SIFTING THROUGH NINE YEARS OF NIST CLOCK DATA WITH TA2*

Marc A. Weiss
Time and Frequency Division
NIST

Thomas P. Weissert
LiteroPhysics

Abstract

We have extended the new TA2 post-process time scale at NIST beyond our previous reports to include all of the period from January of 1984 to the present. Derived from the ensemble of clocks at NIST, this time scale includes the benefits of several recent refinements to the algorithm. By iteratively running the algorithm on the ensemble clock data and characterizing anomalous behavior in the dominant individual clocks of the ensemble between iterations, we obtain an optimized scale which benefits from the informed anticipation of that anomalous behavior and demonstrates an overall decrease in scale disruption. Herein we discuss: changes to the TA2 algorithm that we made while processing the eight-year run, our method of characterizing anomalous behavior in individual clocks, the way unanticipated anomalous behavior is dealt with by the algorithm, and our resulting nine-year time scale.

keywords: time scale, clock ensemble, post-process, stochastic noise, deterministic errors

1. INTRODUCTION

We report here our development of a nine-year time scale derived from the measurements among the best atomic clocks at NIST over the years from 1984 through 1992, using the TA2 algorithm [1,2,3]. In this paper we refer to both our algorithm and the time scale resulting from the application of that algorithm to the NIST clocks as "TA2"; whenever contextual clarity demands it, we differentiate between the TA2 algorithm and the NIST TA2 time scale. The algorithm takes data from an ensemble of clocks and computes a time scale in the form of time and frequency offsets of each clock from ensemble time and frequency.

Because the primary purpose of a post-processed time scale at NIST is to carry forward the evaluation of a primary frequency reference, TA2 was designed to optimize frequency stability in both short and long term. Whereas AT1, the real-time scale at NIST [4,5,6], shares these goals with the post-processed TA2 scale, it must be administered differently due to its greater vulnerability. In real time, the most recent ensemble data must be synthesized into the time scale immediately, leaving little or no opportunity for the anticipation of anomalous behavior. Freed from the constraints of maintaining a real-time scale, we are able to run the TA2 algorithm over a fixed period of data, characterize most of the existent anomalous behavior, and correct for it in subsequent iterations of the algorithm.

TA2 is the name for the process of running the AT2 algorithm both forwards and backwards in time to determine frequency estimates of clocks in each direction, then running a final third pass forward in time to generate ensemble time using the frequency estimates from both the forward and backward passes. The AT2 algorithm is very similar to the AT1 algorithm, but it uses a simple Kalman filter technique to estimate a confidence on the frequency estimate of clocks. This confidence facilitates both automatic frequency step detection and combining the forward and backward passes to obtain a smoothed ensemble frequency.

There is no TA1 algorithm at NIST, rather an algorithm called TA. This algorithm is a pure Kalman Filter [7,8,9,10], derived from the Kalman filter principle of minimizing error in a least squares sense. Since time is not measured, only time difference between clocks, a time-scale algorithm cannot estimate time as a measured quantity in the presence of some measurement noise, as a Kalman filter is usually used. In an ensemble of n clocks, we have n-1 measurements of time difference against a single clock. The algorithm must produce n estimates of clock offsets from ensemble time. This system is underdetermined unless some principle is used to generate another equation to resolve the unknown. In TA the principle used is the Kalman filter principle of minimizing error, in this case time error. TA(NIST) in its attempt to minimize time error eventually puts too much weight on the clock in the ensemble with the best long term stability. This damages the short-term stability of the scale and makes the scale vulnerable to shifts in that clock. Anoher consequence of the design of TA is that the covariance matrix grows without bound. Again, this comes from the lack of measurement of time, the parameter that is being estimated. Operationally this is not a significant problem, since the growth of covariance is slow enough that it can be contained in a double precision floating point variable of eight bytes indefinitely in a practical sense. Yet it is theoretically disturbing. Not, surprisingly, the performance of TA(NIST) is problematic.[1,4]

In addition to our primary goal of optimum stability, we designed the TA2 time scale at NIST to facilitate: the addition and/or removal of clocks from the scale with minimal perturbations to the scale; the detection of steps in either the time or frequency offsets of an individual clock from the scale; and the general ease of use in running the scale. While much of the design of TA2 has been reported before, some corrections and improvements have been made to the algorithm in the process of running the scale over nine years of data.

Using the behavior of the real-time scale AT1 as a guide, we examined all major anomalous features occurring in the ensemble data during the nine years of measurements among the best clocks at NIST. When we encountered frequency steps or drift [11,12,13], we characterized this behavior and inserted anticipatory information into the run-time command file for the scale, and so prevented these effects from pulling the frequency of the scale. We made these characterizations carefully with the conscious intent of maintaining the free-running independence of TA2; thus we say that TA2 has not been steered to any clock in the ensemble or to an external reference. We discuss our method of inserting clock characterizations into the time scale in some detail.

We analyze the performance of the scale by making comparisons both with the AT1 time scale and with time scales and clocks external to NIST. We show that TA2 is comparable in performance to other NIST scales and so provides redundancy and additional diagnostic abilities for characterizing clocks.

2. ANOMALOUS BEHAVIOR CORRECTION

TA2 has the ability to detect frequency steps automatically and correct for them by removing the offending clock from the ensemble until the scale can learn the new frequency. When running the scale in the forward direction a clock is removed chronologically after the frequency step, while in the backward direction, the clock is removed chronologically before the step. Thus the automatic frequency step algorithm keeps the clock in the ensemble at full weight in only one direction in the neighborhood of a frequency step. Whereas this limits the pulling of the ensemble frequency by the clock that has taken a frequency step, it reduces the stability of the ensemble, since the clock is no longer contributing. When we ran TA2 over the nine years, if a high-weight clock took a frequency step we preferred to characterize that step to allow the clock to maintain full weight in the ensemble in both forward and backward directions.

As with other algorithms [7,8,14,15], TA2 incorporates a definition of the ensemble time offset from the common clock using a weighted average of the measured offsets of all the clocks from the common clock. TA2 computes this weighted average beginning with a prediction of the time offset of each clock from ensemble time. Because this prediction represents our expectation for each clock's nominal time offset from the scale, it serves as a natural point of entry for our foreknowledge of anomalous behavior. Just as anomalous behavior breaks from our expectations for a clock, we direct the algorithm to break from its prediction, which is otherwise based solely on a clock's history.

We specifically consider two types of anomalous behavior: frequency steps and frequency drift. We characterize these effects by studying the differential performance of clocks in the ensemble. If a clock has low weight and therefore little correlation with the scale, then we can run the algorithm and characterize that clock's behavior against the ensemble itself. But for higher-weight clocks, the value of frequency steps or drift relative to the ensemble are biased by that clock's correlation with the ensemble, and so better characterizations are obtained by comparing the clock under question with other clocks in the ensemble. When we see a frequency step between two clocks, we need to determine which clock took the step. To do this, we can compare each with a third clock and use majority voting. It also helps to look at more than three, and to consider the possibility of an environmental effect that could perturb several clocks simultaneously at the same site.

We emphasize the significant difference between adjusting a clock's prediction to reflect pre-characterized behavior, and steering the ensemble frequency to an external reference. If a clock in the ensemble takes a frequency step, the ensemble will be pulled off in frequency by that clock because previous frequency estimates contribute to current updates. Thus, it is the prediction of frequency based on previous history which can pull the ensemble frequency if the clock frequency has shifted unpredictably. Since we run the scale after this has occurred, we may characterize this clock's frequency step by comparing the frequency of this clock to other clocks in the ensemble. If we see a consistent value for the frequency step which is considerably larger than the stochastic fluctuations of frequency for this clock, we use that value as the characterization of the clock's frequency step. We insert this value as a shift in the prediction of frequency for this clock against the ensemble at the reference time when the clock stepped. The ensemble frequency is not steered by this effort to any clock. We must assume only that one or more clocks in the ensemble are stable enough in frequency that our characterization of a frequency step is accurate.

The process is similar for frequency drift, but in practice the characterization requires more care. We must see a consistent drift in one clock as compared to several clocks. If we accept this value as the frequency drift of the one clock, we are assuming that the other clocks are neither drifting deterministically nor walking off randomly together. As we shall see, it is possible for clocks to move together. Once we have characterized a frequency drift, we adjust the prediction of a clock to account for it much as with a frequency step, but now the frequency prediction is adjusted continuously over many measurement cycles. Others have studied using maximum likelihood estimation of drift [10]. This requires the assumption that all clocks have a constant drift during the period used for the maximum likelihood estimation.

3. RESULTS

Figure 1 shows the fractional frequency offset of AT1-TA2. Overall the two scales nominally remain within $1 \cdot 10^{-13}$ of each other. There seem to be significant periods of frequency drift between the two scales reaching a magnitude of perhaps 10^{-16}/d for extended periods. Most of this drift appeared when we introduced a robust technique for time steps [16]. This new method tapers clock weights quadratically as prediction error increases from three to four standard deviations, instead of completely removing a clock whose prediction error exceeds three standard deviations. Since AT1 completely removes a clock whose prediction error exceeds three standard deviations, it is not surprising that changing this technique increased differences between the two scales. The magnitude of the change reflects the nonlinear way in which the algorithms respond to small differences.

We emphasize that we cannot know which method produces less drift in an absolute sense. This drift of 10^{-16}/d is much smaller than effects we see in comparing the two scales against

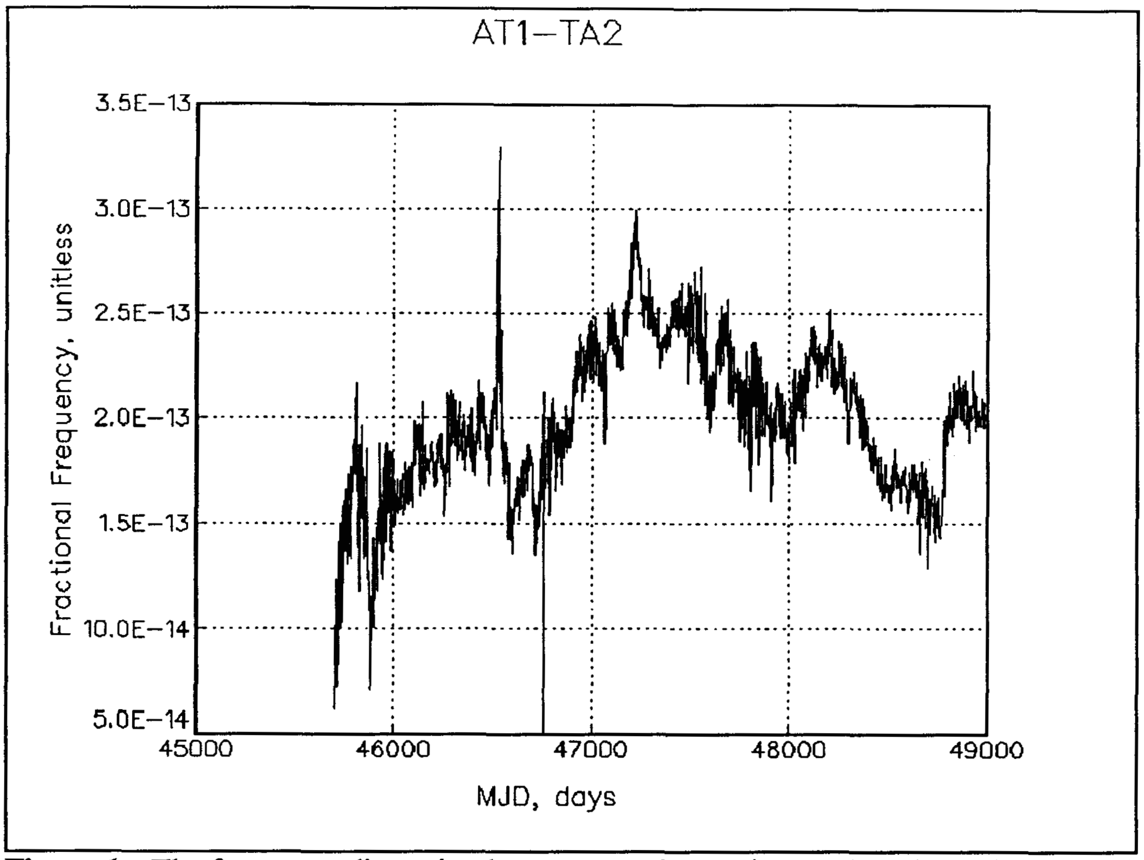

Figure 1: The frequency dispersion between two NIST time scales, the real-time AT1 scale minus the post-processed TA2 scale.

the best available external references. We add the technique to TA2 since it should theoretically improve the scale.

There are occasionally more rapid excursions which existed even without the robust time step technique. These are probably due to using somewhat different clock ensembles and the differences in the algorithms. In TA2 when a clock takes a frequency step it has a smaller effect than in AT1, since TA2 is run in post-processing and AT1 is a real-time scale. Still, the two scales use very similar algorithms on the same major clocks. Again, these differences between the two reveal nonlinearity of the time scales. Small differences and changes can produce large effects. Other significant features include a time step between TA2 and AT1 at MJD 46760, when the NIST (then NBS) measurement system failed temporarily, and a steering in AT1 of $3.44 \cdot 10^{-14}$ applied over a four-day interval centered at MJD 48780, 00:00 hours UT.

Figure 2 shows the fractional frequency offset of TA2 from International Atomic Time, TAI [15]. We see annual deviations over the first five years which seem to attenuate after MJD 47527, the beginning of 1989. The period approximately corresponding to mid 1990 to mid 1992, from MJD 48140 to 48730, however, shows a steady drift of about $-3 \cdot 10^{-16}$/d.

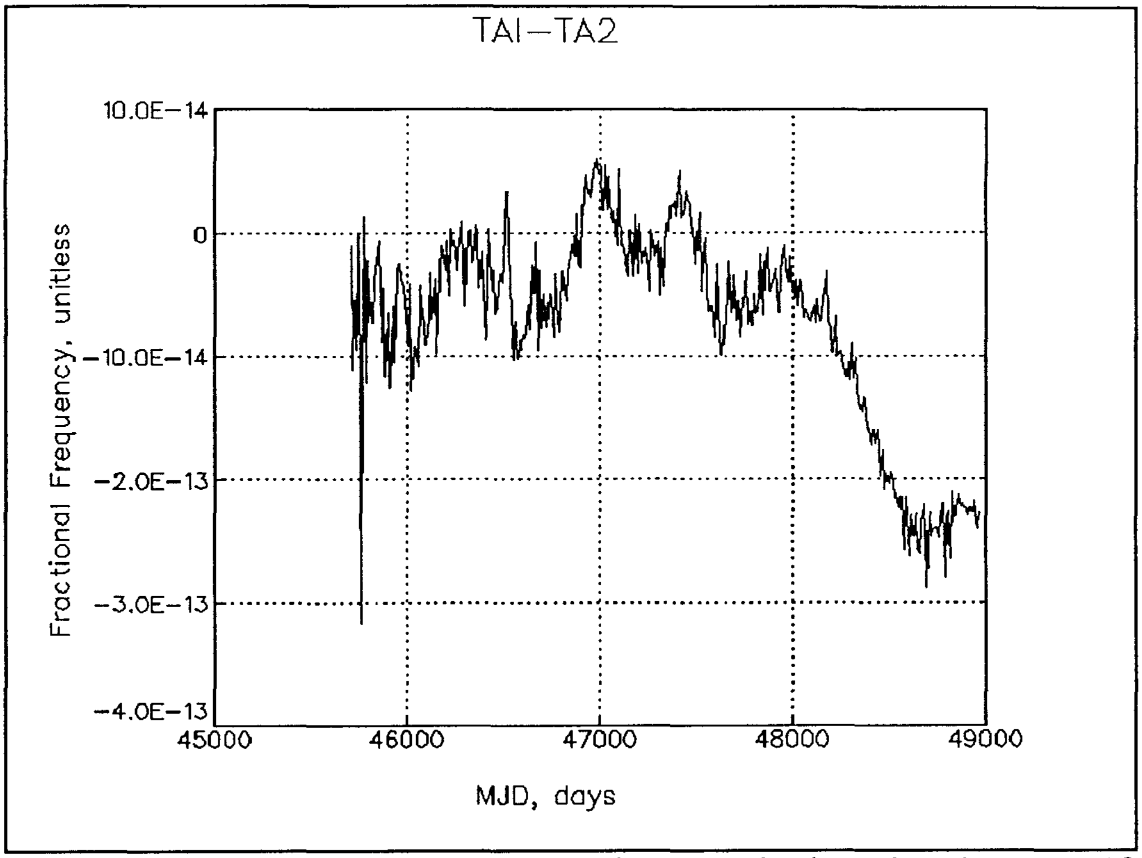

Figure 2: The frequency dispersion of International Atomic Time minus the NIST TA2 time scale.

Figure 3 shows TA2 frequency versus the primary frequency standards of the Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany, PTB CS1 and PTB CS2 [17], and the unsteered master clock of the U.S. Naval Observatory (USNO) [18]. Although each of these clocks is correlated with TAI, there is no reason to expect correlation among them. We might suspect some correlation between the two PTB clocks, because they are in the same laboratory. But both PTB CS1 and the unsteered USNO Master Clock show a similar frequency drift against TA2 over the interval from MJD 48140 to 48730.

We show the five dominant clocks contributing to TA2 during this period in Figures 4 and 5 along with TAI-TA2. For these figures initial frequencies have been set to 0, and frequency steps which have been characterized and inserted into the scale have been removed from the plots. Thus, these are the frequency changes which the algorithm sees, except for modelled frequency drift. Figure 4 shows NIST clocks 1, 7, and 11, which seem to have some similarity in overall frequency drift. Clock 1 seems to have no consistent drift against TA2 over the interval MJD 48140 to 48730, though its significant random walk FM is visible even though it is off line for 176 d. Clock 7 does not seem to drift against TA2 over most of the interval, though it does seem to drop in frequency from about 48550 to the end of the TAI-TA2 drift interval at about 48730. Clock 11 seems to drop slowly, though it is off line

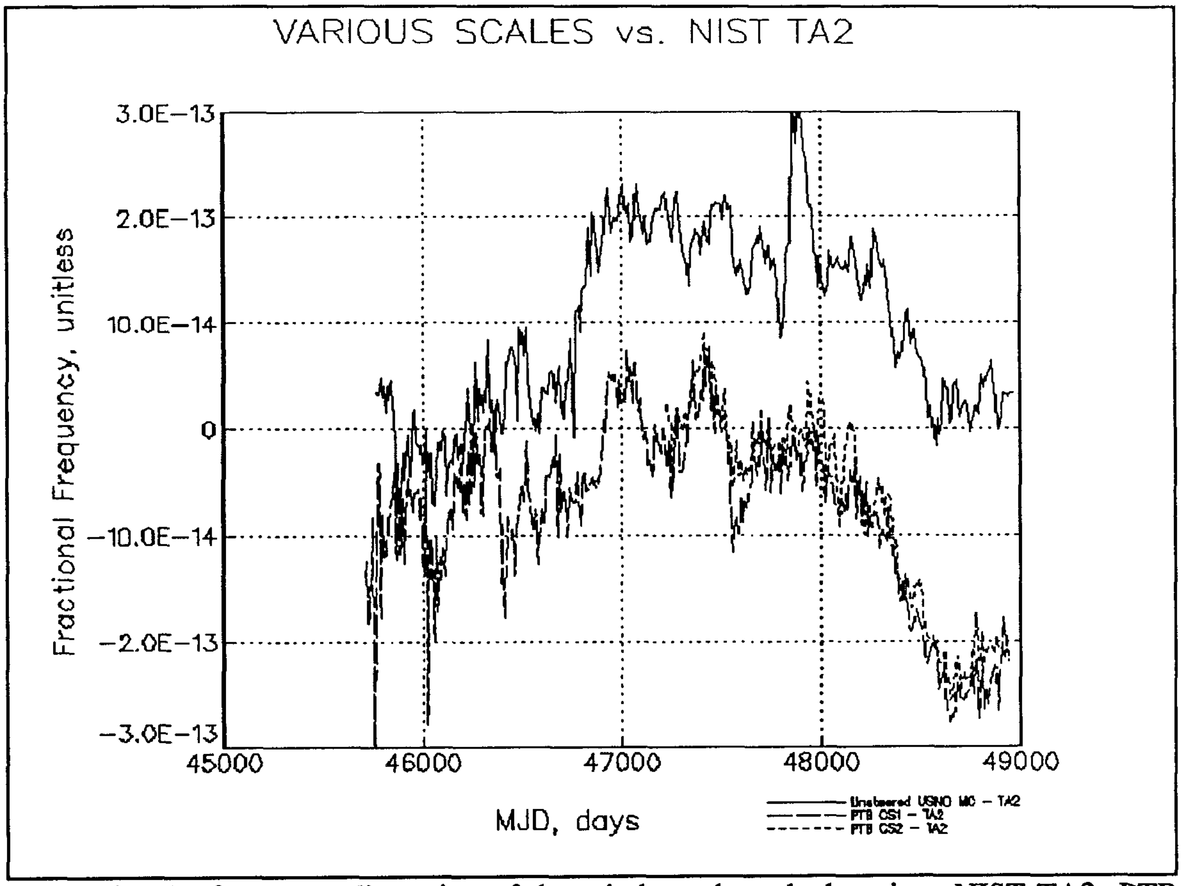

Figure 3: the frequency dispersion of three independent clocks minus NIST TA2, PTB CS1, PTB CS2, and unsteered USNO Master Clock.

for 148d. Figure 5 shows clocks 12 and 24 which seem to be drifting against TA2 opposite to TAI. In these cases we attempted to model drift based on internal comparisons between clocks. We concluded by modelling clock 12 (Figure 5) with a drift of $5 \cdot 10^{-16}$/d from MJD 48340 to 48622, and by modelling clock 7 with a drift of $2.3 \cdot 10^{-16}$/d from MJD 48080 to 48400. It is possible we might have benefited from the maximum likelihood estimation of Jones and Tryon [10], though we would have had to assume a constant drift on all clocks over some interval. We conclude that the drift between TAI and TA2 from MJD 48140 to 48730 is a mean frequency drift among the clocks that dominate TA2 over this period.

4. CHANGES IN TA2

There have been a number of changes in the TA2 algorithm since previous publications [1,2,3]. TA2 is the name given to the three-pass smoothing process of running the AT2 algorithm forwards and then backwards in time, and then combining the clock frequencies generated in those passes into a final third pass forward to generate clock time offsets from

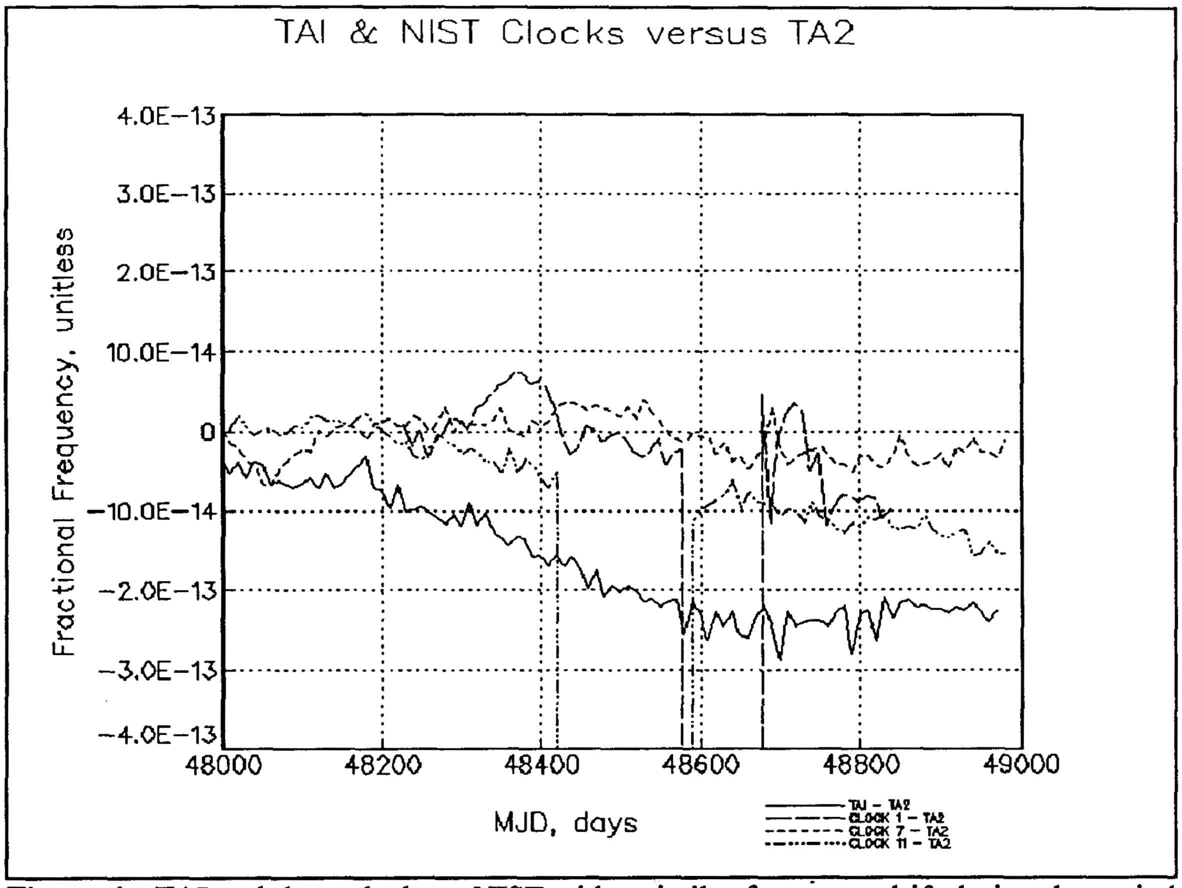

Figure 4: TAI and three clocks at NIST with a similar frequency drift during the period of significant drift for TAI-TA2.

the ensemble time. In the next section of this paper, we present the complete equations of AT2. Here we discuss the motivations for some of the changes.

Several equation changes follow from our recognition of a need to differentiate between two different values of τ, which we call τ_x and τ_y, for each clock in the ensemble. Because data for an individual clock may be missing for various reasons, τ_x is necessary to track the reference time since the last time update. Similarly, we need τ_y to track the reference time since the last frequency update, because we cannot estimate frequency on a cycle in which there is a time step. Hence we may have a time update without a frequency update.

The method we use to estimate the bias of a clock's prediction error due to the clock's correlation with the ensemble has changed to incorporate the recent work of Tavella, Azoubib, and Thomas [19]. Through simulation we have verified that the current formula indeed gives a better value for the bias than the previous one.

We have moved the drift term from the time prediction into the frequency prediction. This better allows us to separate the computations of clocks' time offset, such as time step

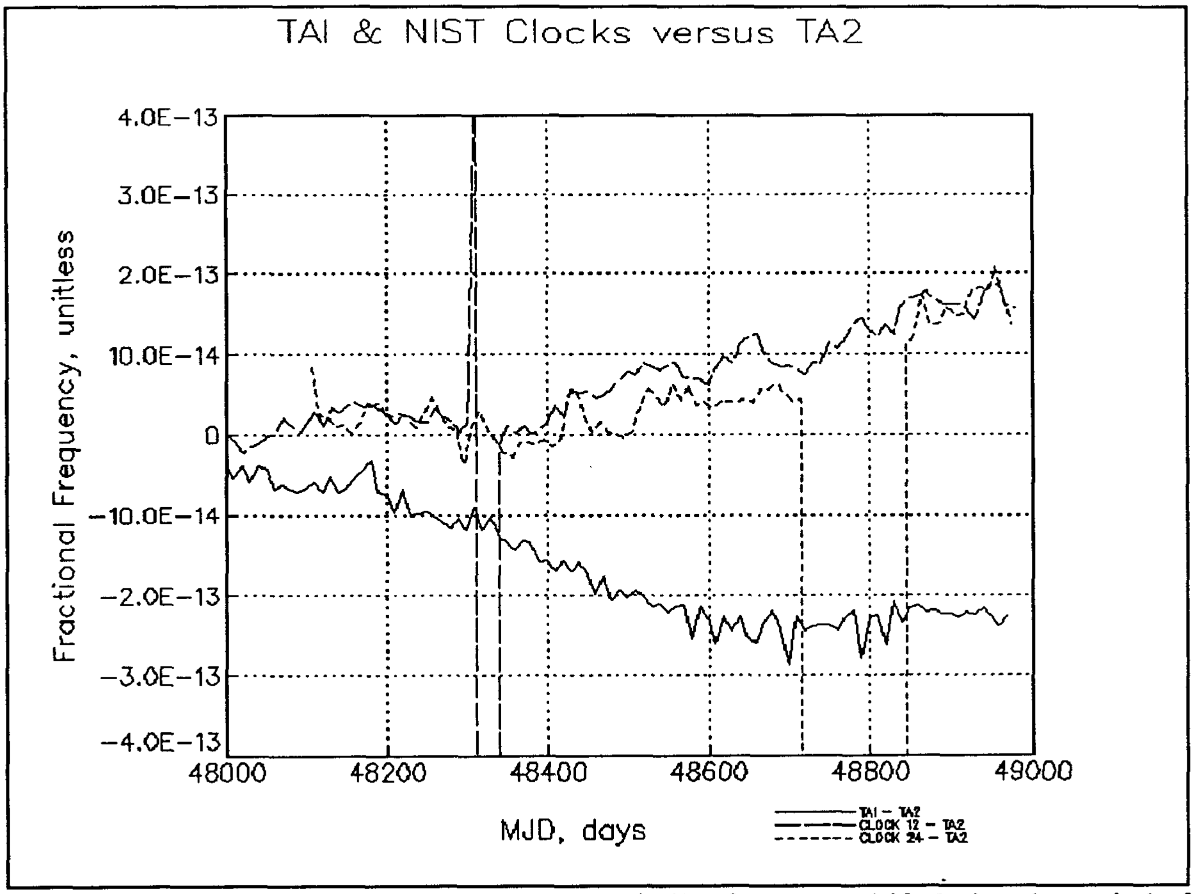

Figure 5: TAI and two clocks at NIST with a similar frequency drift during the period of significant drift for TAI-TA2.

detection and correction, from those with frequency.

We have added the wct (weight control) parameter to enable limiting of a clock's contribution to the ensemble, while the algorithm continues to estimate the clock's prediction error. This is useful in two important situations. First, upon the detection of a time step, whereas previously we abruptly removed the clock from the scale for that measurement cycle, now we gradually deweight a clock as the step size increases from three to four standard deviations from its predicted value. This move from a single-cycle discontinuous gap in a clock's presence in the ensemble, to a ramped-deweighting procedure increases the robust character of the algorithm, in that a small change in a measurement will not produce as large a change in the scale [16]. Second, this parameter allows us to enter a clock in the scale with zero weight while the scale still computes that clock's prediction error variance.

Besides these changes we have corrected some errors in the formulas as published previously, notably in the prediction error update ϵ^2 and in the statistic used for defining frequency steps σ_L. We have also introduced the parameters R and Q to simplify the equations for the Kalman estimate of frequency y, and to clarify the formulas used for converting stochastic parameters among various representations.

5. THE EQUATIONS DEFINING AT2

The AT2 algorithm estimates the deterministic parameters of the clock model, time and frequency offsets of each clock from ensemble time (and frequency). An estimate of frequency drift for each clock can be entered and used for frequency prediction, but it is not updated in the calculations. A time offset is a weighted average, the weight of each clock being determined adaptively by its prediction confidence, which is the normalized reciprocal of the prediction error. There is an upper limit to weight, normally 0.3 of the entire scale. The algorithm allows for times when there are less than four clocks by limiting weight to 0.433 with three clocks and 0.633 with two. Such times should occur only briefly when clocks are temporarily removed due to anomalous behavior. An ensemble with less than three clocks fails mathematically to produce a good scale. AT2 is based on the AT1 time scale algorithm with the addition of a Kalman filtering technique for estimating frequency.

In addition to the deterministic parameters, there are two stochastic parameters in the clock model, white frequency modulation (FM) σ_α and random walk FM σ_β [9,10]. These two parameters are consistent with atomic clock data, though there is evidence that flicker FM commonly exists in cesium clocks also [20-28]. AT2 reduces to AT1 in steady-state if the weights of clocks are chosen properly; hence AT2 inherits the ability in AT1 to model flicker FM also.

Like AT1, AT2 estimates time offsets using predictions from the previous measurement cycle and the current measurements. These predictions allow us to detect time steps in a clock by comparing the difference between its predicted and updated time, the time innovation, with that clock's prediction error. The Kalman technique provides a variance of frequency prediction, which also allows us to detect frequency steps.

In our formalism, we assume there is a constant reference time interval τ_0 between measurements. At each measurement cycle, all clocks are measured against a common clock with negligible measurement noise. Thus, we obtain a measurement between any pair of clocks by differencing the measurements of these two clocks against the common clock. If there are no data for a particular clock, the time interval between measurements can grow to some multiple of τ_0. We denote this interval as $\tau_{x,i}$, because we use x for a clock's time offset, and this interval is the duration of reference time between time updates for clock i. When the algorithm detects a time step, it does not update the frequency for the stepped clock, even though time is still updated. Hence the reference time interval between frequency updates can be different from $\tau_{x,i}$. We denote this second reference time interval as $\tau_{y,i}$, because y denotes the frequency offset of a clock from the ensemble.

5.1 Time prediction

The prediction of the time offset $\hat{x}_i$ of clock i from ensemble time for the current reference time, t, is calculated from the previously updated time offset x_i at reference time $t-\tau_{x,i}$ and the filtered frequency y_i,

$$\hat{x}_i(t) = x_i(t-\tau_{x,i}) + y_i(t-\tau_{x,i})\cdot\tau_{x,i} \tag{1}$$

5.2 Time update

Given measurements of time differences $x_{ji}(t)$ for clock j minus clock i, the time offset of clock i is updated in a weighted average, using the prediction, weight, and time-difference measurement of every clock j, that is,

$$x_i(t) = \sum_{j=1}^{n} w_j \,[\, \hat{x}_j(t) - x_{ji}(t) \,]. \tag{2}$$

The weights w_j, determined from the prediction confidence, are normalized to sum to 1. We assume in this equation that the measurement noise is negligible, so

$$x_{ij} - x_{kj} = x_{ik} \tag{3}$$

for all combinations of i, j, and k. If the validity of this assumption fails, then a pre-filter must be used to obtain estimates of x_{ij} that satisfy (3). We may view equation (2) as the definition of ensemble time, offset from one clock. The ensemble time offset from other clocks is simply

$$x_j = x_i + x_{ji}. \tag{4}$$

5.3 Adaptive Clock Weights

The clock weights used to update the time offsets are calculated from the variances of the time residuals ϵ_i^2 by

$$w_i = \frac{\epsilon_x^2}{\epsilon_i^2} wct_i, \tag{5}$$

where ϵ_x^2 is the ensemble prediction error variance, ϵ_i^2 is the prediction variance since the last time update of clock i, and wct_i is the weight control parameter for this clock. The weights adaptively match the relative stabilities of the clocks in the ensemble because ϵ_i^2 and ϵ_x^2 are estimated from the data. As mentioned above, the weight control parameter affords the algorithm a way of dealing with time steps robustly, by deweighting a clock smoothly as a function of the size of the step [16]. This is discussed further in Section 6. Setting wct to 0 allows a clock to be estimated against the scale without it actually contributing to the scale, which thus keeps that clock uncorrelated.

5.4 Ensemble Prediction Error

The ensemble prediction error is a numerical estimate of the stability of the ensemble over an integration time τ_x. It is calculated as

$$\epsilon_x^2 = \left(\sum_{i=1}^{n} \frac{wct_i}{\epsilon_i^2} \right)^{-1} . \tag{6}$$

The dependence on t is suppressed in this equation. We see here that a poorly performing clock will not destabilize the ensemble; clocks can only improve ensemble stability.

5.5 Prediction Error Estimate

The prediction error of clock i over the interval $\tau_{x,i}$ is estimated from the difference between the time prediction and subsequent update, often called the innovation. This difference is adjusted for the average bias of this value due to the clock's correlation with the ensemble. The equation is

$$\hat{\epsilon}_i^2(t) = \left(x_i(t) - \hat{x}_i(t)\right)^2 \cdot K_i \tag{7}$$

where K_i is the expected proportion of the bias of clock i. Thus $\hat{\epsilon}_i^2(t)$ is a sample estimate of the prediction error variance of the clock, based on the current measurement cycle.

5.6 Bias of the Error Estimate

Because ensemble time is a weighted average of each of the individual clock times, the prediction error of a clock measured against the scale will be biased low, on the average. To correct for this biasing, we include the estimate of the multiplicative bias for each clock [19]

$$K_i = \frac{1}{\left(1 - w_i\right)} \tag{8}$$

Thus the factor K_i is the proportion of the unbiased-to-biased average prediction error, relative to the scale.

5.7 Filtered Prediction Error

We assume a model for the prediction error of white PM and random walk PM. That is, the prediction error estimates vary around some true mean prediction error at the two-hour measurements due to the white FM level of the clocks, but this white FM level varies slowly at about 20 d. Therefore, we exponentially filter the squared prediction error of each clock from the current measurement cycle to obtain a more statistically reliable estimate of an individual clock's variance. The filter is

$$\epsilon_i^2(t) = \frac{\hat{\epsilon}_i^2(t) + N_{\tau_{x,i}} \epsilon_i^2(t-\tau_{x,i})}{1+N_{\tau_{x,i}}}. \tag{9}$$

Because we assume the noise characteristics of a clock are not stationary, past updates are deweighted in the filtering process. The time constant for the filter is typically chosen to be $N_\tau = 20$ d for cesium clocks, representing the time we expect the white FM level to remain constant. The initial value of ϵ_i^2 is estimated as $\tau_0^2 \cdot \sigma_y^2(\tau_0)$.

We identify the estimate of prediction error $\epsilon(t)$ with the white FM level σ_α (in units of time)--the expected amount of time dispersion after an interval τ_0. This is valid if the measurement interval τ is well within the range of integration times for which the white FM is dominant, as is the case with the system at NIST, which makes measurements every 2 h.

5.8 Frequency Estimate

Derived from a simple Kalman filter formalism [29,30], the method we use to calculate the frequency offset of each clock from the ensemble frequency involves an estimate of the nonstationary frequency changes y in the presence of white-noise frequency modulation. We estimate the flicker-plus-random-walk FM components of frequency based on the first difference of time. In this formalism the first difference of time offset is a measurement of y, the white FM is the measurement noise R, and the expected variance in frequency is the process noise Q. The Kalman formalism results in an exponential filter of the first difference of time, much as in the AT1 algorithm. In addition, the Kalman formalism provides a variance P of our estimate of the frequency, which we use to define a confidence on the current frequency. This method allows us to estimate frequency steps in AT2, as well as to combine forward and backward passes of AT2 to obtain the smoothed scale TA2.

The first difference of time, which is the average frequency of each clock at time t over the interval τ_x, based on the latest two time updates (when there are no time steps), is used as an estimate of that clock's current frequency y_{dx}. The calculation is

$$y_{dx_i}(t) = \frac{x_i(t) - x_i(t-\tau_{x,i})}{\tau_{x,i}}. \tag{10}$$

We predict y from previous time taking the last updated value of y and adding any deterministic frequency drift to obtain

$$\hat{y}_i(t) = y_i(t-\tau_{y,i}) + D_i\tau_{y,i}. \tag{11}$$

When a time step is detected, we cannot use the estimate y_{dx}, because the step in time inserts a bias. Hence there can be no frequency update until the next measurement cycle. At that time, the time since the last frequency estimate $\tau_{y,i}$ is larger than τ_x. We estimate y_{dx} over $\tau_{x,i}$, but we predict $\hat{y}$ over $\tau_{y,i}$.

Because we are estimating the non-stationary part of frequency that may be derived from averaging the white FM, the Kalman filter equations prescribe the formulas for combining our measurement y_{dx} with our prediction $\hat{y}$, based on previous estimates. These terms are weighted with the white FM level R as measurement noise and the predicted variance $\hat{P}$ as our confidence on y from the past:

$$y(t) = \frac{\hat{P}(t)\,y_{dx}(t) + R\hat{y}(t)}{\hat{P}(t)+R}. \tag{12}$$

Here, the Kalman formalism provides us with an exponential filter on y, as in AT1. In steady state, AT2 reduces to AT1 if the weights are chosen properly; AT2 thus inherits from AT1 the ability to model flicker FM.

The white FM level R is adjusted from σ_α^2, so that we may interpret it as the expected variation in frequency over $\tau_{x,i}$ due to the white FM process. This interpretation equates R to the average frequency variation over $\tau_{x,i}$ from the mean frequency offset of this clock from ensemble frequency. This mean frequency is defined approximately because white FM is the dominant noise process. We assume that the random walk in frequency is slow compared to the white fluctuations in frequency. Thus, averaging the white FM defines an underlying mean frequency consistent with a pure random walk FM model. This definition of R equates it with the two-point, or Allan, variance and gives it units of frequency according to [9]

$$\begin{aligned} R &= \left\langle \left(\bar{y}(t+\tau) - y_{mean}\right)^2 \right\rangle \\ &= \frac{1}{2}\left\langle \left(\bar{y}(t+\tau) - \bar{y}(t)\right)^2 \right\rangle \\ &= \sigma_y^2(\tau) \\ &= \frac{\sigma_\alpha^2}{\tau_0 |\tau_x|}. \end{aligned} \tag{13}$$

The brackets < > in these equations mean "expected value." Because we estimate σ_α^2 as the "measurement noise," our system constitutes an adaptive Kalman filter, which allows P, the variance of the residuals of y, to evolve with changes in σ_α^2 from an initial value. Thus the integration time of the exponential filter on y, as expressed in equation (12), can change with time. This occurs when initializing a new clock and when a clock's white FM level changes.

5.9 Frequency Variance Prediction

We predict the frequency variance according to

$$\hat{P}(t) = P(t-\tau_y) + Q(\tau_y), \tag{14}$$

where $\hat{P}$ is the prediction of the variance of the residuals of y. The process noise in frequency Q is the variance in y due to random walk FM over the interval τ_y [9],

$$\begin{aligned} Q &= \langle(\bar{y}(t+\tau)-\bar{y}(t))^2\rangle \\ &= 2\sigma_y^2(\tau) \\ &= \sigma_\beta^2 \frac{2n^2+1}{3n}, \end{aligned} \tag{15}$$

where n is the number of τ_0 intervals in τ_y,

$$n = \frac{\tau_y}{\tau_0}. \tag{16}$$

Q grows with τ_y as a function of σ_β^2, the expected variance in y over the interval τ_y due to flicker or random-walk FM.

5.10 Frequency Variance Update

As we obtain a measurement of frequency, so the confidence improves. Hence the variance P decreases as a function of the measurement noise. The Kalman formalism provides for our update of the frequency variance as

$$P(t) = \left[\frac{R\hat{P}}{R + \hat{P}}\right]_t. \tag{17}$$

6. TIME-STEP DETECTION

A time step is defined as an innovation greater than three times the prediction error, that is, when

$$|x_i(t) - \hat{x}_i(t)| > 3\epsilon_i(t). \tag{18}$$

The algorithm responds to a time step by deweighting that clock using the weight-control parameter, wct_i, for the current measurement cycle [16]. If we define the ratio of the innovation to the prediction error as the proportion

$$prop_i(t) = \frac{|x_i(t) - \hat{x}_i(t)|}{\epsilon_i(t)}, \tag{19}$$

then we set wct_i to

$$wct_i = \begin{cases} 1, & \text{if } prop_i \leq 3, \\ 1 - (3 - prop_i)^2, & \text{for } 3 < prop_i < 4, \\ 0, & \text{if } prop_i \geq 4\ . \end{cases} \tag{20}$$

This allows a clock to lose weight smoothly as its time innovation increases from 3 to 4 times the prediction error. A quadratic dependence was chosen as the simplest way to decrease the weight with a zero initial rate of change.

7. FREQUENCY-STEP DETECTION

Frequency steps can be detected only some time after they occur [1,4]. Because our measurements are of phase differences between clocks with inherent white FM, frequency steps appear only after the accumulation of measurements. Flicker or random-walk FM can be understood as a series of stochastic changes in frequency, hence there is no mean frequency. Small frequency steps are indistinguishable from the stochastic noise. We define a frequency step as a shift in frequency greater than four times the standard deviation of the stochastic impulses. Empirically we have found that this detects frequency steps which are clearly significant, while leaving undetected smaller steps which can be interpreted as large stochastic impulses driving the random walk or flicker FM. This implies that the distribution of such impulses is not Gaussian. By contrast, a time step is a shift in phase of a clock that exceeds three times the current white FM characterization.

The detection of frequency steps may be obscured by the presence of time steps and the stochastic noise from white and random-walk FM. If a clock generates a time step, an average frequency spanning the time step will be biased by the size of the time step, divided by the interval between measurements. This bias can be large. Only after two measurements can we conclude that the new frequency after the time step is approximately equal to that before it and so distinguish between time and frequency steps.

It may also be useful to look backward beyond two measurements for the determination of small frequency steps. This procedure allows the white noise to be averaged down to the point where only random-walk or flicker FM remains. Thus the maximum useful extent of a search backwards for frequency steps is where the flicker or random-walk FM begins to dominate over the white FM. We use as a measure of this time interval the time constant weighting the previous frequency estimates in the exponential filter for y (12). If we call this time interval L_{max}, the number of τ_0 intervals is then, from (12),

$$L_{\max} = \frac{R}{\hat{P}}. \tag{21}$$

For $\hat{P}$ in (21), we use the steady-state value P_{ss} of P, then add Q to it as indicated in (14). To obtain P_{ss} we set $P(t+\tau) = P(t)$ and solve, using (13) -(17). Also, substituting in the expressions for P and Q, equations (13) and (15), using $\tau=\tau_0$ we find

$$\begin{aligned} P_{ss} &= Q\left(\sqrt{\frac{1}{4} + \frac{R}{Q}} - \frac{1}{2}\right) \\ &= \sigma_\beta^2\left(\sqrt{\frac{1}{4} + 2\left(\frac{\sigma_\alpha}{\tau_0\sigma_\beta}\right)^2} - \frac{1}{2}\right). \end{aligned} \tag{22}$$

Adding Q to P_{ss} to obtain $\hat{P}_{ss}$, (14), substituting this into (21) and simplifying, we find

$$\begin{aligned} L_{\max} &= \frac{1}{2}\left(\sqrt{1+4\frac{R}{Q}} - 1\right) \\ &= \frac{1}{2}\left(\sqrt{1+4\left(\frac{\sigma_\alpha}{\tau_0\sigma_\beta}\right)^2} - 1\right). \end{aligned} \tag{23}$$

We assume, for evaluating R and Q in (23), that $\tau=\tau_0$, because (23) has been derived from the steady state P.

Because we must measure clock differences, our best estimate of an individual clock frequency will be for a clock against the ensemble. But the ensemble itself must have stochastic processes, and these also need to be considered when testing for frequency steps.

The algorithm tests for frequency steps by iterating backward for each clock a range of measurements from two before the current measurement time up to $L_{\max}$ back. On each measurement cycle, for each clock, we maintain a buffer of the previous $L_{\max}$ values of all the necessary parameters for frequency-step detection. At each measurement, we search for frequency steps for each clock. We define the average frequency of a clock from the first difference over the time interval from the measurement L intervals back ($L < L_{\max}$), x_{-L} at time t_{-L}, to the most recent measurement before the current time x_{-1} at time t_{-1}; that is,

$$y_{avg} = \frac{(x_{-1} - x_{-L})}{(t_{-1} - t_{-L})}. \tag{24}$$

We compare this frequency with the updated estimate of frequency $y(t_{-L})$ saved from time t_{-L}. That is, we compare a filtered estimate of frequency computed at the beginning of a fixed-time interval based on the clock's history with an estimate of the average frequency over that interval. If the frequency difference exceeds a test value, we conclude that the clock generated a frequency step.

The variance for the test is the sum of the expected variance for y_{avg} plus the variance for $y(t_{-L})$. The variance of y_{avg} is the sum of the expected variance of y against the ensemble over the interval from t_{-L} to t_{-1}, plus the variance of the ensemble itself. The expected variation of the ensemble is a function of both $\sigma^2_{\alpha,x}$ and $\sigma^2_{\beta,x}$ where

$$\sigma^2_{\beta_x} = \frac{1}{\sum_i \frac{1}{\sigma^2_{\beta_i}}}. \tag{25}$$

For the variance of y against the ensemble over the interval, we use the maximum $P_{max}(t,L)$ of P either at the beginning or the end of the interval multiplied by L_{max}/L, because the variance decreases by the averaging time L.

The variance of the ensemble frequency due to white FM decreases as L, while the variation in ensemble frequency due to random walk FM increases with L. Finally, the variation in $y(t_{-L})$ is $\hat{P}(t_{-L})$. We use $\hat{P}$ instead of P, since we are testing for a frequency step. Hence we cannot use the measurement of frequency at time t_{-L} to decrease the uncertainty of frequency. For L large, typically the dominant term in σ_L is $\hat{P}$.

The test variance is

$$\sigma_L(t) = \sqrt{\frac{L_{max}}{L}\left[P_{max}(t,L)+\sigma^2_{\alpha_x}(t)\right]+L\sigma^2_{\beta_x}(t)+\hat{P}(t_{-L})}\ . \tag{26}$$

We use four standard deviations of σ_L, simply because we empirically have found it appropriate. Smaller frequency steps have the same effect on the ensemble as random walk FM, which is already modelled.

Hence we define that a frequency step has occurred when

$$|y_{avg}-y(t_{-L})| > 4\sigma_L. \tag{27}$$

If a clock generates a frequency step, usually that same step will appear over a contiguous range of test times. We look for frequency steps over all times in the allowable range and treat that having the largest inequality in (27) as containing the actual frequency step. If we find only one occurrence of a frequency step over a range of L values, we ignore it, assuming the detection to be in error. Any large frequency step must produce steps detected over a range of test times.

When a frequency step is detected, we re-run the scale from the time t_{-L} of that step with that clock deweighted. First we set the frequency and variance estimate of frequency at t_{-L} to

$$y(t_{-L}) = y_{avg}, \qquad P(t_{-L}) = \frac{\sigma_\alpha^2}{L} + \sigma_\beta^2 \cdot L. \tag{28}$$

We do not search for new frequency steps during this secondary run; thus there is no iterative search for frequency steps. The stepped clock remains deweighted until a time of L_{max} measurement cycles after the occurrence of the detected step. In other words, when we find a frequency step, that clock's frequency is re-initialized with an average frequency and the clock is removed from the scale until the algorithm adaptively learns the clock's new frequency.

8. TA2 FROM AT2

The smoothing algorithm TA2 essentially averages the filtered frequencies of clocks, generated by running AT2 forward and then backwards in time over the same data to obtain a "best" estimate of frequency for each clock. We run AT2 first forwards and then backwards in time over a fixed interval of data, saving the frequency values $y_i(t)$ of clocks, and their variances $P_i(t)$. In the third and final, forward pass through the data we average frequencies from the previous two passes for the predictions of x_i, for each clock as in (1). The frequency y_f from the previous forward pass is averaged with the frequency y_b from the backward pass using the reciprocals of their variances P_f and P_b respectively, for weights.

$$y_s(t) = \frac{\dfrac{y_f(t)}{P_f(t)} + \dfrac{\hat{y}_b(t)}{\hat{P}_b(t)}}{\dfrac{1}{P_f(t)} + \dfrac{1}{\hat{P}_b(t)}}. \tag{29}$$

We use the prediction in the backward pass to avoid using the data at time t twice in estimating the smoothed frequency y_s. Then in the third pass, we generate ensemble time using these smoothed frequencies for prediction.

9. INITIALIZATION

When starting the algorithm with no ensemble history or else entering a new clock into an existing ensemble, we need an estimate of the deterministic parameters, offsets of time, frequency, and frequency drift from the ensemble, and also the stochastic parameters σ_α and σ_β of white and random walk FM.

When starting a new ensemble, the deterministic estimates must be consistent. If x_i and x_j

are the initial estimates of time offset from ensemble time for clocks i and j respectively, then their difference must equal the measured difference; that is,

$$x_i - x_j = x_{ij}. \tag{30}$$

Whereas this is the only requirement, any constant value may be added to all x_i, to predetermine the initial time value of the ensemble. Similarly, any constant may be added to all the clocks' frequency or drift estimates to pre-determine the initial frequency or drift of the ensemble. Indeed, at any time, constants may be added (to all clocks) to steer the ensemble to any desired time, frequency or drift. The internal equations of the algorithm are not affected by such steering, which follows from our observation that we never measure the time of any clock, but only time differences between clocks. The algorithm optimizes stability among clocks, but provides no measure of accuracy.

The stochastic parameters can be estimated from an Allan or two-point variance of the clock, which is a variance that estimates fractional frequency stability [20-28]. In our formalism, $\sigma_\alpha(\tau)$ and $\sigma_\beta(\tau)$ are the expected dispersions of a clock after an interval τ, where the dispersion is one of time in ns due to white FM, and of frequency in ns/d due to random walk FM, respectively. We enter the values for $\tau = 1\text{d}$, and internally convert them to $\tau = \tau_0$, two hours in our case. The conversion formula is

$$\sigma^2(\tau') = \sigma^2(\tau)\,\frac{\tau'}{\tau}, \tag{31}$$

for both $\sigma_\alpha(\tau)$ and $\sigma_\beta(\tau)$. This can be derived using the formula relating them to the Allan variance from [9],

$$\sigma_y^2(\tau) = \frac{\sigma_\alpha^2}{n\tau_0^2} + \frac{\sigma_\beta^2(2n^2+1)}{6n}. \tag{32}$$

If we then use the τ dependence of the Allan variance [20-26] for each of the two noise types separately in (32), we obtain (31).

11. CONCLUSIONS

We have shown the performance of the NIST post-processed time scale TA2 over a period of nine years and given the equations defining the algorithm generating these data. The algorithm itself shows properties similar to the real-time scale AT1, which has been running at NIST since the late 1960's. TA2 is a smoother algorithm based on the real-time algorithm AT2. TA2 is generated by running AT2 forward and backward in time to estimate clock frequencies, then running a forward pass of AT2 to combine these frequencies and estimate time. AT2 incorporates some changes from AT1 that should provide some advantages over AT1 for post-processing or for automatic error detection. A third time scale at NIST is TA(NIST), a pure Kalman filter not directly related to TA2. We have also seen here the performance of the NIST ensemble of atomic clocks over this nine-year period.

REFERENCES

[1] Weiss, M.A. and Weissert,T., TA2, A time scale algorithm for Post-Processing: AT1 plus frequency variance, *IEEE Trans. Instrum. Meas.*, 1991, **IM-40,** pp. 496-501.

[2] Weiss, M.A. and Weissert,T., AT2, A new time scale algorithm: AT1 plus frequency variance, *Metrologia*, 1991, 28, 65-74.

[3] Weiss, M. A., Weissert, T., Promise into Practice: Implementing TA2 on Real Clocks at NIST, Proc. 5th European Frequency and Time Forum, Besancon, March 12-14, 1991, pp. 442-448.

[4] Weiss, M. A., Allan, D. W., Peppler, T. K., A Study of the NBS Time Scale Algorithm, *IEEE I&M*, vol. 38, no. 2, April 89, pp 631-635.

[5] D. W. Allan, D. J. Glaze, J. E. Gray, R. H. Jones, J. Levine, and S. R. Stein, Recent Improvements in the Atomic Time Scales of the National Bureau of Standards, Proc. of the 15th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 29-40, Naval Research Laboratory, Washington D. C., Dec. 6-8, 1983.

[6] NBS Monograph 140, 1974.

[7] R. H. Jones and P. V. Tryon, Estimating Time From Atomic Clocks, Journal of Research of the National Bureau of Standards, vol. 88, no. 1, pp. 17-24, Jan - Feb 1983.

[8] Jones, R. H., and Tryon, P. V., Continuous Time Series Models for Unequally Spaced Data Applied to Modeling Atomic Clocks, *SIAM J. Sci. Stat. Comput.*, vol 8, no. 1, Jan 87, pp 71-81.

[9] J. A. Barnes, R. H. Jones, P. V. Tryon, and D. W. Allan, Stochastic Models for Atomic Clocks, NASA Conference Publication 2265-- Proc. of the 14th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 295-306, 1982.

[10] P. V. Tryon and R. H. Jones, Estimation of Parameters in Models for Cesium Beam Atomic Clocks, Journal of Research of the National Bureau of Standards, vol. 88, no. 1, pp. 3-16, Jan.-Feb. 1983.

[11] M. A. Weiss, C. Hackman, Confidence on the Three Point Estimator of Frequency Drift, *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Nov. 1993, McLean, VA.

[12] D. A. Howe, D. W. Allan, J. A. Barnes, Properties of Signal Sources and Measurement Methods, Proc. 35th Symposium on Frequency Control, pp.1-47, 1981.

[13] Barnes et. al. Characterization of Frequency Stability, IEEE Trans. on Instrum and Meas., vol. IM-20, no. 2, pp. 105-120, 1971.

[14] Stein, S.R., Gifford, A., Breakiron, L.A., Report on the Timescale Algorithm Test Bed at USNO, *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Nov. 1989, Redondo Beach, CA, pp. 269-288.

[15] Guinot, B., Thomas, C., Establishment of International Atomic Time, *BIPM Annual Report*, 1988, Part D, pp. D1-D22.

[16] D. B. Percival, Use of Robust Stastical Techniques in Time Scale Formation, 2nd International Symposium on Time Scale Algorithms, Boulder, Colorado, June, 1982, chapter 14.

[17] A. Bausch, et. al., The New PTB Primary Cesium Clocks, special issue of IEEE Trans. on Instrum and Meas. for CPEM '92, in press.

[18] D. B. Percival, The U.S. Naval Obervatory Time Scales, IEEE Trans. on Instrum and Meas., vol. IM-27, no. 4, December, 1978, pp.376-385.

[19] P. Tavella, J. Azoubib, C. Thomas, Study of the Clock-Ensemble Correlation in ALGOS Using Real Data, Proc. 5th European Frequency and Time Forum, Besancon, March 12-14, 1991, pp.435-441.

[20] D. W. Allan, Statistics of Atomic Frequency Standards, Proc. IEEE, vol. 54, no. 2, pp. 221-230, 1966.

[21] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, and G.M.R. Winkler, Characterization of Frequency Stability, IEEE Transactions on Instrumentation and Measurement, IM-20, No. 2, 105-120, 1971.

[22] Characterization of Clocks and Oscillators, Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[23] J. Rutman, Characterization of Frequency Stability: A Transfer Function Approach and Its Application to Measurements via Filtering of Phase Noise, IEEE Trans. on I & M, Vol. IM-23, No. 1, March 1974, pp. 40-48.

[24] J. Rutman, Characterization of Phase and Frequency instabilities in Precision Frequency Sources: Fifteen Years of Progress, Proc. of the IEEE, Vol. 66, No. 9, pp. 1048-1075, September 1978.

[25] J. Rutman and F. L. Walls, Characterization of Frequency Stability in Precision

Frequency Sources, submitted to Proc. of the IEEE, Vol. 79, No. 7, pp. 952-960, July, 1991.

[26] Allan, D.W., Time and Frequency (Time-Domain) Characterization Estimation, and Prediction of Precision Clocks and Oscillators, *IEEE Trans. on UFFC*, vol. UFFC-34, no. 6, November, 1987, pp. 647-654.

[27] D. B. Percival, Reappraisal of Frequency Domain Techniques for Assessing Stability Measures, Proc. of the PTTI, 1987, pp. 69-80.

[28] W. C. Lindsey and C. M. Chie, Theory of Oscillator Instability Based Upon Structure Functions, IEEE 64, 1652-1666 (1976).

[29] A. Gelb, ed., *Applied Optimal Estimation*, M.I.T. Press, Cambridge, MA, 1974.

[30] J. Barnes,Time Scale Algorithms Using Kalman Filters -- Insights from Simulation, Proc. 2nd Symposium on Atomic Time Scale Algorithms, NBS, Boulder, CO, June 23-25, 1982.