Application of Kalman Filters and ARIMA Models to Digital Frequency and Phase Lock Loops

J.A. Barnes
Austron, Inc.
3300 Mitchell Ln.
Boulder, CO 80301

S.R. Stein*
Ball Aerospace Systems Division
P.O. Box 1062
Boulder, CO 80306

Introduction

The problem of controlling the phase and frequency of atomic clocks is analyzed in two stages. The first step requires estimating the state (time and frequency) of the clock at the time of the last measurement and forecasting the future state. The second step is the design of a control algorithm for steering the slave clock to the reference. This division of the overall control problem is useful because different design requirements are utilized for each step.

We apply the Kalman filter to perform optimum state estimation and forecasting. The Kalman filter is advantageous for this application because it provides minimum least squares error estimates and forecasts both during transients and for steady state operation. It also provides internal estimates of the variances of its outputs. Analytic solutions have been found for steady state performance.

The discrete time series analysis described above is concerned only with the values of the variables at the sample times. However, the samples are drawn from an underlying continuous process. The values of the time and frequency of the slave clock between measurements must be considered in order to design a control loop with desirable performance. The control loop can be described by an equivalent ARIMA model which is used to calculate the transfer of frequency stability from the reference to the slave. We compare the performance of simple frequency lock loops, and first and second order phase lock loops of different configurations in terms of the resulting output frequency stability.

There are numerous applications for slave oscillators which are controlled by a reference clock. The benefits can be a signal with the long-term stability of the reference and the short-term stability of the slave. This technique is often used when the reference signal is severly contaminated with short-term noise. Another use of the slave oscillator system is to bridge outages of the reference signal. Such devices are often called "disciplined oscillators" since they "learn" the first few terms of a Taylor series fit to the past corrections to the slave oscillator.

This paper considers various "optimal" solutions to the slave oscillator system for two physical models. Model A is a reference oscillator whose frequency variations are white noise (i.e., random walk phase fluctuations). The slave oscillator, when free running, has pure random walk noise of the frequency variations (i.e., random walk FM), in contrast to the reference. Thus, the reference signal has a higher level of short-term fluctuations than the free running slave, and the reverse is true in long-term. While this physical model is not adequate for many applications it is qualitatively similar to real situations and it is easy to recognize and understand just what is optimized. Model B is a reference oscillator with negligible frequency noise. The slave oscillator is contaminated with both white frequency noise and random walk frequency noise. This model also describes certain situations encountered in practice. One example is the problem of providing a real time output from an ensemble of oscillators, since the ensemble average is more stable than any of the member oscillators.

Asymptotically, Kalman Filters often approach a simple ARIMA model. [Box and Jenkins]. As shown below, these ARIMA models in turn are optimal. One of the primary advantages of Kalman filters over ARIMA models is that they easily handle transient responses such as initial turn-on or irregular data sampling [Gelb]. Still ARIMA models often are adequate for many real systems.

Optimal State Estimation and Forecasting

Physical Model A   Following Gelb, the state vector for the system evolves according to the equation:

$$
\begin{bmatrix} x \\ y \\ u \\ v \end{bmatrix}_k = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ u \\ v \end{bmatrix}_{k-1} + \begin{bmatrix} 0 \\ \eta \\ 0 \\ \epsilon \end{bmatrix}_{k-1} \tag{1}
$$

where $x(k)$ is the time (phase) of the slave oscillator signal, $y(k)$ is the slave frequency, $u(k)$ is the reference time, and $v(k)$ is the reference frequency. The noise terms $\epsilon(k)$ and $\eta(k)$, are independent random normal deviates with mean zero and variances σ_ϵ^2 and σ_η^2 respectively. It is easy to verify that x is the double sum of the η's and hence x is a random walk of frequency, or equivalently, y is a random walk. Similarly, v is the reference frequency, and hence u is the reference time, a random walk. We assume that the time interval between points is unity, just for simplicity. Figure 1 is a plot of the power spectral densities (PSD) of the frequency fluctuations, y and v.

Physical Model B   The alternative model has a noiseless reference and a slave with white frequency noise and random walk frequency noise. A two element state vector is sufficient to describe the evolution of the system:

$$
\begin{bmatrix} x \\ y \end{bmatrix}_k = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_{k-1} + \begin{bmatrix} \epsilon \\ \eta \end{bmatrix}_{k-1} \tag{2}
$$

The reference oscillator is described by $u = v = 0$. Figure 2 shows the power spectral density of frequency fluctuations of the slave oscillator.
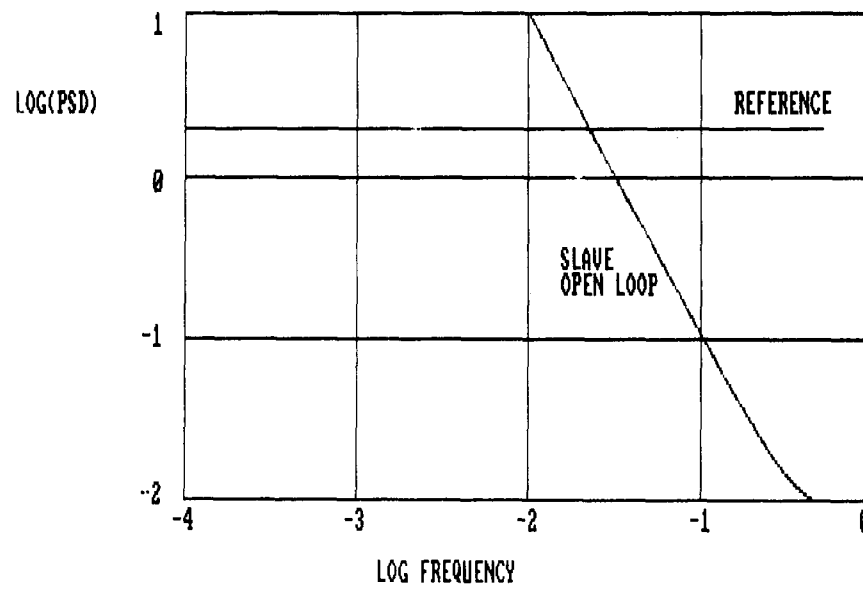
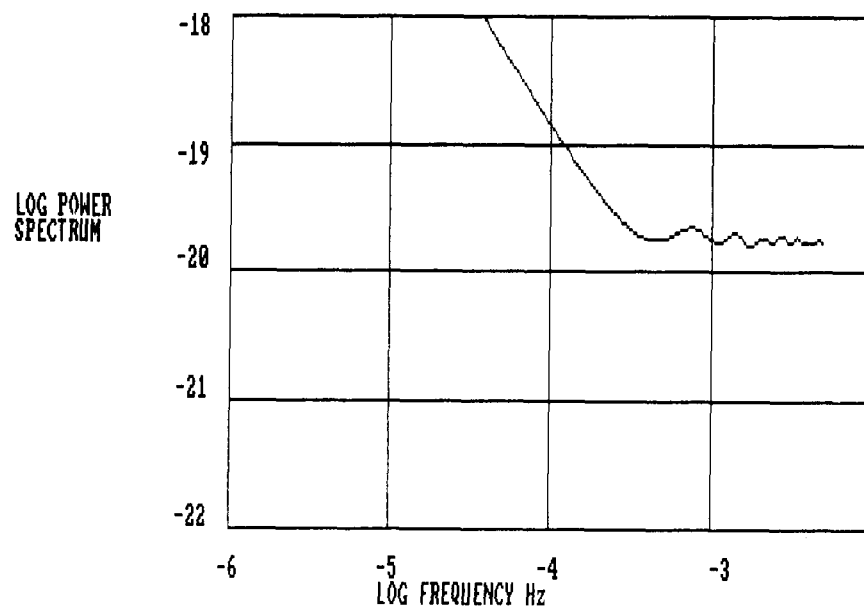Figure 1: PSD of Frequency for Reference and Slave Oscillators, Model A

Figure 2: PSD of Frequency Fluctuations for the Slave Oscillator, Model B

313

Measurement Model With a time interval counter one can readily measure the time difference between master and slave. The measurement model [Gelb], then is just

$$z(k) = x(k) - u(k). \tag{3}$$

Equations 1, 2 and 3 along with some reasonable assumptions about initial conditions allow one to compute the Kalman gains, the covariance matrices, and the estimates of the state vector. (See Appendix A.) One can then pass to the limit and obtain the asymptotic Kalman gains which are the same for both models:

$$
\begin{aligned}
K_x^\infty &= 1 \\
K_y^\infty &= 1 - \theta \\
K_u^\infty &= 0 \\
K_y^\infty &= 0
\end{aligned}
\tag{4}
$$

where θ is given by

$$\theta = 1 + \frac{\sigma_\eta^2}{2\sigma_\epsilon^2}\left(1 - \sqrt{1 + \frac{4\sigma_\epsilon^2}{\sigma_\eta^2}}\right) \tag{5}$$

or equivalently

$$\frac{\sigma_\eta^2}{\sigma_\epsilon^2} = \frac{(1-\theta)^2}{\theta}. \tag{6}$$

Note that the Kalman gains for the reference time and frequency as given in equation 4, above, are zero! Thus, the Kalman filter provides no updates to the initial estimates of the time and frequency of the reference oscillator. On the other hand, the slave time, x, has unit gain, reflecting the assumption of zero measurement noise, that is, after each measurement, the time of the slave is known precisely with respect to the reference. Finally, we find that the asymptotic estimate of the slave frequency, $y(k|k)$, after the kth measurement is given by the recursion

$$y(k|k) = \theta y(k-1|k-1) + (1-\theta)[z(k) - z(k-1)] \tag{7}$$

which is a simple exponential filter applied to $\nabla z(k)$.

We will now examine several of the ways in which the state estimates and forecasts provided by the Kalman filter can be used to control the frequency and time of the slave oscillator. We define three cases which are closely related to analog control loops.

Case I is a first order phase lock loop, designated PLL–1. Such a control loop uses the measured phase error to produce a proportional frequency correction of the slave.

Case II is a frequency lock loop, designated FLL. This control loop uses the estimated frequency of the slave to produce a proportional frequency correction.

Case III is a second order phase lock loop, designated PLL–2. Such a control loop has two poles at the origin rather than the single pole of the PLL–1.

314

Case I: First Order Phase Lock Loop

Figure 3 illustrates the implementation of the digital first order phase lock loop for a tunable slave oscillator. The reference oscillator noise is given by $u(k)$ and the slave oscillator noise is described by $q(k)$. The values of the noises for the two physical models are contained in Table 1
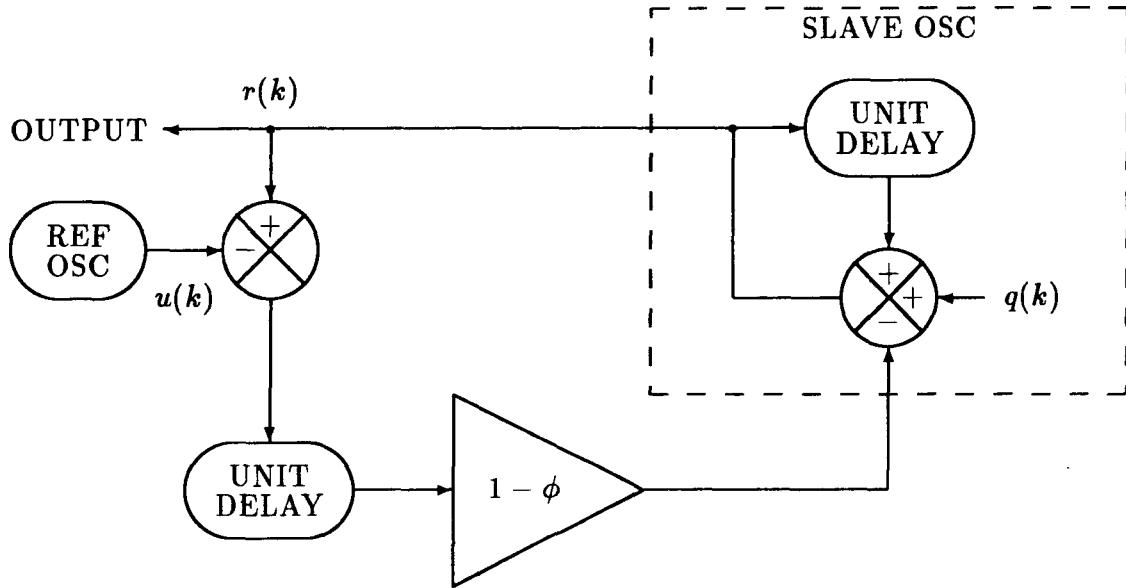

Figure 3: First Order Phase Lock Loop Using a Tunable Oscillator

The Kalman estimate of the slave phase, i.e., the last measurement, is fed back with gain $-1 + \phi$

Table 1: Noise Inputs for Control Loop Analysis

Physical Model A	$\nabla u(k) = \epsilon(k-2)$	$\nabla q(k) = \eta(k-2)$
Physical Model B	$u(k) = 0$	$\nabla q(k) = \eta(k-2) + \nabla\epsilon(k-1)$

to provide frequency correction. The value of the gain determines the behavior of the loop. Appendix B shows that the maximum stable gain, which is achieved when $\phi = 0$ results in minimum variance of the residual phase fluctuations, $r(k)$. However, if $\phi = \theta$ then the variance of the residual frequency fluctuations, $\nabla r(k)$, is minimized. As shown in the appendix, the loop performance is described by the ARIMA model:

$$(1 - \phi B)r(k) = \nabla x(k) + (1 - \phi)u(k-1) + \text{constant.} \qquad (8)$$

The PLL–1 is optimized for phase or frequency performance with different values of the gain. This must be contrasted with the underlying Kalman filter, which provides simultaneous optimum estimates of both the frequency and the phase. There is no paradox. The situation
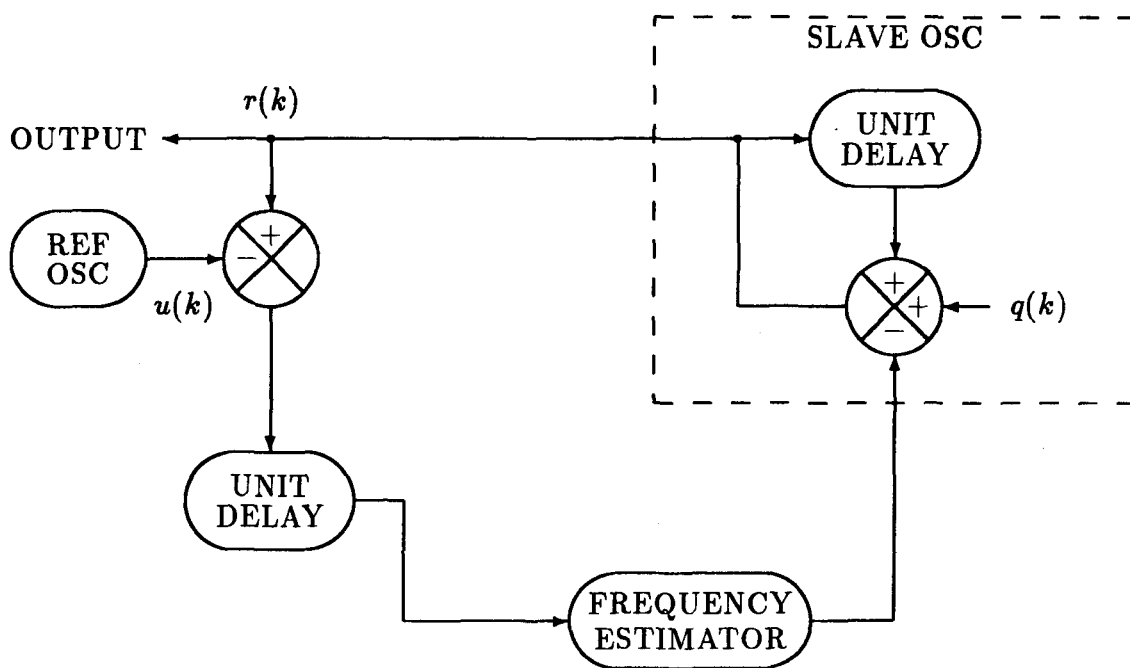
315

Figure 4: Frequency Lock Loop Using Kalman Optimum Frequency Estimate

arises merely because the phase is being controlled by shifting the frequency of the slave. The frequency stability obtained with $\phi = \theta$ is nearly equal to that of the unperturbed slave oscillator. However, a penalty is paid; the steady state phase error due to an initial frequency offset between the two oscillators is increased in proportion to the gain reduction.

Case II: Frequency Lock Loop

It is simple to design a digital filter which implements the Kalman steady state frequency estimator for the open loop frequency given in equation 7. When the loop is closed, as shown in Figure 4, the frequency correction may be written in terms of the closed loop residuals:

$$y(k|k) = y(k-1|k-1) + (1-\theta)[\nabla r(k) - \nabla u(k)]. \tag{9}$$

The phase residuals are described by another ARIMA model:

$$(1 - \theta B)r(k) = \nabla x(k) + (1-\theta)u(k-1) + \text{constant}. \tag{10}$$

A comparison of equations 8 and 10 leads to the conclusion that for the noises considered in models A and B, the PLL-1 with $\phi = \theta$ and the FLL result in the same output frequency stability. Of course, the FLL suffers from unbounded phase errors when there is an open loop frequency difference between the two oscillators and the PLL-1 does not. Figure 5 shows that for low Fourier frequencies, $i.e.$ long periods, the slave noise is higher than the reference noise for both the FLL and the PLL-1. This will manifest itself as unbounded phase errors relative to the reference.
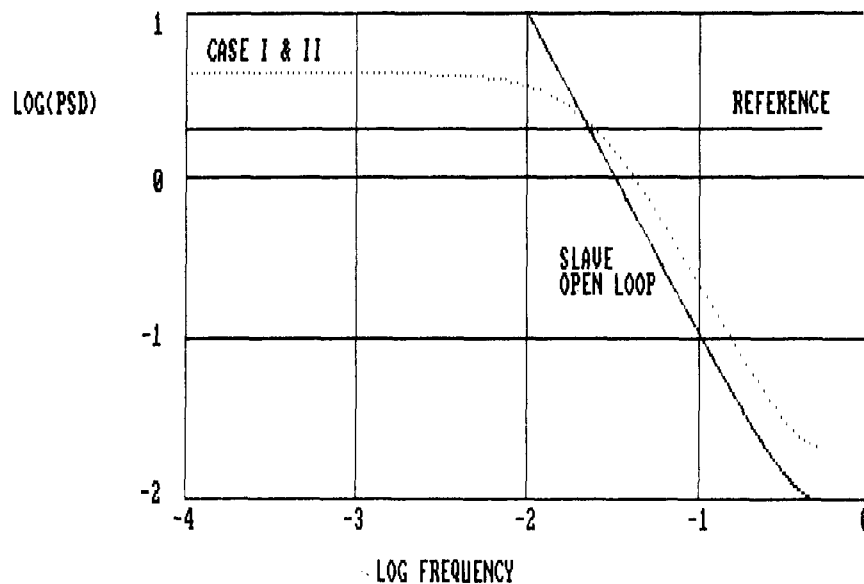
316

Figure 5: Power Spectral Density of Frequency for PLL–1 and FLL

Case III: Second Order Phase Lock Loop

There are many uses of precise clocks which depend on very long-term phase coherence and the servos considered so far would not be adequate. Figure 6 is a block diagram of a second order loop which is a compound of two "optimal" loops. The current estimate of the frequency is fed back along with a fraction $1 - \phi$ of the last phase error. Figure 7 shows the frequency noise power spectral density in comparison with the other alternatives considered. It is interesting to note that when we went to Cases I and II we reduced the high frequency noise of the reference at the expense of higher low frequency noise. The present case provides low levels of the low frequency noises but at an expense of some worsening at the higher Fourier frequencies. In fact, there is now a broad resonance at intermediate frequencies. The tightness of control and the size of the excess noise depend on the parameter ϕ. When $\phi = 0$ the residual phase fluctuations are minimized. As ϕ is increased towards 1, the variance of the residual phase increases while the variance of the residual frequency decreases. When $\phi = 1$, the control loop reduces to a first order phase lock loop with minimum residual frequency variance. As a comparison to the other cases, the slave clock error for this second order loop is given by:

$$(1 - \theta B)(1 - \phi B)r(k) = \nabla^2 x(k) + (1 - \theta)\nabla u(k - 1) + (1 - \theta B)(1 - \phi)u(k - 1) \qquad (11)$$

Summary and Conclusions

The control techniques utilized in the previous sections restrict all control actions to frequency changes. Alternatively, it is possible to step the phase to provide immediate correction of measured phase errors. This would maintain minimum phase error at all times but at a significant
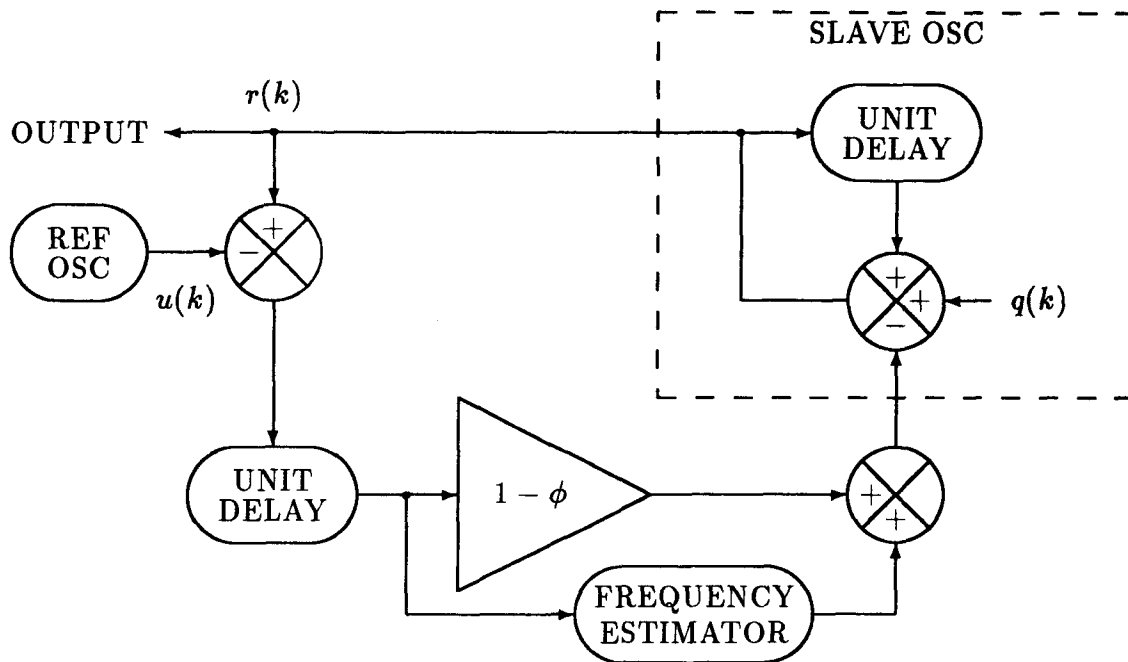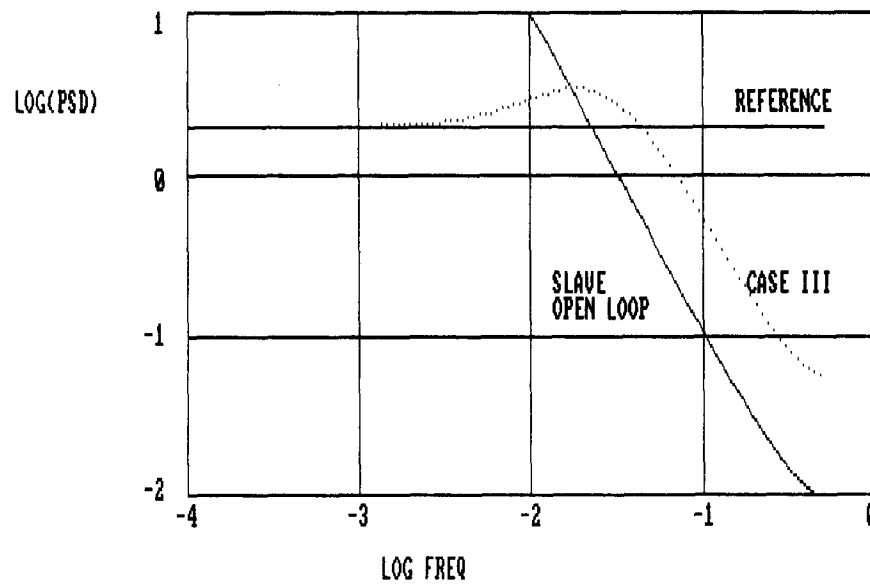
Figure 6: Second Order Phase Lock Loop

Figure 7: PSD for the Second Order Phase Lock Loop

Table 2: Case Comparisons for Physical Model A

Case No.	Description	ϕ	$\mathrm{Var}[\nabla r(k)]$
Case I	PLL–1 optimized for frequency stability	θ	$\sigma_\epsilon^2 \dfrac{1-\theta}{\theta}$
Case I	PLL–1 optimized for phase stability	0	$\sigma_\epsilon^2 \dfrac{1-\theta+\theta^2}{\theta}$
Case II	FLL using the Kalman optimum frequency estimate	N/A	$\sigma_\epsilon^2 \dfrac{1-\theta}{\theta}$
Case III	PLL–2 using the Kalman optimum frequency estimate and optimized for phase stability	0	$\left[(2-\theta)^2 + \dfrac{(2-\theta)(1-\theta)}{\theta(1+\theta)}\right]\sigma_\epsilon^2$

cost in frequency stability. It will not be discussed in more detail here. We have presented three different methods of "optimally" designing a slave clock system for each of two physical models of clock performance. Table 2 compares the three methods when the reference has white frequency noise and the slave has random walk frequency noise (Physical Model A). Table 3 compares the methods when the reference is noiseless and the slave has both white and random walk frequency noises. In this case (Physical model B) it is possible to compute the Allan variance of the slave oscillator. Figure 8 compares the three loop types when $\theta = 0.9$. The square root of the Allan variance of the closed loop residuals, $\sigma_p(m)$, are plotted versus the number of samples.

So far we have considered only technical differences and not the consequences of selecting one system over another for a specific application. It is probably usefull to consider two rather different uses of highly stable clocks and signals:

Consider a slave clock being controlled by a reference signal, and performing according to specifications. At some point in time the reference signal will fail and the slave will continue on its own for some period. During this period, the slave clock will accumulate frequency and time errors. When the reference signal returns there are basically two strategies that can be used: (1) One can reset the time (and frequency) to nearly the correct value obtained from the reference, or (2) one can reset only the slave frequency to the reference and ignore any accumulated phase (time). People working in the precise time field and navigators truely need

Table 3: Case Comparisons for Physical Model B

Case No.	Description	ϕ	$\text{Var}[\nabla r(k)]$
Case I	PLL–1 optimized for frequency stability	θ	$\dfrac{\sigma_\epsilon^2}{\theta}$
Case I	PLL–1 optimized for phase stability	0	$\sigma_\epsilon^2 \dfrac{1 + \theta^2}{\theta}$
Case II	FLL using the Kalman optimum frequency estimate	N/A	$\dfrac{\sigma_\epsilon^2}{\theta}$
Case III	PLL–2 using the Kalman optimum frequency estimate and optimized for phase stability	0	$\dfrac{2\sigma_\epsilon^2}{\theta}$

time as accurately as possible. Clearly they need strategy (1), above. Some telecommunications systems, however, depend on very stable frequencies. Sudden jumps of phase (as small as a few ten's of picoseconds) are sufficient to temporarily disrupt the communications. Clearly they need strategy (2), above and for them it would be foolish to make an unnecessary phase step.

The final conclusion is to recognize that there are optima and there are optima. Sometimes these optima are only optimum locally (such as varying a specific gain setting) and global optima may exist unnoticed by the engineer. Also, one should note that different people need to optimize different measures of performance and it might be wastefull to try to suppress noise components over the entire spectrum.

Appendix A

The model used (see Gelb) is:

$$
\begin{bmatrix} x \\ y \\ u \\ v \end{bmatrix}_k = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ u \\ v \end{bmatrix}_{k-1} + \begin{bmatrix} 0 \\ \eta \\ 0 \\ \epsilon \end{bmatrix}_{k-1} \tag{12}
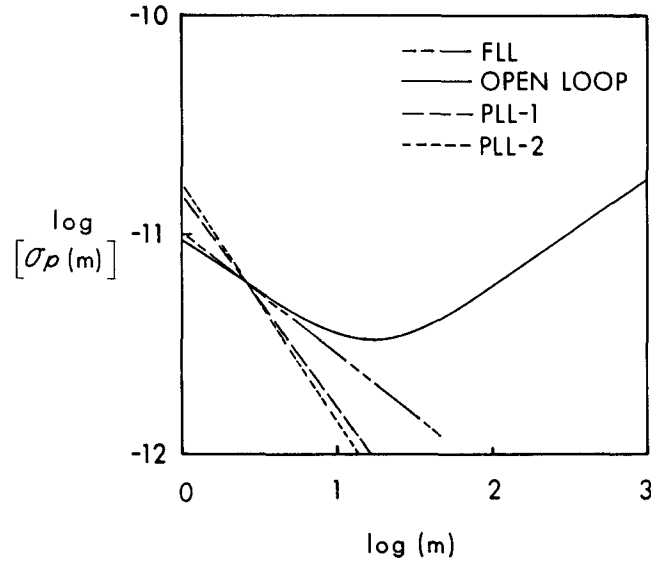$$

Figure 8: Comparison of three control loops for Physical model B and $\theta = 0.9$

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\epsilon^2 \end{bmatrix} \tag{13}$$

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & -1 & 0 \end{pmatrix} \tag{14}$$

Beginning with zero for the covariance matrix, one can compute the forecast covariance matrix, $\mathbf{P}(k|k-1)$, and from that the Kalman gains and the updated covariance matrix, $\mathbf{P}(k|k)$, following Gelb. The asymptotic form of the updated covariance matrix has the form:

$$\mathbf{P}(k|k) = \sigma_\epsilon^2 \begin{bmatrix} \frac{k-1}{1-\theta} & 1 & \frac{k-1}{1-\theta} & 0 \\ 1 & \frac{1-\theta}{\theta} & 1 & 0 \\ \frac{k-1}{1-\theta} & 1 & \frac{k-1}{1-\theta} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{15}$$

for large k. Correspondingly, the Kalman gain is:

$$\mathbf{K}^\infty = \begin{bmatrix} 1 \\ 1-\theta \\ 0 \\ 0 \end{bmatrix} \tag{16}$$

again for large k, where

$$\theta = 1 + \frac{\sigma_\eta^2}{2\sigma_\epsilon^2}\left(1 - \sqrt{1 + \frac{4\sigma_\epsilon^2}{\sigma_\eta^2}}\right) \tag{17}$$

321

Appendix B

With reference to the servo block diagram, Figure 3, one can write down the equation:

$$(1 - \phi B)r(k) = \nabla x(k) + (1 - \phi)u(k - 1) + \text{constant}. \tag{18}$$

Taking the first difference and substituting

$$\begin{aligned} \nabla u(k) &= \epsilon(k - 2) \\ \nabla^2 x(k) &= \eta(k - 2), \end{aligned}$$

we find

$$(1 - \phi B)\nabla r(k) = \eta(k - 2) + (1 - \phi)\epsilon(k - 3) \tag{19}$$

Since the $\eta(k)$ and $\epsilon(k)$ are independent, the right side of equation 19 is equivalent to a single random, normal deviate $b(k)$ with zero mean and variance:

$$\sigma_b^2 = \sigma_\eta^2 + (1 - \phi)^2 \sigma_\epsilon^2 \tag{20}$$

Equation 20 is a simple exponential filter whose impulse response function is just

$$h_n = \phi^n \tag{21}$$

Thus, the variance of $\nabla r(k)$ is given by:

$$\text{Var}[\nabla r(k)] = \sigma_b^2 \sum_{m=0}^{\infty} \phi^{2m} = \frac{\sigma_b^2}{(1 - \phi^2)} \tag{22}$$

The interest is to find that ϕ which minimizes the variance in equation 22; setting the derivative of the right side of this equation to zero, one obtains $\phi = \theta$. The final result is that we can write:

$$(1 - \theta B)r(k) = \nabla x(k) + (1 - \theta)u(k - 1) + \text{constant}. \tag{23}$$

The error signal driving the first order phase lock loop is the measured phase difference between reference and slave which is just the Kalman estimate of the phase difference. Thus the first order PLL with gain $-1 + \phi$ is optimum in two ways: it is based on the optimum state estimate of the slave/reference phase estimate and it minimizes the variance of the residual frequency variations.

References

[1] Box, G.E.P. and Jenkins, G.M. (1970). "Time Series Analysis, Forecasting and Control." Holden Day, San Francisco.

[2] Gelb, A. ed. (1974). "Applied Optimal Estimation." M.I.T. Press, Cambridge

QUESTIONS AND ANSWERS

Brad Parkinson, Stanford University: This is very fascinating. You realize that what you have done can be very simply extended to the phase and frequency locked loops using the GPS system when you create a model for the moving base. That is a very physical acceleration model, eating up phase at varying rates. The thing that always trips the designer up is things like quantizing errors and sampling times. Are you in your studies to the point now that you can talk about these things?

Mr. Stein: We can certainly talk about sampling times. The sampling time is the variable that drives the theta parameter. I didn't write it down here, but one can write down what theta is in terms of sampling time. Then the results fall out. It gets more complicated as you add additional noise types, which you undoubtedly have.

Mr. Parkinson: It is sometimes instructive to run a rather careful hybrid hardware/software simulation of this and you frequently find that things happen that are not what one expects. I don't know if you are at that stage yet.

Mr. Stein: We used a fair amount of simulation along the way in the development of this. I have to admit that both Jim and I produced a rather serious error. This is proof that collaboration produces a heck of a lot better results than working in a vacuum. Jim has done a great deal of simulation on this problem.

Mr. Parkinson: Just as a follow-up—How close are you to actually implementing it with real hardware.

Mr. Stein: It is done both by Austron and by Efratom.

Mr. Parkinson: Would you care to comment on what digital processor you used to do it?

Mr. Stein: One of the processors used was an 80286.

Don Percival, University of Washington: On the physical model "B" where you showed the spectral density function for white noise plus random walk, there seemed to be a "ringing" up at the high frequency end. What was that all about?

Mr. Stein: That was an artifact of slide production over the Thanksgiving Day weekend. I had to use simulated data because I didn't have a plot routine for the theoretical results.