreases as more measurements are averaged. Conversely, the ensemble frequency defined by this algorithm is likely to be limited by flicker and random-walk frequency fluctuations at longer times because those are the statistics of the clocks that contribute to its definition.

These considerations lead us to the usual conclusion: there is an optimum averaging time for estimating the frequency of the working standard with respect to the ensemble. This averaging time is determined by the point at which non-white processes begin to dominate the noise spectrum of f_{me}. By analogy with AT1, we average f_{me} using an exponential filter whose time-constant is set using the considerations we have just discussed:

$$y_{me}(t_k) = \frac{y_{me}(t_{k-1}) + G f_{me}(t_k)}{1+G}. \quad (8)$$

The time constant G is dimensionless and is normally less than 1. Using this formulation, the crossover between white and non-white behavior is set at approximately $1/G$ time intervals (i.e., a time of $\Delta t/G$). The time constant is usually on the order of days for most clocks.

As we discussed above, (8) can also be thought of as defining the frequency of the ensemble with respect to the working frequency standard. It is now a straightforward matter to compute the frequency of each of the other clocks with respect to the ensemble, f_{je}, by adding (8) and (1). As with the working frequency standard, we assume that the fluctuations in the frequency of each clock are dominated by white processes at short times, changing to flicker or random-walk fluctuations at longer periods. We average the frequency of each clock using an exponential filter as defined in (8) above; the time constant for each clock is determined from an evaluation of its noise performance.

We could also think of the point at which non-white processes begin to dominate the statistics of the frequency estimated in (6) as due to the fact that the frequency aging parameters are no longer constant. In this approach, we would use the f_{je} to modify d_{je} rather than y_{je}. It is much more difficult to do it this way because it is the *time integral* of the aging that modifies the long-term frequency rather than the aging itself, and the integral of a white process is no longer white even at short times. We choose instead to modify the aging parameters using the data from the primary frequency standard.

The uncertainty in an evaluation of the primary standard is considerably larger than the noise of the ensemble at short averaging times. If we assume that data from primary frequency standards are characterized by white frequency noise at *all* averaging times, the standard deviation of their mean improves as more measurements are included. At the same time, the frequency fluctuations of the clocks that contribute to the ensemble computation (and therefore the ensemble frequency itself) will be increasingly affected by non-white processes at longer averaging times. The uncertainty in the ensemble frequency therefore *does not* improve with averaging time beyond some point; in fact, it actually begins to increase again at very long times. The uncertainty in the frequency of the primary standard will, therefore, drop below the noise in the scale for some averaging time, and it is at this point that data from the primary standard should be included as a modification to the aging parameters of the ensemble members. The effect of this procedure is to use the aging parameters to model the very long-term change in the frequencies of the clocks in the ensemble.

A natural way to modify these aging parameters would be to define a second exponential filter which steered them using an "error term" proportional to the difference between the frequency of TAI estimated by the ensemble and the estimate from the primary frequency standard. The time constant of this filter would be determined by comparing the performance of the primary frequency standard with the statistics of the ensemble. Assume, for example, that the "flicker-floor" of the ensemble frequency is about 4×10^{-15} and that it is reached at an averaging time of about 6 months. If the primary standard is evaluated once per month, and if the fractional uncertainty of each evaluation is about 10^{-14}, then the standard deviation of the mean of six evaluations would be roughly equal to the flicker floor of the scale after 6 months. Decreasing the interval between evaluations does not improve the overall accuracy very rapidly, since the standard deviation of the mean decreases only as the square root of the number of observations. Increasing the interval between evaluations, on the other hand, means that it takes longer for the standard deviation of the mean frequency of the evaluations to reach the ensemble flicker floor. More importantly, random-walk processes are likely to become increasingly important in limiting the ensemble stability at these longer averaging times, so that the overall performance of the ensemble in "remembering" the evaluated frequency is degraded.

The ensemble frequency itself should have no *deterministic* frequency aging in principle, and this fact can be used to constrain the aging parameter of each of the clocks. The situation in practice is more complicated. In the first place, both the ensemble and the member clocks are likely to exhibit random-walk frequency noise at long periods, and it is not easy to differentiate between large random-walk excursions and a deterministic secular aging. In the second place, TAI itself is steered by the BIPM using data from a worldwide ensemble of primary frequency standards, and this steering will produce an apparent deterministic aging in the ensemble with respect to TAI if it is not included.

IV. Prediction Error and the Weights

The prediction error for each clock is the difference between the frequency of the working standard with respect to the ensemble predicted using its data (5) and the average of these predictions over the ensemble (6). This difference is converted to a time dispersion using the interval between measurements, Δt:

$$e_j(t_k) = \{f^j_{me}(t_k) - f_{me}(t_k)\}\Delta t. \tag{9}$$

As we pointed out above, we expect that e_j will be normally distributed over short time intervals with a mean of 0 and a standard deviation that characterizes the quality of clock j. We can estimate this standard deviation using the standard exponential filter technique with the time constant, G', determined in the same way as G in (8) above:

$$\sigma_j(t_k) = \sqrt{\frac{\sigma_j^2(t_{k-1}) + G' e_j^2(t_k)}{1 + G'}}. \tag{10}$$

The effective time constant in (10) is typically about 10 days. Using a measurement interval of 720 s, values of σ range from about 30 ps to 1 ns.

In order to determine the weight of clock j and to detect a possible glitch in its performance, we form the statistic χ, which is the ratio of the prediction error in this measurement cycle to the average quality of the clock determined from previous data using (10):

$$\chi_j(t_k) = \frac{|e_j(t_k)|}{\sigma_j(t_{k-1})}. \tag{11}$$

1. If $\chi > 4$ for some clock, then we assert that the frequency of that clock has suffered a glitch. We assume that the problem is an error in the measurement of its time difference, since that is the most common type of problem. We set the weight of the clock to 0 in (6) and recompute that average and all of the quantities that depend on it (such as (9) for each of the other clocks). Since our model of the event is that it is a *one-time* measurement error, we do not change any of the parameters of this clock using data from the current cycle, and we hope for better days ahead. The clock will be returned to the ensemble with its current parameters on the next measurement cycle.

2. If $\chi < 3$ then we assert that the clock is behaving normally. We calculate its weight in the ensemble using:

$$w_j = \frac{1}{\sigma_j^2(t_{k-1})} \tag{12}$$

subject to the normalization requirement (7) and possibly to other limitations to be discussed below. (Although σ has the dimensions of time, the weights are dimensionless numbers. The weight calculated above must be normalized by dividing by the sum of the weights before it can be used in (6), and this normalized ratio is dimensionless.)

3. If $3 \leq \chi \leq 4$ then we assert that the clock is not behaving normally but that its error is not large enough to be classified as a glitch. We compute the weight of the clock in the ensemble using (12) multiplied by the deweighting factor:

$$Q_j(t_k) = 4 - \chi_j(t_k), \tag{13}$$

which deweights the clock smoothly from the full weight used for normal performance to the weight of 0 that is used for a glitch. We use this modified weight to recompute the scale parameters as in case 1 above, and we allow the current measurements to modify the parameters of the deweighted clock. This linear deweighting function is chosen because it is simple and smooth—it has not been rigorously justified. It is based on the idea, first suggested by Percival [6], that gradually deweighting a clock was a more reasonable procedure than resetting it abruptly when its error exceeded an arbitrary threshold.

If the prediction errors can, in fact, be characterized as a normal process, then there is a 0.3% probability of making an observation which has a prediction error of 3 standard deviations; the probability that the error is 4 standard deviations is 0.006%. These values are not as small as they seem. Our test ensemble (to be discussed below) has eight clocks whose times are measured every 12 minutes, which is 960 measurements/day. Based on these probabilities, we would expect two or three deweighting events every day and a reset about every two weeks due to statistical fluctuations alone. Since these events conform to the statistical model, the reset procedure we have outlined, which treats them as special non-conforming glitches, will do the "wrong" thing when they occur.

Based on our experience with AT1 (which uses the same sort of reset procedure) and with the AF1 algorithm described in this work, it is quite likely that about 40% of the events that trigger the reset algorithm are actually observations that have large deviations but that conform to the statistics. We consider this balance as reasonable, although it is not possible to justify it rigorously on statistical grounds since it is designed to deal with events that are, by definition, not characteristic of the normal statistical performance of the clock. This procedure is particularly poor in modeling a clock that has large random-walk frequency noise at short periods. The procedure assumes that the noise at short periods is white, and the resulting unmodeled time dispersion will trigger the time-step detector more frequently than is appropriate for the actual performance of the clock.

A more fundamental concern is that the algorithm for determining the weights is biased. The prediction error (estimated by (9)) is always too small because the frequency of a clock is correlated with the frequency of the ensemble, since the ensemble includes a contribution from each device through (6). The weights, which are proportional to the reciprocal of the variance, are, therefore, systematically too large. This problem becomes very important as the weight of a clock increases—in the limit, a clock that starts out as one of the better clocks in a small ensemble can have its weight increased to 100% if the positive feedback that results from this correlation is not limited in some way.

One way of controlling this positive feedback is to use the correction to the variance proposed by Tavella, *et al.* [7] and Tavella and Thomas [8]. They show that the bias in the variance due to the correlation between the clock and the ensemble is proportional to the weight of the clock, and that an unbiased estimate of the variance, σ_u^2, may be calculated using:

$$\sigma_u^2 = \frac{\sigma^2}{(1-w)}, \tag{14}$$

where σ^2 and w are the variance and weight, respectively, of any clock in the ensemble, calculated according to the procedures we have outlined above.

There are a number of difficulties with this procedure. It complicates the error detection algorithm and may introduce a bias if a high-weight clock has a time step. While it attempts to provide an unbiased estimate of the variance of each clock, it does not address the situation in which one clock tries to take over the scale because it *really is* better than most of the others in the short term—a situation that is quite common when masers and cesium clocks are combined in a single ensemble.

Another procedure is to administratively limit the weight of any clock by some means. This is currently done in the computation of TAI by the BIPM and in the NIST time-scale algorithm AT1 [1], [9]. We have adopted the same procedure in the current work as is used in AT1: the normalized weight of any clock is administratively limited to 30% of the scale. The prediction error and standard deviation are calculated for each clock as specified in (9) and (10). If the normalized weight that results from these parameters exceeds 30%, then the corresponding weight is fixed at 30%. Note that this limit is relative—the value of the standard deviation that will cause it to be applied depends on the other clocks in the ensemble. A somewhat different procedure—with an absolute maximum weight—is used in the current computations of TAI by the BIPM, but the overall effect is much the same.

As we will show below, the correction factor proposed by Tavella *et al.* [7] has little effect in our ensemble. The weights of the two masers in our ensemble will be limited by the 30% administrative limit whether the biased or unbiased variances are used, and the correction to the variances of the other clocks that results from applying the correction factor does not make much difference to their weights.

Any procedure that introduces administrative weight limits guarantees that the scale cannot be completely captured by one clock, but this advantage does not come for free. The normalization condition on the weights (7), guarantees that, if the weight of a good clock is administratively limited to less than what its statistical performance would warrant, then the weights of poorer clocks become higher than they should be based on the same considerations. The performance of the scale (under normal operating conditions) is inevitably degraded compared to the theoretical performance that could be achieved if the clocks at hand were combined in an ensemble without such administrative limits.

The weighting procedure that we have defined is statistically optimum if the prediction errors of the clock frequencies with respect to the ensemble can be characterized as a white Gaussian process over the time interval Δt. This is quite likely to be true for the small time intervals we use, but it does not follow that the weights defined in this way result in a scale whose stability is optimum at longer averaging times. In particular, the procedure we have defined assigns a high weight to a clock with good short-term stability even if its long-term stability is poor. This is not as serious as it seems—the scale we have defined is optimized for short-term stability only, with the stability at longer term being provided by the procedure to adjust the aging parameters we outlined above. Nevertheless, it is possible to adjust the performance of the scale in longer term to some degree using the parameters G and G', the time constants used in (8) and (10), respectively.

Parameters G and G' serve the same purpose—they attenuate rapid fluctuations in the estimates of the frequency and the prediction error, respectively, on the assumption that such fluctuations are white noise about underlying mean values which we are trying to estimate [10] (rapid in this context means fast compared to the respective time-constant). The gain of both filters is unity at very long periods (i.e., at times much greater than the corresponding time-constant), and the exact magnitudes of G and G' become irrelevant in this regime. This is not true at intermediate periods, however, where both G and G' affect the partition between deterministic signals and stochastic noise. Increasing G', for example, decreases the filtering effect of (10), increases the effect of e_j on σ_j (making the latter larger) and therefore decreases the effective weight of a clock in the ensemble. The same sort of argument applies to (8)—increasing G puts more of the fluctuations in the short-term frequency, f, into the corresponding average value, y, thereby treating these fluctuations more as signal than as noise (because y is used to predict the performance of the clock between measurement cycles). Note that this effect begins to become important for fluctuations whose periods are on the order of the time constant—the gain of the filter will always be unity at very long periods.

These effects are most important in the analysis of transients and fluctuations whose periods are on the order of the time constants specified by G and G'. Adjusting these constants therefore provides a mechanism for differentiating between clocks with different performance at short and intermediate periods. This may be expedient, but it is usually not optimum, because changing the time constant changes both the performance at intermediate periods and the averaging at short periods, so that the two remain coupled.

V. Results

I have tested the algorithm described above using time difference data from our dual-mixer measurement system. The time differences are read by the hardware every 12 minutes, and a subset of the clocks that make up our official AT1 ensemble are also analyzed by the algorithm described in this paper, which we call AF1. (Note that the AT1 time-scale computation uses measurements made every 2 hours—that is every 10th point of the same data set used in this work.)

The clocks in our test ensemble are described in Table I. The weights in the last column are the formal values calculated using equations (12) and (7) and are rounded to the nearest 0.1%. The weights actually used in the computation are administratively limited to 30% as discussed in the text; the values for the two masers that are larger than 30% are replaced by this limit in the algorithm.

The time constants used in (8) and (10) were set to 60 h and 180 h, respectively, for all of the clocks; the results are not sensitive to changes of up to 30% in these values.

The algorithm was started on MJD 50070 using the data from the BIPM publications as described above and it has been free-running since that time. No administrative adjustments have been made to the parameters of any clock beyond the automatically imposed 30% limit on the weight of any clock that was discussed above.

The first clock in Table I was designated as the working frequency standard in the ensemble. It is also the clock that is used as the local oscillator for the evaluation of NIST-7, and the results of these evaluations are reported to the BIPM with respect to its frequency.

The interval between time-difference measurements is 720 s, and the algorithm computes an estimate of the average frequency of the working standard with respect to the ensemble on the same time mesh using the exponential filter defined in (8). Fig. 1 shows these estimates for a typical 20-day interval. The standard deviation of the data is 1.9e-15; the variance of the frequency fluctuations is a combination of flicker and random-walk spectra as expected—the white fluctuations have been attenuated by the exponential filter. For comparison, Fig. 2 shows the estimates of the frequency difference between the working standard and the second maser, clock 1400222. The significant frequency aging between the two masers has been removed by subtracting a straight line whose parameters are determined from the aging entries in Table I.

The data in Fig. 1 probably represent a pessimistic estimate of the frequency stability of the maser over relatively short time periods because the maser is being compared to an ensemble that is relatively noisy at short times. The values for σ in Table I confirm this. Even with this pes-

TABLE I
THE PARAMETERS OF THE CLOCKS USED TO EVALUATE AF1.

Clock Number and type (BIPM Identifier)	Frequency	Initial Values Aging	σ	Typical Values Weight	
		(s/s)	(s/s^2)	(ps)	(%)
1400201	Maser	−9.17e-14	−1.15e-21	40	186
1340493	Cesium	1.29e-12	−1.68e-19	2000	0.1
1350408	Cesium	1.03e-13	−1.15e-21	170	11
1310569	Cesium	1.50e-12	−2.15e-21	200	6
1350132*	Cesium	1.00e-13	−7.55e-21	170	11
1350182	Cesium	6.98e-14	2.03e-21	170	11
1400222	Maser	8.54e-12	1.90e-21	40	186
1160217	Cesium	−4.87e-13	−1.15e-21	900	0.5
1181007	Cesium	2.60e-12	−2.13e-20	800	0.5

* Clock 1350132 was included in this ensemble for only a portion of the period discussed in the text. Because of the normalization condition, the weights of the other clocks increased when it was not present.

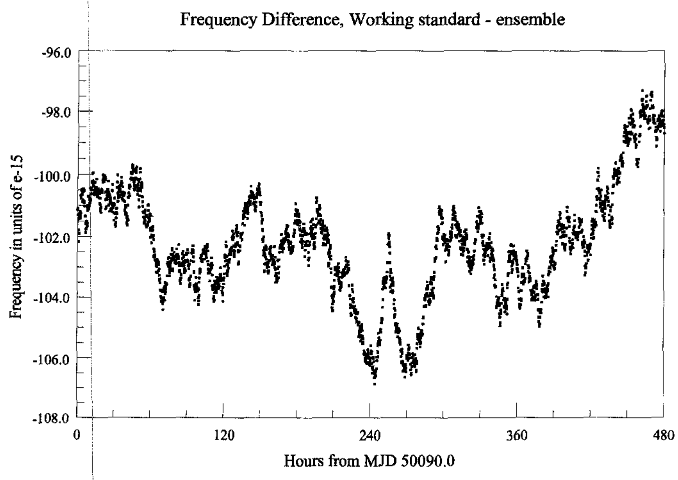


Fig. 1. Average frequency of the working standard with respect to the ensemble from (8).

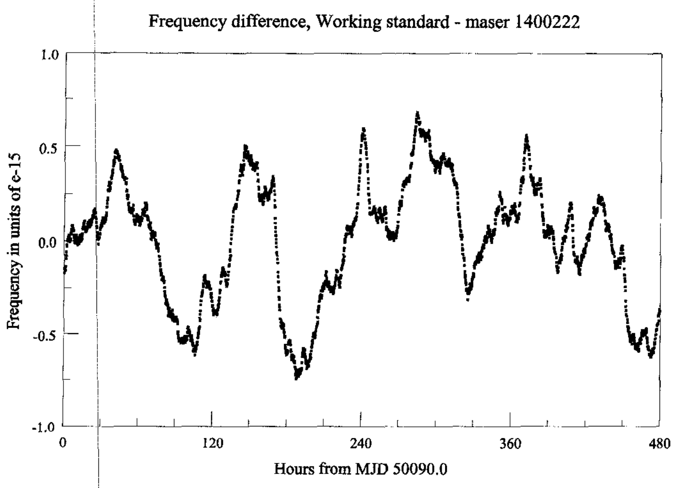


Fig. 2. Average frequency difference between the two masers for the same time period as shown in Fig. 1.

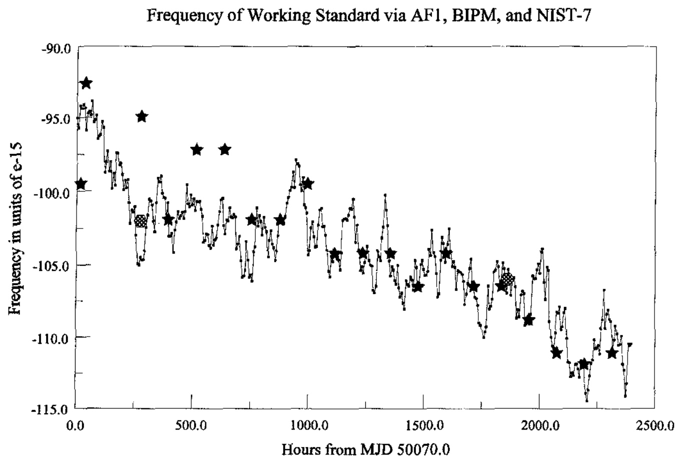


Fig. 3. Average frequency of working standard. Connected points are AF1 estimate; stars are BIPM estimates, and hexagons are NIST-7 evaluations.

simistic view, both the ensemble and the maser have frequency stabilities *substantially better* than 1e-14.

It would be tempting to assert that the actual short-term frequency stability of the maser is $1/\sqrt{2}$ of the variance of the data in Fig. 2 (or about 0.2e-15), but this is probably too optimistic. In the first place, the data of Fig. 2 have a significant frequency-aging term removed (about 0.3e-15/day), and there is an inevitable uncertainty in this process. In the second place, both masers are sensitive to environmental perturbations, so that the data in Fig. 2 may be contaminated with some common-mode frequency noise which would make the plot look too good. Note that even these data show periods when the frequency difference between the two masers changes by almost 1e-15 in a period of a day or so. If these fluctuations are environmentally induced, the data-set may represent a lower bound on what either device is actually doing.

Fig. 3 shows the frequency of the working standard with respect to the ensemble for 100 days, which is all of the data we acquired to date. We have decimated our data to 4 points per day without averaging the skipped data. The lines connecting the points are for ease of identification and are not otherwise significant.

The filled stars show the frequency of the same clock as estimated from the BIPM data. Each value in this series is the first difference of the time of the working standard with respect to UTC divided by the interval between the measurements (10 days before MJD 50079 and 5 days thereafter). Since these data are obtained via GPS, the uncertainty in each time measurement is probably about 2 ns. The uncertainty of a frequency estimated using these time data will be about $(2\text{e-}9 \times \sqrt{2})/(5 \times 86400) = 6.5\text{e-}15$. As can be seen from Fig. 3, the agreement is generally much better than this value—in fact, most of the estimates agree far better than might be expected based on this noise estimate.

Finally, there have been two evaluations of NIST-7 during the period covered by Fig. 3. These evaluations estimate the frequency of the maser used as the working standard in our ensemble. Since these evaluations include a correction for the black-body shift whereas TAI did not during the period in question, we have subtracted 20e-15 from the NIST-7 data to remove this correction. The results are plotted as filled hexagons. The quoted uncertainty of these evaluations is 1e-14, but the agreement among the various estimators is much better than this.

VI. Discussion and Conclusions

I have designed an algorithm for estimating the frequency of a working standard using time-difference data from other devices. The algorithm, which I call AF1, is similar to the AT1 time-scale algorithm that has been used at NIST for many years. I have tested the algorithm using an ensemble of 2 hydrogen masers and 6 cesium standards, all of which are commercially produced. I find very good agreement between the frequency of the working standard estimated using AF1 and the frequency estimated using published BIPM data and using evaluations of NIST-7. In both cases, the agreement is at least as good as what would be expected based on estimates of the uncertainties in the data, and is generally considerably better than this. The technique shows great promise for providing a new tool for monitoring the local oscillator used in the evaluation process of a primary frequency standard and in providing a robust statistical procedure for "remembering" the standard frequency between evaluations. The scale will probably exhibit random-walk frequency fluctuations at long periods, and I have suggested a mechanism for addressing this problem using the primary standard data to modify the aging parameters of the member clocks.

The data also suggest that the current oscillator hardware may not be adequate to support a significant improvement in primary frequency standards. Our results suggest that the best hydrogen maser that we have sometimes shows fluctuations in its frequency on the order of 1e-15 over periods of a few days. This is probably adequate to evaluate a primary standard whose accuracy is only on the order of 1e-14, but it is unlikely to be sufficiently stable to support the evaluation of a primary standard that was a factor of 10 better than this. Evaluation protocols would seem to depend on having a local oscillator whose short-term stability was at least as good—and ideally substantially better—than the overall accuracy of the evaluation. It is possible that this requirement might be satisfied by synthesizing a local oscillator derived from optimally averaging an ensemble of masers, but it is not at all clear how well this will work. We will need to exercise great care to minimize common-mode effects—especially environmental perturbations—if this strategy is to be viable.

References

[1] D. W. Allan, J. E. Gray, and H. Machlan, "The National Bureau of Standards Atomic Time Scale: Generation, stability, accuracy and accessibility," in NBS Monograph 140, *Time and Frequency: Theory and Fundamentals*. Boulder, Colorado: National Bureau of Standards, 1972, chapter 9, pp. 205–230.

[2] T. J. Quinn, "The BIPM and the Accurate Measurement of Time," *Proc. IEEE*, vol. 79, pp. 894–905, July, 1991.

[3] B. Guinot and C. Thomas, "The establishment of International Atomic Time," *BIPM Annu. Rep. Time Section*, vol. 1, part D, BIPM, Sèvres, France, 1989.

[4] S. R. Stein, "Time scale computation system including complete and weighted ensemble definition," U.S. Patent 5,155,695, 1992.

[5] ——, "Time scale computation system," U.S. Patent 5,315,566, 1994.

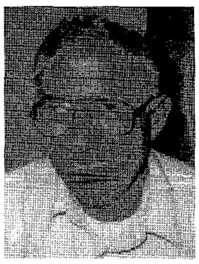
[6] D. B. Percival, "Use of robust statistical techniques in time scale formation," *Proc. 2nd Symp. Atomic Time Scale Algorithms*, NBS, Boulder, CO, 1982, pp. 152–162.

[7] P. Tavella, J. Azoubib, and C. Thomas, "Study of the clock-ensemble correlation in ALGOS using real data," *Proc. 5th European Freq. Time Forum*, 1991, Neuchatel, Switzerland, pp. 435–441.

[8] P. Tavella and C. Thomas, "Comparative study of time-scale algorithms," *Metrologia*, vol. 28, pp. 57–63, 1991.

[9] B. Guinot and C. Thomas, "Establishment of the International Atomic Time, *Annu. Rep. Time Section of the BIPM*, pp. D3–D22, BIPM, Sèvres, France, 1988.

[10] B. Gold, C. M. Rader, A. V. Oppenheim, and T. G. Stockham, Jr., *Digital Processing of Signals*, New York: McGraw-Hill, 1969, pp. 24–25.



Judah Levine was born in New York City in 1940. He received a Ph.D. degree in Physics from New York University in 1966. He is currently a physicist in the Time and Frequency Division of the National Institute of Standards and Technology (NIST) in Boulder, Colorado, and is also a Fellow of the Joint Institute for Laboratory Astrophysics, which is operated jointly by NIST and the University of Colorado at Boulder.

He has worked on time scale algorithms at NIST for about 25 years. His most recent work involves developing the new time-scale algorithm described in this paper. He is also working on methods for distributing time and frequency information using digital networks such as the Internet, and on ways of improving satellite-based time and frequency distribution.

Dr. Levine is a Fellow of the American Physical Society and is a member of the American Geophysical Union and the IEEE Computer Society.