SUR LA DYNAMIQUE DES BOUCLES DE COMMANDE NUMERIQUES COMPORTANT DES SIGNAUX D'ERREURS QUANTIFIES A DEUX NIVEAUX
ON THE DYNAMIC RANGE OF DIGITAL CONTROL LOOPS INCLUDING HARD-LIMITED ERROR SIGNALS

STEVEN BERNSTEIN On leave of absence from M.I.T.
Laboratory, Lexington, Mass., USA

ECOLE NATIONALE SUPERIEURE DES TELECOMMUNICATIONS DEPARTEMENT SYSTEMES ET COMMUNICATIONS
46 RUE BARRAULT 75634 PARIS CEDEX 13

RESUME

Les boucles numériques d'asservissement trouvent maintenant leurs applications dans des domaines variés surtout en communications numériques où elles sont utilisées pour l'asservissement des rythmes des symboles et de la fréquence et de la phase de la porteuse. Dans une telle application, les variations sur un grand intervalle des niveaux du signal et du bruit rendent difficile le choix des paramètres de la boucle. Dans la première partie de l'article, nous présentons à partir d'une simple construction graphique une méthode rapide pour déterminer la dynamique de la boucle. Dans la deuxième partie, nous analysons une méthode comme permettant d'élargir la dynamique de la boucle avec quelque perte en rapport signal à bruit. Le signal d'erreur est quantifié à deux niveaux. Nous présentons en particulier une analyse exacte du comportement des boucles de premier ordre à l'aide de la chaîne de Markov.

SUMMARY

Digital control loops have become very popular in a wide variety of applications. In particular they are often used as time, frequency or angle trackers in digital communication receivers. In such applications the signal and/or noise levels often vary over wide ranges. This leads to complications in choosing loop parameters which will yield acceptable performance (and of course stability) under these varying conditions.

The first part of this paper will present a technique for rapidly determing the useful dynamic range of digital control loops based on a simple graphical construction. The useful range is defined by the criterion of maintaining the r. m. s. error below a design value. It will be shown how such curves can be used as a design aid for setting loop parameters and for determining the efficacy of such techniques as automatic gain control (AGC) preceding the loop.

The second part will present an analysis of a known technique for dynamic range extension (at the expense of some sensitivity), namely, hard-limiting for error signal. In particular, an exact analysis of the behavior of first-order loops based on a Markov chain model, will be given. The results, which are very general, will be compared to linearizing approximations which are often made. In particular the range of validity of the well known 2 db ($2/\pi$) loss for the case of Gaussian noise will be given.

SUR LA DYNAMIQUE DES BOUCLES DE COMMANDE NUMERIQUES COMPORTANT DES SIGNAUX D'ERREURS QUANTIFIES A DEUX NIVEAUX

ON THE DYNAMIC RANGE OF DIGITAL CONTROL LOOPS INCLUDING HARD-LIMITED ERROR SIGNALS

INTRODUCTION

In digital communication receivers and numerons other applications it is necessary to track, i.e., continuously estimate, the value of parameters such as : time, frequency, phase, antenna pointing angle, signal strength, noise background, etc. It is natural to obtain periodic digital (sampled) measurements relating to the tracking errors and then process them in a digital filter to compute updated parameter estimates. In addition to being a natural approach, sampled tracking loops can have performance superior to analog loops. This occurs because only sufficient statistics are used, thus reducing degradations due to non-linearities. This paper concentrates on the dynamic range performance of such digital control loops when the input signal and/or noise levels are allowed to vary. Both linear loops and non-linear loops with hard-limited error signals are considered.

I. LINEAR MODELS

A. General Relationships.

Figure 1 shows a model of the linear system to be investigated first. The parameter to be tracked is x_k with k being the index of discrete time. The tracking system produces $\hat{x}_k$, the estimate of x_k. The only observable available is an error signal e_k which depends on the actual error, $x_k - \hat{x}_k$, and it is corrupted by additive noise. The error signal e_k is the input to the linear tracking filter which has a transfer function given by $F(z^{-1})$.
It is assumed that the measured error signal can be written as :

$$e_k = G(x_k - \hat{x}_k) + n_k \qquad (1)$$

That is, e_k is the sum of a term proportional to the actual error plus an independent zero-mean noise term of variance σ_n^2. The actual error is denoted as φ_k :

$$\varphi_k = x_k - \hat{x}_k \qquad (2)$$

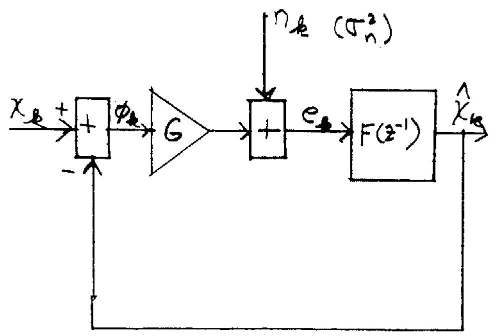

Fig. 1. Linear System Model

The loop's input-output transfer function is given by :

$$\frac{\hat{X}}{X}(z^{-1}) = \frac{GF(z^{-1})}{1 + GF(Z^{-1})} \qquad (3)$$

$$\triangleq H(z^{-1})$$

with a corresponding unit sample response of h_k.
The transfer function relating the error to the noise is given by :

$$\frac{\Phi}{N}(z^{-1}) = \frac{F(z^{-1})}{1 + GF(z^{-1})}$$

$$= H(z^{-1})/G \qquad (4)$$

From this it follows that if the noise samples are assumed independent, the resulting variance of the error (due only to noise) will be :

$$\sigma_\Phi^2 = \sigma_n^2 \, \Sigma (h_k/G)^2 \qquad (5)$$

This relationship leads to the definition, of averaging time N_{avg} (in sample periods) :

$$N_{avg} \triangleq \frac{1}{\Sigma h_k^2} \qquad (6)$$

so that

$$\sigma_\Phi^2 = \frac{1}{N_{avg}} (\sigma_n/G)^2 \qquad (7)$$

The quantity $(\sigma_n/G)^2$ can be considered the reciprocal of the signal-to-noise ratio (SNR) per input signal and is equal to the variance of the estimation error if only one input sample were used instead of an averaging loop.

Normally there could be a maximum allowed value for σ_ϕ^2, $\sigma_{\phi o}^2$ that would be determined by external system constraints, e.g., allowed phase jitter . At some design-point values for σ_n and G, denoted as σ_{no} and G_o, the loop averaging time would be chosen to satisfy by setting $N_{avg,o} = (\sigma_{no}/G_o)^2/\sigma_{\phi o}^2$. This in turn could determine the filter coefficients. However, an important point to see is that the actual value of N_{avg} depends on the actual value of G, which is directly related to the strength of the error signal. Generally, if G increases the averaging time will decrease. However if G increases beyond a certain critical value the loop can become unstable and hence useless as a tracking system. This dynamic range problem is the concern of the remainder of the paper.

B. Determining the Dynamic Range.

As long as $\sigma_\phi^2 = (\sigma_n/G)^2/N_{avg}$ remains below its design value the loop will be considered to operate effectively. The issue is whether σ_ϕ^2 remains less than its design value, $\sigma_{\phi o}^2$, as σ_n^2 and G vary from their design-points. A simple graphical technique is suggested for determining this :

$$\text{Plot } \left(\frac{N_{avg}}{N_{avg,o}}\right) \quad \text{and} \quad \frac{(\sigma_n/G)^2}{(\sigma_{n,o}/G_o)^2}$$

on the same set of axes versus G/G_o. Use logarithmic scaling on both axes to force both curves to go through the origin which corresponds to the design point. Wherever the Navg curve is above the $(\sigma_n/G)^2$ curve the design criterion is met. This will be illustrated below for first and second order loops. (Note that if only σ_n^2 is varying the problem is trivially solved. It is the variation in G that introduces the complication of the variation of loop aneragins time).

C. Firstand Second Order loops

1. First order loops.

For the first order loop $\hat{x}_k$ is updated simply as :

$$\hat{x}_k = \hat{x}_{k-1} + be_{k-1} \tag{8}$$

Without going into detail, the closed loop response will be :

$$h_k = bG(1-bG)^{k-1} \qquad k > 1 \tag{9}$$

and the resulting averaging time will be :

$$N_{avg} = \frac{2-bG}{bG} \tag{10}$$

Figure 2 shows the curves of $(N_{avg}/N_{avg,o})$ for several values of $N_{avg,o}$. It is easy to show that this stays stable only if

$$\frac{G}{G_o} < N_{avg,o} + 1 \tag{11}$$

which corresponds to the rapid fall-off of each curve in Figure 2.

the same. It the signal level were to increase, say, 6 db then the noise could increase 5 db and acceptable performance would still result.

It automatic gain control (AGC) were used before the loop the dynamic range could be extended considerably. For example, if the AGC held $\sigma_n^2 + G^2$ constant then it is not hard to show that :

$$\frac{\sigma_n^2}{G^2} = \frac{1 + (\sigma_{no}/G_o)^2}{(G/G_o)^2} - 1 \tag{12}$$

For any design-value of SNR, $(G_o/\sigma_{no})^2$, this AGC can also be plotted. This was done for a design SNR of - 3 db and is also shown on Fig. 2 labeled "AGC". The addition of AGC renders the loop stable for all values of input signal strength and will give performance within the σ_n^2 constraint as long as $G>G_o$ which corresponds to keeping the input SNR above the design value.

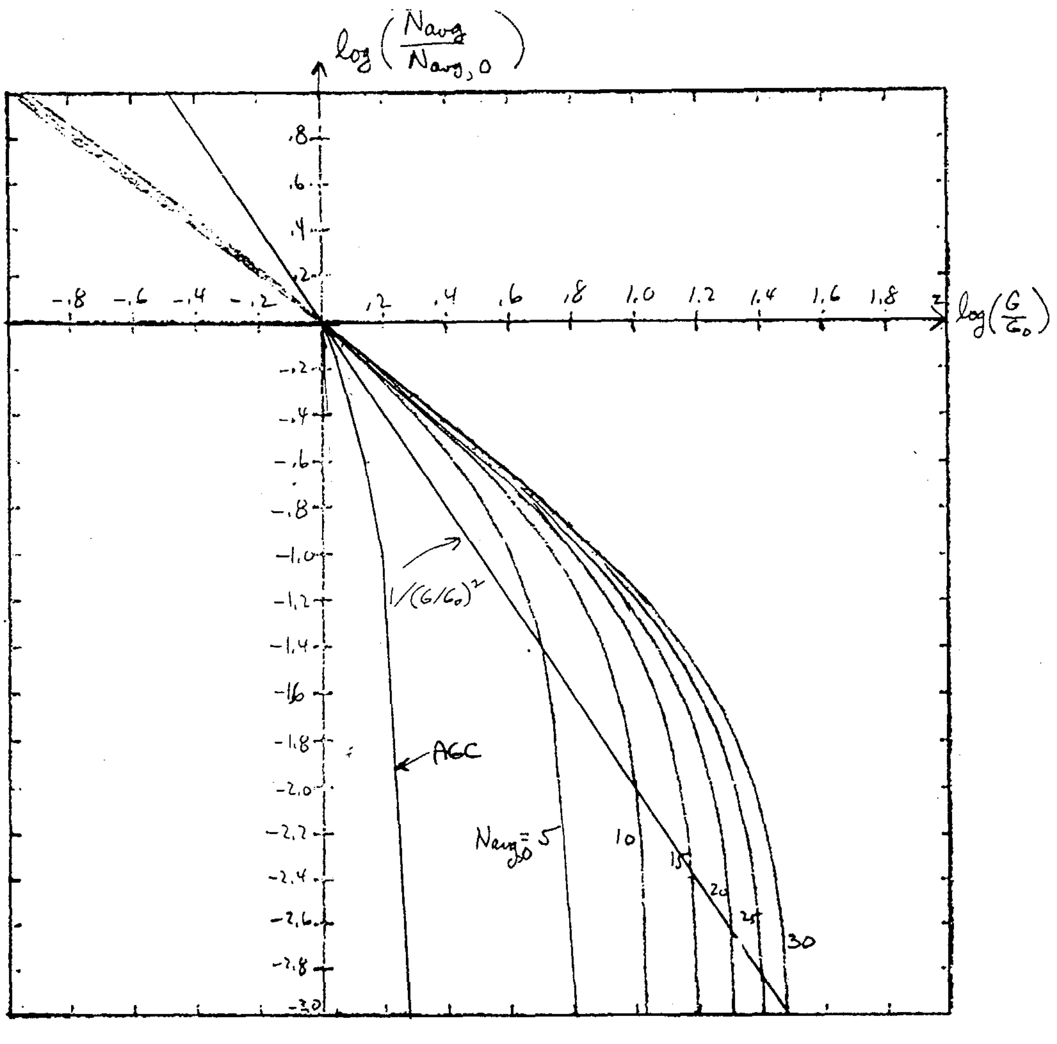

Fig. 2. Variation of Averaging Time with Gain - First Order Loop.

Also shown in Figure 2 is the curve $(\sigma_n/G)^2/(\sigma_{no}/G_o)^2$ corresponding to holding $\sigma_n^2 = \sigma_{no}^2$ and letting only G vary. At any value of G/G_o and $N_{avg,o}$, the vertical distance between the curves indicates by what factor σ_n^2 may be increased while the design constraint is still being met. For example, if $N_{avg,o} = 15$ then the loopcan withstand an increase in signal level of nearly 4 (12 db) beyond its design-point as long as σ_n^2 remains

2. Second order loops.

The second order loop forms its correction to the tracking estimate using two prior values of the error signal and estimated output :

$$\hat{x}_k \quad a_1\hat{x}_{k-1} + a_2\hat{x}_{k-2} + b_1e_{k-1} + b_2e_{k-2} \tag{13}$$

Second order loops can provide better transient tracking behavior than the simple first order loops. In particular, if one sets $a_1 = 2$ and $a_2 = -1$ then it will be able to track all input steps and ramps with no steady state error in the absence of noise. The first order loop only achieves this for step inputs. Using E_q (3) to find the closed loop transfer function the unit sample response can be (and was) found exactly. An expression for N_{avg} can be (and was) also derived. To save space just a few key results will be given. The unit sample response will be of the form $h_k \sim r^k \cos(kb)$ assuming the poles of $H(Z^{-1})$ are complex, where $r^2 = Gb_2 + 1$ and $2r \cos b = 2 - Gb_1$. For the commonly selected value of damping factor equal to $1/\sqrt{2}$ (rate of decay equal to the frequency of oscillation) r should be set equal to e^{-b}. Then h_k can be approximated for small b as :

$$h_k \simeq \begin{cases} 0 & k \le 0 \\ 2be^{-kb}\cos(kb) & k \ge 1 \end{cases} \qquad (14)$$

(for b <<1)

and a good approximation for design purposes is :

$$b = \frac{2}{3(N_{avg} + 0.44)} \qquad (15)$$

which is very accurate for any N_{avg} greater than 1. Knowing b and G_o the values of b_1 and b_2 are found from :

$$b_1 = 2(1-e^{-b}\cos b)/G_o$$
$$b_2 = (e^{-2b}-1)/G_o \qquad (16)$$

Figure 3 shows the curves for the second order loop that correspond to Figure 2 for the first order. The curves are similar, however second order loops offer a somenhat increased dynamic range.

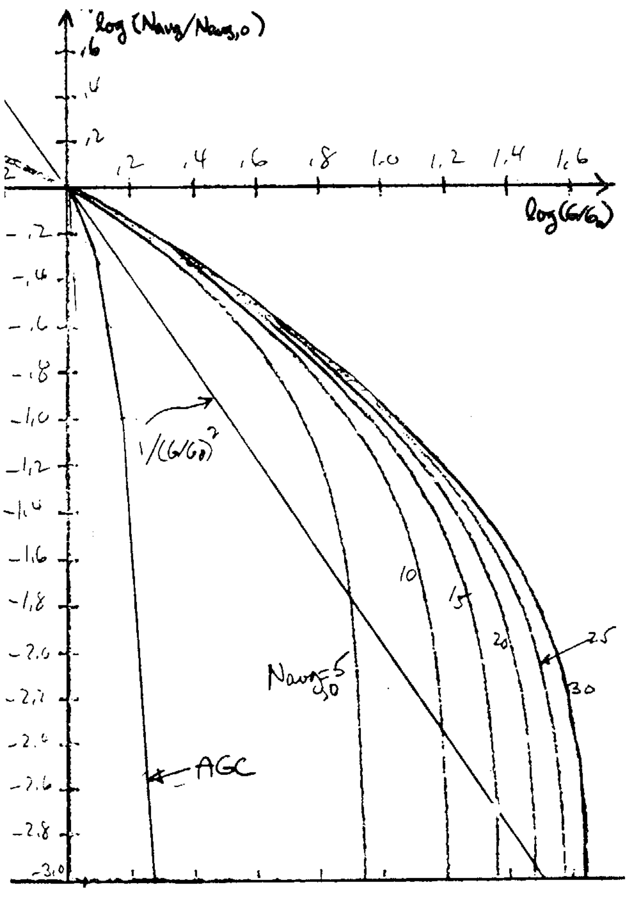

Fig. 3 Variation of Averaging Time with Gain - Second Order Loop.

II. HARD LIMITED ERROR SIGNALS

Another way to deal with the problem of dynamic ronge is to hard limit the error signal as modeled in Figure 4 for the case of additive noise. Only the sign of the error signal is used as the input to the loop filter.

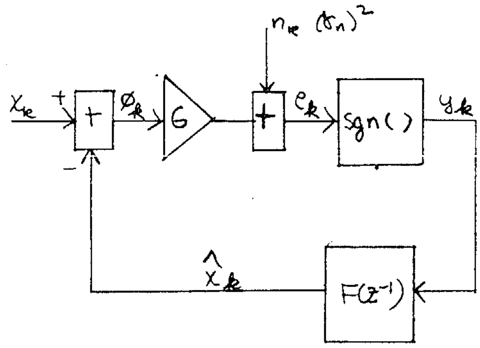

Fig. 4. Loop with Hard Limited Error Signal (Additive Noise).

A. Linearized Analysis. (Additive Noise).

To analyse such a system the mean and variance of the hard limited error signal y_k are often approximated as :

$$\bar{y}_k \simeq 2G\, p_n(0)\, \phi_k \qquad (17)$$
$$\sigma_y^2 \simeq 1$$

when the tracking error is small, when the input SNR also is small, and when the probability density function of the noise, $P_n(x)$, is well behaved at the origin. (The latter assumption must be verified in each case. For example, pulsed noise would not necessarily justify the approximation). The results of the previous section on linear loops can vow be applied with the following effective values for gain and noise variance :

$$G_{eff} = 2Gp_n(0) \qquad \sigma^2_{n,eff} = 1 \qquad (18)$$

These loops generally remain stable with respect to changes in input signal and noise variations although at a sacrifice in sensitivity. Specifically the mean square tracking error will now be

$$\sigma_\phi^2 = \frac{1}{4G^2p_n^2(0)} \quad \frac{1}{N_{avg}} \qquad (19)$$

In the special case of Gaussian noise $p_n(0) = 1/\sqrt{2\pi\sigma_n^2}$ so

$$\sigma_\phi^2 = \frac{\pi}{2}\;\frac{\sigma_n^2}{G^2}\;\frac{1}{N_{avg}} \qquad (20)$$

which shows a degradation of $(\pi/2)$ or 2db in effective SNR.

B. Exact Analysis : First Order loop

The first order loop defined by

$$\hat{x}_k = \hat{x}_{k-1} + by_{k-1}$$

can be analysed exactly and readily using Markov process concepts. (To simplify matters it will be assumed that the input $x_k = 0$ so that the effects of transients will be put aside.

It is important to note that in the analysis to follow the assumption of additive noise is not needed. Indeed in many practical cases, e.g, tracking signal timing, the model here will apply although the noise is not additive. Holmes (1) presents a similar analysis for a specific application in a more complete form. The emphasis here is on the general problem of dynamic range.

The output of the loop $\hat{x}_k$ is seen to describe a random walk with amplitude increments of ± b. The output value itself can be used as the definition of the state of a discrete time, discrete amplitude Markov process. As a matter of fact, it is a discrete time birth-death process since only transitions to adjacent states are allowed. This is indicated in Fig. 5 where each box represents a possible output value, i.e., state of process, and each arrow is labled with the conditional probability in making the indicated trasitions. State j corresponds to a tracking error of bj, j : 0, ± 1, ± 2... . The λ_j's are the probabilities of moving in a direction of increasing absolute error (increasing $|\hat{x}_k|$) from state j ; the μ_j's are the probabilities of moving in a direction of decreasing absolute error ;

$$\mu_j = 1 - \lambda_j .$$

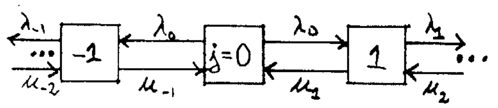

Fig. 5. Markov Process Model of First Order Loop.

If P_j is defined as the probability of being in state j, i.e., error bj, then it is well known (2) that :

$$P_j = P_{-j} = \frac{\lambda_{j-1} \cdots \lambda_0}{\mu_j \cdots \mu_1} P_0 \tag{22}$$

where symmetry has been assumed. P_o is found from the normalizing constraint that $\Sigma P_j = 1$. The mean square tracking error will be :

$$\sigma_\phi^2 = b^2 \sum_{j=-\infty}^{\infty} j^2 P_j \tag{23}$$

Gaussian Noise Example

In the special case when the noise is additive and Gaussian :

$$\lambda_j = \frac{1}{\sqrt{2\pi}} \int_{\frac{Gb|j|}{\sigma_n}}^{\infty} e^{-x^2/2} dx \tag{24}$$

Figure 6 shows the results of numerically finding σ_ϕ^2 from Eqs, (23) and (24) after evaluating a sufficient number of P_j's to get accurate results. Also shown on the figure are the results obtained from the linearized model. Several points are worth noting : (1) The linear model agrees very well with the exact results as long as $b(G/\sigma_n) \lesssim .5$ which can be used to define the "small error" regime. (2) The linear model predicts instability for $\sqrt{2/\pi}(bG/\sigma_n) \geq 2$ whereas the exact solution shows behavior that is always stable. (3) The variance cannot be decreased belox $b^2/2$ even for very large (G/σ_n) since the loop is always kicked off of exact zero error to ± b (but returns immediately).

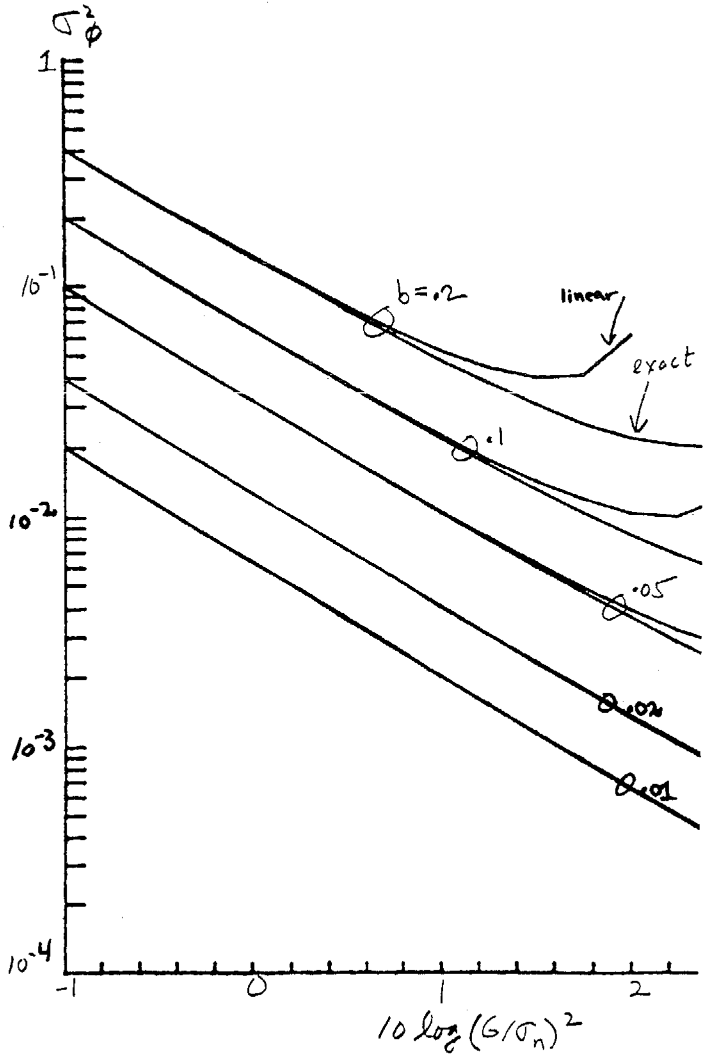

Fig. 6. Gaussian noise, exact and linearised Model.

The transient behavior can be analysed using basic results from markov process, random walk theory.

CONCLUSIONS

Techniques for finding the acceptable dynamic range of digital tracking loops were presented. On the case of linear first and second order loops a graphical construction was give,. For the case of hard-limited error signals, a linearized model was reviewed. For the first order loop with hard limited error signals an exact analysis was show,. To the knowledge of this author the exact solution of the second order loop with hard-limited error signals remains a challenge.

REFERENCES

/1/ Holmes, J.K. "Performance of a First-order Transition Sampling Digital Phase-Locked Loop Using Random-Walk models", IEEE Transactions on Communications, Vol. COM 20, pp 119-131, April 1972.

/2/ Feller, W, "An Introduction to Probability Theory and Its Applications", Vol 1, New York : Wiley 1957.