Metrologia **45** (2008) S23–S33 doi:10.1088/0026-1394/45/6/S04

Realizing UTC(NIST) at a remote location

Judah Levine

Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO 80305, USA

E-mail: jlevine@boulder.nist.gov

Received 25 June 2008
Published 5 December 2008
Online at stacks.iop.org/Met/45/S23

Abstract

I will describe the backup time scale system that I have constructed at the site of the NIST radio stations near Fort Collins, Colorado, and I will compare its performance with the primary ensemble in Boulder. The Fort Collins system is designed to be a backup for the Boulder system and is intended to support all of the NIST time services should the primary scale become unavailable for any reason. The backup time has a number of unique problems and requirements, and I will discuss the design considerations that I used to address these issues. The backup scale tracks UTC(NIST) in frequency with an uncertainty (measured by the Allan deviation of the difference) of about 1×10^{-14} by use of administrative steering that is applied not more often than once per week. The corresponding time deviation is less than 1 ns for all averaging times less than 1 week, and the peak time difference between UTC(NIST) and its backup realization is less than ± 25 ns and is generally much better than this value. This is much better than would be needed for supporting the radio stations, the digital time services (ACTS, a time service that provides time in a digital format using dial-up telephone lines, and the Internet services) and the Frequency Measurement Service. Its frequency stability and time accuracy would not be adequate for the most demanding users of the Global Time Service and for international time and frequency coordination. The primary limitations to the performance of the backup time scale are caused by environmental perturbations, especially temperature and supply voltage, and the existing hardware could probably support all of the NIST services if the environment were improved.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

I will describe a time scale system that I have constructed at the site of the NIST radio stations near Fort Collins, Colorado. The system is designed to be a backup for the primary system in Boulder, and is intended to support all of the NIST time services should the primary scale become unavailable for any reason. Although the system is similar to the primary clock ensemble in Boulder, there are a number of unique considerations in its design, which I will discuss in detail below.

The connection between the backup time scale in Fort Collins and the primary time scale in Boulder must be designed as a compromise between two conflicting goals. The backup time scale must be close enough in time and in frequency to the primary scale so that users will not see a significant discontinuity if the services are switched to the backup scale when the primary scale fails. Satisfying this requirement requires a relatively tight coupling between the primary and backup scales. On the other hand, the backup scale must continue to function when the primary scale is unavailable, so that a lack of data from the primary scale must not cause the backup scale to fail or to become too unstable. Satisfying this requirement is simplified with a relatively loose coupling between the two scales.

The coupling system that I have designed evaluates the difference between the two scales (in time and in frequency) once per week and applies a frequency steer to the Fort Collins scale at that time if necessary. Time steps are never used. The amplitude of the frequency steer is administratively limited to $\pm 5 \times 10^{-15}$; this limit has been exceeded on a few occasions when a sharp drop in the outside temperature resulted in a frequency offset of almost 1×10^{-14}, and this required a steer of equal magnitude to remove the effect. On the other hand, it is quite common for no steering to be applied for several weeks in a row. As I will show in the subsequent discussion, the amplitude of the steering corrections is comparable to the free-running stability of the time scale, so that the effects of

0026-1394/08/060023+11$30.00

the steering should not be visible outside of the time scale itself.

This relatively gentle steering in frequency implies that time offsets as large as 15 ns are possible; more aggressive steering could reduce these peak time offsets but could introduce a more serious problem in case of a failure. For example, if the steering system were to apply a large frequency offset in order to remove a large time offset in a short time interval, and if the system were to fail before this offset could be removed, then the scale would continue to implement this large frequency offset and its response could potentially become unbounded.

The Fort Collins system is implemented with four commercial caesium standards. The standards are interfaced to two identical, independent measurement systems that compute weighted averages of the times of the clocks every 12 min and use these averages to steer two independent phase steppers. The outputs of the phase steppers are connected to two independent GPS receivers that measure the difference between the 1 Hz outputs of the phases steppers and GPS system time by means of up to eight of the satellites that are in view at any epoch. The site has significant multipath reflections, and the data are averaged to attenuate this effect, by the use of the sidereal-day frequency method that I have developed [1].

The receiver software includes an algorithm to remove outliers, which are common at the site because of the significant interference from the transmitters. This interference manifests itself in large ground currents at the radiated frequencies (60 kHz, 2.5 MHz, 5 MHz, 10 MHz, 15 MHz and 20 MHz), in significant out-of-band interference at these frequencies in the front end RF sections of the GPS receivers, and in in-band interference in the measurement system, especially at 5 MHz, which is the output frequency of the caesium clocks that is used in the measurement system. These interference effects can be attenuated to some extent by shielding, but the residual signals that leak through the shields still have significant amplitude. The performance of the system is degraded because of these effects and because of the power transients and diurnal power fluctuations, which are also common at the site.

It is a simple matter to detect most of the outliers, since they are typically greater than 20 ns, which is nearly ten times the measurement noise. The algorithm is not so successful in removing smaller outliers and in modelling the irregular fluctuations in the environmental parameters, especially the local temperature. I will discuss this point in more detail below.

2. Sign convention and units

In the following discussion, all times and time intervals are in units of UTC seconds. A frequency has the units of s/s and is therefore a dimensionless quantity. Frequency aging has the units of s/s^2 or 1/s. All time tags and epochs are based on UTC(NIST)—the local realization of UTC.

The notation for a time difference is x_{ab}, which is equivalent to $x_{\mathrm{a}} - x_{\mathrm{b}}$. A positive value for the time difference implies that the time of the first device is ahead of the time of the second device. The sign convention for frequency and frequency ageing are consistent with this definition: positive values for these quantities imply that the corresponding time difference is increasing.

3. Measurement hardware

The measurement system for the time scale consists of a dual-mixer system that converts the 5 MHz signal from each clock to an intermediate frequency near 10 Hz. (The current hardware cannot accept higher input frequencies, although the clocks can provide them.) The signal used to generate this intermediate frequency is synthesized from the 5 MHz input of the reference clock, which is one of the clocks in the ensemble.

The time stability of the measurement hardware is evaluated by connecting the same clock to two channels. The performance varies depending on which channels are used in the comparison. Figure 1 shows two typical values. All of the channels show a significant peak at one day due to environmental perturbations. However, the contribution of the measurement noise of the hardware is less than the contribution due to the clocks for all averaging times.

Each time a measurement cycle is initiated, the measurement system reports the time difference between the reference clock and each of the other clocks. Since the time differences reported by the hardware are ambiguous modulo 200 ns, the system includes a calibration capability that is used to determine both the integer cycle offset between the data reported by the measurement system and the 1 pps outputs of the clocks and the fractional cycle due to the delays in the cables and the hardware. The fractional-cycle ambiguities change only when the hardware itself is changed, but the integer ambiguities must be redetermined each time the system is turned on from a cold start, and the system can do this automatically.

In order to determine the integer ambiguity following a cold start of the system, the software examines the time-difference data before the outage and linearly extrapolates these measurements to the epoch of the cold start. The difference between this extrapolation and the current measured time differences is examined clock by clock to determine the new integer cycle number for each clock that best matches the previous data. This process is limited by the time dispersion of the clocks, but this dispersion is generally much smaller than 200 ns, so that the integer cycle count can be determined unambiguously even if the interpolated gap is several days long. Since the integer cycle ambiguity of the reference clock affects all of the other time differences, this value is adjusted first, and the new time differences are then recomputed with this new integer for the reference clock.

When the system is turned on for the very first time (or when the hardware is changed), the integer cycle ambiguities and fractional-cycle cable delays are determined administratively by comparing the data reported by the dual-mixer measurement system with measurements made by a standard time interval counter, which has a resolution of 0.1 ns. The counter measures the time difference between the 1 Hz pulse at the output reference plane of the system and the 1 Hz pulse from each of the clocks that is being measured. The

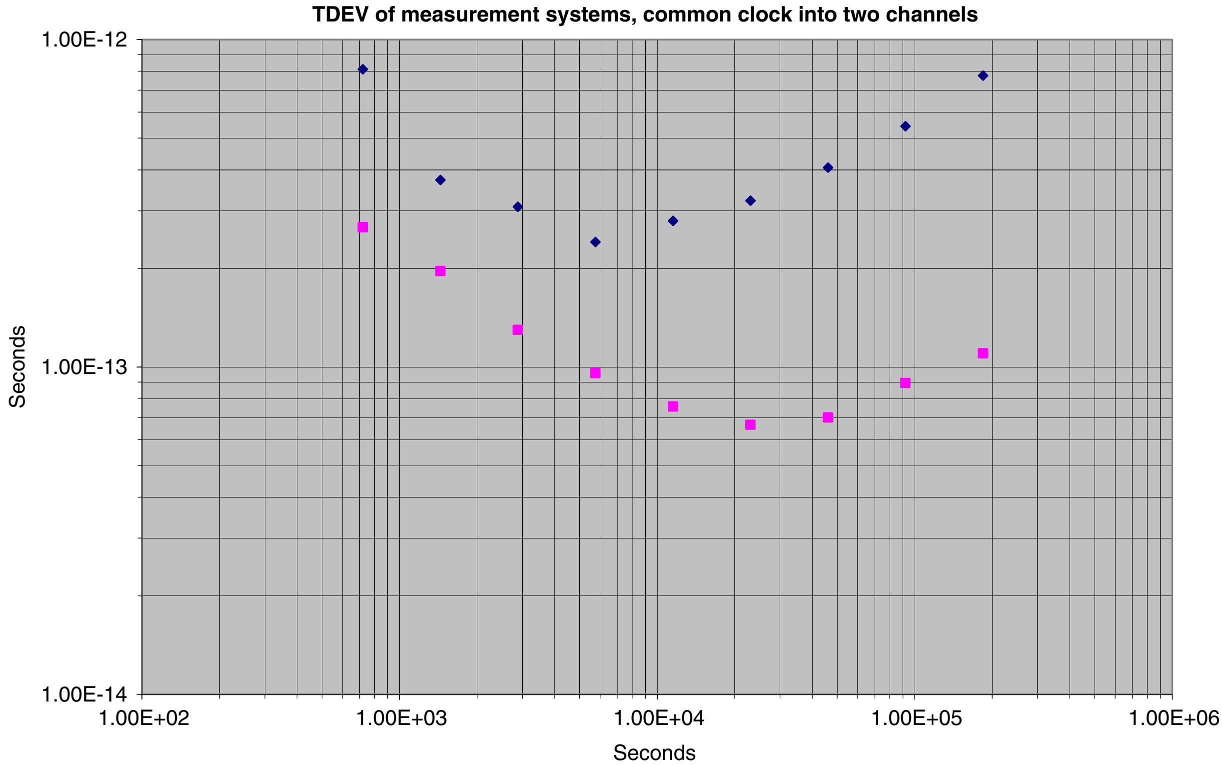

Figure 1. The TDEV of the measurement systems in Fort Collins, evaluated by measuring the time difference between a single clock connected on two different channels. The results depend on which channels are used for the comparison, and the two series show the maximum and minimum values of TDEV that we obtained with two different pairs of channels.

reference plane for each of the time scale systems is the output of the distribution amplifier that is driven by the corresponding phase stepper. The reference plane for each clock is taken as the 1 pps output connector on the front of the device. The cable-delay constants in the software are set so that these two measurements agree. To simplify the redetermination following a subsequent cold start, the cable-delay constants are divided into an integer cycle count and a fractional offset. The integer cycle count has units of 200 ns, while the fractional count has units of picoseconds. (For historical reasons, there is also a third integer offset constant whose units are 2^{23} cycles, or 1.677 7216 s. This constant is used in the Boulder system but not in Fort Collins.) As much arithmetic as possible is done with integers to minimize round-off problems.

4. The AT1 time scale algorithm

The AT1 algorithm has been used for many years at NIST (and previously at NBS) to analyse the data from an ensemble of caesium standards and hydrogen masers. The data input to the algorithm are a series of time-difference measurements between the reference clock, clock r, and the other devices in the ensemble. The time difference between the reference clock and clock j at time t is denoted by $X_{rj}(t)$. The measurements are normally equally spaced in epoch at an interval of τ seconds, so that the measurement times can be expressed recursively by $t_k = t_{k-1} + \tau$. The current implementation uses a value of $\tau = 720$ s (12 min). The exact value is not critical and the value that is used is chosen mostly for computational convenience, since it is an exact decimal fraction of an hour.

I will consider equally spaced measurements to simplify the notation. However, the algorithm does not depend on equally spaced data, and τ in the following equations can be replaced by the actual interval between the current measurement and the previous one. Different values of τ happen occasionally when the hardware fails and some number of measurement cycles are lost. The measurement cycle is resynchronized when the hardware is restarted, so that the gap is an exact integer multiple of τ. The larger interval across the gap is handled with no special processing.

The measured time differences at epoch t_k are expressed as $X_{rj}(k)$. There is nothing special about the reference clock, and its time difference is reported by the hardware as $X_{rr}(k) = 0$.

The time of each clock with respect to the ensemble at epoch t_k is modelled recursively in terms of its time offset, frequency offset and frequency ageing at the previous epoch t_{k-1} by

$$x_{je}(k) = x_{je}(k-1) + y_{je}(k-1)\tau + 0.5d_{je}(k-1)\tau^2, \quad (1)$$

where x, y, and d on the right side of the equation represent the time offset, the frequency offset and the frequency ageing, respectively, of clock j with respect to the ensemble that were estimated at the previous measurement epoch. There are N equations of this type – one for each member of the ensemble, including the reference clock. The frequency ageing term is included here for completeness, but it is set to 0 for all of the clocks in the Fort Collins time scale.

Each one of the measured time differences can be combined with the corresponding model equation for that clock to compute a prediction of the time of the reference clock with respect to the ensemble at the current epoch. Thus

$$\widehat{X}^{j}_{re}(k) = x_{je}(k) + X_{rj}(k) \tag{2}$$

is an estimate of the time of the reference clock with respect to the ensemble at the current epoch based on a measurement of the time difference between the reference clock and clock j. There is one of these equations for each clock in the ensemble, including the reference clock, where it is simply an identity, since the corresponding time difference is 0.

The provisional estimate of the time of the reference clock with respect to the ensemble is the weighted sum of these estimates over all of the members of the ensemble:

$$\widehat{X}_{re}(k) = \sum_{j=1}^{N} w_j(k)\widehat{X}^{j}_{re}(k). \tag{3}$$

The weight of each estimate is computed from the average prediction error of the clock over the previous cycles (defined below):

$$w_j(k) = \frac{1}{\sigma_j^2(k)}. \tag{4}$$

The weights are normalized so that they sum to 1 by the use of:

$$\sigma^2(k)\sum_{j=1}^{N} w_j(k) = 1$$

so that

$$\frac{1}{\sigma^2(k)} = \sum_{j=1}^{N} w_j(k) = \sum_{j=1}^{N} \frac{1}{\sigma_j^2(k)} \tag{5}$$

and

$$w_j(k) = \frac{\sigma^2(k)}{\sigma_j^2(k)}.$$

The prediction error for each clock is the difference between the estimate computed in equation (2) for that clock and the ensemble average of all of these estimates as computed by equation (3):

$$\varepsilon_j(k) = \widehat{X}^{j}_{re}(k) - \widehat{X}_{re}(k). \tag{6}$$

The prediction error on this measurement cycle is compared with the average prediction error over previous cycles. If the average prediction error is limited by the administrative considerations discussed in the next section, then the average prediction error in the denominator of the following expression is replaced by the ensemble average value, $\sigma(k)$, as defined above in equation (5).

$$\kappa_j(k) = \frac{|\varepsilon_j(k)|}{\sigma_j(k)}. \tag{7}$$

Case 1: $k_j(k) \leqslant 3$. Accept the estimate for this clock and continue.

Case 2: $3 < k_j(k) < 4$. Modify the weight of this clock in the ensemble average (equation (3)). In the following expression, the original weight computed in equation (5) is $w_j^0(k)$, and it is replaced in the recomputation of the ensemble average by

$$w_j(k) = (4 - \kappa_j(k))w_j^0(k), \tag{8}$$

which de-weights the clock linearly from its original value for $\kappa = 3$ to 0 when $\kappa_j(k) = 4$. Set a flag to show that the clock has been de-weighted at this epoch. Return to equation (3) and recompute the algorithm.

Case 3: $\kappa_j(k) \geqslant 4$. Set the weight of this clock to 0, return to equation (3) and recompute the algorithm. Set a flag to show that this clock has been dropped from the average at this epoch.

If more than one clock satisfies the conditions of case 2 or case 3, then perform the operation only on the clock with the largest value of κ and recompute the ensemble average by means of equation (3) with the modified weight for that clock. A clock whose weight has been modified by case 2 or case 3 is not tested again on a subsequent loop. When no further failures are detected, continue with the subsequent section on parameter updates.

5. Administrative limit on the weights

The algorithm described above is potentially unstable if one of the clocks is significantly more stable than the others so that its prediction error is consistently smaller than the errors of the other members of the ensemble. The same effect can happen if the combination of the measurement noise and the prediction errors conspire to decrease the prediction error of one of the clocks. Since the weight of a clock in the ensemble average is derived from its prediction error, a clock with a small prediction error has a high weight, and this produces a significant correlation between the time of the clock and the average time of the ensemble. The prediction error for such a clock is always too small, since it contributes to both terms on the right-hand side of equation (6). In an extreme situation, this can result in a positive feedback loop, in which a clock that is initially somewhat better than the others eventually is given a weight close to 100% and effectively takes over the ensemble. This is potentially very troublesome in the backup time scale, since there are only four clocks in the ensemble, so that all of them are going to have significant weight in the ensemble average to begin with.

Since the positive feedback loop results from the correlation between the ensemble average of the prediction errors and the contribution of a high-weight clock, one solution is to compute the effects of this correlation and increase the prediction error to account for it [2]. This method will be discussed below in the parameter update section. A second solution is to limit the maximum weight that any clock can have in the ensemble average. This maximum weight is set at 30%. From equation (5), if the maximum weight is to be limited to 0.3, then, for every clock, the average prediction error is limited by

$$w_j(k) = \frac{\sigma^2(k)}{\sigma_j^2(k)} \leqslant 0.3 \tag{9}$$

Table 1. Average prediction errors and corresponding weights in the Fort Collins ensembles.

Clock	σ_a/ns	Weight-a	σ_b/ns	Weight-b
1	1.6	34%[a]	1.6	28%
2	1.6	34%[a]	1.5	32%*
3	2.5	13%	1.8	22%
4	1.8	27%	1.9	20%

[a] Limited to 30%. See text.

or

$$\sigma_j(k) \geqslant 1.83\sigma(k).$$

If this condition is violated for some clock, then the value calculated in equation (5) is replaced by the constant 0.3 for the weight of the contribution of the clock to the average in equation (3).

Since the weights are normalized so that the sum is 1, reducing the weight of a good clock below what its statistical performance would predict implicitly transfers the weight to poorer clocks and gives them more weight than they deserve. The statistical performance of the scale is then not as good as it could be if the administrative limit were not enforced. This degradation in performance is considered an acceptable price to pay for avoiding potentially having one clock take over the scale.

The administrative limit must always be larger than $1/N$. For example, if one of the clocks in the Fort Collins scale fails, then the normalization condition in equation (5) cannot be satisfied if the remaining three clocks are each limited to a maximum weight of 30%. The software adjusts the administrative limit to 40% in this case.

The interaction between the time scale and the measurement noise is illustrated in table 1, which shows the prediction errors and the corresponding weights in the two time scales in Fort Collins. Although the two systems have the same inputs and are nominally identical, the interaction between the measurement noise and the scale algorithm in the two systems results in very different weights being assigned to the 4 clocks. Note, especially, the very different weights assigned to clock 3 by the two algorithms.

The difference in the weights assigned to the same clocks in the two scales means that the two scales will respond differently to any unmodelled contribution to the time-difference measurements, and that the output times will slowly diverge as a result. Since the frequency stability of the clocks for averaging times of days is on the order of 10^{-14} and since the difference in weights is a few per cent, we might expect that the two time scales in Fort Collins will diverge from each other with a frequency offset of somewhat less than 10^{-15}, and that either scale will diverge from the primary ensemble in Boulder at a similar rate. The time dispersion produced by these frequency offsets will depend on the details of the noise processes, but is unlikely to be worse than a few nanoseconds per week. I will discuss this point in more detail below.

6. Parameter updates

When the algorithm described above is finished, we have an estimate of the time of the reference clock with respect to the ensemble based on the measurements of all of the other clocks whose prediction errors were not too large. We can also consider this datum as the realization of the ensemble time as an offset from a physical clock. Since we have measured the physical time differences between the reference clock and all of the other clocks in the ensemble, we can also realize the ensemble time by combining these physical time differences with the estimated offset of the reference clock from the ensemble.

The time of the reference clock with respect to the ensemble is set to the estimate computed above in equation (3) and the weight of each clock as modified by the administrative limit and the prediction error tests

$$x_{re}(k) = \widehat{X}_{re}(k). \tag{10}$$

And this expression can be used to evaluate the final value of the prediction error of this clock:

$$\varepsilon_j(k) = \widehat{X}^{j}_{re}(k) - x_{re}(k). \tag{11}$$

Both of these updated estimates will be identical to the provisional forms if the prediction errors of all of the clocks were within the acceptable limit of three times the corresponding value for σ.

The first step in the parameter update process is to deal with any clock whose weight was reduced to 0 because its prediction error was too large. We model these clocks as having had a simple time step since the last computation. We adjust the time of the clock with respect to the ensemble so that it matches our expectation of its value based on its current measured time difference. If clock m was reset on this cycle, then

$$x_{me}(k) = x_{re}(k) - X_{rm}(k). \tag{12}$$

We do not update its other parameters on this cycle. If this assumption is accurate, then the clock will return to the ensemble with its parameters unchanged on the next measurement cycle and its prediction error will return to be within the expected range.

If the error is not due to a simple time step, then the previous action is unlikely to fix the problem. If the frequency or the stability of the clock changed since the last measurement cycle, then its behaviour will not be modelled by a single time step, and it will most likely be reset repeatedly on subsequent measurement cycles, which effectively removes it from the ensemble, since its weight is set to 0 repeatedly. This is generally an indication of a hardware failure, and the ensemble simply calls for human assistance.

The next step in the process is to update the parameters of each of the clocks that was not reset on the current measurement cycle.

(1) The time of each clock with respect to the ensemble is set to the computed time of the reference clock with respect to

the ensemble and the measured time difference between each clock and the reference clock.

$$x_{j\mathrm{e}}(k) = x_{r\mathrm{e}}(k) - X_{rj}(k). \quad (13)$$

This updated time is used in the following calculation.

(2) The frequency of the clock with respect to the ensemble is set in two steps. The first step estimates the frequency since the last measurement cycle by means of a simple first-difference of the times of the clock with respect to the ensemble:

$$f_{j\mathrm{e}}(k) = \frac{x_{j\mathrm{e}}(k) - x_{j\mathrm{e}}(k-1)}{\tau}. \quad (14)$$

Since this estimate includes the noise of the measurement process of time differences between each clock and the reference clock, it is passed through an exponential low-pass filter (equation (15), below) to yield the frequency offset to be used for subsequent predictions. The frequency ageing term is included here for completeness, but is 0 for the Fort Collins system. In general, it is used only for the hydrogen masers in the Boulder ensemble. It is determined administratively outside of the time scale algorithm and is treated as a constant in the computation [1].

$$y_{j\mathrm{e}}(k) = \frac{w_y y_{j\mathrm{e}}(k-1) + f_{j\mathrm{e}}(k)}{1 + w_y} + d_{j\mathrm{e}}(k-1)\tau. \quad (15)$$

The time constant used in the exponential filter above is determined administratively based on the statistics of the clock and the measurement noise. Since the time dispersion due to the frequency noise in the clock is a function of the measurement interval whereas the noise of the measurement system is constant, the value used in equation (15) is scaled based on the elapsed time since the last measurement. If T_j is the administrative time constant for clock j in seconds, then

$$w_y = \frac{T_j}{\tau}. \quad (16)$$

The effective time constant for frequency updates of the clocks in the Fort Collins system is about 10 days (1200 measurement cycles).

The update of the prediction error is also computed in two steps. The first step computes the integrated prediction error over the last 24 h:

$$S_j(k) = \sum_{t-24\,\mathrm{hrs}}^{t} \varepsilon_k(t), \quad (17)$$

where the prediction error for each cycle is the difference between the time of the reference clock with respect to the ensemble predicted by this clock and the final value of the ensemble time with respect to the reference clock including any effects of clock resets or weight reductions. This value is corrected for the correlation effect discussed above [3] and the result is passed through an exponential filter with a fixed time constant of 31 days:

$$w_s = \frac{\tau}{86\,400} \frac{1}{1 - w_j(k)}, \quad (18)$$

$$\sigma_j^2(k) = \frac{31\sigma_j^2(k-1) + w_s S_j^2(k)}{31 + w_s}. \quad (19)$$

The weight of any clock in the prediction of the time of the reference clock with respect to the ensemble, $w_j(k)$, is limited to 0.3 by the administrative constraint discussed above. The effect of correlation in equation (18) therefore increases the observed variance by a factor of up to 1.43.

7. Clock steering

The algorithm is used to control a phase stepper so as to produce an output that tracks the ensemble average time. The signal input to the phase stepper is a 5 MHz signal from clock p, which is simply one of the clocks in the ensemble. The output of the phase stepper, which is also a 5 MHz signal, is inserted into the measurement hardware as clock s, just as any other clock would be. Its signal is evaluated by the algorithm exactly as described above, except that its weight is always 0, so that it is characterized with respect to the ensemble, but it has no influence on the ensemble computation.

The steering process consists of three components: the computed offset of the steered clock from the ensemble average, a correction term needed to remove this offset and an additional administrative term that is used to steer the output of the ensemble based on external data. For example, the ensemble in Boulder is steered by the use of data from the BIPM *Circular T*, and figure 2 shows the administrative steering that has been applied to the AT1 time scale to generate UTC(NIST). If needed, these administrative corrections were applied only at 0 UTC on the first day of each month before February 2002 (Modified Julian Day 52 300). A second correction was sometimes applied in the middle of the month (on the day after *Circular T* was received) after that date.

The administrative steering is implemented by means of frequency steering only, so that it is a sequence of piecewise linear functions. The mth equation is defined by

$$\mathrm{X}(t) = x_{\mathrm{a}}(m) + y_{\mathrm{a}}(m)(t - t_m), \quad (20)$$

where x_{a} and y_{a} are the mth administrative time offset and frequency offset, respectively, which are to be applied to the steered clock starting at epoch t_m. The administrative steering never uses time steps, so that the time offset is defined recursively as

$$x_{\mathrm{a}}(m) = x_{\mathrm{a}}(m-1) + y_{\mathrm{a}}(m-1)(t_m - t_{m-1}). \quad (21)$$

The phase stepper supports steering in both time and frequency. However, the primary steering is implemented by means of

[1] The magnitude of the frequency ageing term is generally not larger than $10^{-19}\,\mathrm{s}^{-1}$, and is often a factor of 10 smaller than this value. If the interval between measurement cycles is 720 s, the contribution of the frequency ageing term to the model of the time difference in equation (1) is less than 0.03 ps, which is much smaller than the measurement noise. Therefore, the contribution of the frequency ageing must be determined using data from many measurement cycles so that the contribution of the frequency ageing becomes large enough to be separated from the other contributions to the time differences. This analysis must be performed retrospectively, and this is not consistent with the real-time requirements of the time scale.

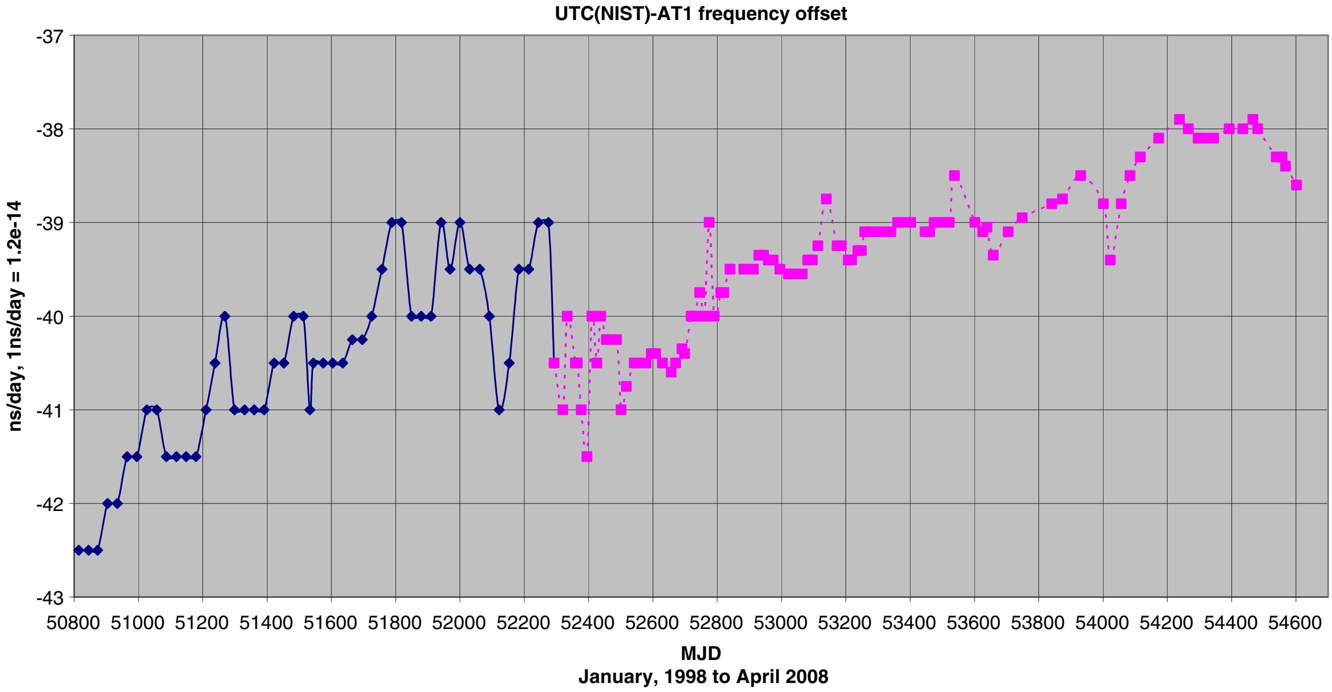

Figure 2. The steering corrections that have been applied to the AT1 time scale in Boulder to generate UTC(NIST). The offsets are given in ns day^{-1}, where 1 ns day^{-1} corresponds to a frequency of 1.2×10^{-14}. The steering corrections were applied only on the first day of every month before February, 2002 (MJD 52 300). An additional steering correction (if needed) was applied in the middle of the month after that date.

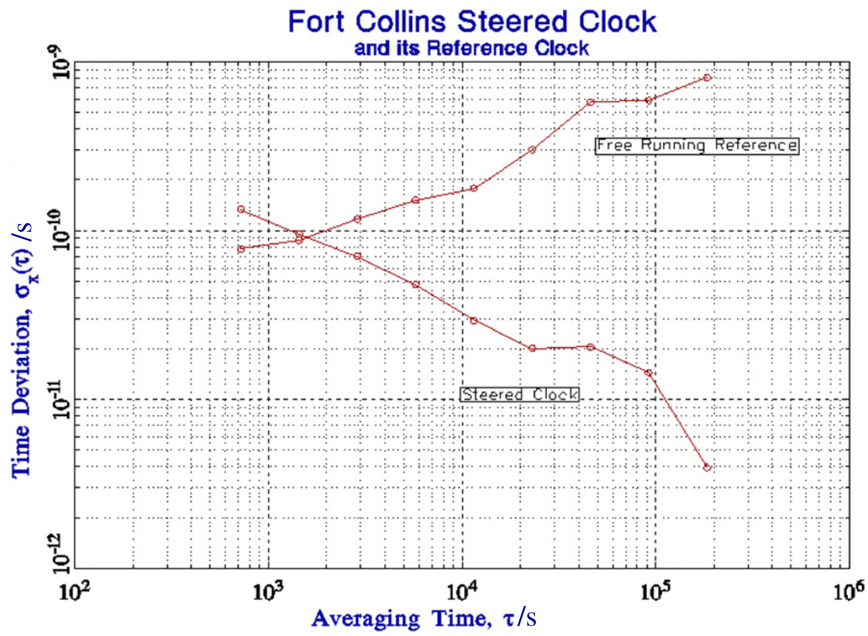

Figure 3. The TDEV of the time of the steered clock with respect to the ensemble compared with the corresponding parameter for the caesium clock that is connected as its reference oscillator. Note the excess noise of the phase stepper at the shortest averaging time (12 min). The performance of the steering system is affected by the diurnal temperature fluctuations, which are not completely cancelled in the time differences that are used to compute the scale.

frequency steering only, and time steering is limited to less than ±25 ps when the system is operating normally. (Larger values are used only following a cold start. The maximum steering rate is limited by the hardware. The typical maximum rate is 10 ns s^{-1}, so that it is possible in principle to remove a time offset of up to 7200 ns in the interval between measurement cycles. The actual maximum steering rate is limited by the software to one-third of the maximum rate supported by the hardware.)

A frequency steering command is applied to the phase stepper after the ensemble parameters are calculated every 12 min. The frequency component of the command is

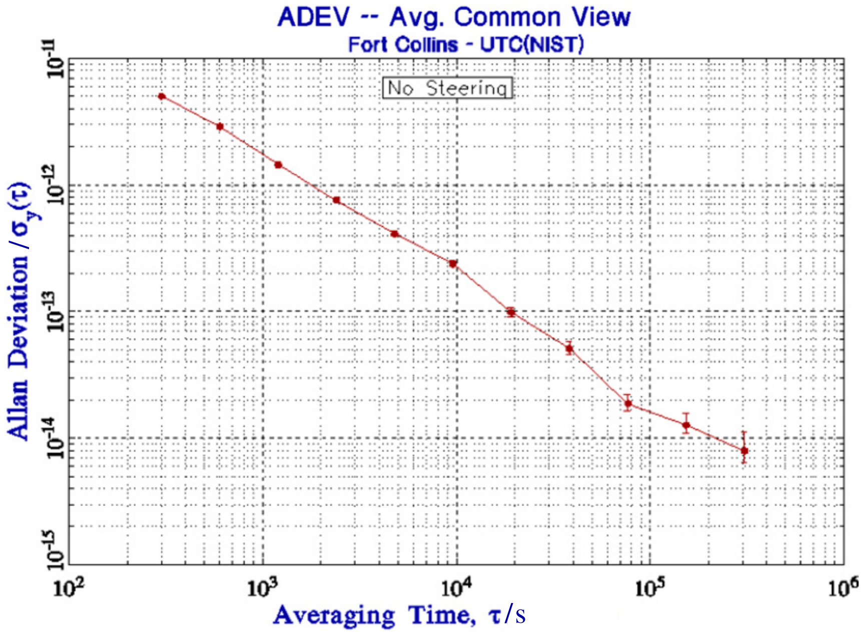

Figure 4. The ADEV of the time differences between the free-running backup time scale in Fort Collins and UTC(NIST) in Boulder. The time differences were computed by averaging the common-view time differences of all of the satellites that were tracked by both receivers at each epoch. The same data are used to compute figure 5.

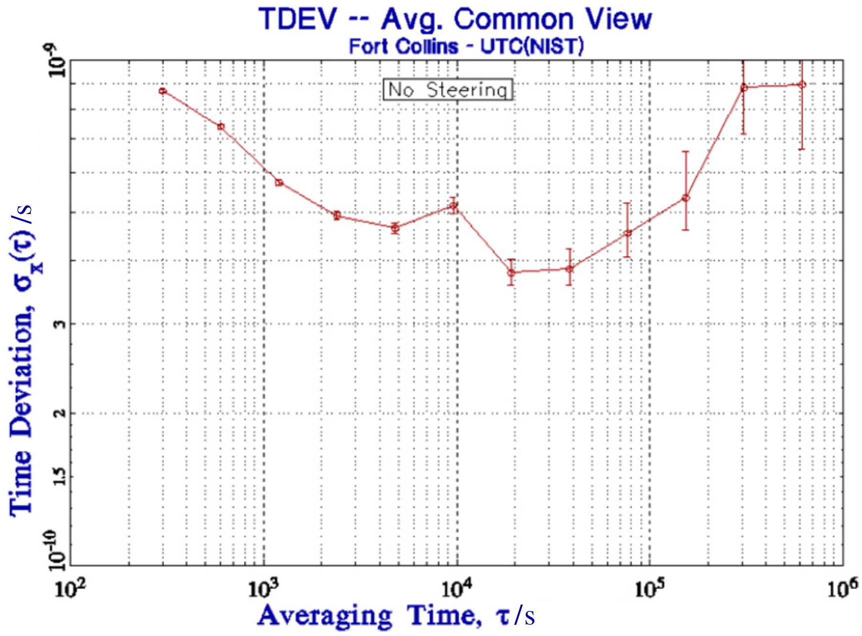

Figure 5. The TDEV of the time differences between the free-running backup time scale in Fort Collins and UTC(NIST) in Boulder. The same data are used to compute figure 4.

given by

$$\Upsilon(k) = -y_{\mathrm{pe}}(k) + y_{\mathrm{a}}(m). \tag{22}$$

The first term on the right-hand side removes the computed frequency offset of the clock that provides the input signal for the phase stepper and the second term is the appropriate administrative offset frequency for the current epoch.

The offset frequency for the Boulder ensemble can be changed up to two times per month: at 0000 UTC on the day following the receipt of a new *Circular T* from the BIPM and at 0000 UTC on the first day of the following month. Most months have only one change, and many have none. See figure 2. As I discussed above, in normal operation the administrative offset frequency for the Fort Collins system is changed no more than once per week, and the maximum adjustment at any time is limited to $\pm 5 \times 10^{-15}$.

A command to steer the time of the steered clock may also be used. The time offset of the steered clock is given by

$$\mathrm{X}(k) = -x_{se}(k) + x_{\mathrm{a}}(t), \tag{23}$$

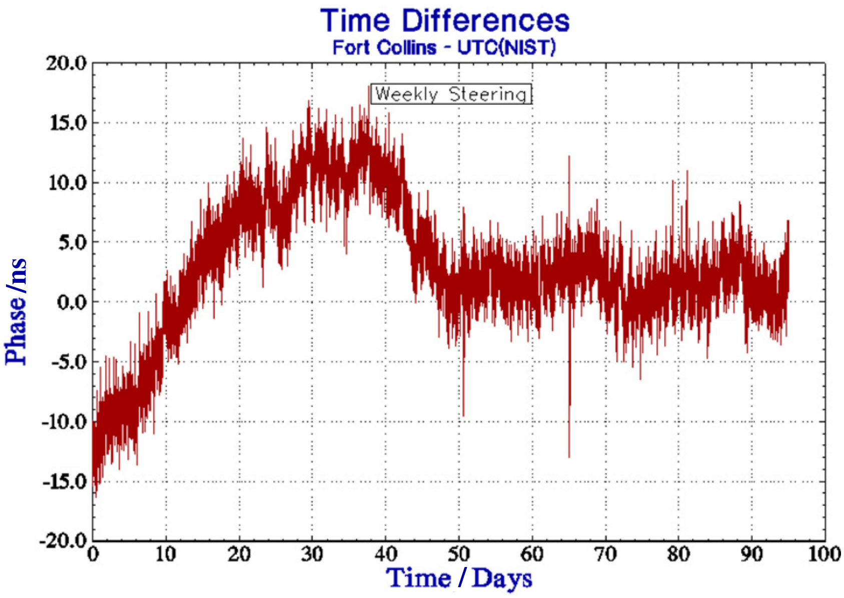

Figure 6. The time differences between the Fort Collins time scale 'a' and UTC(NIST) measured with GPS common-view and computed after the fact. The plot shows 3 months of data with a measurement point every 5 min. Each point is the average common-view time difference computed using the data from all of the satellites that were in common-view at that epoch. The frequency of the Fort Collins systems has been steered as discussed in the text.

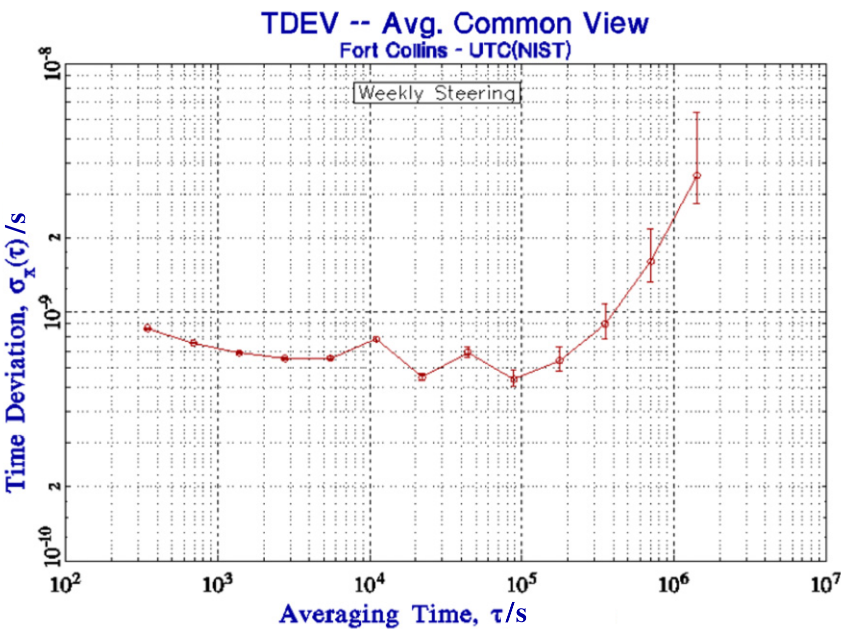

Figure 7. The TDEV of the data shown in figure 6.

where the first term is the time offset of the steered clock with respect to the ensemble as computed by the algorithm described above, and the second term is the additional administrative time offset. If the magnitude of this offset is less than 25 ps (which is the usual case), then it is applied to the phase stepper directly. If its magnitude is larger than this value, then the correction is limited in magnitude to 25 ps and the administrative frequency offset is adjusted to remove the excess with a time constant of 5 days.

The performance of the steering system in realizing the time of the ensemble is shown in figure 3, which compares the time deviation (TDEV) of the steered clock with the TDEV of the free-running caesium clock that is connected as its reference. The comparison shows the contribution of the noise of the phase stepper at the shortest averaging time.

8. Steering to the Boulder scale

The final aspect of the backup time scale is the method used to synchronize the time scale to the primary ensemble in Boulder. The communication link between Fort Collins and Boulder

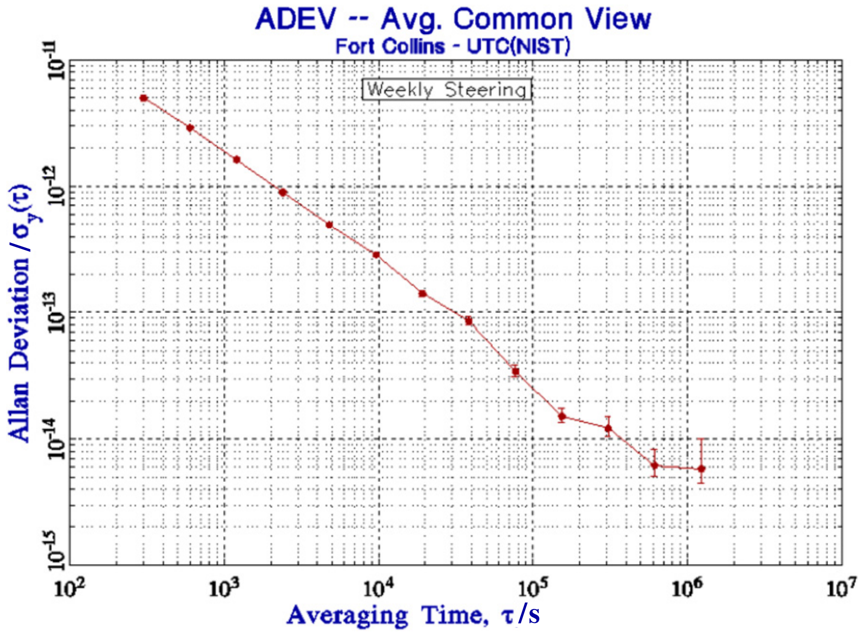

Figure 8. The ADEV of the data shown in figure 6.

Figure 9. A portion of the water bottles added to the time scale room to stabilize its temperature. The water attenuates the short-period fluctuations, but has little effect on the long-period and seasonal fluctuations.

cannot support the standard common-view method in real time, and figure 4 shows the Allan deviation(ADEV) of the average of the common-view time differences computed after the fact; figure 5 shows the TDEV of the same data. The frequency stability of the time scale for an averaging time of a few days is about a factor of five worse than the simple arguments we presented above, which were not sensitive to the common-mode fluctuations of the scale in Fort Collins due to the various environmental perturbations.

Based on these results, I have designed a steering algorithm that applies administrative adjustments to the frequency of the steered clock on a weekly basis. In order to minimize the contribution due to multipath effects, the input data to the algorithm are the integrated sidereal-day frequency differences between the Fort Collins and Boulder time scales. The sidereal-day frequency differences are computed satellite by satellite. The algorithm can use up to eight satellites at any epoch. The result is integrated to produce a time difference, which is then averaged over the day. The output is one value per day giving the average time difference between Fort Collins and Boulder. This value is computed both in Boulder and in Fort Collins. The weekly examination of these data is used to apply administrative frequency steers to the Fort Collins system, observing the administrative limits discussed above. The two independent time scales in Fort Collins often have slightly different administrative steering parameters because of the interaction between the time scale algorithm and the noise processes as discussed above.

Figure 6 shows the time difference between Fort Collins and Boulder for three months (January through March 2008) computed after the fact by means of standard common-view. Each point is a 300 s average of the common-views of up to eight satellites that were in common-view at that epoch. (The time differences are computed satellite by satellite for every satellite that is observed at both sites and these differences are then averaged.) Figures 7 and 8 show the TDEV and ADEV, respectively, computed from these data.

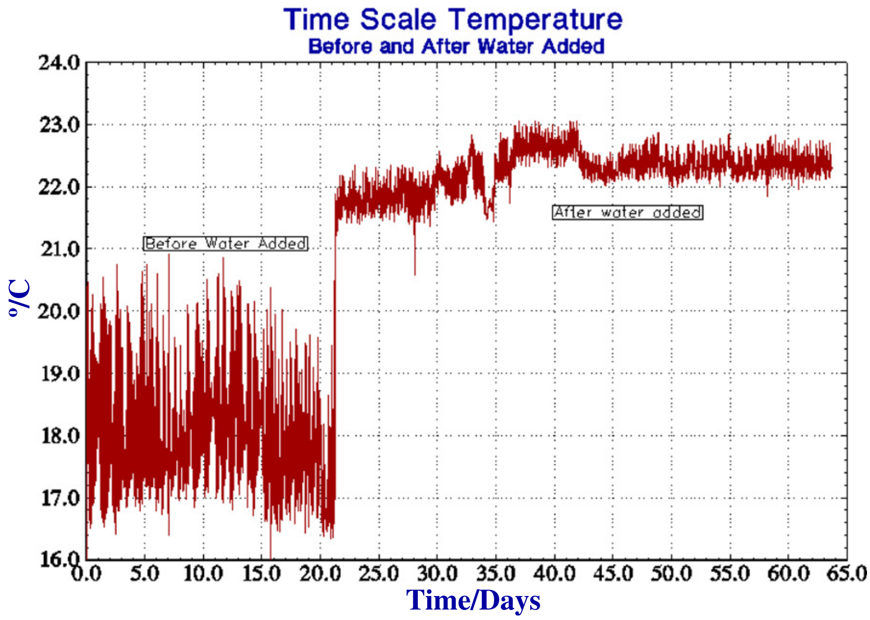

Figure 10. The temperature in the time scale room before and after the water bottles shown in the previous figure were installed. The diurnal temperature fluctuations have been reduced by about a factor of 5 to about 0.5 °C peak-to-peak.

9. Temperature control

When the hardware was first installed, the fluctuations in the environmental temperature had a significant impact on the stability of the time scale. I used a cross-spectral analysis and found that the measurement system had an admittance to temperature fluctuations of about (3 ± 2.5) ps °C^{-1}. The error bars show the approximate variation among the different channels of the measurement system. This admittance was measured with the use of the time differences between the same clock measured in two channels, so that it is not sensitive to any common-mode effects. However, such effects are less important in the measurement system, which reports only time differences. Although the resulting time dispersion is small, it interacts in a complicated way with the time scale algorithm, so that its effects are hard to judge. The temperature admittance of the GPS receivers and the time distribution system is at least an order of magnitude larger than the value quoted above, but this estimate has a large uncertainty because there are many other effects with similar spectra.

I added a large number of water bottles to attenuate the large temperature fluctuations (figure 9). Figure 10 shows the ambient temperature before and after the water bottles were installed.

10. Summary and conclusions

I have constructed a time scale at the site of the NIST radio stations near Fort Collins, Colorado, that is intended to provide a backup to the primary ensemble in Boulder if that scale becomes unavailable for any reason. The time scale is a complete, stand-alone system and can function without any external inputs. It requires no special maintenance and has functioned without attention for weeks at a time.

By means of weekly steering, the scale realizes UTC(NIST) well enough to support almost all of the NIST services. The frequency stability is about 10^{-14} with an averaging time of a few days and the time stability is a few nanoseconds over the same averaging time. These capabilities are more than adequate for controlling the transmissions of the radio stations, and for providing a reference for the Internet Time Services and for ACTS, a time service that provides time in a digital format using dial-up telephone lines, and we plan to implement backup versions of these services in the near future. The stability is marginally adequate for supporting the NIST contribution to international time and frequency coordination or for providing a reference for the most demanding users of the Global Time service. Both of these services depend on common-view GPS and two-way satellite time-difference data with long-term stabilities at the level of 1 ns or better, and this system cannot support this level of service at present. However, its short-term stability would be adequate for these tasks.

The frequency and time stabilities are somewhat worse than estimates derived from the stabilities of the hardware components that make up the system, and I attribute this degradation to the various environmental problems at the site.

References

[1] Levine J 2008 A review of time and frequency transfer methods *Metrologia* **45** S162–74
[2] Tavella P and Thomas C 1991 Comparative study of time scale algorithms *Metrologia* **28** 57–63
[3] Tavella P, Azoubib J and Thomas C 1991 Study of the clock ensemble correlation in algos using real data *Proc. 5th European Frequency and Time Forum (Besançon, France)* pp 435–41