APPENDIX A

THE IMPACT OF THE HP 5071A ON INTERNATIONAL ATOMIC TIME

David W. Allan, Allan's TIME, Fountain Green, UT
Alex Lepek, Tech Projects, Jerusalem, Israel
Len Cutler and Robin Giffard, HP Labs, Palo Alto, CA
Jack Kusters, HP SCD, Santa Clara, CA

Abstract

The international clock ensemble, which contributes to the generation of International Atomic Time (TAI and UTC), has improved dramatically over the last few years. The main change has been the introduction of a significant number of HP 5071A clocks. Of the 313 clocks contributing to TAI/UTC during 1994, 94 of these were HP 5071As. The environmental insensitivity of the HP 5071A clocks is more than an order of magnitude better than that of previous contributing clocks. This environmental insensitivity translates to outstanding long-term stability — with a typical flicker floor of a few $\times 10^{-15}$***. In addition, there are now several hydrogen masers with cavity tuning contributing to TAI/UTC. These not only have outstanding short-term stability, but comparatively low frequency drifts and excellent intermediate-term frequency stability.***

By analyzing the data available from the international ensemble, we have obtained two important results. First, the frequency stability obtainable with an optimum algorithm is about 10^{-15} ***for both the intermediate and long-term regions. It could be as good in the short-term if time transfer measurement instabilities were reduced sufficiently. Second, with cooperation, this performance can be made available on an international basis in near real time. The recent enhancements in the contributing clocks are already providing a significant improvement in the accuracy with which UTC is made available to the world from several of the national timing centers, such as NIST and USNO.***

Introduction

The accuracy improvement rate in atomic frequency standards has been about one order of magnitude every seven years since the first one was built in 1948. Notable events along the way have contributed to that improvement, and more are anticipated that should significantly enhance frequency accuracy. The most recent events have been the operational establishment of NIST-7 optically-pumped primary cesium-beam frequency standard with an accuracy of 1×10^{-14}. Recently, a cesium fountain primary frequency standard was reported to have an accuracy of 3×10^{-15}. The development of the new HP 5071A cesium-beam commercial frequency standard is, as will be shown, a major contributor after its introduction a few years ago.

New frequency standards are imminently expected that could provide accuracies in the vicinity of 1×10^{-15}. One of the purposes of this paper is to suggest techniques for evaluation and comparison of these standards[1]. The standards will often be remote from each other and there will be a local reference adequate to check their performance. The results reported herein should assist substantially toward improving international time and frequency metrology.

The HP 5071A clocks have demonstrated an improvement of more than an order of magnitude in both accuracy and in environmental insensitivity over previous commercial clocks. The excellent long-term stability of these clocks along with the excellent intermediate- and short-term stability of the several contributing hydrogen masers with cavity tuning provides the opportunity of having a combined reference clock with a stability of about 1×10^{-15} for averaging times, τ, ranging from about 1,000 seconds out to the order of a year[2].

Though the above stability is only available in theory, by analyzing the data publicly available from the BIPM ensemble, we have obtained two important results. First, the potential frequency stability obtainable in practice with an appropriate algorithm is about 1×10^{-15} for both the intermediate and long-term regions. It could be as good in the short-term if time transfer measurement instabilities were reduced sufficiently. Second, with cooperation, this performance can be made available on an international basis in near real time. The paper will also demonstrate how 1×10^{-15} is available by using post processing. Some guidelines and suggestions are made in the paper on how a 1×10^{-15} stable real-time frequency reference could be constructed.

The rate of TAI/UTC is syntonized with the primary cesium-beam frequency standards in the world in accordance with the SI (System International) definition. The data analyzed for this paper cover 1991 through 1994, and, during this period, there were three primary standards in continuous operation as clocks. All the contributing clocks act as "flywheels" in preserving the rate or length of the SI second as given by the primary standards. As the rate of TAI/UTC has instabilities, algorithmic procedures are chosen to maintain syntonization within some reasonable limit.

The length of the second is the same for TAI and UTC. In other words, they are perfectly syntonized in the way they are constructed. Their times differ by an exact number of seconds, which changes about annually as leap-seconds are used to steer UTC to stay within 0.9 seconds of UT1 (earth time). As of 1 January 1996, TAI - UTC = 30 s.

Performance of Individual Clocks Contributing to UTC

In this section, the frequency accuracy of the contributing clocks is analyzed. This is important for TAI/UTC, because frequency accuracy almost always translates to long-term frequency stability resulting in the preservation and perpetuation of a best estimate of the SI second.

Factors Contributing to Frequency Accuracy

The goal in making a primary cesium-beam frequency standard is to determine all effects that cause the physical output frequency of the standard to depart from the definition. These effects

are evaluated and removed either 'on paper' or by frequency synthesis in order to provide an estimate of the SI second. The uncertainties introduced by these effects combine to determine the accuracy of the standard.

The factors contributing to cesium-beam frequency standard inaccuracy are: magnetic field, electric field, phase shifts in and between the microwave interrogation regions, velocity of the cesium beam, detuning of the cavity, interference from adjacent quantum transitions to the wanted ground-state transition, stray microwave radiation seen by the atoms, and imperfections in the associated electronics. The current reported accuracy of the second for TAI/UTC is 1×10^{-14}, which is the number reported for NIST-7[3]. The value for the new cesium fountain frequency standard should be included soon.

To obtain accuracy, both the size of the frequency offset caused by each of the above contributors to inaccuracy must be determined as well as the uncertainty associated with this estimate. In addition, for an operating clock to be long-term stable these frequency offsets cannot change with time. That is a difficult challenge in the design of a clock, and has occupied large amounts of effort. Hence, having better operational accuracy almost always guarantees better long-term stability. Having such operational accuracy is one of the main advantages of the HP 5071A.

The accuracies of the commercial clocks contributing to TAI/UTC can be further improved. Two fundamental systematic frequency offsets in well designed and constructed cesium-beam frequency standards (besides that due to the magnetic field (C-field)) are:

1) the relativistic offset due to beam velocity, also known as the second order Doppler shift, and,

2) the offset due to end-to-end microwave cavity phase shift.

The second-order Doppler shift depends on cesium-beam velocities and is typically 1 to 3×10^{-1}. It can be estimated within 20 to 30% using the velocity calculated from the measured atomic-resonance line-width. If the velocity distribution of the detected beam were known in detail, then the second-order Doppler shift could be calculated accurately. The offset due to end-to-end cavity phase shift can be measured in principle by sending the beam through the cavity in the opposite direction, but this is impractical in commercial standards. Both second order Doppler and the end-to-end cavity phase shift offsets, however, lead to small asymmetries in the atomic-resonance line shape. These are independent and additive in lowest order terms. If the velocity distribution of the beam is known, the asymmetry due to second-order Doppler can be calculated and subtracted out. The residual asymmetry can then be attributed to end-to-end phase shift and the resulting frequency offset can be calculated. This is only possible if asymmetries due to neighboring quantum mechanical transitions are negligible.

One of the authors (Cutler) has developed an accurate technique for determining the velocity distribution of the detected beam in the HP 5071A from a measurement of the beam current versus applied microwave frequency while the standard is off-line. An iterative technique is used to solve the integral equation for the velocity distribution assuming rectangular microwave distribution in the cavity ends. The detailed shape of the excitation, however, is not critical in this technique for determining the velocity distribution. Results appear to be accurate to

better than one percent. In the HP 5071A, the microwave interrogation technique permits one aspect of the total line asymmetry to be measured while the standard is in operation. This allows a long averaging time to reduce the measurement noise to an acceptably low value. The only equipment required is a personal computer with appropriate software tied to the standard through its RS232 port. This overall technique should allow both offsets to be calculated accurately enough so that the absolute frequency of the cesium standard will be known to within 1×10^{-13}. This is an improvement of an order of magnitude to the current specification, and could effectively turn all of the HP 5071As contributing to TAI/UTC into primary standards.

Histogram Study of Frequency Accuracies

In this sub-section, only the most recent data available (for the year 1994) from the BIPM for the clocks contributing to TAI/UTC were used. The frequency reference used was the length of the second for TAI/UTC. The data are reported in histogram form to simplify the presentation. For convenience of comparison, all values in this subsection are given in units of 1×10^{-15}. For histogram Figures 1 through 3a the units for the bins are 500×10^{-15} and 100×10^{-15} for Figures 3b and 4.

The EAL - clock(k) data were provided by Mr. Jacques Azoubib of the BIPM for the years 1991 through 1994, inclusive. The data listed each clock contributing to TAI/UTC against EAL, which is the BIPM free time scale. As needed these data were related to TAI/UTC as well as to the SI second as given by the primary frequency standards. During 1994 the frequency difference y(EAL) - y(TAI) was 740×10^{-15}.

The average frequency of TAI/UTC for the year 1994 was -3×10^{-15}. However, the frequency calibration data supplied by the PTB did not reflect the black-body radiation correction. There were no steering corrections applied to TAI/UTC during this year. In the BIPM annual report for 1994 the listed uncertainty for the SI scale unit for TAI/UTC is 20×10^{-15}. The deviations of the scale unit with respect to the weighted average of the primary standards stayed well within this uncertainty. One may note that the mean frequency of the primary frequency standards is not equal to the weighted mean used as the BIPM best estimate of the SI second.

Figure 1 shows the frequency distribution of all the contributing clocks during 1994. In partial agreement with the skewness, the mean was -174×10^{-15}. The standard deviation was 1050×10^{-15}, and the standard deviation of the mean (estimated error of mean) was 60×10^{-15}. The standard deviation of the mean, under the assumption of independence among the standards, is computed in the usual way by dividing the standard deviation by the square root of the number of clocks. This assumption may not be valid for commercial clocks as there may be frequency biases that are not adequately dealt with in the production process. In Figure 1, for example, the mean is nearly three times the standard deviation of the mean.

Figure 2 is a similar plot for the contributing hydrogen masers. The abscissa scales have been kept the same for Figures 1 through 3a for convenience of comparison. The mean, the standard deviation and the standard deviation of the mean were, respectively, -130, 319, and 52×10^{-15}. It must be noted that the hydrogen masers are generally calibrated and are not independent frequency sources better than their intrinsic accuracies of about 1000×10^{-15}.

Figure 3a is the histogram for the 94 Hewlett Packard model 5071A standards – including both the high-performance option as well as the standard performance units. These two options were studied separately, but the distribution curves did not seem to be significantly different. The mean, the standard deviation and the standard deviation of the mean were, respectively, 48, 131, and 14×10^{-15}. For the third time, the assumption of independence may not be valid as the mean differs by more than 3.5 times the standard deviation of the mean. However, in this case the standard deviation of the mean is of the same order as that given by the best primary frequency standards in the world. The published accuracy specification for HP 5071A is 1000×10^{-15}, and the mean is 20 times better than the specification.

Figure 3b is the same data as in Figure 3a, but plotted with a different abscissa to compare with the distribution of frequencies as given by the laboratory primary frequency standards shown in Figure 4. In the case of the primary standards, the mean, the standard deviation, and the standard deviation of the mean are, respectively, 32, 57, and 33×10^{-15}. Though the sample size is only three, the standard deviation of the mean is consistent with the mean. This result, however, is somewhat artificial since the second for TAI is steered to be in agreement with the primary standards.

If future experiments confirm the theory that allows the HP 5071A to be considered a primary frequency standard, this will considerably increase the data base for Figure 4. In addition, the mean, the standard deviation, and the standard deviation of the mean could be decreased dramatically for the data shown in Figure 3. It should not be anticipated that the accuracy of N standards will improve the combined accuracy by $1/\sqrt{N}$ because the frequency inaccuracies among the standards may be correlated.

Independent Estimates of Frequency Instabilities

The same data from the international clock ensemble was used to calculate the stability, $\sigma_y(\tau)$, for each of the Hewlett Packard 5071A high performance cesium beam frequency standards over the year 1994. Since each standard contributes at most only a few percent to the computation of TAI/UTC, the resulting stabilities will be optimistically biased at most by a few percent. The results for 78 of the HP 5071As are shown as $\sigma_y(\tau)$ scatter diagram in Figure 5. Four of the 82 clocks were not used because there was either insufficient data or the data were pathological. The BIPM reports data every 10 days, so calculations were made for $\sigma_y(\tau)$ for each clock for frequency averaging times, τ, of 10, 20, 40, 80, and 160 days. The confidence of the estimate for $\sigma_y(\tau)$ (τ = 160 days) of each clock is poor due to the small number of data samples available at this averaging time. The maximum overlapping technique was used in analyzing the data to obtain a good confidence of the estimate[4].

Figure 5 also demonstrates the great improvement in the stability of TAI/UTC with the inclusion of the HP 5071As, since each clock is relatively independent of the international set of clocks. Over the four year period, while the HP 5071As were being added into the computation of TAI/UTC, the stability performance of international timing has improved well over an order of magnitude.

Figure 6 shows the RMS/$\sqrt{N}\sigma_y(\tau)$ values for the 78 clocks. These values give an estimate of ensemble performance under the assumption that the clocks are about the same in their stability

performance. Figure 5 shows a fairly wide distribution of stabilities. Hence, the RMS/$\sqrt{N}$ values will be a pessimistic estimate. As can be seen, the flicker floor is about 8×10^{-16}, which is an order of magnitude better than the stability of TAI/UTC prior to 1991.

Figure 6a also shows an optimum weighted estimate for the ensemble for each τ value. In other words, an optimum weighted combination of the clocks could not be better than these values. The degree to which these values can be approached will be both a function of the algorithm employed as well as of the consistency of the stability behavior of each contributing clock. There is nothing with which to measure this performance. The best laboratory clocks in the world have not demonstrated this level of long-term stability. The problem of measuring the performance will be dealt with in the next section.

Some additional very important messages from this data are: the stability performance of EAL is very impressive – at about 1×10^{-15} in the long-term. Several of the HP 5071As perform at similar levels. The diversity of the long-term frequency stability of these standards varies by more than an order of magnitude. Hence, an optimum weighting approach for combining their readings is essential to take advantage of and to properly utilize such diversity. This is illustrated in the next section.

A long-term frequency drift exists on several of the clocks contributing to EAL. If not properly dealt with, these drifts could cause significant long-term instabilities in EAL. Fortunately, the ALGOS algorithm used in generating EAL de-weights drifting clocks.

Ensemble Performance of Contributing Clocks

It is well known that combining algorithmically the contributing-clock readings in an optimum way has several distinct advantages. The computed time can have better stability than any of the contributing clocks both in the short-term and in the long-term. Detection of and immunity against errors of individual contributing clocks is part of the process. The ideal algorithm should have adaptive characteristics so as to respond to improvements or degradations in the individual contributors – gradual or otherwise. The better a clock performs, the better it must perform or it will get de-weighted in the algorithm's computation. In other words, each clock's errors are tested each measurement cycle; if the errors are consistent with its weight, that error value is used to perpetuate its weighting factor. If the error is too large, the clock is rejected. If the error degrades with time, the weighting factor is degraded. If the error improves with time, the weighting factor increases. It can further be shown that even the worst clock enhances the algorithm's output and adds robustness to the ensemble time calculation. The algorithm generates a real-time estimate of the figure of merit of each clock, which is not only used in the optimum weighting procedures, but also as a diagnostic measure. Knowing the weighting factors for all the contributing clocks provides the necessary information to calculate an estimate of the ensemble's performance against a perfect clock. The algorithm needs to handle the addition and removal of clocks optimally.

An often overlooked point is that if an optimum weighting procedure is not used, then the worst clocks in the ensemble contribute adversely. This effect has been observed very dramatically over the last few years with TAI/UTC. The ALGOS algorithm generating these time scales has

an upper-limit of weight, which is arbitrarily set. In the past, because it was set too low, the worst clocks contributing to TAI/UTC caused a significant annual term to be introduced. As the upper-limit of weight is increased, the annual term in TAI/UTC decreases. This is because ALGOS weightings approach the theoretical ones for the HP 5071As, which are environmentally insensitive and have almost no detectable annual terms.

Since each clock, in principle, contributes to the algorithm's output, if that clock is compared with the output, it is being compared, in part, with itself. This biases the measured stability toward zero. Carefully designed algorithms can remove these biases -- prohibiting the best clock from getting too much weight or from taking over the time scale's output. Not accounting for such biases would make the time scale less robust.

Such algorithms have been tested and utilized for many years with significant success. Taking advantage of the marked improvements in the contributing clocks to TAI/UTC and of appropriate algorithms allows the significant reference frequency improvements reported in this paper.

The results documented in this paper lead to another significant potential improvement in international time and frequency metrology. Heretofore, UTC has only been available about a month or two after the fact. The procedures outlined herein provide a real-time estimate of UTC at an accuracy level of better than 10 ns along with also providing a real-time stable frequency reference good to the order of 1×10^{-15}.

Because the character of clocks contributing to TAI/UTC has changed so dramatically over the last few years, using adaptive algorithms is very important. The basic algorithm used for this paper is based on the AT1 algorithm, initially written in 1968 for the NBS time scale system. Updates and revisions of this algorithm have been made over the years and a significant level of experience and improvements have now been obtained with these approaches to time keeping. A PC version of this algorithm with further adaptive characteristics has been written and is now employed in the generation of the Israeli time scale, UTC(INPL). The results shown below are the output of this PC version of the algorithm[5, 6].

Independent Estimate of Stability of Algorithm Outputs

Because of the different character of the clocks, they were divided into four different groups:

- 1) the primary standards contributing to the definition of the SI second;
- 2) the hydrogen masers;
- 3) the HP 5071A standards, which just became available in 1991; and
- 4) all of the rest of the clocks which contributed over this four year period.

The number of clocks that participated in each of the four sets for the following analysis were: 3 primarys, 40 masers, 80 HP 5071As and 100 other clocks. Each group of clocks was used to produce an ensemble using the algorithm described above.

Because the number of HP 5071A standards available during the beginning of the data period was small, the best stability results were obtained by excluding the first part of the data.

There are three contributions to the variations in the data:

i. the noise of the individual clocks being measured,

ii. the measurement noise or transmission or processing errors, and

iii. the noise of the reference time scale, in this case, EAL (which is TAI without steering).

Since the measurements made at each site to provide the data are made at the same time, the reference (EAL) was subtracted out and comparisons made between the individual clocks – leaving only the first two kinds of contributions. The measurement noise is non-negligible and contributes noticeably at the shorter times as evidenced by the $1/\tau$-like slope at $\sigma_y(\tau = 10days)$. The performance at longer times is affected by apparent phase jumps, some of which were removed in the original data and some in the data processing. Probably, not all were caught. In addition, some of the clocks showed frequency drifts which were not removed and this shows up as a τ^{+1} dependence at the longest sample times in a $\sigma_y(\tau)$ or Mod.$\sigma_y(\tau)$ stability diagram.

In order to obtain an independent estimate of stability, the three-cornered hat technique was used: $\sigma^2(i) = (\sigma^2(i,j) + \sigma^2(i,k) - \sigma^2(j,k))/2$, where i, j and k represent three independent clocks. This technique has the disadvantage that the worst of the three can be observed with the best confidence, and the best of the three has the worst confidence of the estimates. The longer the data length the better the estimates of stability; the last 710 days of data was used. Because of the above mentioned disadvantage the "other" clocks were not used. Figure 6b is the Mod.$\sigma_y(\tau)$ plot of the results of this analysis. Mod.$\sigma_y(\tau)$ was used because the measurement noise is significant. These independent estimates are consistent with the estimates in Figure 6a for the reasons outlined above and documenting that the long-term stability of the HP 5071A ensemble is best, followed by that of the hydrogen masers and then the primary cesium clocks. The primary clocks undergo some disturbances in order to maintain their accuracy. This may contribute to the long-term instabilities, but eventually such disturbances should average out and there is indication in the Mod.$\sigma_y(\tau)$ slope as τ increases that this is the case.

Real-Time Estimate of Time and Frequency

Given that the optimally combined frequency stability of the clocks contributing to TAI/UTC is about 1×10^{-15} for averaging times longer than about two hours, there are two basic problems in making this available in real-time at any location desired on the earth. First, the current time and frequency transfer techniques are inadequate to sustain this level of performance for either the short-term (seconds) or the intermediate-term (days) stability regions. Only in the long-term (months and years) are the comparison methods adequate. This inadequacy problem will be addressed in the next section. Second, TAI/UTC is calculated more than one month after the fact; hence, to have a real-time traceable reference to UTC requires prediction to the current time over an interval of about a month and a half.

The optimum predictor in the presence of white-noise FM (the classical noise for cesium-beam clocks) is to use the last time available from the clock, and the mean frequency over the life of the clock as the rate with which to predict forward. Because of the excellent environmental

immunity of the HP 5071As, the white-noise FM model fits over a very large region of prediction times ranging from about 10 seconds to about a month or longer for some of the best performing units. As can be seen from Figure 5, for sample times of the order of a few weeks to a few months, some of these units begin to exhibit flicker-noise and/or random-walk FM like behavior. If such is the case, near optimum prediction techniques have been developed to deal with these more dispersive noise processes.

Over the last six months (since 22 April 1995), the USNO has had an RMS prediction error for their UTC(USNO MC) with respect to UTC of 6 ns. Their prediction algorithm takes advantage of optimum estimation techniques – using the simple mean frequency assumption (white-noise FM). The RMS optimum prediction error for white-noise and random-walk-noise FM cases is given by $\tau \times \sigma_y(\tau)$. If the prediction interval is about 45 days (1 1/2 months), then this implies $\sigma_y(\tau = 45 \text{ days}) = 1.5 \times 10^{-15}$. This represents the relative instability between the USNO ensemble of clocks used for prediction and that of TAI/UTC. Since the clocks at USNO contribute about 40% to the generation of TAI/UTC, this stability number is biased low by the factor 1/(1 - weight), where the weight is that part the USNO clocks have in the ALGOS computation of TAI/UTC. The actual percentage of weight varies over the course of the data since clocks come in and go out of ALGOS computation. As an example, if this weight is 40%, then the measured stability needs to be multiplied by 1.67 to obtain an unbiased estimate. This unbiased estimate is about $\sigma_y(\tau = 45 \text{ days}) = 2.6 \times 10^{-15}$ for the combined instabilities of EAL and USNO. It is safe to say that one of these two scales has a stability better than this number divided by $\sqrt{2}$ or 1.8×10^{-15}. USNO has about 40 of the HP 5071A clocks in their ensemble.

The current specification level for the stability of the high-performance HP 5071A is approximately $\sigma_y(\tau) = 8 \times 10^{-12}\tau^{-1/2}$. If a timing center has an ensemble of these standards, a simple equation, relating the number of these clocks in the ensemble and the integration or averaging time necessary to reach a stability of 1×10^{-15}, is $N\tau = 6.4 \times 10^7$. Hence, if N = 1, almost two years of averaging would be necessary. Most of these clocks would exhibit non-white-noise instabilities before 1×10^{-15} could be reached. If N = 4, six months are needed, which is not impractical. If N = 10, then two and one-half months are needed. And if N = 40, then less than three weeks are needed.

Having 10 or more of these clocks optimally used would allow a laboratory to have a real-time predicted frequency reference with about 1×10^{-15} traceability to TAI/UTC. Having 4 or more of these clocks would probably provide a real-time frequency estimate of about 2×10^{-15}.

Accuracy and Stability of Methods of Distributing Time and Frequency

A perfect clock is limited by the means of distributing its time and frequency. This is, of course, true both within a laboratory setting as well as for remote distribution and comparisons.

As atomic clocks have improved at a rate of about one order of magnitude every seven years, this rate of improvement has placed significant demands on the methods of distribution and comparison. Natural limits have been reached for many different methods so that they are no longer useful for state-of-the-art clocks. HF broadcasts, such as WWV, are limited at the

millisecond level due to propagation path delay variations as the ionosphere moves up and down. LF and VLF transmissions, such as Loran-C, are limited at the microsecond level due also to propagation path delay variations.

New techniques are needed and satellite timing systems have opened up opportunities. There are systems which will work at the 10 picosecond level, but it is often a question of operational complexity and cost. Today's operational time scales have sub-nanosecond day-to-day predictabilities. To meet these needs a systems approach should be taken. In addition to satellite techniques, the potential of using optical-fiber communications is very promising. Now that the communications industries are laying fibers extensively and they also need time and frequency, a closer cooperation between them and the time and frequency community could be very beneficial.

Figure 7 is a summary plot of some of the best methods of time and frequency comparisons. The stability measure used is called the time variance, TVAR, and is given by $\sigma_x^2(\tau) = < (\Delta^2 x)^2 > /6$, where Δ^2 is the second difference operator, x is the time differences averaged over an interval τ, and the brackets "<>" denote the expectation value. What is plotted is $\sigma_x(\tau)$ for each of the different techniques. Since $\sigma_x(\tau) = \tau \mathrm{Mod}.\sigma_y(\tau)/\sqrt{3}$, the $\mathrm{Mod}.\sigma_y(\tau)$ stability values are also shown for each decade. This is of particular value since the confidence on the estimate of the frequency difference measured over an interval τ is given by $2 \times \mathrm{Mod}.\sigma_y(\tau)$ if the residuals are modeled by white-noise PM. If this model is not valid, the $2 \times \mathrm{Mod}.\sigma_y(\tau)$ value is still an approximate estimate of the confidence for using a particular technique for frequency transfer.

Figure 7 includes both time and frequency distribution techniques as well as time and frequency transfer techniques. In all cases, the clocks can be remote from each other, but in some cases there are limitations. Loran-C is plotted as a well-known stability reference. The Loran-C values are limited by the ground-wave propagation path, which is about two to four Megameters. The ground-wave signals vary because of distance, terrain variations, and atmospheric conditions; the effects of diurnal and annual variations are also shown. Loran-C can be used as a real-time time and frequency distribution system or in the common-view mode. The latter provides better accuracy and stability. The disadvantage of the common-view method is that it is an after-the-fact computation. The range of stabilities plotted covers both methods.

The GPS common-view technique depends not only on the baseline distance, but also on the receiver hardware and processing techniques. The maximum baseline distance is about 13 Mm (the circumference of the earth is about 40 Mm). Over the longest baselines the tropospheric and ionospheric delays are the limiting uncertainties. In this case the satellite's emphemeris must also be known well. For short baselines, this technique provides a lot of common-mode cancellation of errors. The common-view technique was a major break-through for international time and frequency comparisons and is still today the main means of communicating the times of most of the contributing clocks in the generation of TAI/UTC[7].

Originally, day-to-day stabilities of as good as 0.8 ns were obtained between baselines as far apart as Boulder, Colorado and Ottawa, Canada. Global accuracies of about 4 ns have been obtained with careful post-processing.

Since the original GPS common-view receivers were designed, built and experiments conducted,

there seems to have been a gradual degradation in the performance of this technique. Day-to-day stabilities of from 2 to 8 ns are now more typical, and significant temperature coefficients have been measured due to antenna and lead-in cable sensitivities. Problems have crept into the common-view technique at about the 10 ns level. The source of these is being investigated at this time.

The GPS advanced common-view (ACV) technique is a systems approach. With digital GPS multichannel receivers, new opportunities become available that could provide major advancements in time and frequency metrology among clocks remote from each other. The basic benefits of the common-view technique can be built upon because of the large increase in the available data.

The ability to track several satellites continuously at a single observing location means that a more productive common-view schedule can be used, particularly between less-distant sites. The increased diversity of the measurement should reduce multipath effects since the multipath tends to average across the sky. Ionosphere and troposphere modeling errors will be basically the same for both techniques; except, comparing multiple tracks allows a comparison of an individual track from day-to-day. Since the geometry stays the same from one sidereal day to the next for a fixed site, the scatter in a given common-view track could be used in a time series weighting procedure to minimize errors for the remote clock comparison.

From a simple "degrees of freedom" argument there is significant advantage to the GPS ACV technique. If the measurement noise is white PM, then the confidence on the estimate of the frequency difference between two ideal reference clocks, as determined from a linear regression to the time-difference residuals taken between the two clocks, is: $\sqrt{12} \times \sigma/(\tau_0 \times n^{3/2})$, where σ is the standard deviation of the white-noise residuals, τ_0 is the measurement interval, and 'n' is the degrees of freedom (the number of independent measurements).

Since GPS common-view instrumentation errors have apparently gotten worse as time has gone on, and these original techniques involved a lot of analog circuitry, the question arises that perhaps the delays and delay stabilities may be better in these new miniaturized digital circuits. Hence, understanding and documenting instrumentation errors would be useful.

With some of the new digital multi-channel GPS timing receivers, it is possible to track several satellites at a time and to obtain a solution each second for each satellite. This has the potential of increasing the data density more than three orders of magnitude over the original common-view technique outlined above. If this white-noise model persisted, this would allow 1×10^{-15} confidence interval to be reached on a one-day regression line between two standards. The question is: can this GPS ACV approach significantly and efficiently increase the effective number of degrees of freedom over the original common-view technique?

An experiment was set up at HP labs in Palo Alto, CA using two eight-channel GPS receivers with the same reference clock feeding both and the same antenna providing the signal for both. All errors should cancel except for instrumentation errors and cable length differences. A multi-channel digital filter was designed to provide 10 s averages in order to decrease the volume of data and still take advantage of the available degrees of freedom. Each receiver and associated counting and computing system produced the 10 s average time difference between

each satellite clock and the common local-reference clock. These time series were differenced satellite by satellite and averaged across the satellites for each 10 s interval to produce the time difference between the two common times of the reference clock.

A plot of the time stability of these measurements is shown in Figure 7. The above equation for the confidence on the frequency estimate is equivalent to $2 \times \mathrm{Mod}.\sigma_y(\tau)$. If the white-noise FM, $\tau^{-1/2}$ level were to persist, then indeed the 1×10^{-15} level could be reached for an averaging time of 1 day. The little hump for τ just longer than 10 s is probably due to our digital filter. For longer τ values the stability level reached below 100 ps for τ greater than about 1 hour. Environmental effects need still to be evaluated.

Also in Figure 7, The upper and lower values for the two-way satellite time and frequency transfer (TWSTFT) technique are the stabilities for continuous operation. The bottom curve is a measured instrumentation stability limit achievable. The upper curve is a more typical performance stability observed between two sites remote to each other. The upper limit is dotted on the right end as a reminder that TWSTFT is not typically used in the continuous mode for this range of sample times, but rather three times per week, and the frequency transfer uncertainty will not be as good as that shown, but would be nominally given by 1 ns/τ.

The TWSTFT technique both transmits and receives, and cannot be used for dissemination, but for after-the-fact time and frequency transfer. Because the typical mode is for intermittent operation, it is more amenable for after-the-fact time transfer. The baseline distances between clocks being compared is limited by the position of the geostationary communications satellite being used. Distance up to about 9 Mm have been realized.

The enhanced GPS (EGPS) technique can be used both for time and frequency transfer and for real-time distribution. It is also a systems approach and is highly dependent on the reference clock used; hence, the different levels of performance when a quartz oscillator is used, or a rubidium frequency standard, or a cesium-beam frequency standard. Because EGPS employs an SA filter and is phase-locked to GPS, the long-term stabilities all approach the GPS stability regardless of the reference clock used. The upper curve is dotted on the right end as a projection of theoretical behavior. The other values are based on experimental analysis. If SA is removed, the stability of EGPS will be significantly improved – especially in the intermediate-term region of averaging times.

Because GPS is global, there is no limit on the baseline separation of the clocks being compared or which are receiving the distributed time and frequency information. Hence, this approach is excellent as a telecommunication network-node synchronization and syntonization technique. With an excellent reference clock the instrumentation residual errors have been documented at about 1.5 ns. It is also extremely cost effective – producing in real time a simple 1 pps output that is very stable. An EGPS receiver can lock either to GPS system time or to the broadcast estimate of UTC(USNO MC). The latter is usually kept within about 20 ns of the master-clock at the observatory.

A potentially very useful experiment that has not been conducted would be to treat GPS system time (the composite clock) as a common-clock. Assume there are two perfect clocks remotely located with respect to each other anywhere on the earth. If these two clocks were

to perform the same kind of optimum regression analysis estimate of the frequency of GPS system time using EGPS receivers over the same integration interval, and then the difference in these frequency estimates were calculated, the uncertainty in this estimate should improve with the length of integration. In other words, each of the two sites is measuring the same clock in the same way within some noise band. Subtracting their measured values from each other subtracts out the common clock. If for long integration times, the time-difference residuals had a white-noise spectrum, the uncertainty on frequency transfer could improve as fast as $\tau^{-3/2}$. It should only take a few weeks of averaging time to reach 1×10^{-15} very cost effectively and with straight-forward data processing.

The GPS Carrier-phase technique has the smallest uncertainty for frequency comparison of remote clocks. Experiments have been conducted comparing hydrogen masers remote to each other by having geodetic type receivers at both sites. By locking to GPS common carrier-phase at the two sites, RMS residuals of 30 ps have been measured. The data plotted in Figure 7 are between Goldstone, California and Algonquin Park, Canada. The baseline distance is about 3.4 Mm (2,000 miles). About 35 monitor stations were involved in determining accurate ephemeridies for the satellites. Both sites have to view the same satellites at the same time. The main problems with this technique are the difficulty in the data processing and the expense of the receivers. Both of these problems could be overcome, and this technique could be among the best for minimizing remote frequency comparison uncertainty.

Time accuracies of the order of 10 ns may eventually come out of the FAA's Wide Area Augmentation System (WAAS) via the signals from the INMARSAT satellites. These could also be very useful to the timing community.

Conclusions

Intrinsically, the proper algorithmic combination of the global set of clocks contributing to the composition of TAI/UTC provide a reference as good as 1×10^{-15} or better for sample times, τ, longer than about two hours. More than a month after the fact, the long-term stability of TAI/UTC is available from the BIPM Circular-T and it approaches the ideal stability intrinsically available. The measurement noise limits the stability of TAI/UTC at $\sigma_y(\tau = 10 \text{ days})$ to about 1×10^{-14}. The paper discusses ways to improve the intermediate stability (τ = 1 day to a month) measurement noise for international comparisons to better than 1×10^{-15}. The paper also suggests ways to transfer optimally the stability of the international clock set to a local clock set so that frequency standards at two different locations can compare frequencies with uncertainties at or below 1×10^{-15}. This can be done for both the intermediate and long-term stability ranges and in real-time or in post processing. The post processed data intrinsically have better uncertainties.

Existing local clock sets properly utilized and optimally predicted forward can project to the current time much of the intrinsic stability of the international clock set. This can be done at the 1×10^{-15} or better level for frequency comparisons or to better than 10 ns of UTC timing accuracy. The outstanding long-term stability of a new commercial cesium-beam clock contributes a key element to the 1×10^{-15} comparison ability now available. The well known short-term stability of hydrogen masers can contribute substantially to the frequency comparison

effort – especially if they also have excellent intermediate and long-term stabilities.

The BIPM data analyzed in this paper were for the period 1991 through 1994. The data base has only improved since then, and an international cooperative, using the Internet, for example, could make available much of the intrinsic stability of the international clock set for both the intermediate as well as the long-term stability comparisons of remote clocks. This could be done at or below the 1×10^{-15} level and in near-real time.

Acknowledgements

The authors are very grateful for the extended effort of Mr. Jacques Azoubib for providing much of the data used in the analysis. Mr. Mihran Miranian of USNO has been very helpful with some of the comparisons and in making data available. Mr. Mike King of Motorola has been extremely helpful toward optimally utilizing the new multi-channel digital GPS receivers. The service and support of the time and frequency staff at NIST are much appreciated.

Figure Captions

Figure 1 Histogram of offsets from TAI for all clocks, 311 total, reporting in the international time scale for 1994. There are several outliers plotted in both the -30 and $+30 \times 10^{-13}$ bins. The large number of units in the 0 bin are mainly due to the primary standards and the HP 5071As.

Figure 2 Histogram of offsets from TAI for all hydrogen masers, 37 units, reporting in the international time scale for 1994.

Figure 3a Histogram of offsets from TAI for all HP 5071As, 94 units, reporting in the international time scale for 1994.

Figure 3b Same as Fig. 3a but with bin size reduced to 1×10^{-13}.

Figure 4 Histogram of offsets from TAI for the three primary standards reporting in international time scale for 1994.

Figure 5 Scatter plot of $\sigma_y(\tau)$ for high performance HP 5071As (78 units) reporting in the international time scales for 1994. The ideal theoretical white-noise FM slope on this plot should be proportional to $\tau^{-1/2}$. With some outliers, many of the units tend to follow this slope within the confidence of the estimates. The slope tends to be slightly steeper between τ = 10 days and 20 days. This may be caused by residual measurement noise. For the longest τ values, some of the clocks tend to be flatter than the $\tau^{-1/2}$ behavior. In these clocks there appears a slight frequency drift, flicker or random-walk noise. Those indicating drift are of the order of a few parts in 10^{-16} per day.

Figure 6a RMS/$\sqrt{N}$ of sigmas for the data in Fig. 5 . The RMS/$\sqrt{N}$ values are also plotted for the hydrogen masers. The RMS/$\sqrt{N}$ will typically give a pessimistic estimate for ensemble stability because the values with larger sigmas are weighted heavier. The results using optimum weighting for the HP 5071A ensemble are also plotted.

Figure 6b The Mod.$\sigma_y(\tau)$ stability results from an independent three-cornered hat analysis. The three independent ensembles were the primary clocks, the hydrogen masers and the HP 5071As. The $\tau^{-3/2}$ behavior is consistent with white PM measurement noise at a level of 1.3 ns. Apparently, the HP 5071A ensemble is sufficiently better than the measurement noise or the other two ensembles that its stability cannot be measured with confidence because of only having 71 data points.

Figure 7 A plot of the time stability of state-of-the-art techniques for time and/or frequency comparison at locations remote to each other, and showing Loran-C as a well-known stability reference. The Loran-C values are for ground-wave signals and vary because of distance and terrain and atmospheric conditions; the effects of diurnal and annual variations are also shown. GPS common-view technique also depends on the baseline distance, but more importantly on the receiver hardware and processing techniques. The GPS advanced common-view (ACV) technique shows the first experimental results for the hardware only. The upper and lower values for the two-way satellite time and frequency transfer (TWSTFT) technique are the stabilities for continuous operation. The bottom curve is a measured instrumentation stability limit achievable. The upper curve is a more typical performance stability observed between two sites remote to each other. The upper limit is dotted on the right end as a reminder that TWSTFT is not typically used in the continuous mode for this range of sample times, but rather three times per week, and the frequency transfer uncertainty will not be as good as that shown, but would be nominally given by 1 ns/τ. The enhanced GPS (EGPS) technique can be used both for time and frequency transfer and for real-time distribution. It is a systems approach and is highly dependent on the reference clock used; hence, the different levels of performance when a quartz oscillator is used, or a rubidium frequency standard, or a cesium-beam frequency standard. Because EGPS employs an SA filter and is phase-locked to GPS, the long-term stabilities all approach the GPS stability regardless of the reference clock used. The upper curve is dotted on the right end as a projection of theoretical behavior. The other values are based on experimental analysis. All GPS methods basically assume that SA will stay at the current level.

References

[1.] D. W. Allan, A. Lepek, L. Cutler, R. Giffard, and J. Kusters, *"A 10^{-15} International Frequency Reference Available Now,"* Proceedings of Fifth Symposium on Frequency Standards and Metrology, Woods Hole, MA, October 1995.

[2.] A.A. Uljanov, N.A. Demidov, E.M. Mattison, R.F.C. Vessot, D.W. Allan, and G.M.R. Winkler, *"Performance of Soviet and U.S. Hydrogen Masers,"* submitted to Proc. of 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 1990.

[3.] Jon H. Shirley, W.D. Lee, and R. Drullinger, *"The Evaluation of NIST-7 A New Era,"* Proceedings of the Symposium on Frequency Standards & Metrology, Woods Hole, Massachusetts, 15-19 Oct. 1995 (World Scientific Publishing)

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

[5.] A. Lepek, A. Shenhar, and D. W. Allan, *"INPL Virtual Time Scale, UTC(INPL),"* to be published, Metrologia.

[6.] M.A. Weiss, D.W. Allan and T.K. Peppler, *"A Study of the NBS Time Scale Algorithm,"* IEEE Transactions on Instrumentation and Measurement, 38, 631-635, 1989.

[7.] D.W. Allan, D.D. Davis, M. Weiss, A. Clements, B. Guinot, M. Granveaud, K. Dorenwendt, B. Fischer, P. Hetzel, S. Aoki, M.-K. Fujimoto, L. Charron, and N. Ashby, *"Accuracy of International Time and Frequency Comparisons Via Global Positioning System Satellites in Common-View,"* IEEE Transactions on Instrumentation and Measurement, IM-34, No. 2, 118-125, 1985.

HISTOGRAM OF FREQUENCY OFFSETS (TAI - CLOCK) FOR 1994
(ALL CLOCKS)

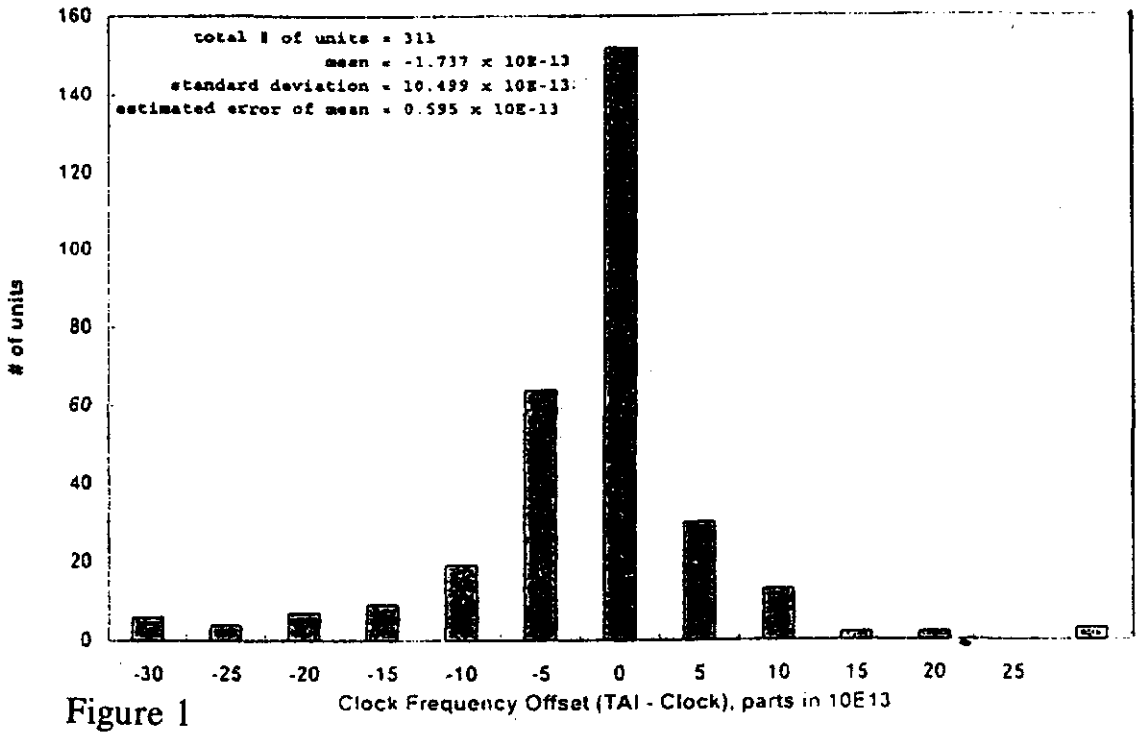

Figure 1

HISTOGRAM OF FREQUENCY OFFSETS (TAI - CLOCK) FOR 1994
(HYDROGEN MASERS)

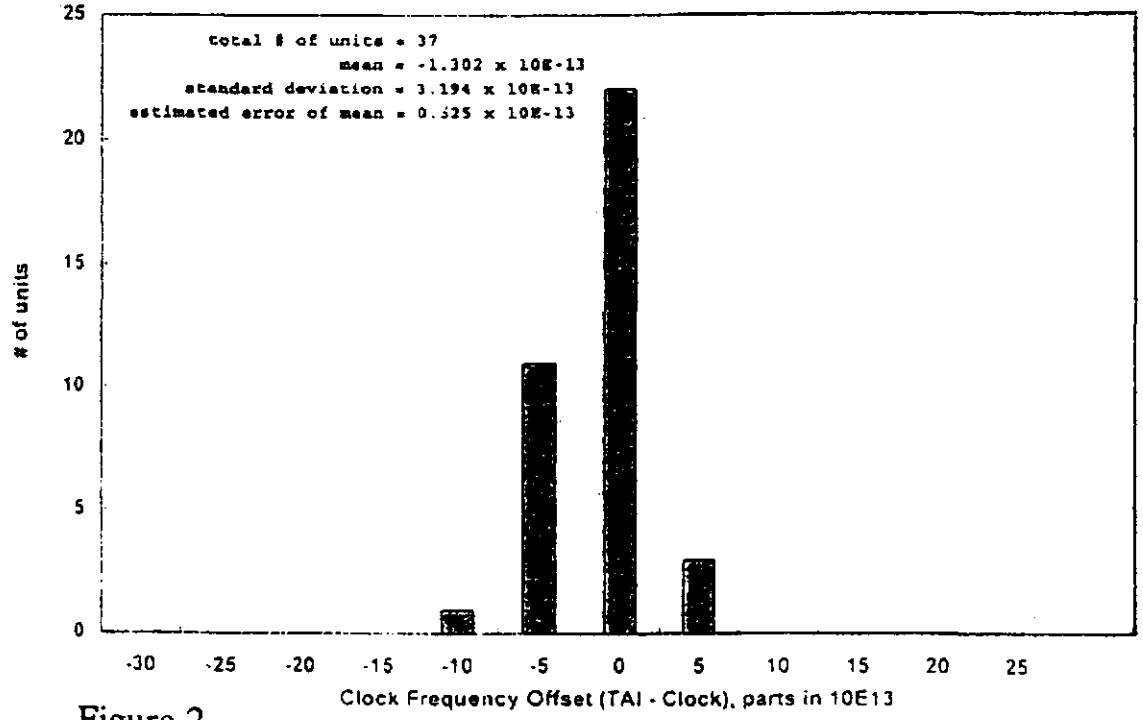

Figure 2

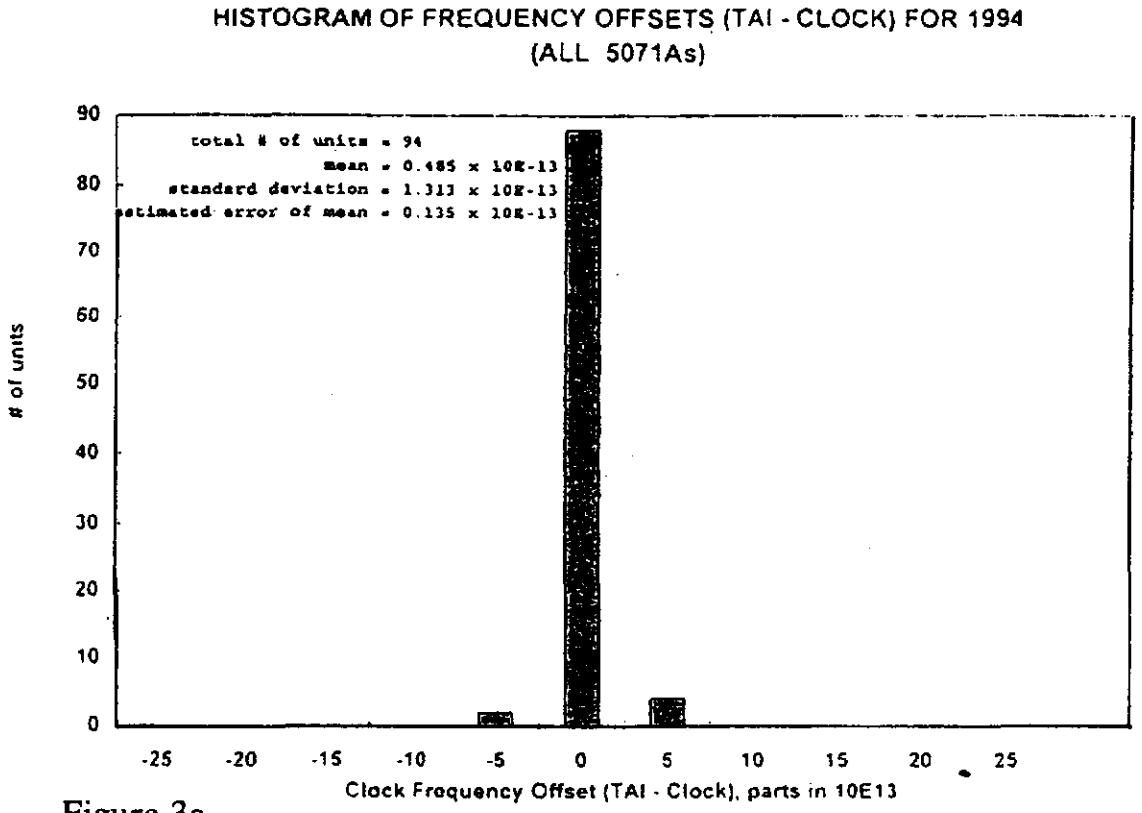

Figure 3a

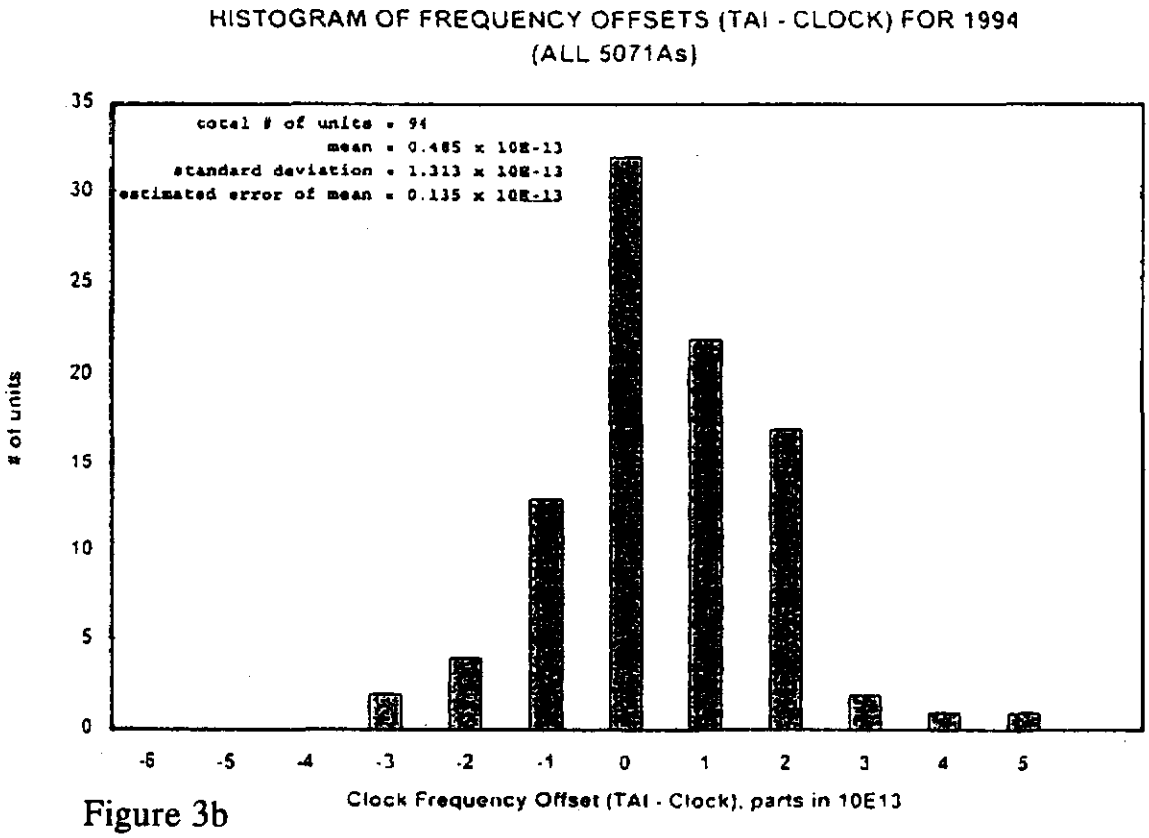

Figure 3b

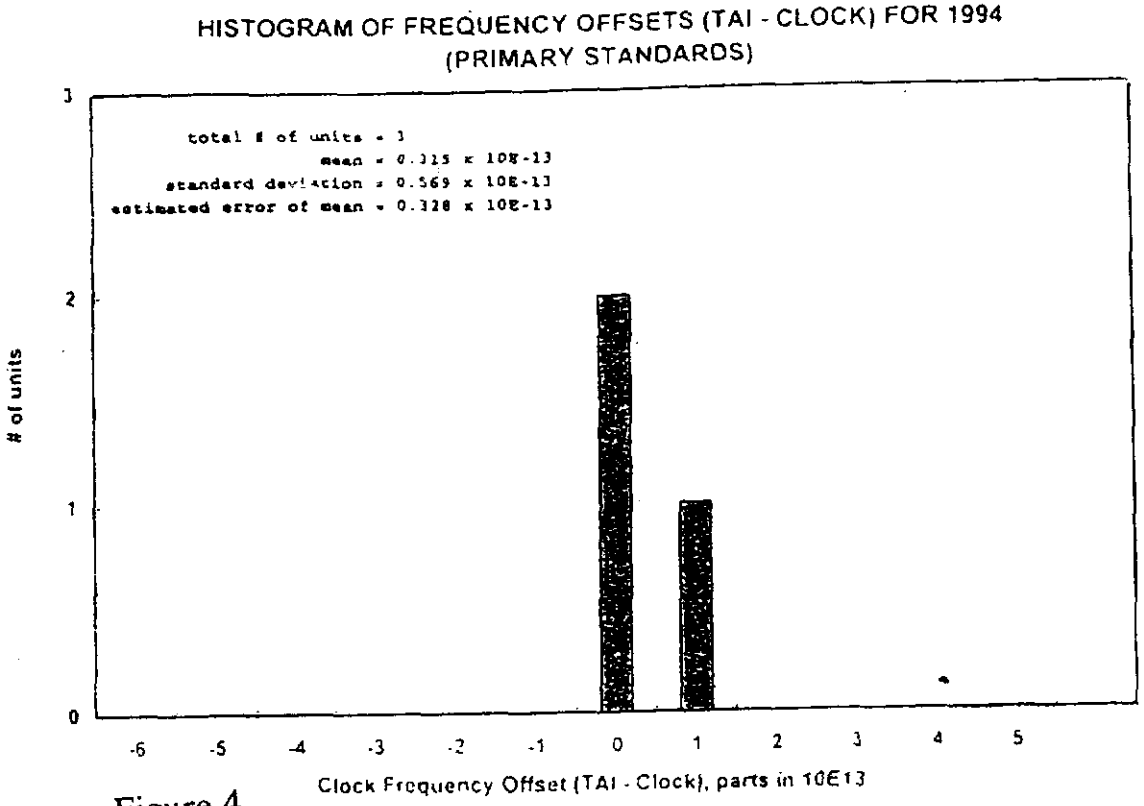

Figure 4

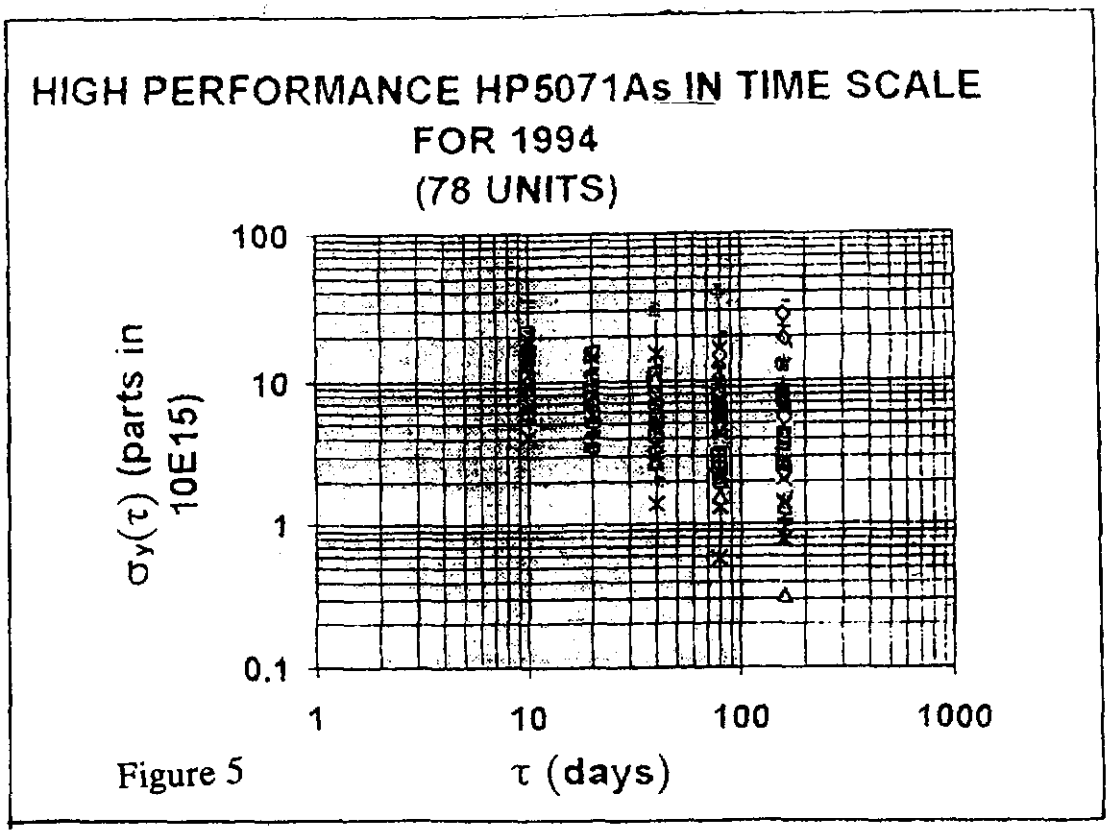

Figure 5

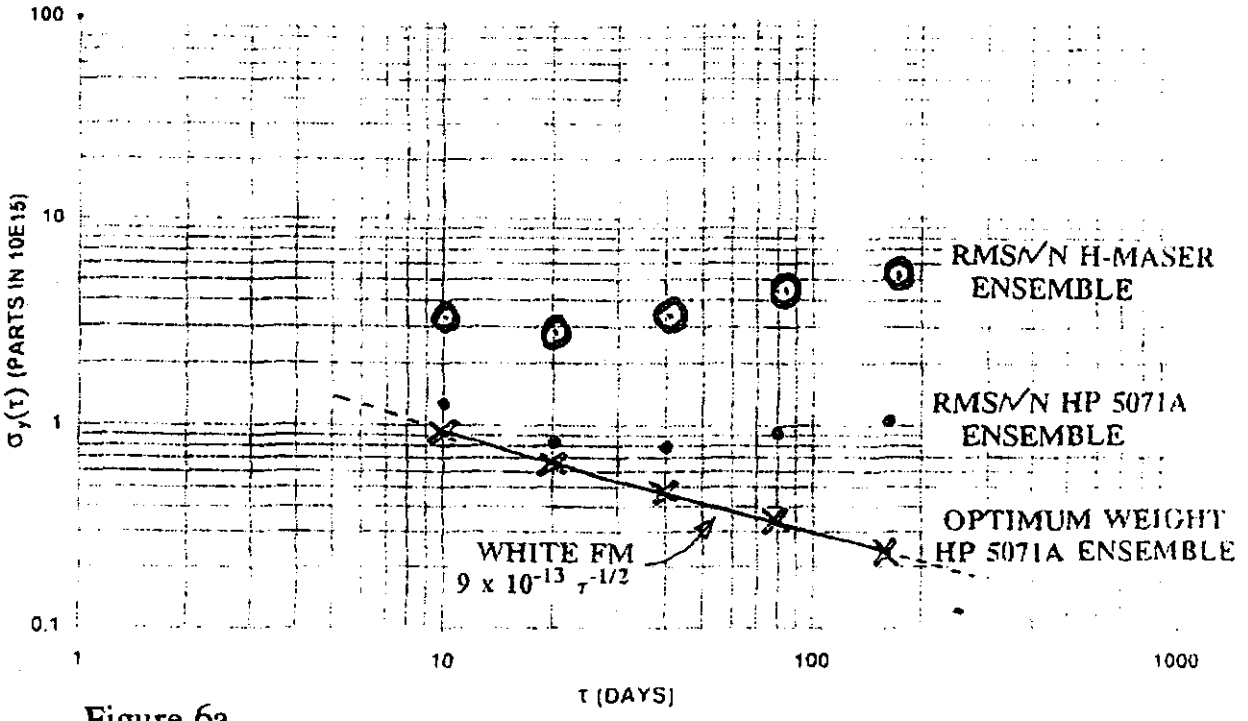

Figure 6a

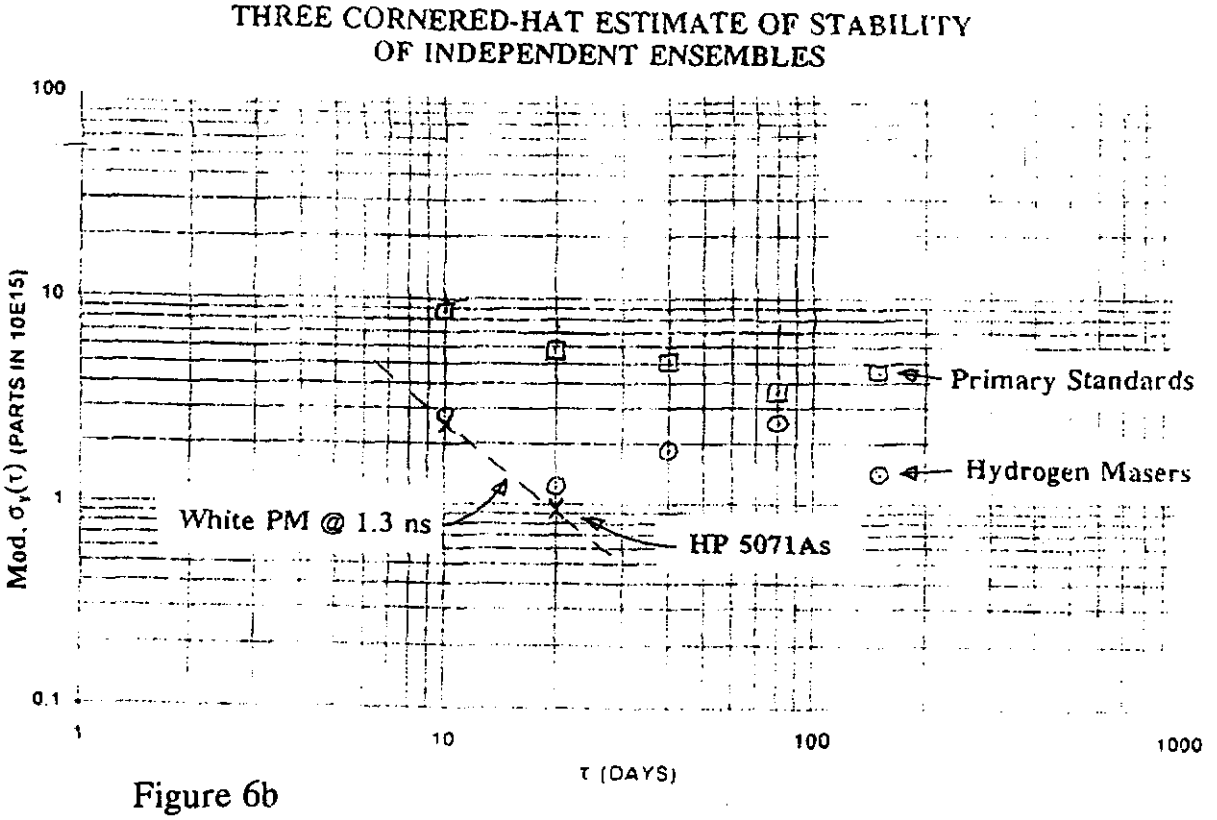

Figure 6b

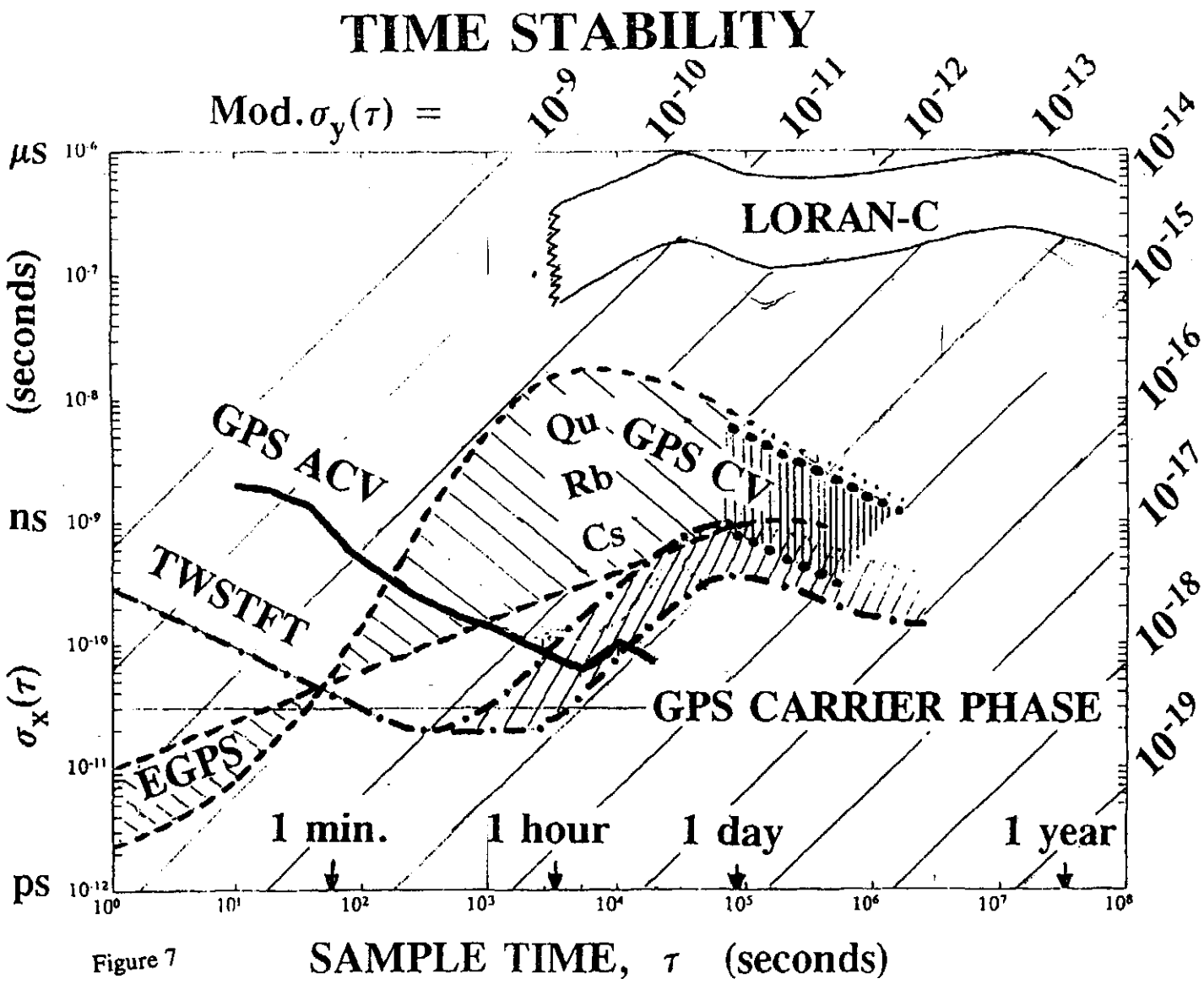
TIME STABILITY
Mod.$\sigma_y(\tau)$ =
10^{-9}
10^{-10}
10^{-11}
10^{-12}
10^{-13}
10^{-14}
10^{-15}
10^{-16}
10^{-17}
10^{-18}
10^{-19}
LORAN-C
GPS ACV
Qu
Rb
Cs
GPS CV
TWSTFT
EGPS
GPS CARRIER PHASE
1 min.
1 hour
1 day
1 year
μs
ns
ps
(seconds)
$\sigma_x(\tau)$
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
SAMPLE TIME, τ (seconds)

Figure 7