STEERING A TIME SCALE

Judah Levine
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80305, USA
jlevine@boulder.nist.gov

Abstract

A time scale is a procedure for combining the data from an ensemble of clocks or frequency standards. The input data to the ensemble algorithm are generally the time (or frequency) differences between each of the members and the reference device for the system. Therefore, the overall time and frequency of the ensemble are free parameters and must be continuously adjusted using external data and a steering algorithm. In addition, the outputs of the procedure are the average time and frequency of the group and the characteristics of each device (time, frequency, frequency aging, prediction error, etc.) with respect to this average, and a second job of the steering algorithm is to realize this computed average time and frequency in a physical device. I will discuss the considerations that govern the design of steering algorithms in general. I will illustrate these considerations using the algorithms that realize UTC (NIST) from the ensemble average of the time differences of the cesium standards and hydrogen masers that are located at the NIST laboratory in Boulder. I will also discuss the design of the steering of the backup for UTC (NIST), which is based on an ensemble of cesium standards located at the NIST radio station in Fort Collins. The backup time scale is intended to support the NIST services should the primary time scale become unavailable, so that it must track UTC (NIST) as closely as possible, which implies a tight coupling between the two scales. At the same time, it must remain independent of UTC (NIST) so that it does not fail if its external reference becomes unavailable, which implies a loose coupling. I will discuss the actual design, which is a compromise between these two incompatible requirements.

INTRODUCTION

Applications that depend upon precise time and frequency usually use an ensemble of several clocks to provide the time reference. The ensemble is implemented using several independent devices, and the output time or frequency signals are based on the weighted average of the contributors. The method of determining the weights varies from one algorithm to another, but a common method is to base the weight of each contributor on its stability (often called the prediction error) over some previous number of measurement cycles. When a physical output signal is required, the computed ensemble average time and frequency are commonly realized using a phase stepper, which offsets the output of the one of the clocks in time and frequency so that the steered output implements the computed average. (This steered output is in addition to the unsteered output of the same clock, which is used in the ensemble computation.) Using the ensemble average to control the phase stepper improves the reliability of the system, since the output is still available even if one of the contributing devices fails. In addition, an ensemble of different types of devices can exploit the best statistical features of each type, so that the weighted average can have statistical performance that can be difficult to realize by any one of the contributors acting alone.

Although the algorithm used to compute the weighted average of the members of the ensemble can attenuate the instabilities of the individual contributors, it cannot totally remove these stochastic effects. Therefore, a free-running ensemble will eventually walk away from the initially correct values of the time and frequency even if those values were set arbitrarily well to begin with, and even if the contributing members of the ensemble were characterized perfectly at the outset. All real ensembles must therefore be steered using external data. The details of the steering will depend on the requirements of the application that uses the ensemble data, on the statistical performance of the ensemble, and on the quality and availability of the external data that are used to compute the steering corrections. I will illustrate these considerations using several ensembles at NIST – the clock ensemble that realizes UTC (NIST) in Boulder and a backup ensemble, which uses a smaller number of clocks and is located in Fort Collins, Colorado. Some of the considerations that govern the Fort Collins configuration can be useful in other contexts, and I will briefly discuss these possibilities.

THE AT1 TIME SCALE ALGORITHM

The AT1 algorithm **[1,2]** has been used for many years at NIST (and previously at NBS) to compute the weighted average of time differences acquired from an ensemble of cesium standards and hydrogen masers. The data input to the algorithm are a series of time-difference measurements between the reference clock and the other devices in the ensemble. Since 1980, these measurements have been made using a dual-mixer system, which measures the phase difference between the 5 MHz output of each clock and the corresponding signal from the reference device, which is simply one of the clocks in the ensemble. The phase-difference measurements are made at an intermediate frequency of approximately 10 Hz **[3].**

The performance of the measurement system can be evaluated by measuring the time difference between the same clock connected to two measurement channels. The time deviation (TDEV) of these time differences, measured using two pairs of channels are shown in Figure 1. The results depend on which pair of channels is chosen for the comparison, and the two traces show the best and worst pair in our current configuration.

The divergence at longer periods is primarily a result of the sensitivity of the measurement hardware to fluctuations in the ambient temperature, but the time deviation is less than 1 ps even for averaging times out to 1 day, and the best channel pair have a TDEV of less than 0.1 ps at that averaging time. Since the hardware and the algorithm use time differences, any common-mode temperature sensitivity does not affect the stability of the system. The current implementation measures these time differences every 720 s, but the time interval between measurements is not critical and the value that is used is chosen mostly for computational convenience, since it is an exact decimal fraction of an hour.

The algorithm characterizes each contributing member using three deterministic parameters: a time offset, a frequency offset, and a frequency aging. The time offset and the frequency offset are estimated by the ensemble algorithm on each measurement cycle; the frequency aging term is computed outside of the scale and is set administratively; it is treated as a constant by the algorithm. It is 0 for the cesium standards and is of order 10^{-16}/day for the masers. The frequency aging term is very important for masers, since aging makes a significant contribution to the measurement variance of the maser data at all but the very shortest periods.

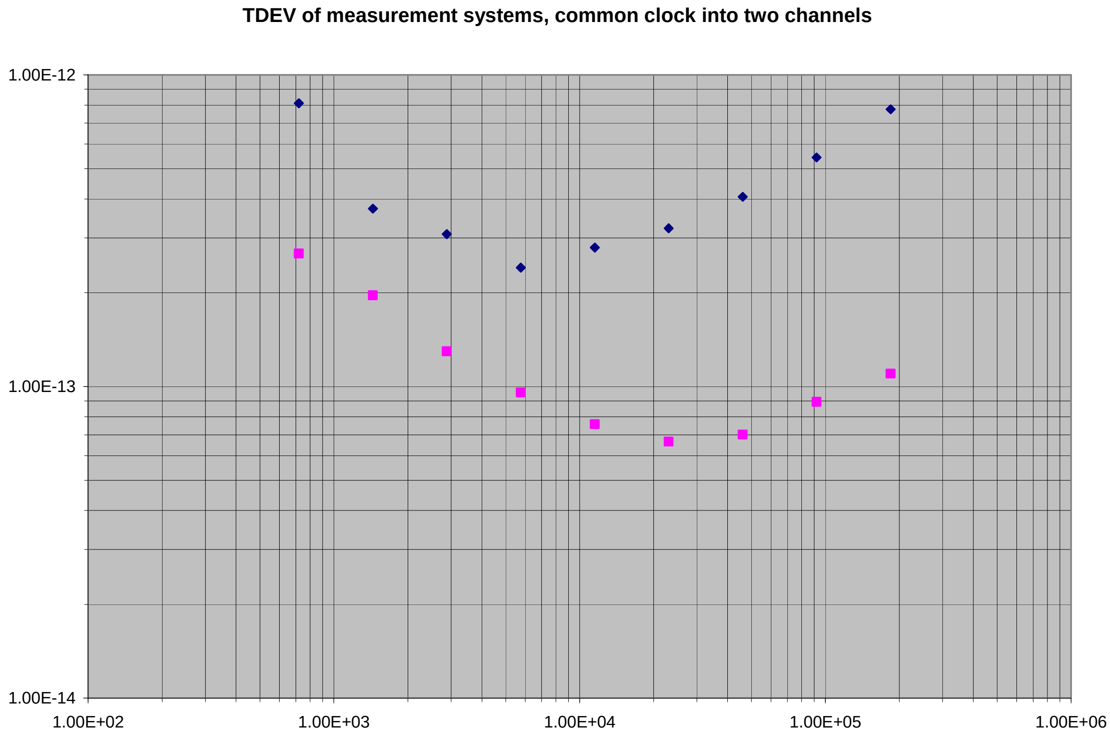


Figure 1. The best and worst TDEV (s) vs. averaging time (s) of time difference between same clock in two channels.

The AT1 time scale is not steered, so that its frequency slowly walks away from the definition of the duration of the SI second. This is due to imperfections in the modeling of the long-term variation in the frequencies of the clocks that contribute to the ensemble and to the stochastic fluctuations in these frequencies, which can be modeled only statistically. At the present time, the frequency of the AT1 ensemble differs from the SI frequency by about 38 ns/day.

THE UTC (NIST) TIME SCALE

The UTC (NIST) time scale is derived from AT1 by applying a series of administrative steering offsets using the data from Circular T, which is published by the BIPM every month. Figure 2 shows the steering corrections that have been applied to AT1 to realize UTC (NIST).

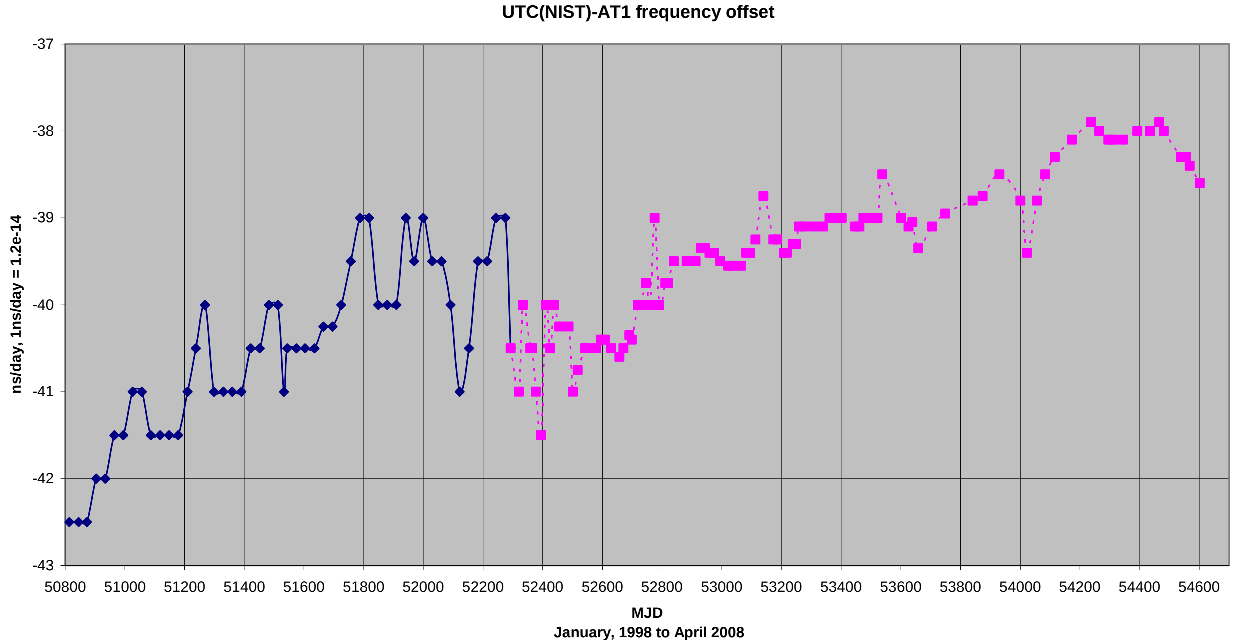


Figure 2. Steering corrections UTC (NIST) - AT1. Blue is slow steering; pink is moderate steering.

The Circular T data for any month are not available until the middle of the next month. The performance of UTC (NIST) can therefore be divided into three averaging regimes:

1. The short term. The stability of UTC (NIST) for times less than 1 month is determined by the free-running stability of the AT1 time scale, since we have no new external calibration data during this period. The statistics of UTC (NIST) in this regime are important for many of our users, but they are outside of the scope of this discussion, which is focused on steering algorithms.

2. The long term. Every steering algorithm will always be designed to steer UTC (NIST) towards UTC as computed by the BIPM, so that we would expect that the time difference between UTC (NIST) and UTC would be bounded over long averaging times so that any frequency offset would decrease to an arbitrarily small value as the averaging time increases. Therefore, we would expect that the stability of UTC (NIST) at very long averaging times would be the same as the stability of UTC itself, and that this conclusion would be substantially independent of the details of the algorithm used to realize the steering control. (UTC itself is a steered scale, but it has no physical realization, and the considerations that are important in its steering algorithm are very different from those of a timing laboratory that must provide signals in real time.)

3. The intermediate regime. Different steering algorithms will perform very differently here, and this regime is the focus of this discussion.

The design of the steering algorithm that realizes UTC (NIST) from the free-running AT1 time scale must be a compromise between two conflicting requirements. One the one hand, minimizing the difference between UTC (NIST) and UTC would imply aggressive steering, so that any time offset is removed as quickly as possible. On the other hand, maintaining the frequency smoothness of UTC (NIST) would suggest very gentle steering, so that the steering did not unduly degrade the frequency stability of the underlying scale.

We can illustrate the difference between these two steering philosophies by examining the statistics of the time differences between UTC and UTC (NIST) during two different periods as shown in Fig. 2. We have used data from Circular T published by the BIPM. We have excluded data from the most recent 2 years from this study, because UTC (NIST) has been significantly degraded during this period by construction at the NIST Boulder site and by various environmental perturbations.

We use two measures to characterize these differences: The standard Allan deviation (ADEV), which illustrates the frequency stability of the UTC (NIST) time scale, and the histogram of the time differences. The histogram is more useful in some contexts than the statistical time deviation, since it provides information on the worst-case performance of the scale.

We considered three steering algorithms: A very slow steering algorithm that used the time-difference data from Circular T to steer UTC (NIST) relative to AT1 only on the first day of the month following the receipt of the Circular. This method resulted in a delay of about 2 weeks between when we received Circular T and when we applied the steering correction. The maximum steering adjustment was limited to ±2 ns/day in frequency. The moderate steering used these same data with up to two steering adjustments per month: one in the middle of the month when we received Circular T and the second on the first day of the following month. The steering adjustments used a time constant of 6 months. The aggressive steering algorithm used the same two adjustments per month, but used a shorter time constant of 2 months. The results are shown in the following figures.

As we would expect, the histogram of the time differences is both smaller and narrower when aggressive steering is used, and this would be the strategy of choice if minimizing the time differences is the most important consideration. The Allan deviation is somewhat worse relative to moderate steering, but the frequency stability at intermediate periods is better than 2×10^{-15} for either method. Both methods are clearly superior to the gentle steering algorithm both with respect to frequency stability and time accuracy.

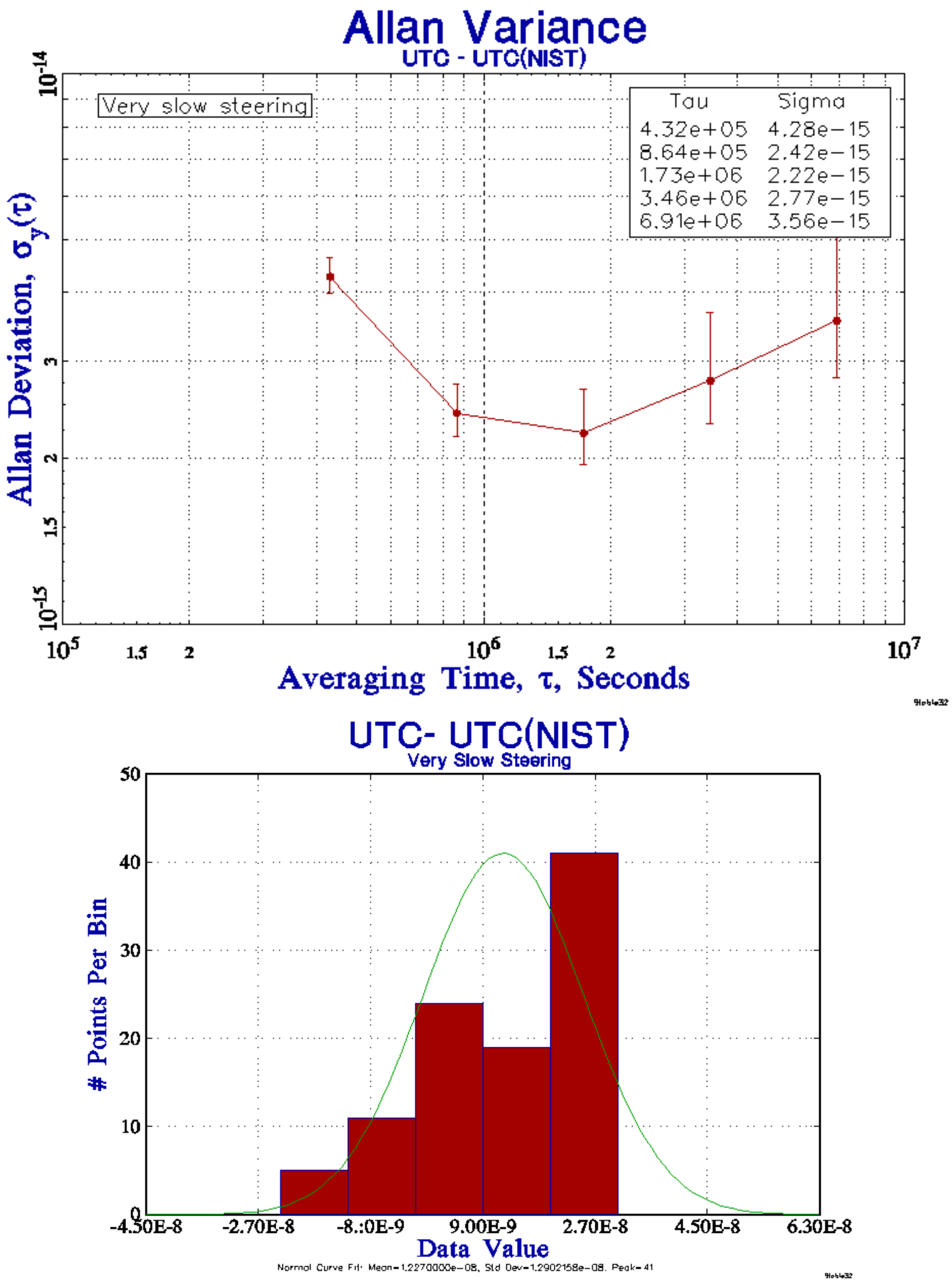
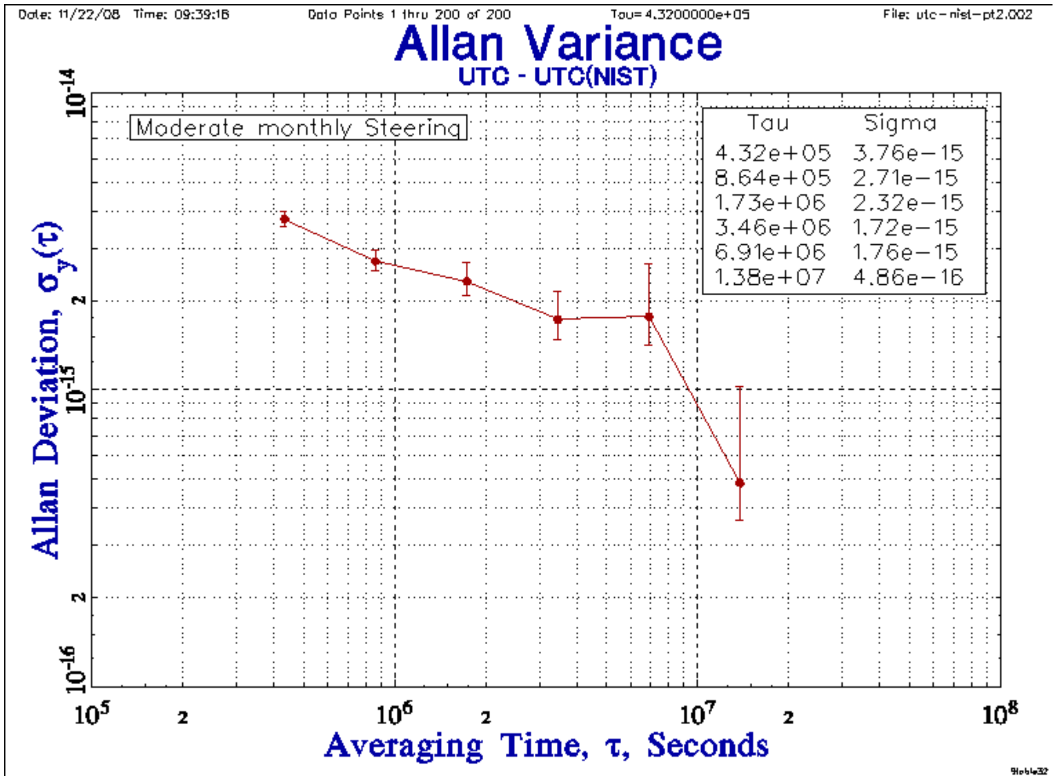


Figure 3. Allan Deviation and histogram of time differences with slow steering.



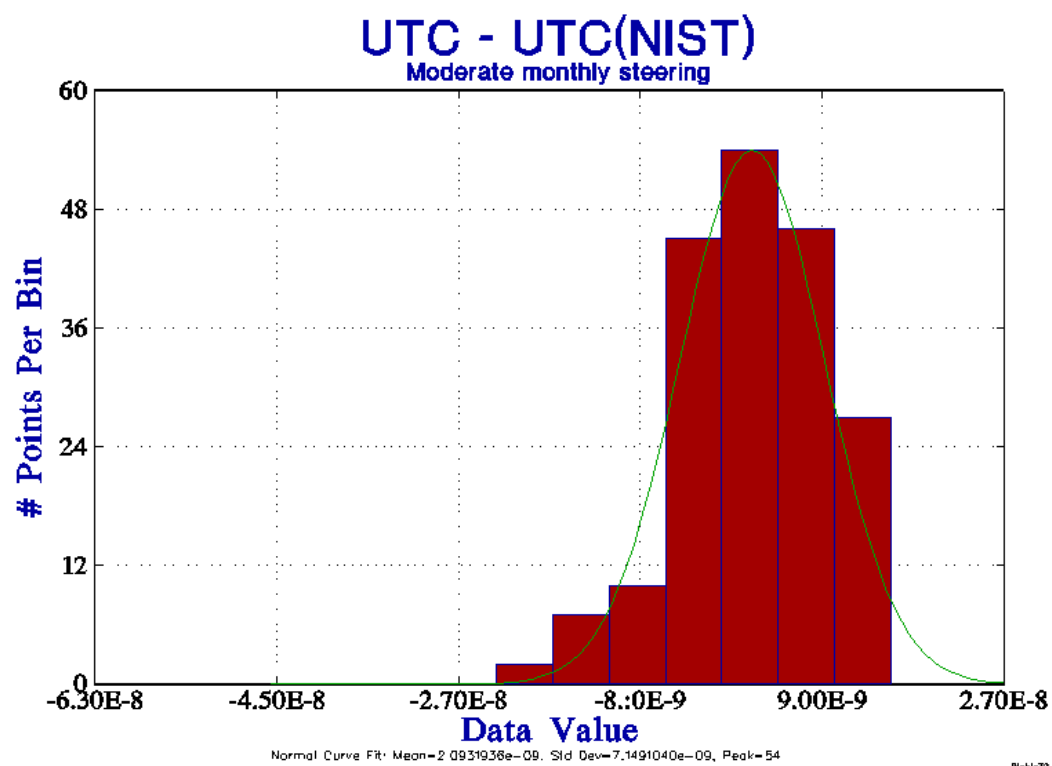


Figure 4. Allan deviation and histogram of time differences for moderate monthly steering.

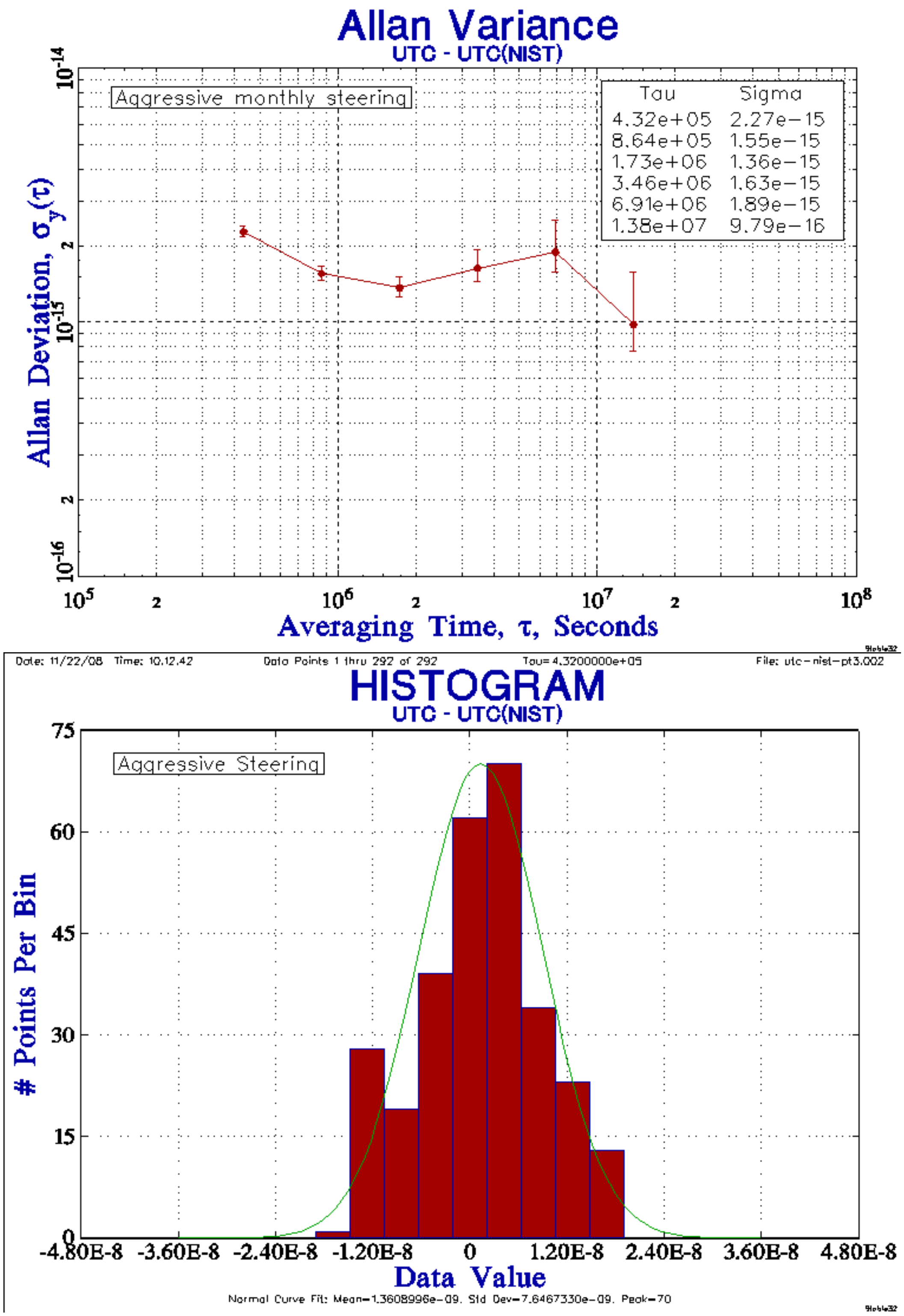


Figure 5. Allan deviation and histogram of time differences with aggressive steering.

THE BACKUP FOR THE UTC (NIST) TIME SCALE

This time scale is located at NIST radio stations WWV and WWVB. The transmitters for these stations are near Fort Collins, Colorado, approximately 60 km from Boulder. The system is designed to be a backup for the primary system in Boulder, and is intended to support all of the NIST time services should the primary scale become unavailable for any reason.

The hardware at Fort Collins is similar to the Boulder system – redundant dual-mixer systems measure the time differences between three high-performance commercial cesium standards and a fourth one, which is designated as the reference clock for the system. The ensemble average is used to discipline a phase stepper, which applies an offset to the output of one of the cesium clocks. The unsteered output of the clock contributes to the ensemble average in the same way as the other devices. The output of the phase stepper is somewhat noisier than its reference clock at the shortest averaging time, but it is driven by the average of the ensemble, and is therefore more stable than its reference at longer times. This comparison is shown in Figure 6, which shows the TDEV of each clock with respect to the ensemble average time.

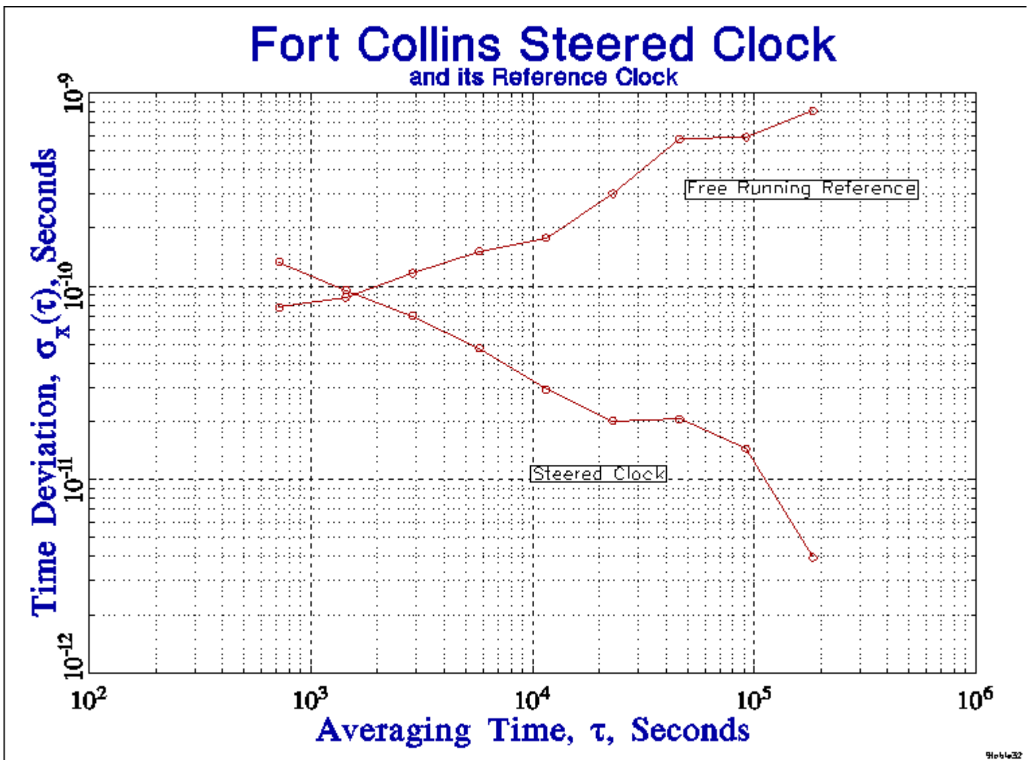


Figure 6. The time deviation of the steered clock compared to the clock used as its reference.

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

The only services that depend on the time scale in Fort Collins at this time (November, 2008) are the short-wave and low-frequency radio broadcasts. These have only modest accuracy requirements, so that the time scales there can be used as a laboratory to study various steering algorithms without compromising the services that use them for a reference. The experiments can be more extensive than the previous discussion, since both the time scales and the comparison procedures are under our control.

We have used common-view GPS measurements to compare the time scale in Fort Collins with UTC (NIST) in Boulder. The GPS data are somewhat noisier than at other sites because of the interference from the transmitters. In addition, these data are not available in real time because of limited communications facilities at Fort Collins. The results of the most aggressive steering algorithms are, therefore, based on simulations using the actual measured time differences but computed after the fact.

Steering algorithms that are too slow or too aggressive have similar time dispersions, but the causes are different. When the algorithm is too slow, the time dispersion due to the noise in the clocks themselves and the temperature-driven noise of the measurement system dominate the histogram of the time differences. When the steering is too aggressive, the contribution due to the measurement noise becomes important. The time stability is degraded in both of these situations. Figure 7 shows an example of these problems when the steering is too slow.

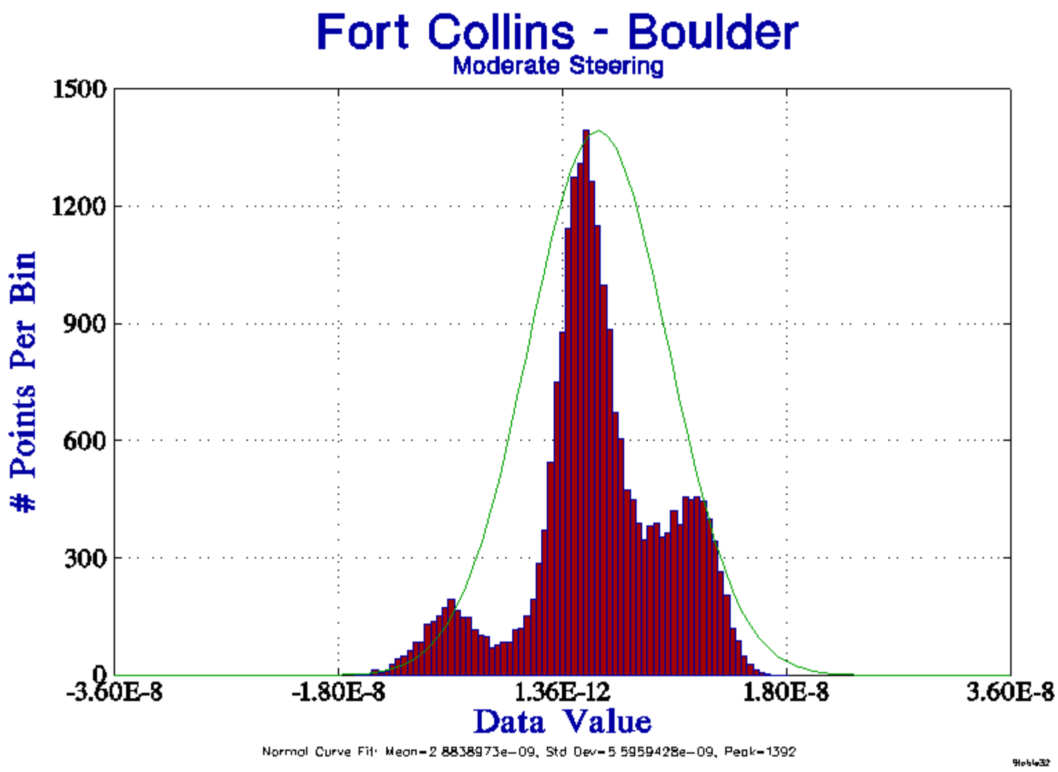


Figure 7. The histogram of the time differences when very slow steering is used.

The optimum steering uses weekly steering corrections whose magnitudes are limited to $\pm 8 \times 10^{-15}$ in frequency. Using this steering, the difference of Fort Collins – UTC (NIST) is shown below.

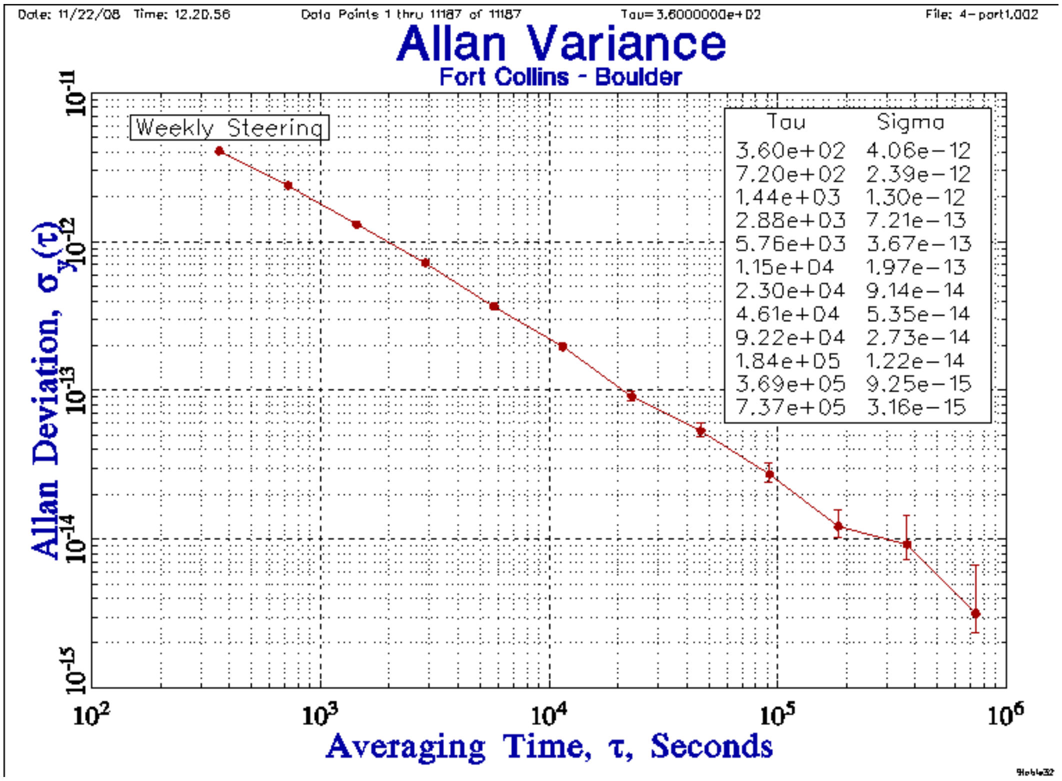


Figure 8. The Allan deviation of the difference (Fort Collins – Boulder), weekly steering at Fort Collins.

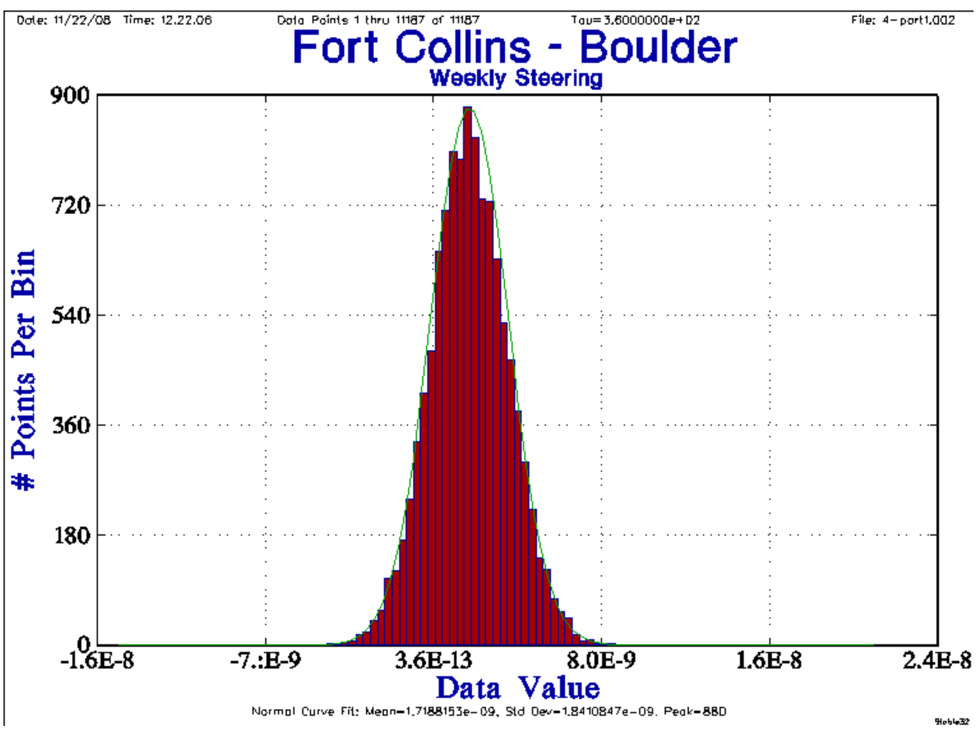


Figure 9. The histogram of the difference (Fort Collins – Boulder) using weekly steering at Fort Collins.

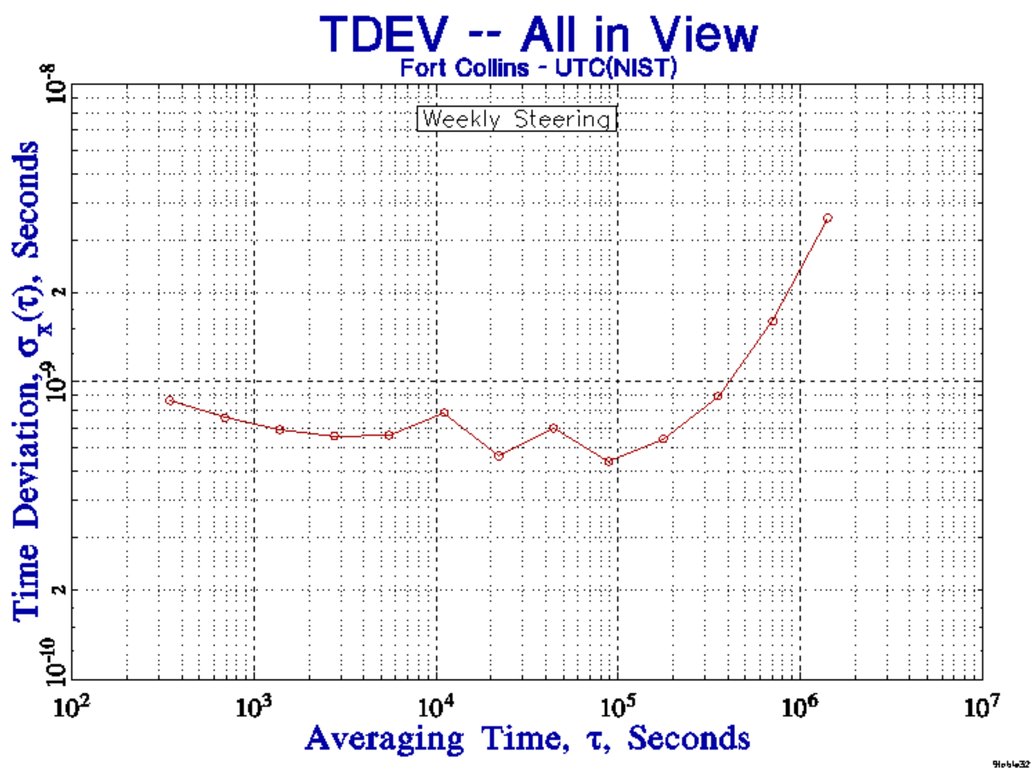


Figure 10. The time deviation of the difference (Fort Collins – Boulder) using weekly steering.

DISCUSSION AND CONCLUSIONS

The results of both studies show the inadequacy of a steering algorithm that is too slow, where "too slow" means that the free-running stability of the scale being steered is poorer than the stability of the calibration data. The stability of the calibration data may be degraded by any fluctuations in the characteristics of the channel used to compare the local and remote scales, and this limitation can become important when the steering corrections are too aggressive, since the contribution of the noise in the channel often becomes increasingly important in this regime. For example, the steering of the Fort Collins time scale is limited because of the noise in the common-view GPS comparisons. This is particularly serious for averaging times close to 1 day, since there are many perturbations to the time-difference data that have periods of this order. The two most important are fluctuations in the ambient temperature at Fort Collins and multipath reflections at both sites. These effects are clearly seen in Figure 10, which shows the marked increase in TDEV beginning at averaging times of about 0.5 days.

This noise might be reduced with more sophisticated postprocessing, but this is not consistent with the real-time requirements of the system. In spite of this limitation, the optimum steering algorithm has a time deviation of less than 1 ns for averaging times less than about 4 days. Furthermore, the histogram shows that the time difference data are well behaved, with no outliers. This is important for applications that depend on a worst-case analysis rather than the average values computed by the TDEV and ADEV statistics.

The Fort Collins time scale operates in an unfavorable environment – it is in the near field of transmitters that broadcast at 5 MHz, which is the frequency used by the dual-mixer measurement system, and its ambient temperature environment is not well controlled. Nevertheless, its performance characteristics might be useful in other contexts. For example, this type of ensemble might be useful in synchronizing the stations of navigation systems like LORAN, since the performance of the system for a navigation application depends on the time synchronization of multiple transmitters. Based on our results, a station with a hardware configuration similar to the Fort Collins system could operate "in holdover" (that is, without any external synchronization) for several days if the required time accuracy was 1 ns and for a much longer period if the timing requirement was relaxed. This performance could be improved if the hardware were located in a better temperature environment. In such an improved environment, the system could probably operate without any external reference with a timing accuracy of about 1 ns for about 10 days. Other navigation overlay systems might benefit from similar configurations.

REFERENCES

[1] J. Levine, 1999, *"Introduction to time and frequency metrology,"* **Review of Scientific Instruments, 70,** 2567-2596.

[2] P. Tavella and C. Thomas, 1991, *"Comparative Study of time scale algorithms,"* **Metrologia, 28,** 57-63.

[3] S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe, and L. Erb, 1982, *"Performance of an Automated High Accuracy Phase Measurement System,"* in Proceedings of the 36th Annual Frequency Control Symposium, 2-4 June 1982, Philadelphia, Pennsylvania, USA (IEEE Publication), pp. 314-320.