THE EVALUATION OF NIST-7
A NEW ERA

Jon H. Shirley, W.D. Lee, and R. Drullinger
National Institute of Standards and Technology
325 Broadway, Boulder, Colorado 80303
E-mail: rdrullinger@boulder.nist.gov

Abstract

We have developed a set of evaluation tools whereby all of the known systematic effects are managed by means of some leveraged experiment not involving, or limited by, precision frequency measurements on the standard. The result is reduced uncertainty in the evaluated biases and their possible correlations. This will greatly reduce the "combined standard uncertainty" and extend the life of thermal-beam frequency standard technology.

I. Introduction

The previously accepted technique for evaluating many small (often zero) frequency bias terms in primary, cesium-beam frequency standards has been to observe the parametric dependence of the frequency on some operating parameter.[1] Examples involve questions about the magnetic field inhomogeneity, line overlap shifts and various imperfections in the electronics. However, this technique is not suitable when pushing to the limits of available accuracy. It involves painfully long, precision frequency measurements which, *at best*, return information about the potential bias limited by the frequency measurement precision. When information of this type is combined to give an overall estimate of the accuracy of the standard (1) many of the terms are of comparable size and (2) there are difficult questions about correlated terms. This both limits the attainable overall accuracy and leads to serious concern about how the many uncertainty terms should be combined: arithmetic sum or square root of the sum of the squares (RSS)? We have developed a set of evaluation tools that allow all the systematic effects we know to be evaluated in leveraged experiments that do not involve, and are not limited by, precision frequency measurements. Knowledge of the values and independence of the various bias terms is vastly improved.

In section II, we investigate the way accuracy limits arise when parametric frequency measurements are used for part of an evaluation. In section III, we describe some of the new evaluation tools we have developed and the way they combine to give a much better look at the systematic errors. Finally, in section IV, we discuss the ramifications of this new evaluation methodology.

II. Traditional Evaluation

When evaluating a standard, we make frequency difference measurements between the standard and a reference clock. We choose to represent these measurements as

$$F \equiv \nu_{cs} - \nu_r \pm \sigma + \sum b_i \ ,$$

where F is the measured frequency, ν_{cs} is the frequency of the *unperturbed* cesium hyperfine resonance, ν_r is the frequency of the reference clock, σ is the uncertainty in the frequency measurement imposed by the averaged measurement noise, and $\Sigma\, b_i$ is the sum over all frequency biases. Implicit in this representation are that: (1) ν_r is constant over the measurement period (hours to days to weeks), (2) the dominant noise process is white (atomic shot noise), and (3) the list of systematic effects is complete and includes any individual variation (environmental effects) and their associated uncertainty; that is, each b_i is, in fact, $b_i(t,\sigma_i)$.

From this representation, it is easy to see the problems and limits to an evaluation based, even in part, on measurements involving the parametric dependence of the frequency on some operating parameter like microwave power or C-field. First, the bias of interest may not vary strongly with the clock operating parameter. Second, many other biases may also change with the same operating parameter. And, finally, all of the measurements are limited by σ, the frequency measurement precision. This not only seriously limits the uncertainty with which bias effects studied by this technique can be evaluated. It also leads to limited knowledge about correlations between uncertainties of different bias effects and, hence, questions about how best to combine uncertainties in the final, overall error budget, that is, arithmetic sum or RSS.

III. New Evaluation Tools

For space reasons, the full details of the evaluation tools cannot be given here. Instead, we will give only a brief outline and some specific examples of leveraged measurements. The larger biases are evaluated by conventional means and have been described in the literature.[2] The end-to-end cavity phase shift is measured by beam reversal but drops out of the final result (see below).

The smaller physical biases which actually shift the position of the line are evaluated through a number of techniques. Shifts resulting from the magnetic field inhomogeneity, cavity pulling and overlap of neighboring Zeeman lines are evaluated by measuring the offset of each Ramsey fringe from its corresponding Rabi pedestal as a function of Zeeman state (m_F), microwave power and C-field value.[3] Fluorescence light shift can be quantitatively amplified by changing the optical pumping transition and geometry. Similarly, distributed-cavity phase shift can be forced quantitatively by movable beam masks placed in front of the cavity beam windows. Microwave leakage is an important item. Leakage outside the standard and its routes into the standard are hunted down and stopped.[4] Leakage within the standard is accounted for in the beam reversal only if it remains constant in phase and amplitude every place within the standard during the evaluation. We have done experiments wherein we injected radiation into the standard to study these effects, and we have taken pains in the design and operation of the standard to insure this effect is under control.[5]

Imperfections in the electronics can result in shifts to the apparent, as-measured line position. RF spectral purity and some problems associated with switching transients have been published elsewhere.[6] Many of the potential effects (modulation asymmetry or feedthrough into the signal channel, synchronous AM on the laser or the RF, for example) can be studied by interrupting the servo loop at some point and observing the integrator output over time. Pure noise will result in a random walk while a coherent, systematic effect will accumulate linearly. One has to average long enough to see the linear drift above the integrated noise. The leverage in this type of measurement comes from the fact that one can usually configure the standard to reduce the equivalent noise into the integrator during the test; e.g., block the atomic signal while looking for integrator offset or reference signal cross-talk, block the optical pumping beam while looking for AM on the detecting laser. Of course, care must be taken to insure the effect under investigation is not altered and non-linear or digital aliasing are properly treated.

In this way, we are able to measure (or calculate) every bias we know (except end-to-end phase shift) with uncertainty small compared to our normal frequency measurement precision. Hence, their effects can be removed from the frequency measurement, producing a reduced frequency measurement f,

$$f = \nu_{cs} - \nu_r \pm \sigma + V\Phi \quad ,$$

in which all of the otherwise known biases have been removed. Here, $V\Phi$ is the frequency bias resulting from the effective, end-to-end phase difference. We have separated the effective phase difference (Φ) from the velocity, microwave power, and modulation dependence (V) of the frequency shift. The value of V can be calculated using the data from the velocity measurements and known modulation parameters. The effective, end-to-end phase difference (Φ) is constant if microwave leakage and distributed-cavity, phase-shift have been properly handled.[5] A pair of frequency measurements with beam reversal then yield

$$f = \nu_{cs} - \nu_r \pm \sigma + V\Phi \qquad \text{and} \qquad f' = \nu_{cs} - \nu_r \pm \sigma' - V'\Phi \ ,$$

where the sign of the phase difference has reversed, the measurement precisions (σ) are of the same magnitude and the values of V differ slightly as a result of the different oven temperatures and beam alignments in the two directions. From these measurements, we can extract the effective, end-to-end, cavity phase shift with an uncertainty limited by that of the differenced frequency measurements:

$$\delta f = \pm\sqrt{\sigma^2 + \sigma'^2} + [V + V']\Phi \simeq \pm\sqrt{2}\sigma + 2V\Phi \ .$$

However, if we combine the two frequency measurements in a weighted average, we get

$$\bar{f} \equiv \frac{V'f + Vf'}{V + V'} \simeq \nu_{cs} - \nu_r \pm \frac{\sigma}{\sqrt{2}} \ ,$$

where the bias from end-to-end phase shift has dropped out *and* the individual measurements have combined to give a reduced net measurement noise. This result holds provided our initial conditions (ν_r is constant, the set of b_i is complete, and their uncertainties are small compared to σ) are met. This process can also be extended to a larger set of frequency measurements to further average down the measurement noise.

IV. Results and Conclusions

This view of the evaluation process is new to us, and there remains a great deal of work to do to achieve the ultimate evaluation of NIST-7. However, a number of things are already clear. (1) In the list of potential systematic effects of which we are aware, it seems we can evaluate each of them to a contribution level below the present stability of our maser ensemble. (2) At the very least, the combined standard uncertainty of the ultimately evaluated standard will be greatly reduced from our originally estimated limit of 1×10^{-14}. (3) The fact that the noise type of individual frequency measurements is white, means that long runs are not necessary to lower measurement noise. Shorter frequency measurements can be combined with confidence. This allows more frequent looks at bias terms that may have environmental coupling. (4) Correspondingly, the way the statistics from individual frequency measurements actually do combine is a powerful diagnostic for the validity of the process. If N runs of precision σ do not combine to yield an overall statistical uncertainty of $\sigma/\sqrt{N}$, some bias is not under control. We have used this technique to find a problem resulting from unregulated cesium oven heaters. As the power line voltage changed with the daily load variation in the building, we had a 2×10^{-15} fractional frequency variation. We were able to clearly see the effect on a series of frequency measurements whose individual, statistical uncertainty was only 1×10^{-14}. (5) An accuracy exceeding the stability of our reference timescale also means that it exceeds the capability of present technology to relay it around the world for comparison with other laboratories. (6) Hence, until timescale and time transfer technology improve, the functional life of thermal atomic beam technology has been extended, even in light of emerging fountain technology.

V References

1. see, for example, J. Vanier and C. Audoin, *The Quantum Physics of Atomic Frequency Standards* (Adam Hilger, Bristol, 1989), p. 845.
2. W.D. Lee, J.H. Shirley, J.P. Lowe and R.E. Drullinger, *IEEE Trans. Instr. Meas.*, **IM-44**, (1995) 120.
3. J.H. Shirley, W.D. Lee, G.D. Rovera and R.E. Drullinger, *IEEE Trans. Instr. Meas.*, **IM-44**, (1995) 136.
4. W.D. Lee, J.P. Lowe, J.H. Shirley and R.E. Drullinger, in *Proc. European Forum on Time and Frequency*, 1994, p. 513.
5. This topic is very involved and will be the subject of subsequent papers.
6. W.D. Lee, J.H. Shirley, F.L. Walls and R.E. Drullinger, in *Proc. 1995 IEEE International Frequency Control Symposium*, p. 113.