Metrologia **42** (2005) 423–430 doi:10.1088/0026-1394/42/5/013

Operation of the NIST-F1 caesium fountain primary frequency standard with a maser ensemble, including the impact of frequency transfer noise*

T E Parker, S R Jefferts, T P Heavner and E A Donley

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

Received 17 May 2005
Published 13 September 2005
Online at stacks.iop.org/Met/42/423

Abstract
The operation of a caesium fountain primary frequency standard is greatly influenced by the characteristics of two other important facilities. The first is a stable frequency reference and the second is the frequency-transfer system. A stable frequency reference such as a hydrogen maser is a virtual necessity since essentially no fountain dead time can be tolerated without it. The frequency stability of this reference has a significant impact on the procedures for evaluating certain systematic biases in the fountain. State-of-the-art frequency transfer technology is also necessary if the fountain is intended to contribute to TAI or to be compared with other remotely located frequency standards without excessive degradation of stated uncertainties. We discuss the facilities available at the National Institute of Standards and Technology (NIST) and how they impact the operation of NIST-F1, the primary frequency standard at NIST.

1. Introduction

A systems approach is used in the operation of NIST-F1, the caesium fountain primary frequency standard at the National Institute of Standards and Technology (NIST) [1]. The fountain is of course the heart of the operation, but two other important technologies play a significant role in the details of how the fountain is operated. One is the high-stability hydrogen maser ensemble that is used by the fountain as a reference. This includes not only the individual frequency references but also the time-difference measurement equipment that allows different frequency references at NIST to be related to each other and enables the creation of a time scale. The stability of this frequency reference system, which at NIST is based primarily on an ensemble of five cavity-tuned, active hydrogen masers [2], impacts the amount of fountain dead time that can be tolerated without significantly degrading the uncertainty of the fountain measurements. It also influences how some systematic biases in the fountain are evaluated. In this report we will look in some detail at the procedures we use to evaluate the fountain spin-exchange bias and its uncertainty.

The other important technologies are long-distance time and frequency comparison techniques including GPS (Global Positioning System) common view, GPS carrier phase and two-way satellite time and frequency transfer (TWSTFT). The stability of these techniques contributes to the uncertainty of the comparison of widely separated fountains and to the uncertainty of a comparison to TAI (International Atomic Time). Uncertainty in frequency transfer generally decreases in a manner inversely proportional to the length of the time interval of the comparison, which makes long comparisons desirable [3]. Long fountain runs consequently mean that high atom densities are not required to reach an acceptable statistical (type A) uncertainty. Using low atom densities results in a smaller spin-exchange bias and therefore a smaller spin-exchange uncertainty.

Consequently, the choice of operating parameters in a caesium fountain depends on the stabilities of the available frequency reference and the frequency transfer techniques being used. This is illustrated in table 1, which summarizes the

* US government work, not protected by US copyright.

0026-1394/05/050423+08$30.00

Table 1. Summary of fractional frequency uncertainties for the June 2004 evaluation of NIST-F1.

1	NIST-F1 systematic uncertainty (type B)	3.3×10^{-16}
2	NIST-F1 statistical uncertainty (type A), includes spin-exchange uncertainty	5.1×10^{-16}
3	Combined fountain uncertainty	6.1×10^{-16}
4	Dead-time uncertainty	4.0×10^{-16}
5	Combined uncertainty including dead time	7.3×10^{-16}
6	Frequency transfer uncertainty	5.0×10^{-16}
7	Total uncertainty into TAI	8.8×10^{-16}

fractional frequency uncertainty values for a 60-day NIST-F1 evaluation that was submitted to the Bureau International des Poids et Mesures (BIPM) in June 2004 and published in the BIPM publication *Circular T*, number 198. Rows 1, 2 and 3 in the table give, respectively, the type B (systematic), type A (statistical) and quadrature sum of these uncertainties. The type B uncertainties have been discussed elsewhere [1, 4, 5] and will not be covered here. Row 4 is the dead-time uncertainty and row 5 is the quadrature sum of rows 3 and 4. The report interval for this evaluation was 60 days, but the fountain was in operation for only 34.7 days, which results in the dead-time uncertainty. Of the 25.3 days of dead time, 91% was intentional, including 7 days added before and 11 days after the fountain run in order to increase the length of the report period. This increased report period decreased the frequency transfer uncertainty. As seen in the sixth row the frequency transfer uncertainty was 5×10^{-16}, and a 60 day run was required to achieve this (see section 3). However, stretching the report period in this way increased the uncertainty due to dead time. A roughly equal level of dead-time uncertainty and frequency transfer uncertainty was chosen in order to give a total uncertainty (row 7) of 8.8×10^{-16}. This was the first report into TAI of a primary frequency standard with a total uncertainty of less than 1×10^{-15}.

2. Stable frequency reference

NIST operates an ensemble of five commercial hydrogen masers and four commercial caesium thermal beam standards to generate a real-time time scale, AT1, that is used to produce UTC(NIST). This ensemble is also used to generate a post-processed paper time scale identified generically as TP171. This scale is also known as AT1E, which is offset in fractional frequency from TP171 by a constant -483×10^{-15} [6]. Though the caesium standards help improve the long-term stability of the ensemble, it is, by far, the masers that dominate the ensemble performance. All of the clocks are maintained in environmentally controlled chambers. Optimum maser performance can be obtained only if both temperature and humidity are controlled, and the standards must also be kept in an environment with a relatively stable magnetic field [7]. The fractional frequency stability of a good maser is in the mid-10^{-16} range from about 1 day to 10 days [7]. The post-processed maser ensemble, which performs better than any single maser, plays an important role in the operation of NIST-F1.

NIST-F1 is not operated as a clock. In normal operation NIST-F1 measures the frequency of one of the hydrogen masers

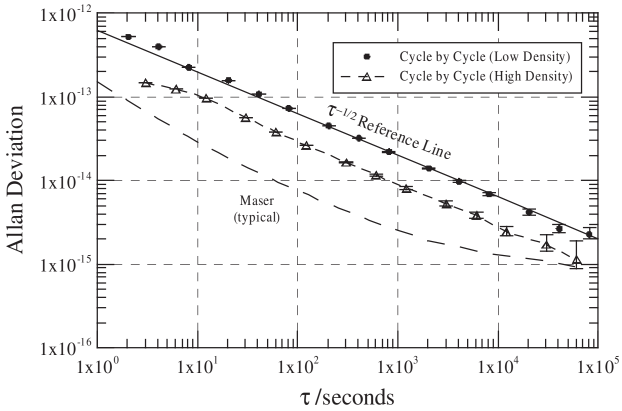

Figure 1. Allan deviation of NIST-F1 for low and high atom density as measured against a hydrogen maser. The solid line is a $\tau^{-1/2}$ reference line. The lower dashed line illustrates the noise of a typical maser used as a reference for NIST-F1.

(usually a maser with low drift and better than average short-term stability), and the output from NIST-F1 is a series of frequency-offset measurements made at intervals of every 2 s to 3 s. Barring unexpected interruptions these data are usually reduced to a 24 h average starting and ending at 0 : 00 UTC. The average maser frequency is then related to the frequency of TP171 through internal time-difference measurements. Our ability to do this depends, in part, on the stability of the instrumentation that measures the time (phase) difference of the 5 MHz signals coming from each maser and caesium standard. These clock difference data are used to create the time scale. We use a dual-mixer system and have recently upgraded the equipment. The time deviation of this equipment is at or below 1 ps out to about 10 days. The Allan deviation due to measurement noise is about 3×10^{-17} at 1 day and approaches 3×10^{-18} at 10 days. This is insignificant compared with the uncertainty of the fountain at 1 day or 10 days. Even our older measurement system, which is noisier by about a factor of 3, does not contribute significantly to the measurement uncertainty. The conversion of each of the 24 h NIST-F1 measurements of the reference maser to TP171 introduces a type A fractional frequency uncertainty of less than 2×10^{-16}. This uncertainty arises because the start or stop of the fountain measurement does not always coincide exactly with the automated clock difference measurement.

Figure 1 shows the Allan deviation of NIST-F1 versus the reference maser in both a normal low atom-density mode (1 in arbitrary laboratory units) and also at a density higher by about a factor of five. The stability of the maser itself is shown by the dashed line. The smallest value of τ in this plot is determined by the cycle time of the fountain, which is typically about 2 s or 3 s. At low density the Allan deviation of NIST-F1 is well above that of the maser for all values of τ shown. In the high-density mode the fountain approaches the stability of the maser only at large τ values. This enables us to characterize the stability of the fountain with very little impact from the masers. The white FM noise level obtained from the Allan deviation plot for each individual 24 h run is used to determine the statistical uncertainty u_i of that run, which is the Allan deviation at the run length for the ith run.

Most systematic biases and their associated uncertainties in NIST-F1 [1, 4, 5] are measured before and/or after a formal evaluation, but some can be measured during the evaluation. The high stability of our local frequency reference gives us a unique flexibility in how we measure these biases. This is particularly true of the spin-exchange bias, which will be discussed in detail here as an example of a bias measured during the formal evaluation.

2.1. Spin-exchange bias

There are several possible approaches to measuring the spin-exchange bias and its uncertainty. One is to measure the bias before and/or after the formal evaluation period and to use these numbers to correct the results of the evaluation [8]. The assumption here is that nothing has changed to affect the bias during the evaluation. Another approach is to measure the spin-exchange bias and uncertainty during the formal evaluation period, which is what we do. This minimizes the risk that something has changed and is also an efficient use of the fountain run time. There are two ways of implementing this approach. One way is to quickly alternate between different atom densities with a period of minutes to hours. This minimizes the impact of any maser noise because the cycle time is in the range where the maser noise is predominantly white FM, and consequently it averages down approximately as the square root of the number of cycles [9]. However, the mechanics of quickly changing atom densities usually limits the change in density to about a factor of two with NIST-F1. The other approach, used with NIST-F1, is to use large density changes, but this requires adjustments to the laser power, caesium oven temperature and molasses time, which may take 1 h or more to accomplish. It has been verified that these changes do not affect the velocity and spatial distributions of the atom cloud by more than a few per cent [1]. The frequency shift due to spin-exchange collisions is essentially constant with respect to variations in the collision energy at the collision energies in NIST-F1 [10]. Further, we impose a symmetric (in time) sequence of low, medium, high, medium, low density variation in the evaluation of this shift, which helps reject long-term drifts in the spatial and velocity distribution of the launched cloud. As a result, variations in the spatial and velocity distributions result in frequency uncertainties which are much smaller than the statistical uncertainty of the spin-exchange frequency bias.

In order to minimize the dead time from these large density changes they can be made only every few days. A larger density variation gives a smaller uncertainty in the frequency versus atom density slope, but the slower cycle time puts more demand on the long-term stability of the maser frequency reference. Our own independent estimates of the stability of the maser ensemble [2] indicate that it is sufficiently stable for this purpose. However, we now have enough fountain data to quantify the impact of the long-term maser instabilities on the calculation of the spin-exchange uncertainty. In principle, we could always run in the high density configuration and use the state-selection cavity to quickly reduce the atom density. However, this would require a large change in microwave power which may change the profile of the atom cloud [5].

Figure 2 shows, in chronological order, the average fractional frequency difference between NIST-F1 and TP171 for the 38 runs that made up the June 2004 evaluation.

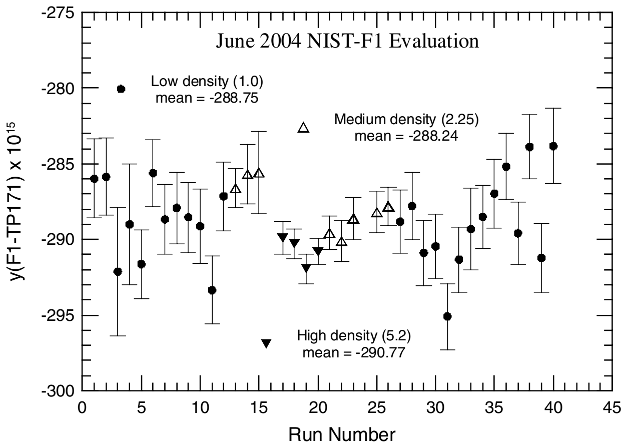

Figure 2. Series of NIST-F1 measurements (nominally 24 h each) of TP171 in chronological order and at different densities for the June 2004 evaluation. Fractional frequency units are 10^{-15}. The error bars represent the statistical (type A) uncertainty, u_i, for each run and were determined from Allan deviation data.

Most runs are 24 h in length, but a few were shorter due to intentional or unintentional interruptions. Excluding intentional interruptions, the fountain ran 94% of the time. There were a total of 40 runs, but two were not used. One was excluded because of equipment problems and the other because it was made at a high microwave power. The evaluation began with 12 runs at low density (1.0 in laboratory units). Next we made 3 runs at medium density (2.25). This was followed by four 24 h runs at high density (5.2). Finally we returned to medium density for 5 runs and low density for 14 runs. The error bars represent the statistical uncertainty, u_i, for each run and were determined from the Allan deviation data. Note the smaller error bars on the higher-density runs. Not all runs lasted a full 24 h and this accounts for some larger than normal error bars. A reasonable degree of time symmetry was used for the various densities in order to minimize the impact of any linear frequency drift in TP171. No post-processing is performed on TP171 during an evaluation unless one of the clocks malfunctions, and none was required during any of the evaluations discussed in this report. The large frequency offset between TP171 and NIST-F1 has no significance.

A weighted linear least-mean-square fit of frequency versus atom density, by means of the equations of [11], is used with data such as those in figure 2 to determine the frequency at zero atom density and its uncertainty. The weight for each point is determined from the statistical uncertainty, u_i, for the corresponding run. Figure 3 shows fractional frequency versus atom density for three NIST-F1 evaluations: December 2003, June 2004 and January 2005. For clarity of presentation all the data at each density in the individual evaluations have been combined into one point with an appropriate error bar. The fit results are shown by the three solid lines. The slopes and uncertainties are given in fractional frequency change (units of 10^{-15}) per unit of atom density. The slopes agree within their uncertainties. Note that the frequency shift from a density of 1 to 0 is only about 5×10^{-16}.

In the fitting routine that is used the uncertainties of the intercept, u_b, and slope, u_{slope}, are determined by u_i and the atom densities, x_i, [11] and not by the actual scatter of

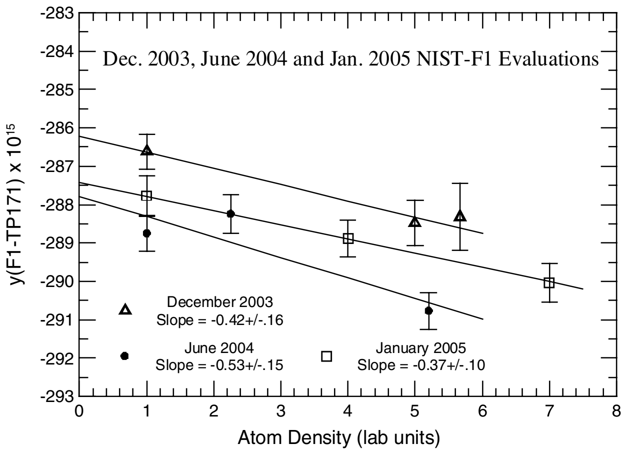

Figure 3. Weighted least-mean-square fit to atom density for three NIST-F1 evaluations. The slopes and uncertainties are given in fractional frequency change (units of 10^{-15}) per unit of atom density. The error bars represent the statistical (type A) uncertainty for each density.

the data. Note that u_i comes only from the fountain noise and not the maser noise. If the white FM noise characteristic seen in figure 1 extends out to the duration of the evaluation these calculated uncertainties will be correct. However, the uncertainties from the weighted fit will be underestimated if the noise of the maser ensemble begins to degrade the long-term stability of the measurements in figure 2. The uncertainty of the intercept, u_{b}, is the statistical (type A) uncertainty of the evaluation and is calculated using the standard expressions of [11]. u_{b} can be viewed as being made up of two components. One part, u_{stat}, is based only on the noise of the fountain and is the combined statistical uncertainty of all the runs such that

$$\frac{1}{u_{\mathrm{stat}}^2} = \left[\sum_{i=1}^{n} \frac{1}{u_i^2}\right]. \qquad (1)$$

The other part of u_{b} is the uncertainty of the spin-exchange bias, u_{e}, which is also type A [5]. This uncertainty is the product of the weighted mean atom density, D_{atom}, (each density is weighted by $1/u_i^2$) and the uncertainty of the slope, u_{slope}. The uncertainty of the intercept, u_{b}, can then be viewed as

$$u_{\mathrm{b}}^2 = u_{\mathrm{stat}}^2 + u_{\mathrm{e}}^2, \qquad (2)$$

where

$$u_{\mathrm{e}} = D_{\mathrm{atom}} u_{\mathrm{slope}}. \qquad (3)$$

For example, in the June 2004 evaluation u_{b} was 0.51×10^{-15}. This was composed of a u_{stat} of 0.27×10^{-15} and u_{e} of 0.43×10^{-15}, where D_{atom} was 2.9 and u_{slope} was 0.15×10^{-15} per unit atom density. It can be easily shown that equation (2) above (using equations (1) and (3)) is equivalent to the expression for u_{b}^2 in [11].

Figure 4 shows the residuals of the fits to atom density for the evaluations of (a) December 2003 in *Circular T* #192, (b) June 2004 in *Circular T* #198 and (c) January 2005 in *Circular T* #205. Each data point is a one-day average (in chronological order), and some short runs have been combined. These data now constitute three time series of fractional frequency values covering a total of nearly 78 days

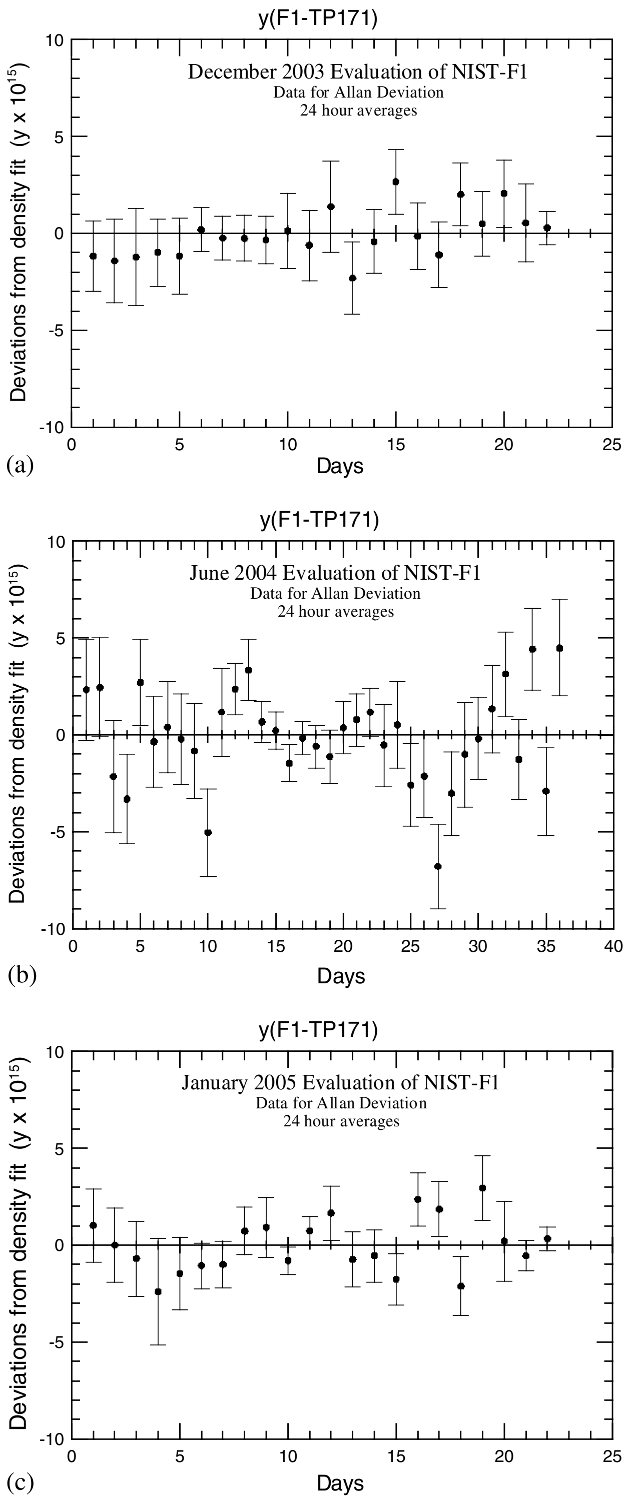

Figure 4. Residuals of weighted least mean square fit to atom density, in chronological order, for the (*a*) December 2003 evaluation, (*b*) June 2004 evaluation and (*c*) January 2005 evaluation. Each data point is nominally a 24 h average. The error bars represent the statistical uncertainty, u_i, for each run and were determined from Allan deviation data.

from which the spin-exchange bias has been removed. Though there is some dead time in the data for each evaluation, an Allan deviation calculation can be performed on each series since we are dealing primarily with white FM noise. We show

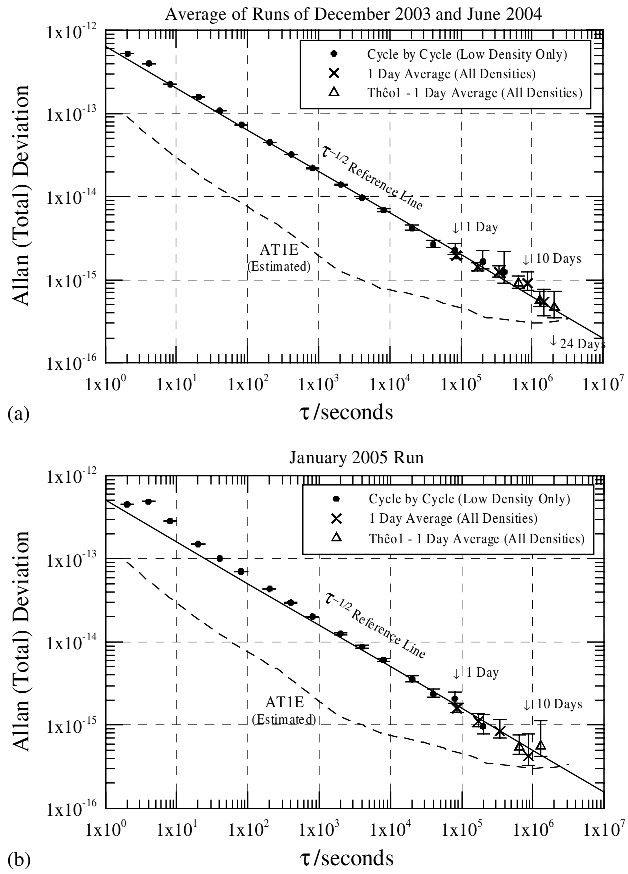

Figure 5. Allan deviation of NIST-F1 versus AT1E. (*a*) Combined December 2003 and June 2004 evaluations and (*b*) January 2005 evaluation.

data from three evaluations to illustrate the reproducibility from evaluation to evaluation. Figure 5(*a*) is a composite Allan deviation plot using 14 days of two-second low-density data (cycle by cycle measurements) from the June evaluation (solid circles), along with the daily averages from both the December 2003 and June 2004 runs. (These runs were made under nearly identical conditions and thus had very similar noise levels.) The 24 h averages include all densities (in order to minimize the dead time) and this biases the Allan deviation values slightly low (there is less frequency scatter between high density runs). However, 72% of the 24 h data is from low-density data so the bias is only about 12%, which is negligible on the scale of figure 5(*a*). The Allan deviation values from the December 2003 and June 2004 runs were averaged together to give better confidence levels at large τ (the runs could not be concatenated because of the long interval between them). TOTAL and Theo1 deviations [12] were used for the larger τ values to further improve the confidence levels. The solid line is a $\tau^{-1/2}$ reference line representing white FM noise. For comparison an estimate of the stability of AT1E (or TP171) is shown by the dashed line. Note that the stability of AT1E is better in the long term than an individual maser because the maser frequency drift is accounted for in AT1E. Figure 5(*b*) is a similar plot for the January 2005 run. Because of an improved microwave synthesizer [5] this run had a somewhat lower white FM noise level and therefore was not combined with the data in figure 5(*a*). It is shown to again illustrate the reproducibility of the noise characteristics of the evaluations. The cause of the deviation from white FM at small values of τ in figure 5(*b*) is not clear. It is related to the phase-locked-loop that locks a low noise quartz oscillator to the maser [5], but not in a simple manner. We are continuing to investigate this issue.

Two important observations to be made from figures 5(*a*) and (*b*) are (1) the noise of the fountain at low density (where most of the data are collected) is white FM all the way out to time intervals on the order of the length of an evaluation and (2) the fountain noise level is well above that of the maser ensemble except at the largest τ values. The first observation above verifies that our assumptions about the stability of the maser ensemble are correct.

A quantitative estimate of the impact of maser noise on our calculation of the uncertainty of the intercept can be obtained with a chi-squared analysis [11] of the data in figure 4. The reduced chi squared is given by

$$\chi_{\mathrm{r}}^2 = \frac{1}{d}\left[\sum_{i=1}^{n}\frac{(y_i-(sx_i+b))^2}{u_i^2}\right], \qquad (4)$$

where n is the number of data points, $d = n-2$ is the number of degrees of freedom, y_i is the ith 24 h fractional frequency average, with a frequency uncertainty of u_i, for atom density x_i. s and b are, respectively, the slope and intercept from the weighted linear least-mean-square fit. The term in the numerator in the brackets is the square of the residuals of the fit. We can define R as

$$R = \sqrt{\chi_{\mathrm{r}}^2} \qquad (5)$$

and it represents the rms ratio of the standard deviation of the fit (many days of data) to the uncertainty expected from the fountain white FM noise level measured each day. The average value of R should be 1.0 if the white noise of the fountain is the dominant noise source over the course of an entire evaluation. For the December 2003, June 2004 and January 2005 evaluations the R values were 0.94, 1.19 and 1.02, respectively. The weighted mean of the three is 1.07, where the weighting is based on the duration of the evaluations. The weighted average of R from the nine most recent NIST-F1 evaluations is 1.06. This indicates that, on average, we have been underestimating the uncertainty of the intercept by at most 6%. (The maser noise affects only the u_{e} component of u_{b}.) Given that the uncertainty of the intercept typically makes up about one quarter (added in quadrature) of the total uncertainty reported into TAI, the additional uncertainty due to the maser noise is a negligible contribution. By using a large range of atom densities (which necessitates a slow cycle time between different densities) we ultimately reduce the uncertainty due to the spin-exchange bias, but this can be done only because of the exceptionally high stability of our maser ensemble. For the June 2004 evaluation the uncertainty of the intercept would have been 70% larger if the four high density runs were replaced by medium density runs of equivalent length. The uncertainty of the slope would be more than three times larger.

Fountain stability and accuracy are likely to improve in the future, whereas there is no immediate prospect of significant improvements in maser frequency stability. Therefore, it is likely that the procedures used for determining the spin-exchange shift on NIST-F1 may not be optimal for future, improved fountains. In this case, one of the other approaches discussed above may be a better choice.

2.2. Dead time

In addition to influencing the procedures used to evaluate frequency biases, the stability of the reference masers also impacts how much dead time can be tolerated [6]. Though we have made great progress in reducing unintentional dead time in NIST-F1 [5], we can still use intentional dead time to reduce the overall uncertainty of a long-distance frequency comparison when the uncertainty of frequency transfer is taken into consideration. As will be discussed in section 3, the frequency transfer uncertainty decreases as the time interval increases. Consequently a longer evaluation gives a smaller transfer uncertainty. On the other hand, increasing dead time results in a larger value for the uncertainty of the maser frequency measurement. Figure 6 shows how the uncertainty of a 30-day maser frequency measurement increases as the amount of live time decreases (dead time increases) [6,13]. It is assumed that all of the live time falls within the 30 day interval. This curve is based on typical maser noise characteristics shown in the figure. Note that it makes a difference whether the live time is located in the centre of the 30 day interval or at either end.

The uncertainty of neither the frequency transfer uncertainty nor the dead time is known with great precision. The frequency transfer uncertainty for the purpose of calibrating TAI is based on a standard formula from the BIPM used by all laboratories when reporting primary frequency standards. It is a reasonable approximation of the actual transfer uncertainty but does not reflect the actual conditions at the time of the evaluation (see section 3). The dead-time uncertainty is perhaps slightly better known since the maser ensemble provides some indication of how each maser is behaving during the evaluation. However, the calculated dead-time uncertainty is based to some extent on the past performance of the masers. Since neither is known with great precision, it is reasonable to trade off increased dead-time uncertainty with reduced frequency transfer uncertainty to obtain the lowest total uncertainty. This occurs when the report period of a fountain evaluation is arbitrarily increased symmetrically with dead time to where the uncertainties from dead time and frequency transfer are approximately equal. This was done in the June 2004 evaluation (table 1) and to a lesser extent on all NIST-F1 evaluations. As a practical matter, no matter how long the fountain is run, it is almost always desirable for us to extend the report interval beyond the period of operation of the primary standard in order to obtain a lower overall uncertainty.

3. Frequency transfer

As mentioned earlier, the uncertainty of frequency transfer can play a significant role in the operation of a fountain. In this section a short review will be presented of the stability characteristics of various time and/or frequency transfer techniques. Frequency transfer noise is generally not a major consideration within a laboratory, where standards can be compared over relatively short distances of coaxial cable or optical fibre. However, for widely separated standards (hundreds of kilometres or more) the transfer noise can be a major contributor to the total uncertainty of a comparison.

Figure 7 shows the time deviation of a comparison between the maser ensembles at NIST and the United States Naval Observatory (USNO) using three different time (frequency) transfer techniques [14]. The three techniques are common-view GPS (single channel receiver), Ku-band TWSTFT and carrier-phase GPS. Since the maser ensembles at NIST and USNO are very quiet the values of $\sigma_x(\tau)$ represent the noise of the time-transfer systems for all three cases at values of τ smaller than a few days. The noise in common view dominates out to about 20 days, but even for TWSTFT and carrier phase, which are more stable, the transfer noise dominates over the clock noise out to about 5 days. In this figure we can see that the instabilities in common view are characterized by flicker phase modulation (FPM) (little dependence on τ) at a level of about 1 ns from just under a day out to beyond 10 days. TWSTFT is intermediate in stability at about 200 ps with evidence of a diurnal (daily) cycle at τ close to half a day. As with common view, the time deviation of two-way is predominantly FPM out to the larger values of τ, where clock noise dominates. GPS carrier phase is the most stable technique at short time intervals, but its noise increases with τ at a rate approximating random-walk PM (white frequency modulation). Though this resembles clock noise the level is too high to be caused by the maser ensembles for τ less than about 5 days. Figure 7 is just an example of the stability characteristics of these three time transfer techniques

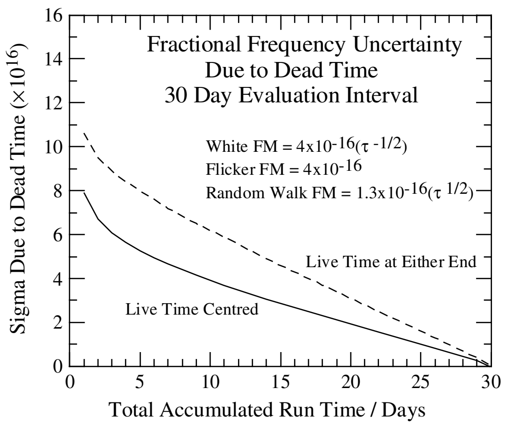

Figure 6. Measurement uncertainty due to dead time for a 30 day measurement interval.

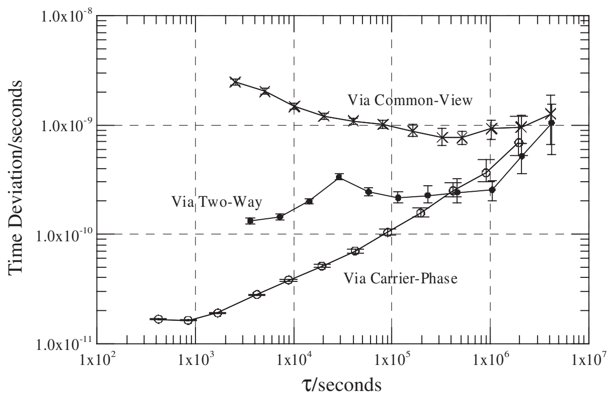

Figure 7. Time deviation between maser ensembles at NIST and USNO for different types of time transfer.

for τ less than 5 days to 10 days (beyond that the clock noise dominates). Many factors influence the stability of a transfer technique and the data in figure 7 should not be viewed as the best possible stability. The use of (1) multi-channel GPS receivers, (2) two frequency receivers that measure ionospheric delays, (3) the International GPS Service (IGS) ionosphere maps, or (4) the P-code for timing can all contribute to better stability in common view than that shown in figure 7. TWSTFT is also improving through the use of better electronics and environmental controls and, for some links, reaches the 100 ps level at 1 day. Some of the best carrier-phase results are reported in [15].

The use of more than one time and frequency transfer technique is needed to evaluate the characteristics of transfer noise at intervals longer than a few days since multiple techniques allow the elimination of the clock noise. Ideally you would like to have three independent (uncorrelated) techniques, but in practice this is very difficult to accomplish. Currently the best that can usually be done is to have the use of two (mostly independent) techniques. These typically are TWSTFT and one of the GPS based techniques. (Two-way and GPS may not be totally independent since the equipment for both techniques usually exists in the same environment, and the signals pass through the same atmosphere. However, this is generally the best that can be done.) It is even less likely that the instabilities in different GPS techniques (i.e. code versus carrier phase) are independent and at this time they should not be treated as uncorrelated. Clock noise is cancelled by observing the difference between two independent transfer techniques when they are used to compare the same pair of clocks. The resulting long-term stabilities are the combined instabilities of the two transfer techniques. Such observations indicate that transfer instabilities are generally flicker PM in nature out to as many as 100 days [3, 14, 15]. For this type of noise the Allan deviation (and hence the frequency transfer uncertainty) decreases as τ increases. Currently, the fractional frequency transfer uncertainty for TWSTFT at 1 day is in the range of (2–6) $\times\ 10^{-15}$. For GPS carrier phase the level can be as low as 2×10^{-15} and for common view it is in the range of (7–20) $\times\ 10^{-15}$. The actual values observed depend on a number of factors, including the length of the baseline, the equipment used and even the weather. Assuming that the transfer noise type stays as flicker PM beyond 1 day, simulations have shown that the uncertainty in frequency transfer (the Allan deviation) will decrease approximately as $\tau^{-0.85}$.

When directly comparing two remote fountains it is clearly desirable to use two-way or carrier phase, or preferably both, rather than common view. Even though it may not be possible to separate out the instabilities of the two individual frequency transfer techniques, a reasonably rigorous measure of the transfer uncertainty is obtained by averaging the frequency comparisons from both techniques. In this case the uncertainty of the comparison is to a good approximation one half of the Allan deviation of the combined noise of both techniques [16]. It is assumed here that the instabilities in the two transfer techniques are largely uncorrelated.

A comment is appropriate here on the use of the modified Allan deviation, Mod $\sigma_y(\tau)$, as an estimate of the frequency transfer uncertainty. In the presence of white or flicker PM noise Mod $\sigma_y(\tau)$ gives a lower value than $\sigma_y(\tau)$ and it is tempting to use Mod $\sigma_y(\tau)$. However, it must be remembered that Mod $\sigma_y(\tau)$ is obtained by averaging the phase (time) difference over τ. For example, if a frequency comparison were made over 30 days one would have to average the time difference over the first 15 days and compare it with the time difference averaged over the last 15 days in order to use Mod $\sigma_y(\tau)$ as a measure of the comparison uncertainty. Hence, it would be Mod $\sigma_y(\tau)$ at $\tau = 15$ days that is used and not $\tau = 30$ days. However, if one uses the time difference at the beginning compared with the time difference at the end without any averaging then $\sigma_y(\tau)$ at $\tau = 30$ days can be used as an estimate of the frequency comparison uncertainty. In the presence of flicker PM, $\sigma_y(\tau)$ at $\tau = 30$ days is generally a better choice than Mod $\sigma_y(\tau)$ at $\tau = 15$ days. For white PM noise Mod $\sigma_y(\tau)$ is usually the better choice. Depending on the details of the transfer noise, the optimum approach is often to average the time difference over 1 day or 2 days (assuming there is more than one data point in 1 day or 2 days) and then to use the Allan deviation [16] on these averaged data. In this situation the Allan deviation at τ will be lower than Mod $\sigma_y(\tau)$ at $\tau/2$ as long as τ is less than about 60 days.

One problem with using $\sigma_y(\tau)$ or Mod $\sigma_y(\tau)$ to estimate frequency uncertainty is that both are second difference statistics. They are therefore insensitive to a linear rate offset in the time difference data. This is highly desirable for clock differences, but in situations where clock differences have been removed these statistics will not observe some real frequency errors in frequency transfer techniques [16]. In situations where the difference between two transfer techniques is available, a first difference statistic would be a better choice since it will see slow processes that look like a linear rate offset, which represents a real frequency error. As long as the noise characteristics of the transfer techniques are well-behaved power law processes, Mod $\sigma_y(\tau)$ or $\sigma_y(\tau)$ give good estimates of frequency uncertainty, but this is not always the case. Sometimes the estimates from Mod $\sigma_y(\tau)$ or $\sigma_y(\tau)$ can be too low [16].

Frequency (time) transfer into TAI (as compared with a direct comparison between two remote fountains) is complicated by the fact that TAI does not exist in a single physical location. TAI is a paper time scale that is calculated from clock data supplied by about 46 laboratories around the world. Various time transfer techniques including GPS common view and TWSTFT are used to transfer the clock data. However, by calculating the stability of NIST's AT1E versus TAI one can get some insight into what the magnitude of the frequency and time transfer instabilities are for getting NIST-F1 data into TAI without having to know the transfer characteristics of each of the links. Figure 8 shows the Allan deviation, $\sigma_y(\tau)$, and time deviation, $\sigma_x(\tau)$, of AT1E versus TAI calculated for the 1000 days prior to 28 February 2005 (data on TAI are obtained from the BIPM publication *Circular T*). Clock noise dominates beyond about 30 days, but for smaller τ time/frequency transfer noise begins to dominate because AT1E and the clock ensemble in TAI are very stable. For τ less than 20 days we again see that the time deviation is flicker PM in nature at a level just below 1 ns. This is a clear indication that we are observing transfer noise and not clock noise. The solid line in figure 8 represents the expression $u_{\text{link/TAI}} = 3 \times 10^{-14}/\tau$ (for τ in days) currently

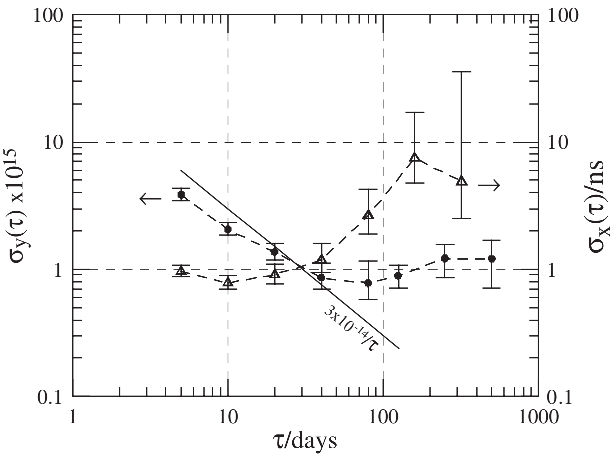

Figure 8. Stability of AT1E versus TAI. The standard uncertainty for frequency transfer into TAI used by the BIPM is shown by the solid line.

used by the BIPM for the uncertainty of fractional frequency comparisons with TAI. We see that it is a little high with respect to the AT1E minus TAI Allan deviation in the range of τ from 5 days to 10 days. (Recent improvements in many time transfer links may be resulting in a gradual improvement in frequency transfer into TAI.) Using the BIPM expression a report duration of 30 days is required to reduce the frequency transfer uncertainty to 1×10^{-15}. If flicker PM characteristics persist beyond about 10 days then $\sigma_y(\tau)$ for frequency transfer will decrease approximately as $\tau^{-0.8}$ and the current BIPM estimate at 30 days and 60 days will be close to being correct.

4. Summary

The successful operation of a caesium fountain primary frequency standard requires a significant amount of time and frequency infrastructure, and the nature of this infrastructure impacts on the details of how the fountain is operated. When frequency comparisons are made between remote fountains, or when the fountain is compared with TAI, one must take into account the stabilities of both the local frequency reference and the frequency transfer process if a minimum total uncertainty in the comparison is to be achieved. This requires a thorough understanding of the instabilities in these systems. A high-stability local frequency reference allows improved evaluation of systematic biases, and careful use of intentional dead time can be made to minimize the total uncertainty of a long-distance comparison.

Acknowledgments

The authors gratefully acknowledge Jim Gray, Judah Levine, Trudi Peppler, Marc Weiss and Victor Zhang, who are responsible for the time scale and the time and frequency transfer systems at NIST.

References

[1] Jefferts S R *et al* 2002 Accuracy evaluation of NIST-F1 *Metrologia* **39** 321–36
[2] Parker T E 1999 Hydrogen maser ensemble performance and characterization of frequency standards *Proc. 1999 Joint Meeting of the IEEE Int. Frequency Control Symp. and the European Frequency and Time Forum (Besançon)* pp 173–6
[3] Parker T E, Hetzel P, Jefferts S R, Weyers S, Nelson L M, Bauch A and Levine J 2001 First comparison of remote cesium fountains *Proc. 2001 IEEE Int. Frequency Control Symp. (Seattle, WA)* pp 63–9
[4] Heavner T P, Jefferts S R, Donley E A, Shirley J H and Parker T E 2005 Recent improvements in NIST-F1 and a resulting accuracy of $\delta f/f = 0.61 \times 10^{-15}$ *IEEE Trans. Instrum. Meas.* **54** 842–5
[5] Heavner T P, Jefferts S R, Donley E A, Shirley J H and Parker T E 2005 NIST-F1: recent improvements and accuracy evaluations *Metrologia* **42** 411–22
[6] Parker T E 2001 Comparing and evaluating the performance of primary frequency standards: impact of dead time *Proc. 2001 IEEE Int. Frequency Control Symp.* pp 57–62
[7] Parker T E 1999 Environmental factors and hydrogen maser frequency stability *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **46** 745–51
[8] Weyers S, Hübner U, Schröder R, Tamm C and Bauch A 2001 Uncertainty evaluation of the atomic caesium fountain CSF1 of the PTB *Metrologia* **38** 343–52
[9] Pereira Dos Santos F, Marion H, Bize S, Sortais Y, Clairon A and Salomon C 2002 Controlling the cold collision shift in high precision atomic interferometry *Phys. Rev. Lett.* **89** 233004
[10] Leo P J, Julienne P S, Meis F H and Williams C J 2001 Collisional frequency shifts in ^{133}Cs fountain clocks *Phys. Rev. Lett.* **86** 3743–6
[11] Bevington P R 1969 *Data Reduction and Error Analysis for the Physical Sciences* (New York: McGraw-Hill)
[12] Howe D A and Peppler T K 2003 Very long-term frequency stability estimation using a special-purpose statistic *Proc. 2003 Joint Meeting of the IEEE Int. Frequency Control Symp. and the European Frequency and Time Forum (Tampa, FL)* pp 233–8
[13] Douglas R J and Boulanger J S 1997 Standard uncertainty for average frequency traceability *Proc. 11th European Frequency and Time Forum (Neuchâtel)* pp 345–9
[14] Parker T E, Zhang V S, McKinley A, Nelson L M, Rohde J and Matsakis D 2002 Investigation of instabilities in two-way time transfer *Proc. 2002 Precise Time and Time Interval Meeting (Reston, VA)* pp 381–90
[15] Larson K M, Levine J, Nelson L M and Parker T E 2000 Assessment of GPS carrier-phase stability for time-transfer applications *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **47** 484–94
[16] Parker T E, Howe D A and Weiss M 1998 Accurate frequency comparisons at the 1×10^{-15} level *Proc. 1998 IEEE Int. Frequency Control Symp. (Pasadena, CA)* pp 265–72