DIGITAL PHASE LOCKED LOOPS WITH SEQUENTIAL LOOP FILTERS: A CASE FOR COARSE QUANTIZATION

J. R. CESSNA
Space Environment Laboratory
Environmental Research Laboratories
National Oceanic and Atmospheric Administration
Boulder, Colorado

Summary A class of digital phase locked loops exists for which performance can be demonstrated to be comparable to that of the linear analog model, yet the use of coarse quantization seems to be favored. This class is characterized by the use of (1) a phase detector with quantized output, (2) a sequential loop filter, and (3) a discrete phase adjustment clock. The sequential filter is an inherently nonlinear device which operates on a sequence of inputs and provides a sequence of outputs, usually fewer in number and with a variable interval between outputs.

Operation of these loops is decidedly nonlinear and a conventional linear loop bandwidth cannot be defined. Although higher than first order loops of this class may be constructed, their analysis, so far, has been limited to first order types by random walk techniques. A quasi-bandwidth measure can be defined for these loops in terms of their transient response which serves as an absolute basis for comparison of different digital loops. Using this tool, examples of 3- and 4-level quantization in these loops are found to have no obvious advantages over the binary quantized model and in some cases striking disadvantages.

Introduction The purpose of this paper is to illustrate the case for coarse quantization in a class of digital phase locked loops. Material presented here will of necessity be basically summary in nature and the stimulated reader will find a more complete development of this material in the author's thesis [1]. The basic binary model is described in a recent publication [2]. For brevity sake, equations will be given without derivation which can be found in the above two references.

The digital loops to be discussed here should be distinguished from the more familiar class of loops which are merely sampled data versions of analog loops which, for the most part, may be analyzed as if they were analog loops. Rather, these systems represent a direct synthesis with logic elements of what seems intuitively to be a desirable loop operation. Holmes [3,4] and Simon [5] have described somewhat similar digital loop models in which

the loop filters are simply length-M accumulators and produce discrete loop phase adjustments every M cycles.

Review of Binary Loop Operation Figure 1 shows the loop model for the binary case. The input signal is assumed to be a train of alternating sign square pulses with energy E_s and duration of ψ seconds and is perturbed by additive white Gaussian noise with zero mean and spectral density $N_0/2$. For the first order case, the symbol rate is assumed to be fixed and near enough to the loop clock frequency that the loop phase error is at most a slowly varying (relative to the loop transient time) quantity. δ is defined to be the difference between the phase of the input signal and the closest clock position available.

Phase error measurements are obtained with an "integrate-and-dump" type filter in which the ψ second integration period is centered on a positive-to-negative transition of the signal for zero phase error. If the loop clock is leading the signal, the signal component, $R(\phi)$, of the integrator output is positive, since the integration interval is shifted in the direction of the positive pulse. When $\phi = 0$, $R(\phi) = 0$. In fact, $R(\phi)$ can be interpreted as the cross correlation of the unperturbed signal with a square wave in phase quadrature. Since the noise is uncorrelated over successive integration intervals, its contribution to the integrator output is independent of ϕ. Hence, the input to the quantizer may be represented as

$$r(t,\phi) = R(\phi) + n_o(t),\ t = 0,\tau,2\tau,... \tag{1}$$

where the $n_o(t)$ are Gaussian random variables with zero mean and variance $N_o/2$. The binary quantizer gives an output according to whether r(t) is positive or negative with probabilities u_1 and u_{-1}. respectively.

$$u_1 = P_r[r > 0] \tag{2}$$
$$u_{-1} = P_r[r \leq 0] \tag{3}$$

These are evaluated as follows:

$$u_1 = 1 - Q\left[\frac{R(\phi)}{\sqrt{N_o/2}}\right] \tag{4}$$

$$u_{-1} = Q\left[\frac{R(\phi)}{\sqrt{N_o/2}}\right] \tag{5}$$

where

$$Q(\alpha) = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\beta^2/2}\, d\beta \tag{6}$$

u_1 and u_{-1} are, respectively, the probabilities of lead and lag pulses which are inputs to the sequential filter.

Examination of (4) and (5) shows that when ϕ is near zero (hence, $R(\phi)$ is near zero) or N_o is large, u_1 and u_{-1} tend to be near 0.5 and the phase error estimate is poor. If the loop phase corrections were made on the basis of those outputs directly, many incorrect adjustments would be made, causing undesirable phase-error fluctuations. Consider now the probability of making correct phase-error measurements. It will help to write the definition of $u_1(\phi)$ and $u_{-1}(\phi)$ in the following form:

$$u_1(\phi) = P_r[\text{“lead” indication} \mid \phi] \tag{7}$$

$$u_{-1}(\phi) = P_r[\text{“lag” indication} \mid \phi] \tag{8}$$

It follows that the probabilities of “correct” phase error measurements can be written in terms of the u_1 and u_{-1} as:

$$P_r[\text{correct "lead" indication}] = P_r[\phi > 0 \mid r(\phi) > 0] = \frac{\sum_{\phi > 0} u_1(\phi)}{\sum_{\text{all } \phi} u_1(\phi)} \tag{9}$$

$$P_r[\text{correct "lag" indication}] = P_r[\phi < 0 \mid r(\phi) < 0] = \frac{\sum_{\phi < 0} u_{-1}(\phi)}{\sum_{\text{all } \phi} u_{-1}(\phi)} \tag{10}$$

Of course, we would like these probabilities to be as close to 1.0 as possible which is the case when $u_1(\phi)$ is near zero for $\phi < 0$ and $u_{-1}(\phi)$ is near zero for $\phi > 0$.

Although a loop filter is not strictly needed to obtain closed-loop operation, let us consider the effect of a sequential filter whose operation can be viewed as a transformation of the lead and lag probabilities, $u_1(\phi)$ and $u_{-1}(\phi)$ into retard and advance probabilities, $U_1(\phi)$ and $U_{-1}(\phi)$. The objective of such a filter would be to maximize (9) and (10) with the new probabilities substituted. Hence, the following requirements are placed on the transformation

$$U_1 > u^1,\ 1.0 > u_1 > 0.5 \tag{11}$$

$$U_1 < u_1,\ 0.0 < u_1 < 0.5 \tag{12}$$

$$U_1 = u_1,\ u_1 = 0.0,\ 0.5,\ 1.0 \tag{13}$$

A complementary set of conditions is imposed on u_{-1} and U_{-1} The greater the inequalities in (11) and (12), the closer to unity are the probabilities of correct clock phase adjustments.

In order to satisfy (11) and (12), each filter output must depend on a sequence of inputs from the phase detector. A partial measure of the filter's "goodness" is how few inputs are required, on the average, to produce an output. That is, the fewer input cycles required for a given improvement satisfying (11) and (12), the larger is the loop bandwidth (or inverse transient time).

The sequential filters to be described below exhibit the property that the number of inputs required to produce an output is not constant, but depends on the particular sequence of random inputs. The expected value of this time, $T(\phi)$, is a function of ϕ through the dependence of the input sequence on $R(\phi)$ and is a random variable through the sequence dependence on input noise. These filters can also be made to reset or "Start over" when input sequences contain approximately equal numbers of lead and lag inputs. Hence, $T(\phi)$ tends to reach a maximum value at $\phi = 0$ where this is the case. The resulting loop operation is improved over one in which $T(\phi)$ = constant, since, in the former, fewer phase corrections will be made when the loop phase error is near zero where erroneous corrections are likely.

Examples of sequential filters satisfying the above criteria are shown in Figures 2 and 3. In the N-before-M filter, if an input sequence contains a sufficient number of lead or lag inputs that the appropriate N register fills up before the sequence length reaches M, then an output occurs and a clock phase adjustment is made. Otherwise, all counters are reset, no output occurs, and another input sequence is operated upon. The random walk filter achieves a very similar result. The rate at which the counter reaches boundary 0 or 2N is dependent on the relative frequency of occurrence of lag and lead inputs, respectively, and when they occur in nearly equal numbers, the count tends to hover near N for extended periods.

The transformation of input to output probabilities are evaluated for the N-before-M filter by

$$U_1 = \frac{\alpha}{\alpha + \beta} \tag{14}$$

$$U_2 = 1 - U_1 \tag{15}$$

where

$$\alpha = \sum_{i=N}^{M} \binom{i-1}{N-1} u_1^{N} \cdot u_{-1}^{i-N} \tag{16}$$

and

$$\beta = \sum_{i=N}^{M} \binom{i-1}{N-1} u_{-1}^{N} \cdot u_{1}^{i-N} \tag{17}$$

and for the random walk filter by

$$U_1 = \frac{1 - (u_1/u_{-1})^N}{(u_1/u_{-1})^{-N} - (u_1/u_{-1})^N} \tag{18}$$

T is given by

$$T = \frac{(1-\alpha-\beta)\cdot M + \sum_{i=N}^{M} \binom{i-1}{N-1} \left[u_1^N \cdot u_{-1}^{i-N} + u_{-1}^N \cdot u_1^{i-N} \right] \cdot i}{\alpha + \beta} \tag{19}$$

and

$$T = \begin{cases} N^2 & , u_1 = u_{-1} = 0.5 \\ \dfrac{N}{u_1 - u_{-1}} - \dfrac{2N}{(u_1 - u_{-1})[1 + (u_1/u_{-1})^N]} & , u_1 \neq 0.5 \end{cases} \tag{20}$$

for the two filters, respectively. Figure 4 shows representative examples of U_1 and T for N-before-M filters. Plots for random walk filters are very similar.

Each filter output causes a discrete phase adjustment of the clock. The magnitude of the adjustment is Δ. Hence, there are $K = 360°/\Delta$ possible phase error states (module 2π). The minimum phase error state occurs when $\phi = \delta$, where $|\delta| \leq \Delta/2$. Since the system changes state every time there is a filter output, U_1 and U_{-1} are the state transition probabilities. If we ignore the state duration inequality (i.e., $T(\phi) \neq$ constant) the state selection probabilities, L(i)[1], of the imbedded Markov chain formed by observing the loop only at state transition times is given by

$$L(i) = U_{-1}(i-1) \cdot L(i-1) + U_1(i+1) \cdot L(i+1), \quad i = 0, 1, \ldots K-1$$
$$\sum_{i=0}^{K-1} L(i) = 1 \tag{21}$$

The absolute state probabilities, P(i), are related to the L(i) and T(i) as follows

$$P(i) = \frac{L(i) \cdot T(i)}{\sum_{j=0}^{K-1} L(j)\, T(j)} \tag{22}$$

[1] Hereafter, the functional dependence of loop parameters on ϕ will be denoted through the index of the K possible ϕ-states.

Note that (22) reduces to the identity P(i) ≡ L(i) if T(i) = constant as is the case in which filters with fixed time intervals between outputs are used or if no filter is used.

Digital computer solutions to (21) and (22) were obtained for many loop configurations, a representative example of which is shown in Figure 5 for E_s/N_o = -20 dB. From this loop error probability mass function, parameterized in E_s/N_o , the loop rms phase error versus E_s/N_o curve can be constructed. However, this is of only small interest unless some knowledge of the loop's bandwidth or transient behavior is at hand.

Insight into the loop transient behavior can be obtained through the following definition. Let the loop be assumed to have just reached state i. Then $T_o(i)$ is defined to be the expected value of the time to reach the minimum phase error state $\phi = \delta$ for the first time. $T_o(i)$ can be evaluated in terms of the state transition probabilities and duration times by the following difference equation.

$$T_o(i) = T(i) + U_1(i) \cdot T_o(i - 1) + U_{-1}(i) \cdot T_o(i + 1) \qquad (23)$$

$$i = 1, 2, \ldots K - 1; \; T_o(0) \equiv 0$$

Figure 6 shows representative solutions to (16) where the circles indicate the actual computer solutions and the solid lines illustrate the approximate linear nature of the locus of these points. That is, $T_o(\phi)/\phi \approx$ constant. From this, it is quite apparent that the loop operation is inherently nonlinear. The loop phase adjustment is basically slew-rate limited. Hence, a conventional linear bandwidth cannot be defined for these loops.

Out of a conviction that a bandwidth-like measure must be found for these loops in order that any meaningful evaluation of their performance be made, the following "quasi-bandwidth" measure will be defined. Since in a first order linear system, the noise bandwidth can be defined in terms of its exponential time constant as $B = 1/4\pi$ and since detailed transient information for these digital loops is available, a bandwidth definition in terms of this response is an obvious choice. But if a time constant is to be defined in this case, it must be a time for traversing a reference angular displacement. Let this reference angle be 45° and let the digital loop bandwidth be defined as the reciprocal of the time for the loop to traverse this angle.

$$B' = \frac{1}{(T_o(i)/\phi_i) \cdot 45^\circ \cdot \tau} \qquad (24)$$

where $T_o(i)/\phi_i$ has the units in cycles/degree and τ is in seconds/cycle. Unfortunately, this definition must depend upon the arbitrary choice of the reference angle. It was chosen to be equal to one Δ for the coarsest quantization of phase adjustment felt to be of any practical value. Equation (24) does afford an absolute comparison of the performance of

different binary digital loops as shown in Figure 7 where rms phase error is plotted against signal-to-noise ratio in equal “quasi-bandwidths,” B'. It is significant to note that for low P_s/N_oB', all four digital loop curves fall on the same line while for high P_s/N_oB' a limiting phase error is attained. It appears that this limiting value can be reduced to whatever level is acceptable simply by choosing a small enough phase correction quantum, Δ.

Also shown in Figure 7 for reference, is the curve for the first order linear analog model. Such a comparison should, however, be used with caution due to the only approximate correspondence between B_L and B' and the arbitrary scale factor in B'. Further discussion of this comparison and the superior bandwidth adaptability of the digital loops can be found in reference [2].

We have hastily reviewed the description of and taken a glimpse at the performance of a class of digital phase locked loops characterized by discrete phase adjustment, sequential loop filtering, and binary quantization of the phase error measurement. In spite of their coarse quantization and, hence, component simplicity, their performance can be surprisingly good. Let us now apply the above techniques and definitions to some examples of loops with higher than binary quantization.

Higher Levels of Amplitude Quantization One would intuitively expect that an improvement in performance might be obtained with the use of more than two quantization intervals in the loop phase error amplitude. In the case of the phase detector output, and, hence, the loop filter input, complete flexibility exists for the choice of quantization intervals for each output. On the other hand, the loop filter outputs are limited to only even numbers, assuming each filter output causes the clock to be adjusted by $\pm \Delta$ or some multiple thereof. Figure 8 illustrates two examples of possible extensions of the basic digital loop to ternary and quaternary quantizations, respectively. In the ternary loop, a null zone has been added to the phase detector providing outputs which indicate phase error measurements which are within some interval of zero. The filter is configured in such a way that whenever the null zone outputs are frequent, clock phase adjustments are infrequent, since overflow of the M register causes filter operation to be reset without an output occurring.

In Figure 8b,the phase detector output is quantized into four intervals so that each output provides information about both the sign and magnitude of the phase error measurement.

Note that the quantization boundaries are defined as functions of the signal amplitude. This is a requirement in order that the quantization be made in terms of phase error measurement amplitude which scales linearly with $\sqrt{E_s}$. This, in turn, infers the use of AGC in the actual hardware.

Evaluation of the performance follows from the same formulations as in the binary case. Filter transformation relations are similar to (14) through (17) with a double summation being required in the ternary case and a tripple summation in the quaternary case. Equations (21), (22), and (23) apply directly to the ternary loop but (21) and (23) each have two additional terms for the quaternary loop.

Detailed examination of the operation of these loops is not possible here due to the large number of combinations of independent loop parameters (η, Δ, N, L, and M) and the subtile ways in which they influence the loop operation. The basic performance, however, is, except for a few striking examples of anomalous behavior, similar to that of the binary loops. The most important variances from that basic behavior will be illustrated in the remainder of this section.

Figure 9 shows the rms phase error vs E_s/N_o for a ternary loop for several values of δ. For the case of $\delta = 0$, there is a definite advantage to having added a third level of quantization since the phase error can approach zero. Unfortunately, δ is generally not a parameter under our control and may be anywhere between $\pm\Delta/2$. The region in E_s/N_o4 where this phenomenon occurs is dependent on the choice of η and is produced by the fact that for states lying within the null zone boundary ($R(\phi) < \eta \sqrt{E_s}$, T(i) is excessively large and transitions across the minimum phase error state are unlikely.

Of more importance, however, is the ternary loop transient behavior shown in Figure 10. Here, the importance of the parameter η is more obvious. For large η and large E_s/N_o , the null zone output probability, u_o, approaches 1.0 for more than the minimum phase error state, and the loop transient times become excessively large. An optimal choice for η is somewhere near $R(\Delta)/\sqrt{E_s}$, the value below which this condition cannot occur.

Figure 11 depicts the relative performance of three ternary loops with chosen to equal the above optimum value. The binary loop N = 8, M = 15, $\Delta = 22.5^\circ$ and $\delta = 0$ is shown for comparison since this is the asymtotic limit for a ternary system with N = 8, $\Delta = 22.5^\circ$, and $\eta \rightarrow 0$. Generally speaking, these curves lie slightly above (higher phase noise in equal B') those of the binary loops of Figure 7 with the exception of the sharp dips for large E_s/N_o. However, this behavior is somewhat artificial because of its dependence on δ being small.

The primary anomaly in the quaternary,loop operation is shown in Figure 12. In the quaternary case, there is no peaking of T(i) for large E_s/N_o as was observed in the ternary case but instead a region of E_s/N_o exists for which $\pm 2\Delta$ phase corrections are strongly favored and the minimum phase error state is nearly unattainable from all odd numbered states. This makes $T_o(i)$ excessively large for these states. The boundary of this region

shifts to lower E_s/N_o as η increases and in the limit as η increases the region is eliminated and the quaternary loop operates in essentially a binary mode.

Four quaternary loops are compared in Figure 13. All of these curves seem to be displaced to the right from that shown for the binary loops of Figure 7.

On the basis of the examination of these two classes of extensions of the binary loop model to ternary and quaternary loops, it seems that the added complexity is not warranted. In the ternary case, no characteristics that could clearly be labeled as better performance could be identified. In the quaternary case, there, in fact, seems to be some loss of quality of performance. Although far from exhaustive in investigating possible quantization extensions, this study tends to suggest that binary quantization is actually preferred.

Extensions to Higher Order Loops In the foregoing, it was assumed that δ was fixed or at most slowly varying. This is a requirement if the simple analysis techniques used above are to be applied. However, this condition frequently does not exist and a phase error bias is introduced when the signal frequency and the loop clock frequency are not identical.

The loop shown in Figure 14 is an extension of the basic "first order" binary loop to one which will correct for frequency offsets. Although no analysis of this model has been performed, the loop has intuitive appeal based on the performance which has been exhibited for the basic binary model.

An additional filter is used to monitor the relative frequency of retard and advance pulses. Obviously, if there were a phase error bias due to a frequency offset, the loop would favor phase adjustments of one sign in order to remain in equilibrium. If this bias persists, the appropriate N register of the N-before-M filter will give an output which adjusts the tap on the clock divider and, hence, adjusts the effective clock frequency.

It is interesting to note that if δ = constant or is slowly varying, in the steady state the additonal sequential filter of this loop rarely produces an output and the operation is essentially identical with that of the first order model considered earlier.

Conclusion It has been demonstrated that a class of digital phase locked loops exist for which binary amplitude quantization of the loop phase error can be employed without a corresponding significant sacrifice in performance. Insofar as it was possible to make comparisons of this performance with that of the linear analog model on an equal basis, the digital loop performance is roughly equal to that of the analog loop except for large E_s/N_o. Even this deviation can be suppressed by choosing Δ sufficiently small.

Insofar as a comparison was carried out with similar digital loops employing ternary and quaternary quantization, the binary case seems to be preferred.

Such conclusions are significant because they show that good loops can be constructed that not only enjoy the usual advantages of digital systems but also exhibit a component economy yielding lower cost and size and greater system reliability.

References

[1 J. R. Cessna, "Steady state and transient analysis of a class of digital phase-locked loops employing coarse amplitude quantization and sequential loop filters," Ph.D. dissertation, Dep. Elec. Engr., Univ. Iowa, Iowa City, Jan. 1970.

[2] J. R. Cessna and D. M. Levy, "Phase noise and transient times for a binary quantized digital phase-locked loop in white gaussian noise," IEEE Trans. Commun. Technol. COM-20, no. 2, pp. 94-1o4, April 1972.

[3] J. Holmes, "Coding and synchronization research: Performance of a firstorder digital phase-locked loop," Jet Propul. Lab., Pasadena, Calif., Rep. SPS 37-63, Vol. 3, pp. 28-34, June 1970.

[4] J. K. Holmes, "Performance of a first-order transition sampling digital phase-locked loop using random-walk models," IEEE Trans. Commun. Technol. COM-20, no. 2, pp. 119-131, April 1972.

[5] M. K. Simon, "The steady state performance of a data-transition type of first-order digital phase-locked loop," Jet Propul. Lab., Pasadena, Calif., Rep. SPS 37-66, pp. 59-68, Dec. 1970.

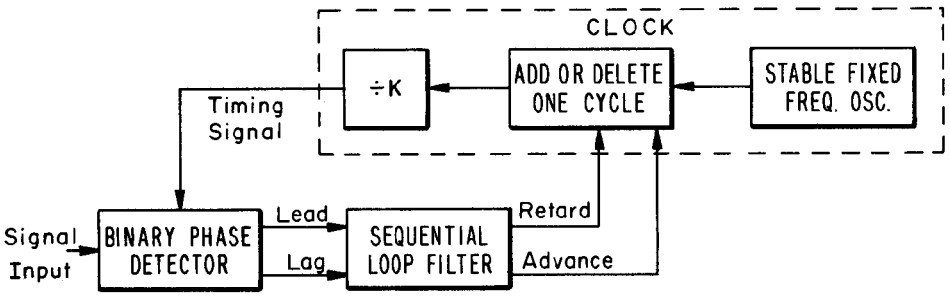

Fig. 1 - Binary digital phase-locked loop.

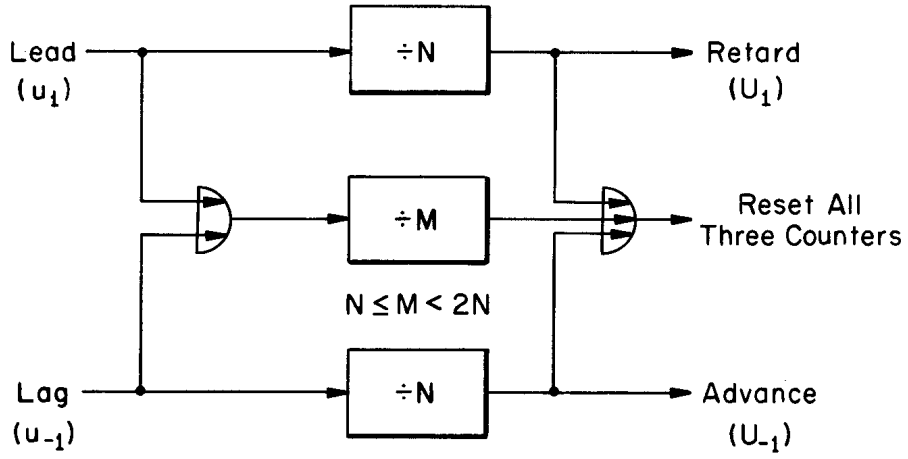

Fig. 2 - N-before-M filter.

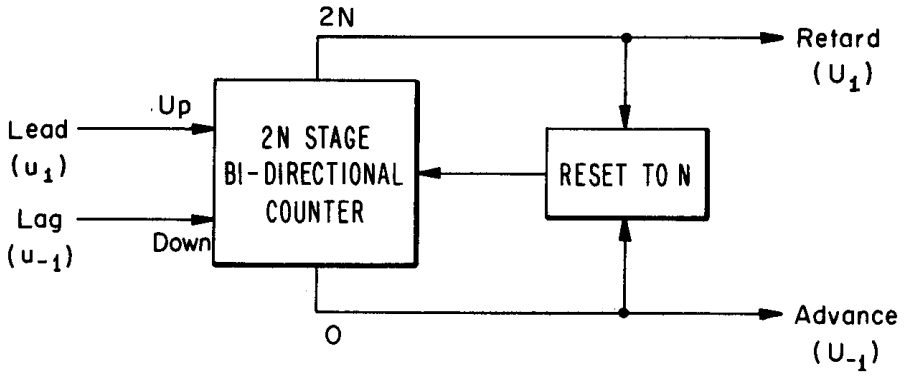

Fig. 3 - Random walk filter.

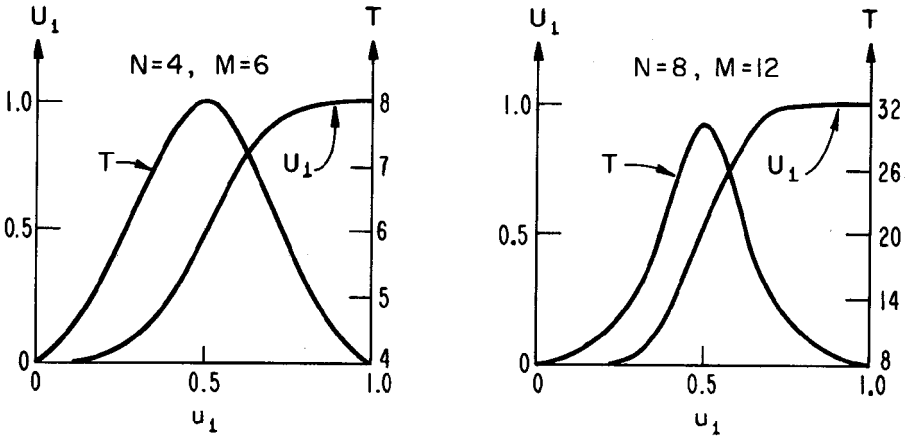

Fig. 4 - N-before-M filter probability transformation.
(a) N = 4, m = 6. (b) N = 8, M = 12.

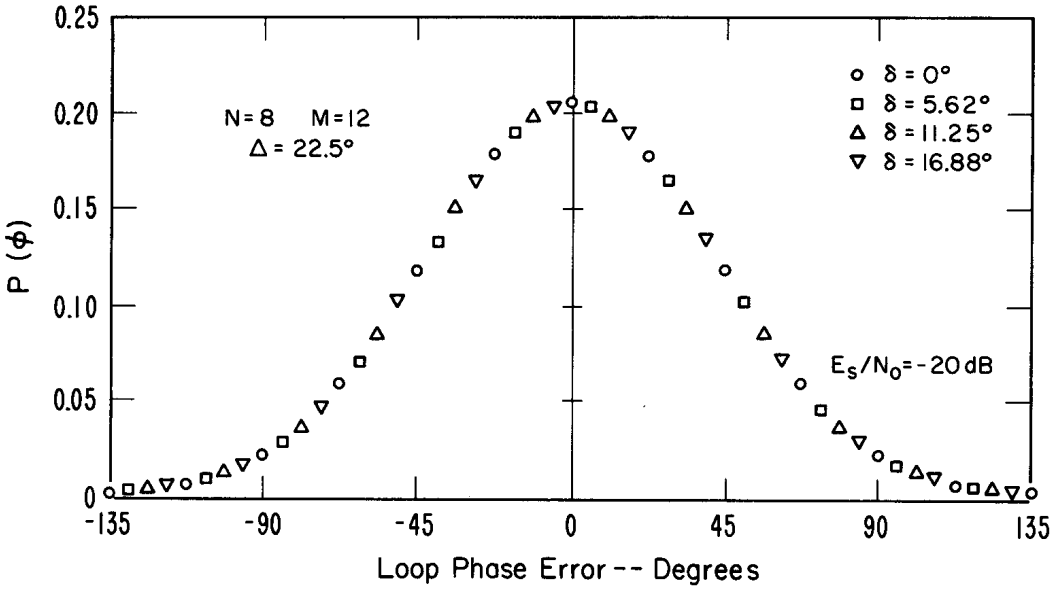

Fig. 5 - Binary loop phase error probability mass function; E_s/N_0 = -20 dB.

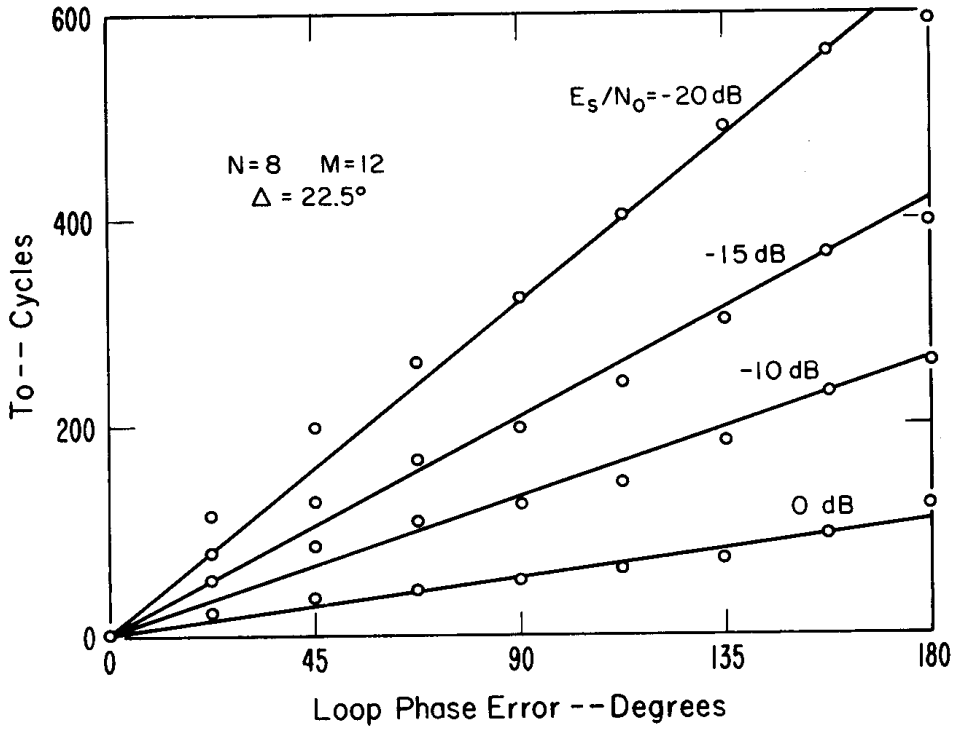

Fig. 6 - Binary loop transient times.

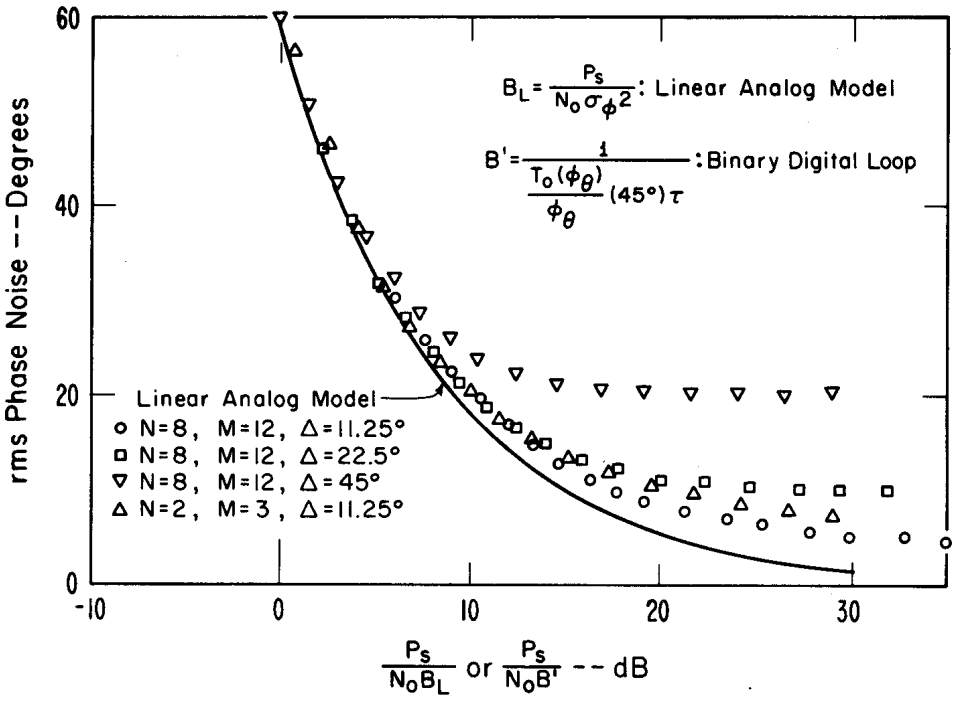

Figure. 7 - Binary loop performance comparison.

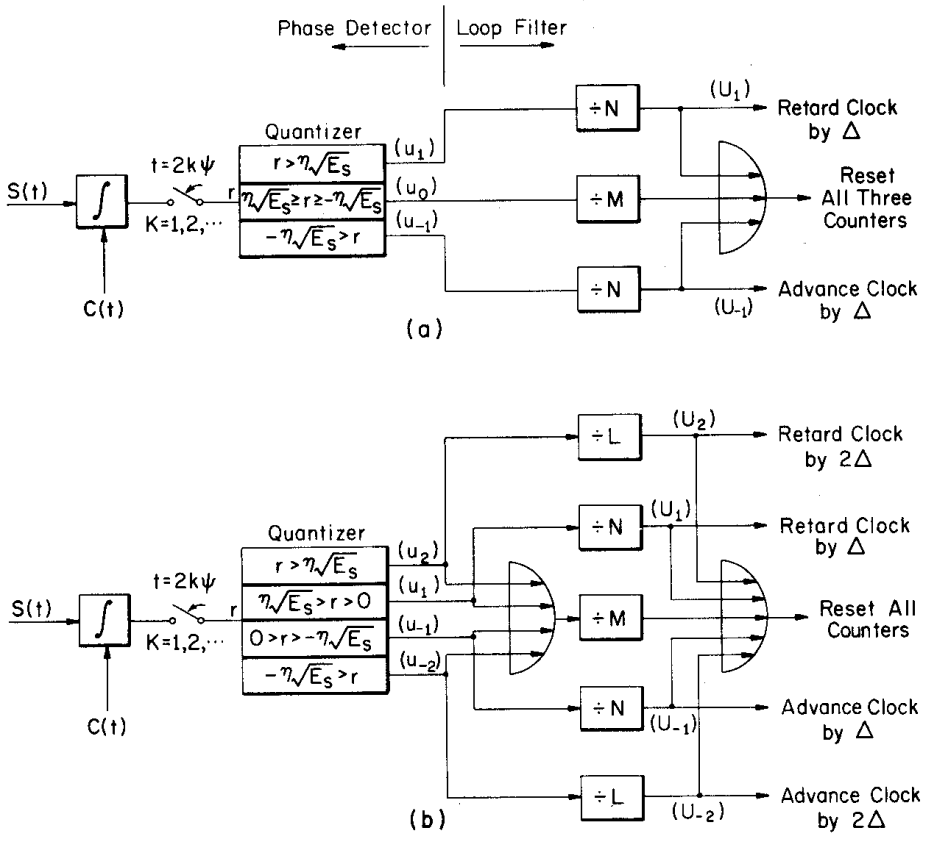

Fig. 8 - Loop modifications for higher orders of quantization.
(a) ternary case. (b) quaternary case.

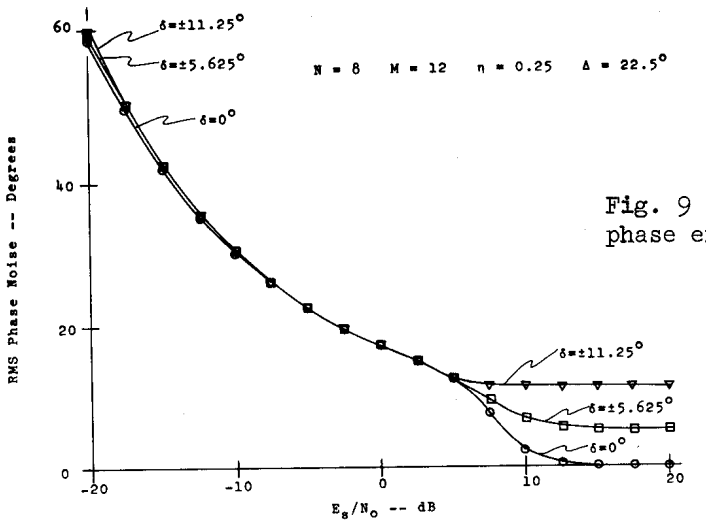

Fig. 9 - Ternary loop rms phase error dependence on δ.

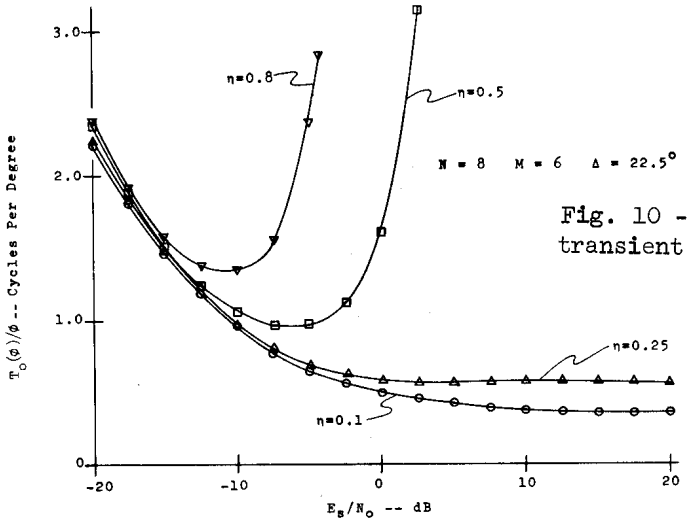

Fig. 10 - Ternary loop transient rate dependence on η.

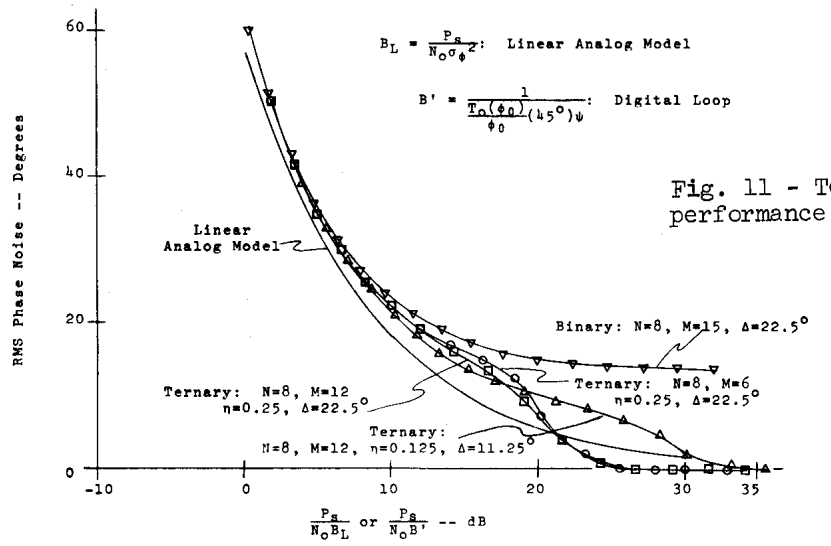

Fig. 11 - Ternary loop performance comparison.

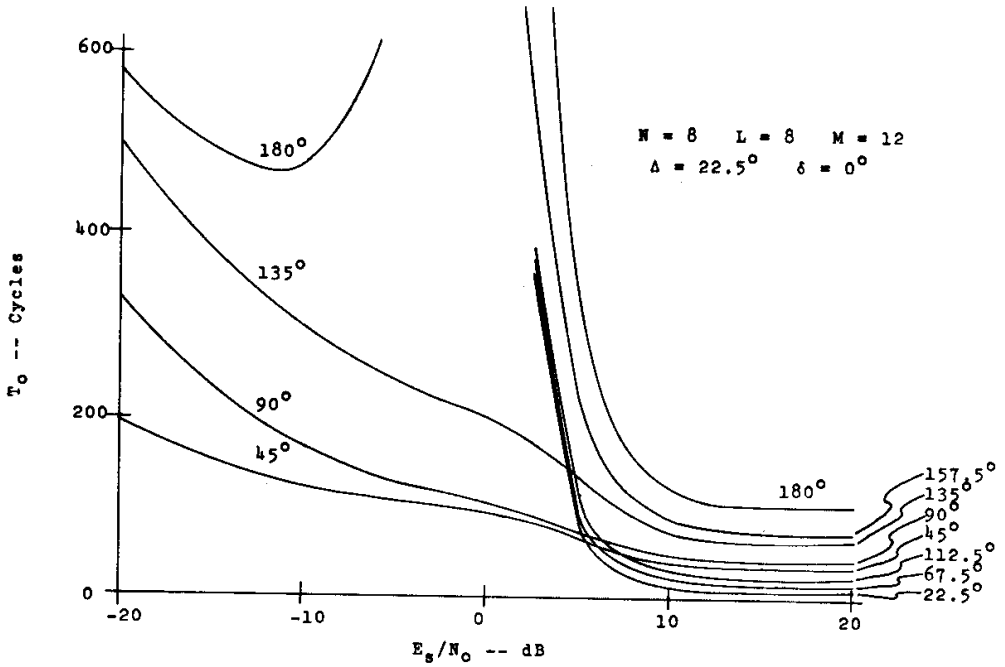

Fig. 12 - Quaternary loop transient times.

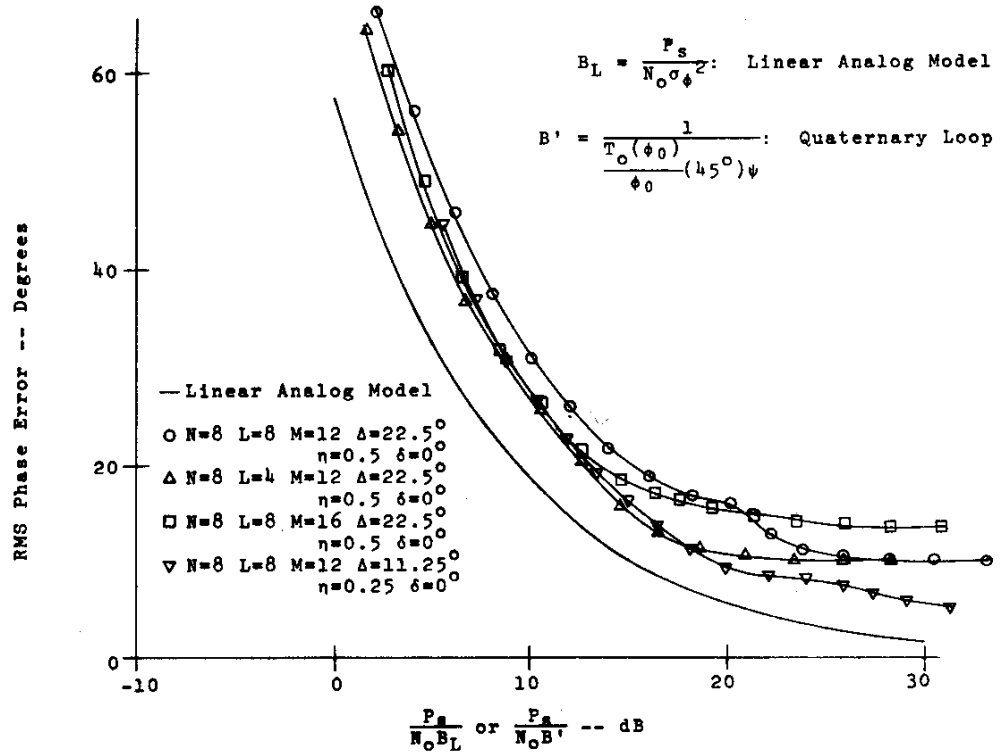

Fig. 13 - Quaternary loop performance comparison.

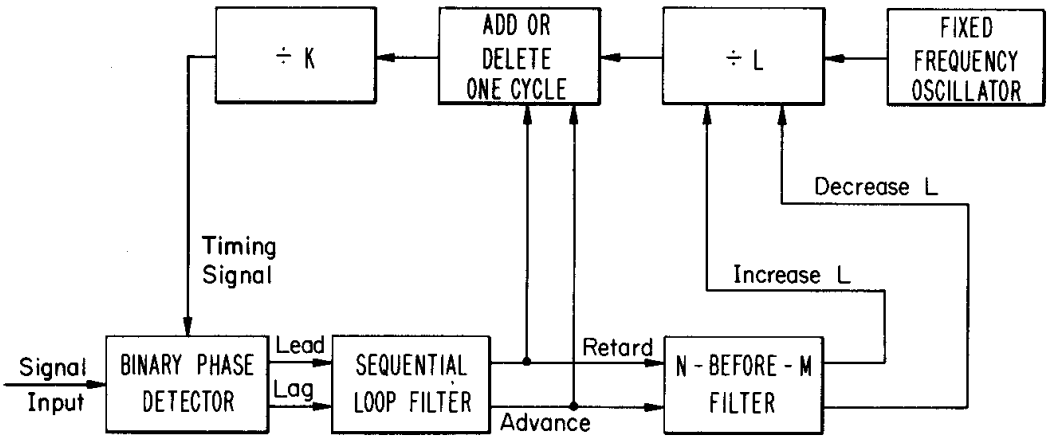

Fig. 14 - Binary second order loop.