Theory of AFC Synchronization*

WOLF J. GRUEN†, MEMBER, IRE

Summary—**The general solution for the important design parameters of an automatic frequency and phase-control system is presented. These parameters include the transient response, frequency response and noise bandwidth of the system, as well as the hold-in range and pull-in range of synchronization.**

I. Introduction

AUTOMATIC FREQUENCY and phase-control systems have been used for a number of years for the horizontal-sweep synchronization in television receivers, and more recently have found application for the synchronization of the color subcarrier in the proposed NTSC color-television system. A block diagram of a general AFC system is shown in Fig. 1.

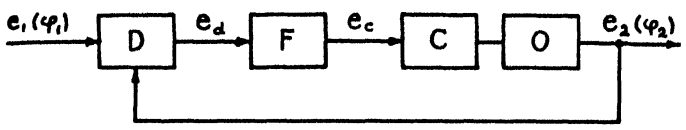


Fig. 1—Block diagram of A.F.C. loop.

The phase of the transmitted synchronizing signal e_1 is compared to the phase of a local oscillator signal e_2 in a phase discriminator D. The resulting discriminator output voltage is proportional to the phase difference of the two signals, and is fed through a control network F to a frequency-control stage C. This stage controls the frequency and phase of a local oscillator O in accordance with the synchronizing information, thereby keeping the two signals in perfect synchronism. Although in practice the transmitted reference signal is often pulsed and the oscillator comparison voltage non-sinusoidal, the analysis is carried out for sinusoidal signal voltages. The theory, however, can be extended for a particular problem by writing the applied voltages in terms of a Fourier series instead of the simple sine function. An AFC system is essentially a servomechanism, and the notation that will be used is the one followed by many workers in this field. An attempt will be made to present the response characteristics in dimensionless form in order to obtain a universal plot of the response curves.

II. Derivation of the Basic Equation

If it is assumed that the discriminator is a balanced phase detector composed of peak-detecting diodes, the discriminator-output voltage can be derived from the vector diagram in Fig. 2. For sinusoidal variation with time, the synchronizing signal e_1 and the reference signal e_2 can be written

$$e_1 = E_1 \cos \phi_1 \tag{1}$$

and

$$e_2 = E_2 \sin \phi_2. \tag{2}$$

ϕ_1 and ϕ_2 are functions of time and, for reasons of simplicity in the later development, it is arbitrarily assumed that ϕ_1 and ϕ_2 are in quadrature when the system is perfectly synchronized, that is when $\phi_1 = \phi_2$.

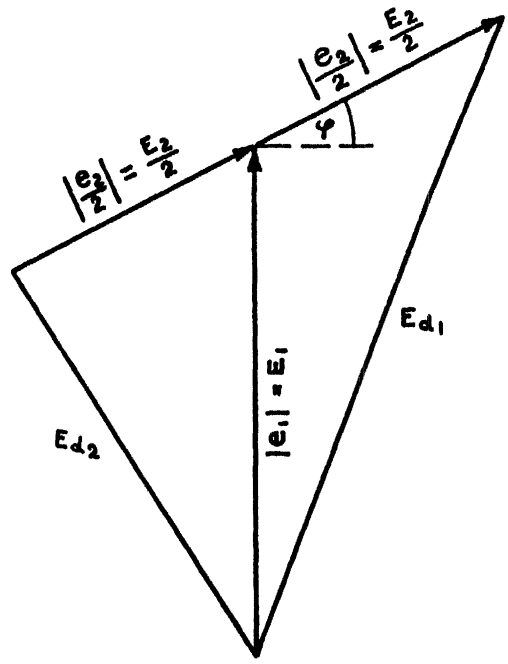


Fig. 2—Discriminator vector diagram.

While one of the discriminator diodes is fed with the sum of e_1 and $e_2/2$, the other is fed with the difference of these two vectors as shown in Fig. 2. The resulting rectified voltages E_{d1} and E_{d2} can be established by simple trigonometric relations. Defining a difference phase

$$\phi \equiv \phi_1 - \phi_2, \tag{3}$$

one obtains

$$E_{d1}^2 = E_1^2 + \frac{E_2^2}{4} + E_1 E_2 \sin \phi \tag{4}$$

and

$$E_{d2}^2 = E_1^2 + \frac{E_2^2}{4} - E_1 E_2 \sin \phi. \tag{5}$$

The discriminator output voltage e_d is equal to the dif-

* Decimal classification: R583.5. Original manuscript received by the Institute, August 21, 1952; revised manuscript received February 25, 1953.

† General Electric Co., Syracuse, N. Y.

ference of the two rectified voltages, so that

$$e_d = E_{d1} - E_{d2} = \frac{2E_1E_2}{E_{d1} + E_{d2}} \sin \phi. \tag{6}$$

If the amplitude E_1 of the synchronizing signal is larger than the amplitude E_2 of the reference signal, one obtains

$$E_{d1} + E_{d2} \cong 2E_1. \tag{7}$$

The discriminator output voltage then becomes

$$e_d = E_2 \sin \phi \tag{8}$$

and is independent of the amplitude E_1 of the synchronizing signal. As ϕ_1 and ϕ_2 are time-varying parameters, it should be kept in mind that the discriminator time constant ought to be shorter than the reciprocal of the highest difference frequency $d\phi/dt$, which is of importance for the operation of the system.

Denoting the transfer function of the control network F as $F(p)$, the oscillator control voltage becomes

$$e_c = F(p)E_2 \sin \phi. \tag{9}$$

Assuming furthermore that the oscillator has a linear-control characteristic of a slope S, and that the free-running oscillator frequency is ω_0, the actual oscillator frequency in operational notation becomes

$$p\phi_2 = \omega_0 + Se_c. \tag{10}$$

Substituting (3) and (9) into (10) then gives

$$p\phi + SE_2F(p) \sin \phi = p\phi_1 - \omega_0. \tag{11}$$

The product SE_2 repeats itself throughout this paper and shall be defined as the gain constant

$$K \equiv SE_2. \tag{12}$$

K represents the maximum frequency shift at the output of the system per radian phase shift at the input. It has the dimension of radians/second.

Equation (11) can be simplified further by measuring the phase angles in a coordinate system which moves at the free-running speed ω_0 of the local oscillator. One obtains

$$\boxed{p\phi + F(p)K \sin \phi = p\phi_1}. \tag{13}$$

This equation represents the general differential equation of the AFC feedback loop. $p\phi$ is the instantaneous-difference frequency between the synchronizing signal and the controlled-oscillator signal and $p\phi_1$ is the instantaneous-difference frequency between the synchronizing signal and the free-running oscillator signal.

Equation (13) shows that all AFC systems with identical gain constants K and unity d.c. gain through the control network have the same steady-state solution, provided that the difference frequency $p\phi_1$ is constant. If this difference frequency is defined as

$$\Delta\omega \equiv p\phi_1 = \omega_1 - \omega_0, \tag{14}$$

the steady-state solution is

$$\sin \phi = \frac{\Delta\omega}{K}. \tag{15}$$

This means the system has a steady-state phase error which is proportional to the initial detuning $\Delta\omega$ and inversely proportional to the gain constant K. Since the maximum value of $\sin \phi$ in (15) is ± 1, the system will hold synchronism over a frequency range

$$|\Delta\omega_{\text{Hold-in}}| \leqq K. \tag{16}$$

Equations (15) and (16) thus define the static performance limit of the system.

III. Linear Analysis

An AFC system, once it is synchronized, behaves like a low-pass filter. To study its performance it is permissible, for practical signal-to-noise ratios, to substitute the angle for the sine function in (13). Then, with the definition of (3), one obtains

$$p\phi_2 + KF(p)\phi_2 = KF(p)\phi_1. \tag{17}$$

This equation relates the output phase ϕ_2 of the synchronized system to the input phase ϕ_1. It permits an evaluation of the behavior of the system to small disturbances of the input phase, if the transfer function $F(p)$ of the control network is specified.

a. $F(p) = 1$

This is the simplest possible AFC system, and represents a direct connection between the discriminator output and the oscillator control stage. Equation (17) then becomes

$$p\phi_2 + K\phi_2 = K\phi_1. \tag{18}$$

If the initial detuning is zero, the transient response of the system to a sudden step of input phase $|\phi_1|$ is

$$\frac{\phi_2}{|\phi_1|}(t) = 1 - e^{-Kt}. \tag{19}$$

Likewise, the frequency response of the system to a sine-wave modulation of the input phase is

$$\frac{\phi_2}{\phi_1}(j\omega) = \frac{1}{1 + j\dfrac{\omega}{K}}. \tag{20}$$

The simple AFC system thus behaves like an RC-filter and has a cut-off frequency of

$$\omega_c = K \text{ [radians/sec]}. \tag{21}$$

George[1] has shown that the m.s. phase error of the system under the influence of random interference is proportional to the noise bandwidth, which is defined as

[1] T. S. George, "Synchronizing systems for dot interlaced color TV," Proc. I.R.E.; February, 1951.

$$B = \int_{-\infty}^{+\infty} \left| \frac{\phi_2}{\phi_1}(j\omega) \right|^2 d\omega. \tag{22}$$

The integration has to be carried out from $-\infty$ to $+\infty$ since the noise components on both sides of the carrier are demodulated. Inserting (20) into (22) then yields

$$B = \pi K \text{ [radians/sec]}. \tag{23}$$

It was shown in (15) that for small steady-state phase errors due to average frequency drift, the gain constant K has to be made as large as possible, while now for good noise immunity, i.e., narrow bandwidth, the gain constant has to be made as small as possible. A proper compromise of gain then must be found to insure adequate performance of the system for all requirements. This difficulty, however, can be overcome by the use of a more elaborate control network.

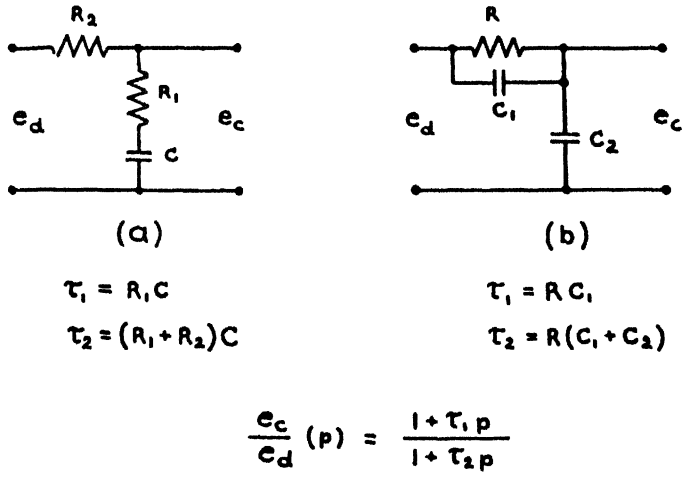


Fig. 3—Proportional plus integral control networks.

$$b.\ F(p) = \frac{1 + \tau_1 p}{1 + \tau_2 p}$$

Networks of this type are called proportional-plus integral-control networks[2] and typical network configurations are shown in Fig. 3. Inserting the above transfer function into (17) yields

$$p^2\phi_2 + \left(\frac{1}{\tau_2} + K\frac{\tau_1}{\tau_2}\right) p\phi_2 + \frac{K}{\tau_2}\phi_2 = K\frac{\tau_1}{\tau_2} p\phi_1 + \frac{K}{\tau_2}\phi_1. \tag{24}$$

ϕ_1 and ϕ_2 are again relative phase angles, measured in a coordinate system which moves at the free-running speed of the local oscillator. To integrate (24), it is convenient to introduce the following parameters

$$\omega_n^2 \equiv \frac{K}{\tau_2} \tag{25}$$

and

$$2\zeta\omega_n \equiv \frac{1}{\tau_2} + K\frac{\tau_1}{\tau_2}. \tag{26}$$

ω_n is the resonance frequency of the system in the absence of any damping, and ζ is the ratio of actual-to-critical damping. In terms of the new parameters the time constants of the control network are

$$\tau_1 = \frac{2\zeta}{\omega_n} - \frac{1}{K} \tag{27}$$

and

$$\tau_2 = \frac{K}{\omega_n^2}. \tag{28}$$

With these definitions (24) becomes

$$p^2\phi_2 + 2\zeta\omega_n p\phi_2 + \omega_n^2\phi_2 = \left(2\zeta\omega_n - \frac{\omega_n^2}{K}\right) p\phi_1 + \omega_n^2\phi_1. \tag{29}$$

The transient response of the system to a sudden step of input phase $|\phi_1|$ is found by integration of (29) and the initial condition for the oscillator frequency is obtained from (10). The transient response then is

$$\frac{\phi_2}{|\phi_1|}(t) = 1 - e^{-\zeta\omega_n t}\left[\cos\sqrt{1-\zeta^2}\,\omega_n t - \frac{\zeta - \frac{\omega_n}{K}}{\sqrt{1-\zeta^2}} \sin\sqrt{1-\zeta^2}\,\omega_n t\right]. \tag{30}$$

For $\zeta < 1$ the system is underdamped (oscillatory), for $\zeta = 1$ critically damped and for $\zeta > 1$ overdamped (non-oscillatory). In order to avoid sluggishness of the system, a rule of thumb may be followed making. $4 < \zeta < 1^2$. The transient response of (30) can be plotted in dimensionless form if certain specifications are made for the ratio ω_n/K. As the time constant τ_1 of the control network must be positive or can at most be equal to zero, the maximum value for ω_n/K is found from (27), yielding

$$\left.\frac{\omega_n}{K}\right|_{\max} = 2\zeta. \tag{31}$$

In this case the control network is reduced to a single time constant network ($\tau_1 = 0$). On the other hand, if for a fixed value of ω_n the gain of the system is increased towards infinity, the minimum value for ω_n/K becomes

$$\left.\frac{\omega_n}{K}\right|_{\min} = 0. \tag{32}$$

Fig. 4 shows the transient response of the system for these two limits and for a damping ratio of $\zeta = 0.5$.

[2] G. S. Brown and D. P. Campbell, "Principles of Servomechanisms," John Wiley & Sons Publishing Co., New York, N. Y.; 1948.

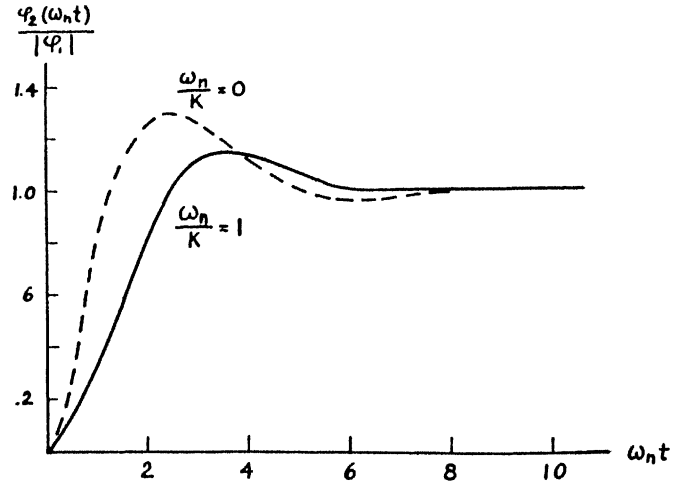


Fig. 4—Transient response for $\zeta=0.5$.

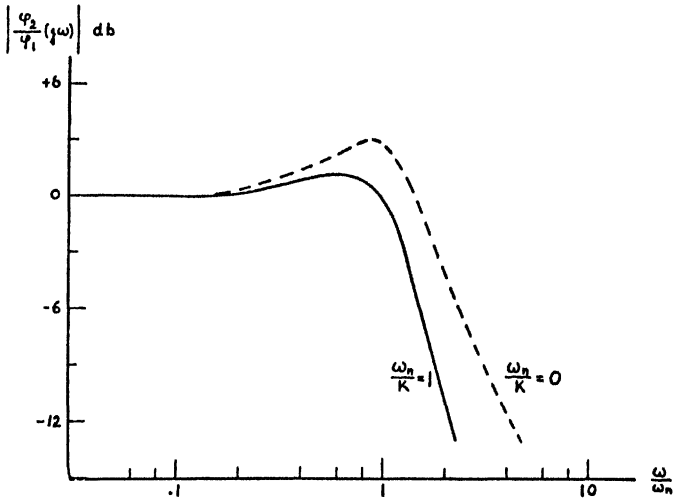


Fig. 5—Frequency response for $\zeta=0.5$.

The frequency response of the system is readily found from (24) and one obtains

$$\frac{\phi_2}{\phi_1}(j\omega)=\frac{1+j2\zeta\frac{\omega}{\omega_n}\left(1-\frac{\omega_n}{2\zeta K}\right)}{1+j2\zeta\frac{\omega}{\omega_n}-\left(\frac{\omega}{\omega_n}\right)^2}. \tag{33}$$

Its magnitude is plotted in Fig. 5 for the two limit values of ω_n/K and for a damping ratio $\zeta=0.5$. The curves show that the cut-off frequency of the system, for $\zeta=0.5$, is approximately

$$\omega_c \cong \omega_n \text{ [radians/sec.]}. \tag{34}$$

If ϕ_1 and ϕ_2 in (33) are assumed to be the input and output voltage of a four-terminal low-pass filter, the frequency response leads to the equivalent circuit of Fig. 6.

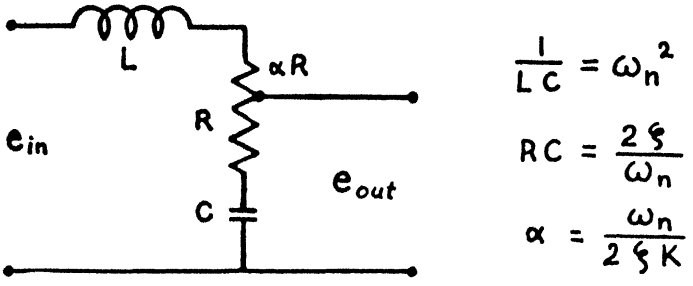


Fig. 6—Equivalent low-pass filter.

The noise bandwidth of the system is established by inserting (33) into (22) and one obtains

$$B=\omega_n\int_{-\infty}^{+\infty}\frac{1+4\zeta^2\left(\frac{\omega}{\omega_n}\right)^2\left[1-\frac{\omega_n}{2\zeta K}\right]^2}{1-(2-4\zeta^2)\left(\frac{\omega}{\omega_n}\right)^2-\left(\frac{\omega}{\omega_n}\right)^4}\,d\left(\frac{\omega}{\omega_n}\right). \tag{35}$$

The integration, which can be carried out by partial fractions with the help of tables, yields

$$B=\frac{4\zeta^2-4\zeta\frac{\omega_n}{K}+\left(\frac{\omega_n}{K}\right)^2+1}{2\zeta}\pi\omega_n. \tag{36}$$

For small values of ω_n/K, it is readily established that this expression has a minimum when $\zeta=0.5$. Hence, the noise bandwidths for the limit values of ω_n/K and $\zeta=0.5$ become

$$B\,|_{(\omega_n/K)\to 1}=\pi\omega_n=\pi K \text{ [radians/sec]} \tag{37}$$

and

$$B\,|_{(\omega_n/K)\to 0}=2\pi\omega_n \text{ [radians/sec]}. \tag{38}$$

The above derivations, as well as the response curves of Figs. 4 and 5, show that the bandwidth and the gain constant of the system can be adjusted independently if a double time-constant control network is employed.

c. Example

The theory is best illustrated by means of an example. Suppose an AFC system is to be designed, having a steady state phase error of not more than 3° and a noise bandwidth of 1,000 cps. The local oscillator drift shall be assumed with 1,500 cps.

The required gain constant is obtained from (15), yielding

$$K=\frac{\Delta\omega}{\sin\phi}=\frac{2\pi\cdot 1{,}500}{\sin 3^\circ}=180{,}000 \text{ radians/sec.}$$

Since K is large in comparison to the required bandwidth, the resonance frequency of the system is established from (38).

$$\omega_n=\frac{B}{2\pi}=\frac{2\pi\cdot 1{,}000}{2\pi}=1{,}000 \text{ radians/sec.}$$

The two time constants of the control network, assuming a damping ratio of 0.5, are determined from (27) and (28) respectively

$$\tau_1=\frac{2\zeta}{\omega_n}-\frac{1}{K}=\frac{1}{1{,}000}-\frac{1}{180{,}000}\cong 10^{-3} \text{ sec,}$$

and

$$\tau_2=\frac{K}{\omega_n^2}=\frac{180{,}000}{1{,}000^2}=0.18 \text{ sec.}$$

These values K, τ_1, and τ_2 completely define the AFC system. A proper choice of gain distribution and control-network impedance still has to be made to fit a particular design. For example, if the peak amplitude of the sinusoidal oscillator reference voltage is $E_2=6$ volts, the sensitivity of the oscillator control stage must be $S=30{,}000$ radians/sec/volt to provide the necessary gain constant of 180,000 radians/sec. Furthermore, if the capacitor C for the control network of Fig. 3(a) is assumed to be 0.22 *uf*, the resistors R_1 and R_2 become 4.7 kΩ and 820 kΩ respectively, to yield the desired time constants.

IV. Non-Linear Analysis

While it was permissible to assume small phase angles for the study of the synchronized system, thereby linearizing the differential (13), this simplification cannot be made for the evaluation of the pull-in performance of the system. The pull-in range of synchronization is defined as the range of difference frequencies, $p\phi_1$, between the input signal and the free-running oscillator signal, over which the system can reach synchronism. Since the difference phase ϕ can vary over many radians during pull-in, it is necessary to integrate the nonlinear equation to establish the limit of synchronization.

Assuming that the frequency of the input signal is constant as defined by (14), (13) can be written

$$p\phi + F(p)K \sin \phi = \Delta\omega. \tag{39}$$

Mathematically then, the pull-in range of synchronization is the maximum value of $\Delta\omega$ for which, irrespective of the initial condition of the system, the phase difference ϕ reaches a steady state value. To solve (39), the transfer function of the control network again must be defined.

a. $F(p) = 1$

The pull-in performance for this case has been treated in detail by Labin.[3] With $F(p)=1$ (39) can be integrated by separation of the variables and it is readily found that the system synchronizes for all values of $|\Delta\omega| < K$. The condition for pull-in then is

$$|\Delta\omega|_{\text{Pull-in}} < K. \tag{40}$$

Large pull-in range and narrow-noise bandwidth thus are incompatible requirements for this system.

b. $F(p) = \dfrac{1+\tau_1 p}{1+\tau_2 p}$

Inserting this transfer function into (39) and carrying out the differentiation yields

$$\frac{d^2\phi}{dt^2} + \left[\frac{1}{\tau_2} + K\frac{\tau_1}{\tau_2}\cos\phi\right]\frac{d\phi}{dt} + \frac{K}{\tau_2}\sin\phi = \frac{\Delta\omega}{\tau_2}. \tag{41}$$

This equation can be simplified by inserting the coefficients defined in (25) and (26), and by dividing the resulting equation by ω_n^2. This leads to the dimensionless equation.

$$\frac{d^2\phi}{\omega_n^2 dt^2} + \left[\frac{\omega_n}{K} + \left(2\zeta - \frac{\omega_n}{K}\right)\cos\phi\right]\frac{d\phi}{\omega_n dt} + \sin\phi = \frac{\Delta\omega}{K}. \tag{42}$$

A further simplification is possible by defining a dimensionless difference frequency

$$y \equiv \frac{d\phi}{\omega_n dt} \tag{43}$$

and one obtains a first order differential equation from which the dimensionless time $\omega_n t$ has been eliminated. It follows

$$\frac{dy}{d\phi} = \frac{\dfrac{\Delta\omega}{K} - \sin\phi}{y} - \frac{\omega_n}{K} - \left(2\zeta - \frac{\omega_n}{K}\right)\cos\phi. \tag{44}$$

There is presently no analytical method available to solve this equation. However, the equation completely defines the slope of the solution curve $y(\phi)$ at all points of a $\phi - y$ plane, except for the points of stable and unstable equilibrium, $y=0$; $\Delta\omega/K=\sin\phi$. The limit of synchronization can thus be found graphically by starting the system with an infinitestimal velocity Δy at a point of unstable equilibrium, $y=0$; $\phi=\pi-\sin^{-1}\Delta\omega/K$, and finding the value of $\Delta\omega/K$ for which the solution curve just reaches the next point of unstable equilibrium located at $y=0;\phi=3\pi-\sin^{-1}\Delta\omega/K$. The method is discussed by Stoker[4] and has been used by Tellier and Preston[5] to find the pull-in range for a single time constant AFC system.

To establish the limit curve of synchronization for given values of ζ and ω_n/K, a number of solution curves have to be plotted with $\Delta\omega/K$ as parameter. The limit of pull-in range in terms of $\Delta\omega/K$ then can be interpolated to any desired degree of accuracy. The result, obtained in this manner, is shown in the dimensionless graph of Fig. 7, where $\Delta\omega/K$ is plotted as a function of ω_n/K for a damping ratio $\zeta=0.5$. Since this curve represents the stability limit of synchronization for the system, the time required to reach synchronism is infinite when starting from any point on the limit curve. The same applies to any point on the $\Delta\omega/K$-axis, with exception of the point $\Delta\omega/K=0$, since this axis describes a system having either infinite gain or zero bandwidth, and neither case has any real practical significance. The practical pull-in range of synchronization, therefore, lies inside the solid boundary. The individual points

[3] Edouard Labin, "Theorie de la synchronization par controle de phase," *Philips Res. Rep.*, (in French); August, 1941.

[4] J. J. Stoker, "Non-linear vibrations," *Interscience;* New York, 1950.

[5] G. W. Preston and J. C. Tellier, "The Lock-in Performance of an A.F.C. Circuit," Proc. I.R.E.; February, 1953.

entered in Fig. 7 represent the measured pull-in curve of a particular system for which the damping ratio was maintained at $\zeta = 0.5$. For small values of ω_n/K this

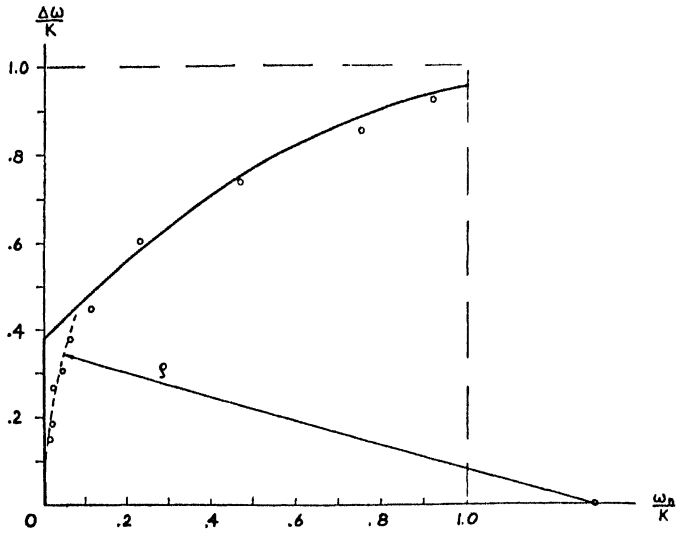


Fig. 7—Pull-in range of synchronization for $\zeta = 0.5$.

pull-in curve can be approximated by its circle of curvature which, as indicated by the dotted line, is tangent to the $\Delta\omega/K$-axis and whose center lies on the ω_n/K-axis. The pull-in range thus can be expressed analytically by the equation of the circle of curvature. If its radius is denoted by ζ, the circle is given by

$$\left(\frac{\omega_n}{K} - \zeta\right)^2 + \left(\frac{\Delta\omega}{K}\right)^2 = \zeta^2. \tag{45}$$

Hence, for $(\omega_n/K) \to 0$, the pull-in range of synchronization is approximately

$$\left|\Delta\omega_{\text{Pull-in}}\right|_{(\omega_n/K)\to 0} < \sqrt{2\zeta\omega_n K - \omega_n^2} \cong \sqrt{2\zeta\omega_n K}. \tag{46}$$

ζ can be interpreted as a constant of proportionality which depends on the particular design of the system, and which increases as the system gets closer to the theoretical limit of synchronization.

Equation (46) shows that the pull-in range for small values of ω_n/K is proportional to the square root of the product of the cut-off frequency ω_n and the gain constant K. Since the bandwidth of a double time constant AFC system can be adjusted independently of the gain constant, the pull-in range of such a system can exceed the noise bandwidth by any desired amount.

V. Conclusions

The performance of an AFC system can be described by three parameters. These are the gain constant K, the damping ratio ζ and the resonance or cut-off frequency ω_n. These parameters are specified by the requirements of a particular application and define the over-all design of the system. It has been shown that among the systems with zero, single and double time constant control networks, only the latter fulfills the requirement for achieving good noise immunity, small steady-state phase error and large pull-in range.

Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television*

DONALD RICHMAN†, SENIOR MEMBER, IRE

Summary—**The results of an evaluation of the capabilities of the NTSC color-carrier reference signal (the color burst) show this new color television synchronizing signal to be more than adequate; information inherent to the signal permits performance far in excess of that achieved by conventional circuits.**

Phasing information inherent to the burst is considered first with particular regard to measures of accuracy, the required amount of integration, and the extent of the spectral region necessary to translate the burst information.

Properties of elementary passive and active circuits for using the burst in receivers are described along with a determination of the limits of burst synchronization performance for these circuits.

Fundamental considerations in the theory of synchronization show that better performance is obtainable with two-mode systems.

Properties of two-mode systems are considered and lead to an evaluation of the limits of synchronizing performance permitted by the color burst.

The mathematical derivations necessary to support the discussion are presented in the Appendixes.

NTSC COLOR television adds color to a monochrome picture by means of a narrow-band, frequency interleaved carrier color signal which carries one component of the color information in its phase, and another component in its amplitude. It is customary to provide a phase reference in the transmitted signal in order that receivers shall be able to measure the instantaneous phase angle of the carrier color signal so as to reproduce the desired color. This is accomplished by transmitting a short burst of oscillations at color subcarrier frequency during line retrace intervals,[1] at a reference phase which corresponds to the (*Y-B*) axis.[2]

The color burst carries phasing information. This paper shows how much phasing information is contained in the color burst, and how it may be used.

Analysis of the factors limiting performance shows that, even under extreme conditions of interference and of stabilization requirements, the burst contains adequate information to provide a reliable color-carrier reference signal; in fact, the amount of phasing information in the color signal appears adequate enough so that a customer-operated control relating to color sync should be unnecessary on NTSC color television receivers. Analysis shows that presently used sync instrumentation systems appear capable of meeting but not necessarily exceeding a reasonable measure of the above requirements. However, information existing in the *signal* permits substantially better performance. The real limits of performance and sync systems which more fully utilize the signal information are discussed in this paper. Because of the excess of existing information, a variety of types of circuits can be used.

Several questions may be asked with regard to the amount of phasing information contained in the color burst and its application to provide a reference signal for color demodulation. These are: (a) How closely can the color-carrier reference signal be maintained to the true value, when signals are strong (and hence noise-free) and after transient effects have subsided? (b) How closely can the color-carrier reference signal be maintained in the presence of noise interference? (c) How long will a system or circuit designed to give satisfactory operation on (a) and (b) require to reach a stable mode of operation when stations are switched or a receiver is turned on? (d) How much performance is required in (a), (b), and (c)?

Of these questions, (d) is the most difficult to answer precisely; it depends on many subjective factors and may be obscured by temporary equipment difficulties. In order to provide a standard of comparison for use in this paper, a conservative (pessimistic) estimate has been made, based on past experience.

The answers to the questions are as follows:

(a) With a strong (clean) sync signal, the color-carrier reference signal may be maintained as closely accurate as desired, independent of other factors; in the presence of noise, the *average* phase may be maintained as closely as desired, independent of the required integration and transient characteristics; for example, designs presented later show how the static or average phase of the color-carrier reference signal may be controlled to within five degrees of the true value. Expressed as a time value this is an accuracy of approximately six mμsec. This phase accuracy implies a color fidelity probably substantially better than can be distinguished by the observer.[3]

(b) *The real limitation on performance* is thermal-noise interference, since this type of interference is the most difficult type to reject. It *is* rejected, however, *to any selected measure of reliability* by integration of the synchronization timing information over a suitably long period. Either of two basic types of integrators may be used. These are, one, passive integrators, and two, frequency-and-phase-locked self-oscillating integrators. The analysis presented in this paper shows that, under severe assumptions on the requirements of phase stability and signal-to-noise ratio, the required integration time for passive integrators is of the order of mag-

* Decimal classification: R583. NTSC Technical Monograph No. 7, reprinted by permission of the National Television System Committee from "Color System Analysis," report of NTSC Panel 12.

† Hazeltine Corp., Little Neck, N. Y.

[1] "Recent developments in color synchronization in the RCA color television system," RCA Labs. Report, Princeton, N. J.; Feb., 1950.

[2] Fig. 1 of "Minutes of the Meeting of Panel 14," NTSC; May 20, 1952.

[3] D. L. MacAdam, "Quality of color reproduction," PROC. I.R.E., vol. 39, pp. 468–485; May, 1951.

Reprinted with permission from *Proc. of IRE*, D. Richman, "Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," vol.42, pp. 106-133, January 1954.

nitude of 0.005 second, or less than a sixth of a frame period. Locked integrators on the same assumptions require 0.01 second for the integration to take place.

(c) The third requirement, of pull-in or stabilization time, is also limited by the signal-to-noise ratio and the requirement for integration. This may vary considerably with the method of instrumentation, but the limiting or optimum performance with regard to stabilization time is determined by the information carried in the signal; the limit imposed by signal information is found to be (for a reasonable measure of reliability) a few times the integration time discussed above. Later in this report this is shown to be approached under certain conditions by fairly simple passive integrators. It is also shown how locked integrators, characterized by some new forms of automatic frequency- and phase-control loops, may be made to achieve the upper limit of performance. Typical present APC (automatic phase control) circuits fall somewhat short of this limit, but when properly designed can be made to pull in quickly enough so as to appear virtually instantaneous, while permitting most of the burden of frequency stability to be borne by the transmitter.

These facts lead to the conclusion that there is adequate information in the color burst for completely automatic operation, without need for a customer control. The factors leading to this conclusion are presented in the following sequence:

Performance limitations for sync systems which are already synchronized are discussed first, in the section on "Synchronization Accuracy." The reliability of phase difference measurements, and factors relating to the integration time necessary to obtain a specified measure of reliability in the presence of noise are considered.

Then performance limitations of instrumentation systems are discussed with particular regard to the process of synchronization. The basic characteristics of passive and locked integrators are discussed in the section on "Elemental Sync Systems."

Evaluation of ultimate limitations for the signal, and factors leading to new sync systems capable of fully utilizing the signal information are presented in the section on "Theory of Synchronization." Factors of interest are mechanisms of pull-in, the reliability of frequency difference measurements, and the exchange of integration time for a specified measure of reliability in the presence of noise.

Effects of echoes and stability of the gate are briefly discussed.

The conclusions drawn regarding the adequacy of the signal are stated.

Mathematical derivations, which substantiate and illustrate the facts presented in this paper, are presented in several appendexes.

The NTSC Color Synchronizing Signal

Fig. 1 shows the NTSC color synchronizing signal in relation to the video and synchronizing wave form, in the vicinity of one line-retrace interval. It consists of a burst of approximately 9 cycles of sinusoidal wave form at the color-carrier frequency of 3,579,545 ($\pm$.0003%) cps,[4] approximately centered on the portion of the line blanking pulse following each horizontal sync pulse. It is omitted during the nine lines in each field in which the field synchronizing information is transmitted.

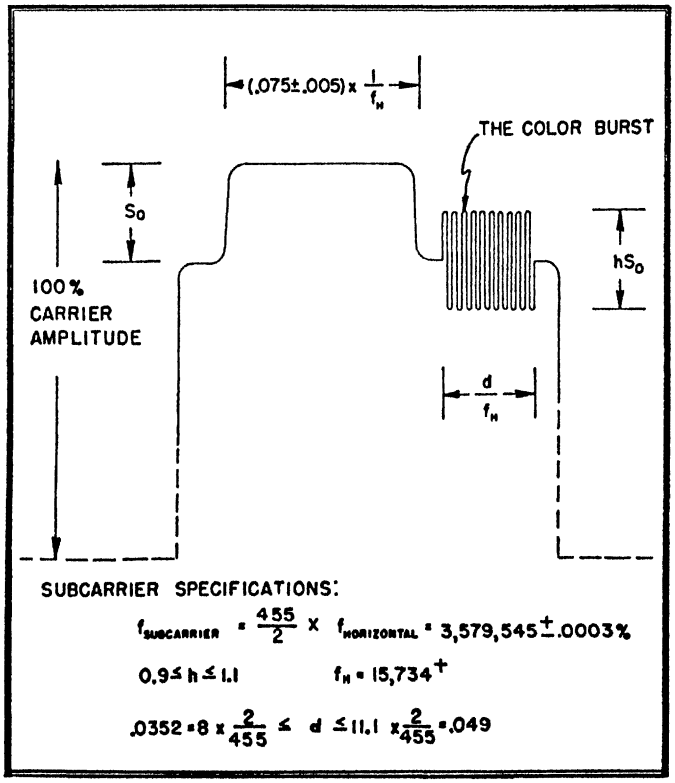


Fig. 1—Wave form during line retrace interval showing horizontal sync pulse and the NTSC burst reference signal.

Parameters of interest which are shown on the figure are:

S_0 = the amplitude of the line and field sync pulses, normally 25% of peak carrier amplitude measured in the video signal.

hS_0 = the peak-to-peak amplitude of the burst, measured in the video signal.

f_H = the line scanning frequency.

d = the duty cycle of the burst.

The color burst is used in the color television receiver to provide a control signal for the generation of a local continuous wave signal at the nominal burst frequency and locked to it in phase.

Synchronization Accuracy

Synchronizing Information

Any time-varying signal can carry timing information, the character of which depends on the distribution of signal energy throughout the frequency spectrum. In

[4] As specified by NTSC in February, 1953. The analysis is not critically dependent on the exact value of the color carrier frequency.

the case of a continuous sine wave, this timing information consists only of phase reference information because it is impossible to identify cycles of the carrier from each other. The same is essentially true of the pulse modulated sine wave which constitutes the burst; envelope information in the burst is not used. It is this phase reference information which is of interest with regard to color-carrier reference phase synchronization.

A signal which passes only through linear noiseless channels may be located in time (or phase) with theoretically unlimited precision. In the presence of noise the data obtained by a time (or *phase*) meter from the signal will fluctuate. This occurs because the timing information which can be extracted from the combination of signal-plus-noise in any specified interval is limited by the signal-to-noise ratio as well as by the statistical characteristics of the signal and noise.

Integration for Signal-to-Noise Ratio Improvement

The fluctuations in the phase data may be smoothed by integration. For example, the instantaneous output of the phase meter may represent the average of all data obtained over some preceding integration period T_M in duration.

Any measuring device which uses any form of integration or memory averages some effective number of independent measurements. One such integrator directly obtains a suitably weighted average (such as the least square error average) of all the data obtained in the preceding period T_M. Such an integrator provides a standard of comparison. Other forms of integrators may then be characterized by their effective integration times, T_M; several practical integrators are described later.

A section of a signal existing in an interval of duration T_M may be expressed as a sum of harmonics of the fundamental frequency $(1/T_M)$; the noise bandwidth associated with each component is equal to the spacing between components, or $1/T_M = f_N$. This means, for example, that if all of the timing information obtained in a period T_M from a signal consisting effectively of a signal sinusoidal component is averaged, that an improvement in reliability is obtained equivalent to that produced by passing the signal through a filter having a noise bandwidth of f_N.

Noise Interference

Noise is specified by its energy content and statistical characteristics. For a flat energy spectrum, taken as an example, impulse noise and white thermal noise represent opposite extremes, since for white thermal noise the relative phases of the several frequency components are completely random and incoherent; for impulse noise the relative phases of all components are related and are not random, although the time of occurrence of any impulse is a random variable.

Noise may be measured in terms of any convenient co-ordinate system into which the signal-plus-noise may be transformed, such as frequency, phase, amplitude, time of arrival, or more complex parameters.

Thermal noise is the most difficult to reject. It may be discriminated against only by averaging; this makes the effective error due to noise vary inversely as the square root of the number of measurements; hence, (for systems with fixed bandwidth) the error varies inversely as the square root of the integration time.

Impulse noise, or noise intermediate between thermal and impulse noise, may be rejected more easily than thermal noise since it represents a signal which can be recognized with a high measure of reliability and removed from the transmission channel.

A synchronizing system is a form of predictor which bases its estimates on past experience. When the input to the system has such a character (such as an improbable amplitude) that it is recognized with high reliability to be a disturbance, it is usually much better to use (at least approximately) the predicted signal as the input to the system for the duration of the disturbance. An equipment system for performing these operations is called an *aperture*. (Aperture systems are now widely used for line and field sync; the same principles are involved in the application to burst sync.)

Since thermal noise represents the most serious (as well as perhaps the most common) limitation to color synchronization performance, it is used in this paper as the measure of interference which must be overcome.

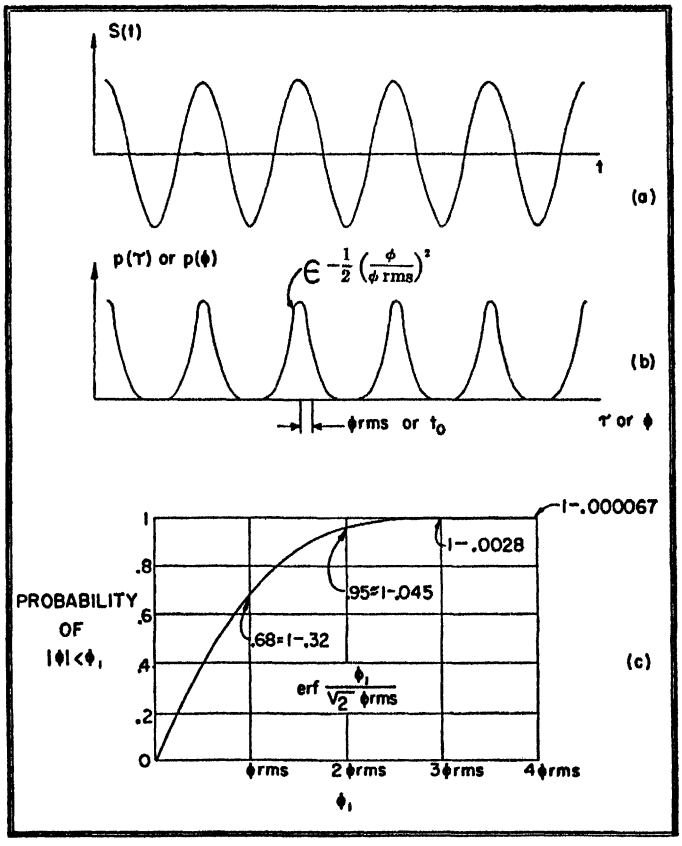


Fig. 2—Timing error distribution.

Measures of Reliability

A section of the burst reference signal is represented as $S(t)$ in Fig. 2(a). The time scale associated with the synchronizing signal may be identified with some representative point in a cycle which is selected as a reference.

The timing accuracy which is obtained for a given signal-to-noise ratio may be expressed in terms of a relative probability density function $p(\tau)$ such as is plotted in the curve of Fig. 2(b). The relative probability density curve permits the determination of the probability that the sync timing answer which results from a single measurement of the sync signal, using all of the information derived from the preceding period T_M, will occur within a specified time or phase interval. This probability is proportional to the area under the curve $p(\tau)$ or $p(\phi)$ within the specified interval. Due to the cyclic nature of the information, the time scale may be replaced by a phase scale. The curve for $p(\phi)$ defines the probability laws for the noise at the output of the synchronizing system. The curve is repetitive at the sync frequency. (The output noise from the sync measuring device has the same basic character from cycle to cycle.) For many signal energy distributions, and particularly for burst synchronization at the levels of output noise which give satisfactory performance, the curve $p(\tau)$ or $p(\phi)$ has very nearly the shape of a normal or Gaussian probability curve represented by the expression

$$\epsilon^{-1/2(t/t_0)^2} \quad \text{or} \quad \epsilon^{-1/2(\phi/\phi_{rms})^2}$$

in which case the phasing information may be completely described by the rms time error, t_0, or the rms phase error ϕ_{rms}, which may be expected for a specified set of measurement conditions.

For this case of the normal law the absolute probability that any measurement will yield an answer within a specified measure of the true answer may be represented in terms of the rms error. Fig. 2(c), which represents the integral of one lobe of the curve of Fig. 2(b) for the normal law, represents the probability that the magnitude of the phase error at any time is less than some selected phase error ϕ_1. ϕ_1 is measured in multiples of ϕ_{rms}. The curve illustrates that the probability is nearly unity only when ϕ_1 approaches $4\phi_{rms}$, which means that the effective peak value of Gaussian noise is near four times the rms value.[5]

The Sync Accuracy Equation

The parameters which determine the rms time error, seconds, for burst sync are:

The signal amplitude $\frac{1}{2}hS_0$ volts.
The duty cycle of the gated sine wave d as a fraction.
The rms noise (assumed flat over the band) N_W volts.
The video bandwidth occupied by the signal and noise f_W cycles per second.
The subcarrier frequency f_{SC} cycles per second.
The effective integration time T_M seconds.
The rms phase error ϕ_{rms} in degrees.
Equation (1) relates these parameters

$$\frac{S_0}{N_W} = \frac{1}{\sqrt{df_WT_M}} \frac{1}{t_0f_{SC}} \frac{1}{\pi h} = \frac{1}{\sqrt{df_WT_M}} \frac{360}{\phi_{rms}} \frac{1}{\pi h} \tag{1}$$

This equation is derived in Appendix A.[6] The physical significance of the several factors in (1) is as follows:

The factor S_0/N_W represents (for example) the smallest ratio of line sync amplitude to rms noise for which t_0f_{SC} will not exceed a selected arbitrary value. It may be visually estimated if the composite video signal is viewed with a wide band oscilloscope. When $S_0/N_W = 1$ the rms noise is equal to sync pulse amplitude. Since S_0 represents 25% carrier amplitude, and since the effective peak value of Gaussian noise is approximately four times the rms value, the condition $S_0/N_W = 1$ also corresponds to the "peak" noise being approximately equal to 100% of carrier amplitude.

The factor t_0f_{SC} represents the fraction of a cycle of phasing error at frequency f_{SC} corresponding to the timing error, t_0. Thus

$$t_0f_{SC} = \frac{\text{rms phase error in degrees}}{360^\circ} = \frac{\phi_{rms}}{360^\circ}.$$

The factor df_WT_M is the number of effectively independent measurements yielding phase information which may be made in the interval T_M on a signal which is present for only a fraction d of time, and which occupies portions of the bandwidth f_W. The signal is actually present for a period dT_M; the effect of integrating over the period T_M is therefore to reduce the rms error by

$$\sqrt{df_WT_M} = \sqrt{d\frac{f_W}{f_N}},$$

where $f_N = 1/T_M$ is the effective noise bandwidth.

The factor $1/\pi h$ is a constant.

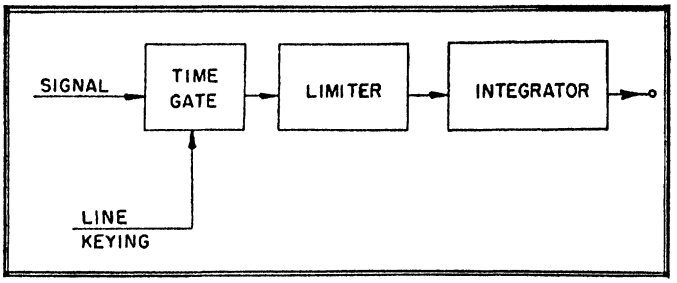


Fig. 3—Typical color-carrier phase reference generation system.

The Required Sync Accuracy

Equation (1) represents the theoretical upper limit of the phasing accuracy which may be derived from the subcarrier burst. A variety of circuits are available which can approach closely to this limit; these circuits are often of the form shown in Fig. 3. The composite

[5] V. D. Landon, "The distribution of amplitude with time in fluctuation noise," PROC. I.R.E., vol. 29, pp. 50–55; Feb., 1941.

[6] D. Richman, "Theoretical limit to time difference measurements," *Proc. NEC*, vol. 5; pp. 203–210; 1949.

video signal is fed to a time-gate which is keyed from line flyback to select the burst, which is then amplitude limited and integrated. Practical integrators are described later.

The sync accuracy equation permits the determination of how much integration is required in order to obtain satisfactory performance under extreme conditions. However, due to the many subjective factors involved it is not possible to specify exactly what is the lowest level of signal-to-noise ratio which will be tolerable from a visual viewpoint;[7] it is equally difficult to specify exactly the largest value of rms phasing error which will not cause visible degradation of the picture. Accordingly, Fig. 4, which is a plot of (1), presents graphically the relations between the relevant factors over a range which probably includes the limiting case of interest. Fig. 4 is based on adverse tolerances presented below.

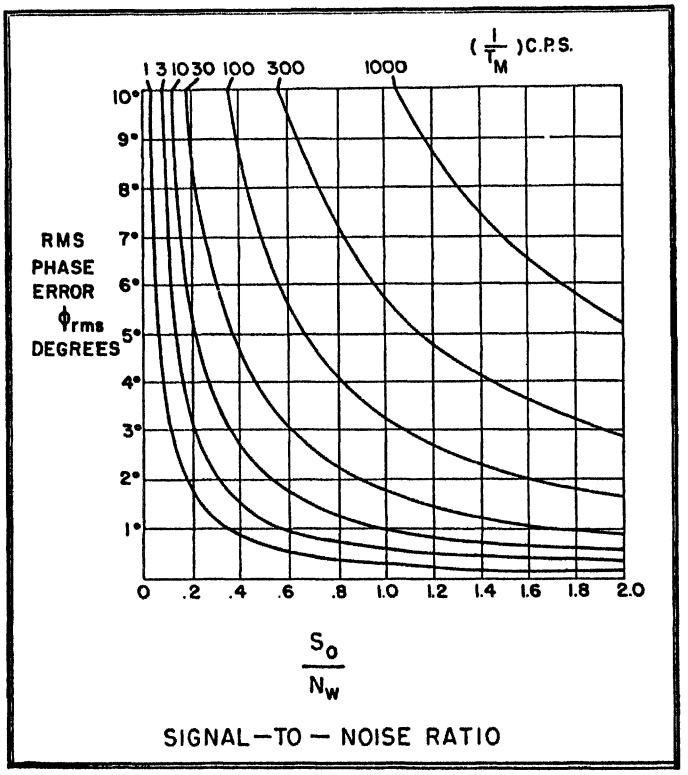


Fig. 4—Phasing accuracy relations for NTSC burst synchronization.

Fig. 4 presents the relation between the rms phase error, ϕ_{rms}, (in degrees) and the signal-to-noise ratio, S_0/N_W, with the integration time, T_M, (in seconds) as a parameter.

For the case corresponding to the most adverse tolerances, $h=.9$, $d=.0352$, and $f_W=4.3$ mc. Equation (1) then reduces to

$$\phi_{rms}\frac{S_0}{N_W} = .33\sqrt{\frac{1}{T_M}} = \frac{1}{3}\sqrt{\frac{1}{T_M}} \qquad (2)$$

which is shown graphically in Fig. 4.

[7] P. Mertz, A. D. Fowler and H. N. Christopher, "Quality rating of television images," Proc. I.R.E., vol. 38, pp. 1269–1283; Nov., 1950.

These curves show that any selected phase accuracy ϕ_{rms} can be obtained with decreasing signal-to-noise ratio S_0/N_W if more time T_M is taken for integration of the signal timing information; i.e., if more measurements are integrated in each complete measurement.

(The facts presented later in this paper with regard to the relations between noise integration and other properties of sync systems indicate that the conclusions reached regarding the reliability of the signal are *not* critically dependent upon the assumed values of S_0/N_W and ϕ_{rms}.)

System Efficiency and the Distribution of Timing Information

The relationships presented above describe the performance of the system when all of the information of the signal is applied usefully. Another parameter which needs to be introduced in order to determine the actual noise bandwidth required is the decoding efficiency, which represents the fraction of the timing information of the signal which is used. Systems with equal noise bandwidths but different decoding efficiencies will give different performance.

In the burst system practical considerations relating to tolerances and to the stability of the gate derived from horizontal sync may result in a gate width r times wider than the narrowest sync burst. Factors relating to this are described later. It results in a requirement of noise bandwidth and integration time such that

$$T_M = \sqrt{r}\,T_{M\text{LIMIT}}$$

$$f_N = \frac{1}{\sqrt{r}}f_{N\text{LIMIT}}$$

where

$(1/\sqrt{r})$ is a system efficiency

r = ratio of actual gate width to minimum burst width.

There is another cause of loss of decoding efficiency in sync systems which is of interest. This relates to the relative distribution of timing information in the frequency spectrum. For burst sync systems which are properly designed, effectively all of the information may be used; (common attainment in horizontal sync systems has not been so high).

Fig. 5(a) shows the relative distribution of timing information in the frequency spectrum occupied by the burst. The basis for this curve is discussed in Appendix A. The effective accuracy which can be obtained if only a portion of the information is used may be measured in terms of the ratio of the noise bandwidth required (at any specified signal-to-noise ratio) to the noise bandwidth required if all of the information is used. For example, a problem of interest in receiver design is the relationship between bandwidth in the burst amplification channel and efficiency. If a passband symmetrically tuned about subcarrier frequency is used in this channel then the system efficiency resulting is represented by

the curve sketched in Fig. 5(b). The curve depends, of course, on the width of the burst. Even for the narrowest burst a total bandwidth of approximately 600 kc translates nearly all of the timing information.

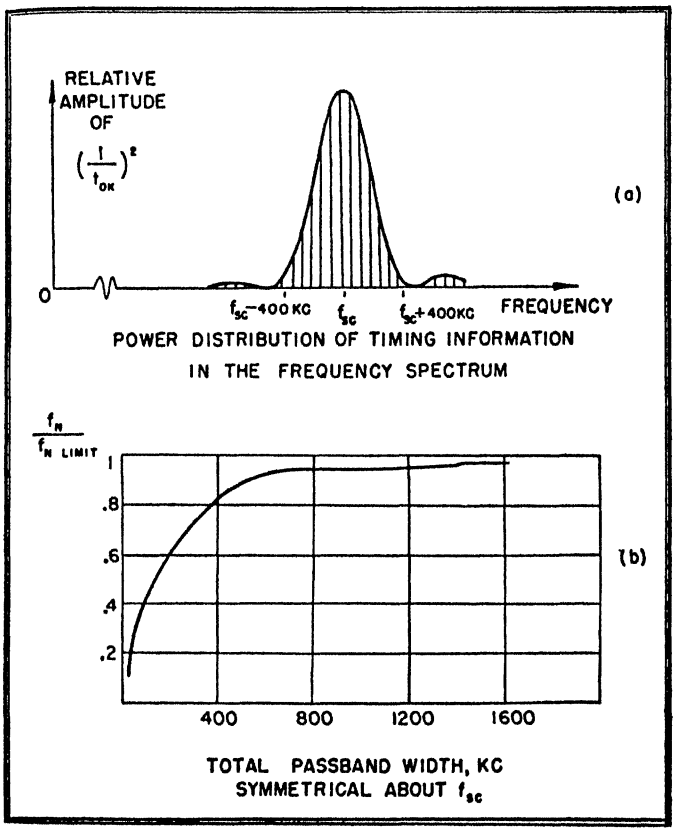


Fig. 5—Frequency distribution and system efficiency of burst sync timing information.

Example: As an illustration suppose the limiting parameter values of interest are approximately $\phi_{rms} = 5°$ and $S_0/N_W = 1$; these conditions correspond to the point in the center of Fig. 4; then from (2) $T_M \geqq 0.0045$ second. *The required noise bandwidth* for a gate width ratio $r = 1.2$ is then approximately $f_N = 200$ cycles per second. *This figure is used as a basic design parameter for the practical forms of integrators which will be discussed in this paper.*

Elemental Sync Systems

The function of combining signal information derived over an extended interval of time is accomplished by use of circuits which may broadly be classified as integrators. The performance characteristics of two basic forms of integrators are discussed below. The parameters of interest are:

1. The noise bandwidth and integration time of the system.
2. The static phase accuracy. In general, in systems involving feedback, this varies inversely with a circuit gain parameter and may be made nominally as small as desired.
3. The frequency pull-in range of the system. This is the maximum (single peak) frequency detuning for which the system will automatically achieve the desired final operating condition.
4. The stabilization time T_S; or the time required for all operating characteristics to reach effectively their stabilized conditions. This may consist of one or more definable segments.
5. The phase pull-in time T_ϕ; or the transient time required for the output phase of the system to reach some definable measure of its final conditions.
6. The frequency pull-in time T_F, applicable to systems in which a local signal oscillator must be controlled, or the time necessary for the oscillator frequency to be changed from its initial frequency to some selected reference frequency such as a frequency from which the net differential phase change between sync signal and reference oscillator will not exceed one whole cycle. This overlaps the phase pull-in time T_ϕ.

The first integration system discussed is the Passive Integrator. For this system stabilization consists effectively of a phase transient. The limitations of this system are: practical limitations on how high the circuit Q may be and the possibility of detuning.

These limitations are overcome in the second form of integrator called a Standard APC (Automatic Phase Control) System. In this system the signal is heterodyned against a local carrier at the same frequency permitting the desired filtering to be accomplished by means of a low-pass filter which thus effectively provides unlimited Q. The limitations of this system relate to the difficulties of obtaining synchronization and the long pull-in times which result when narrow noise bandwidths are required.

The real limitations imposed by the signal, and some system fundamentals related to using all of the information in the signal, are presented later.

Passive Integrator

The circuit of Fig. 6(a) shows one form of practical integrator. This is a passive integrator in which the required integration is obtained by use of a high-Q filter. The input signal to the filter consists of time-gated amplitude-limited bursts of sine waves at subcarrier frequency f_{SC}. Because of the gating and limiting, sidebands near f_{SC} (which are separated by integral multiples of f_H) as well as harmonics of f_{SC} which are generated in the preceding limiter, all have effectively the same phase modulation due to noise. The noise bandwidth of the filter needs to be less than or equal to the value of f_N which was computed above. If the filter is approximately equivalent to a single resonant circuit, the noise bandwidth is $f_N = (\pi/2)f_3$ where f_3 is the 3 db bandwidth. The bandwidth f_N is indicated in Fig. 6(b). Thus the filter bandwidth should be approximately $(2/\pi)(200) = 127$ cps between 3 db points. The Q desired is $f_{SC}/f_3 \approx 28{,}000$. This requires the use of a crystal filter. Practical crystals in the frequency range of the color subcarrier can achieve the required Q, but up to the present time apparently cannot exceed it by a

large factor.[8,9] The sum of transmitter frequency tolerance of ±11 cps and the frequency tolerance of the crystal is comparable with the filter bandwidth. Fig. 6(c) shows how undesirably large static phase shift might result from normal detunings. This is prevented in the system shown in Fig. 6(a) by use of feedback for automatic

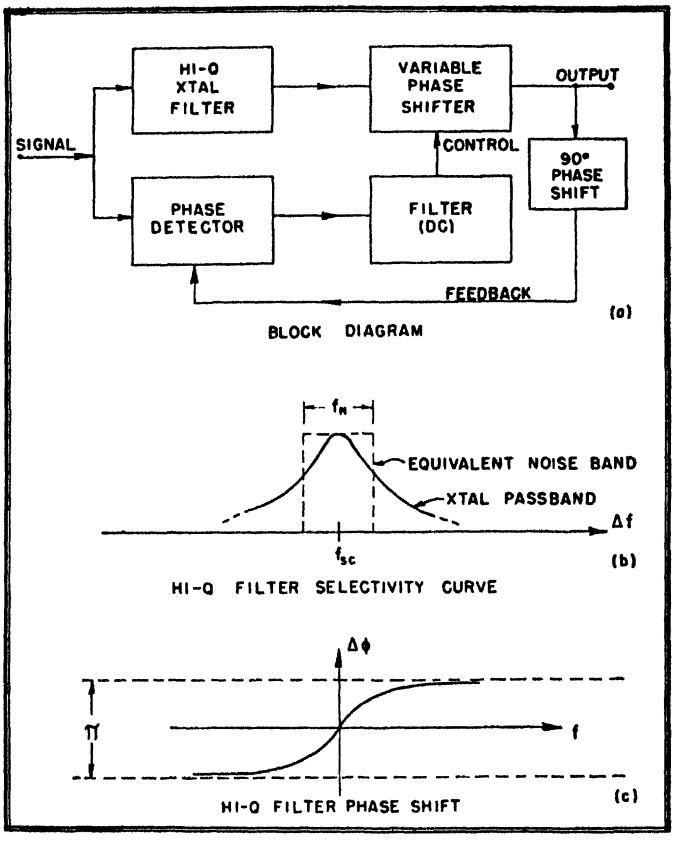


Fig. 6—Passive integrator.

static phase correction. The circuit includes in addition to the high-Q crystal filter a variable phase shifter, a phase detector (which has associated with it a 90° phase shift in one of the signal paths) and a low-pass (dc) filter in the feedback loop for correcting the average phase of the system. Other arrangements are possible; for example, a post-corrector might be used with the feedback signal derived directly from the output of the crystal filter, or a controllable reactance might be coupled to the crystal filter to insure optimum tuning.

The static phase may be maintained as closely accurate as desired by putting a suitably large amount of dc gain in the feedback loop. The signal-to-noise ratio at the output of the system will not be measurably changed if the dc filter is such that the bandwidth of the phase feedback loop is narrower than that of the crystal filter. Design considerations are discussed in Appendix B.

If the crystal stability is comparable to the transmitter frequency stability, the frequency error will be small enough so that rapid phase stabilization will occur when channels are switched. The switching transient is a phase transient and the stabilization time for small detunings will be or the order of a few times the transient time constant of the phase feedback loop. For the crystal bandwidth required, this time is essentially instantaneous. It may be noted however that if appreciable mistuning could occur the gain versus frequency characteristic of the high-Q filter would substantially reduce the amplitude of the correction signal, resulting in considerably increased stabilization time, and effectively reduced loop gain.

Standard Automatic Frequency and Phase Control Locked Integrator

Fig. 7(a) shows the block diagram of a standard automatic frequency and phase control loop. It includes a local reference oscillator, a phase detector which compares the relative phase difference between the sync signal and the oscillator, a filter which partly determines the transfer characteristic of the APC loop as an integrator, and a reactance tube for controlling the oscillator frequency. The loop gain for this system has the dimen-

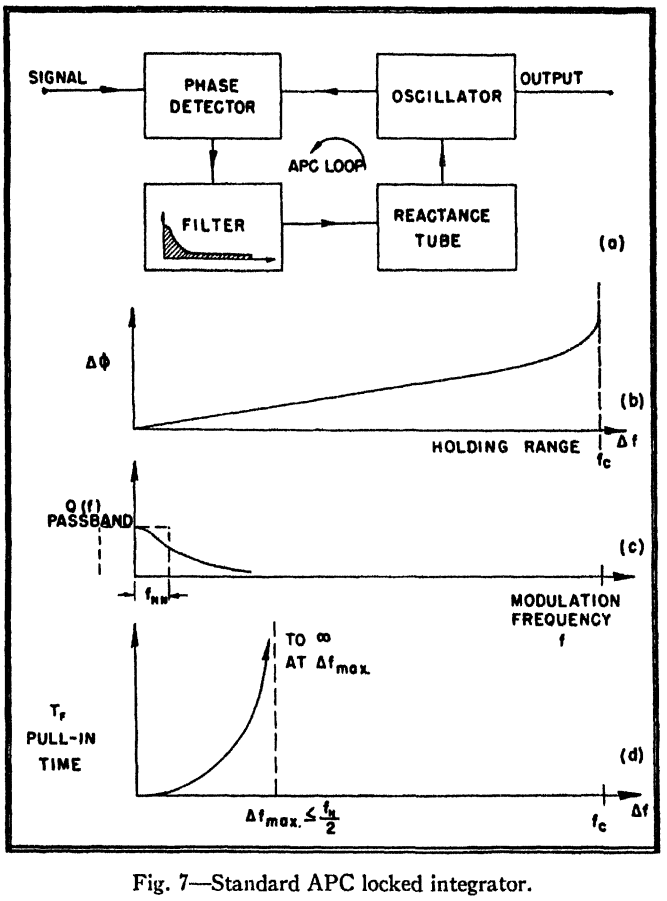


Fig. 7—Standard APC locked integrator.

sions of a frequency, f_c, which is equal to the frequency holding range of the APC system. Included in this characteristic is the dc transmission of the filter. Fig. 7(b) shows the relationship between the static phase error, $\Delta\phi$, and initial oscillator detuning, Δf. By making the holding range much larger than the normal operating range the static phase may be controlled as tightly as desired; here again the price of this control is high loop

[8] W. G. Cady, "Piezoelectricity," McGraw-Hill Book Co., Inc., New York, N. Y.; 1946.
[9] A. W. Warner, "High-frequency crystal units for primary frequency standards," Proc. I.R.E., vol. 40, pp. 1030–1033; Sept., 1952.

gain. Fig. 7(c) shows effective passband characteristic $Q(f)$ of the APC loop as a function of modulation frequency. This is determined largely by the ac transmission of the filter in conjunction with the feedback characteristics of the loop. The noise bandwidth f_{NN} is defined in the normal fashion and indicated on the figure. Since an APC loop phase detector is essentially a synchronous detector and does not distinguish between those noise components which are above or below the local oscillator frequency, then $f_{NN}=f_N/2$, and the effective integration time $T_M=1/2f_{NN}$; the noise bandwidth of the APC loop should not exceed approximately 100 cps for equivalent performance with the high-Q filter.

Fig. 7(d) is a sketch of pull-in time for this loop as a function of Δf. The pull-in range cannot exceed half the gating frequency, i.e. $f_H/2$, and for many designs is substantially smaller. The pull-in mechanism of this loop is not the most efficient one possible. Pull-in times are particularly long near the limit of the pull-in range. The APC loop of Fig. 7(a) is of the same basic type[10] which has achieved essentially universal use in television receivers as an integrator for line frequency synchronizing information. A detailed analysis of the characteristics of this loop is presented in Appendix C and a derivation of the pull-in time relationships is presented in Appendix D.

The pull-in range and time are a function of some design parameters discussed later. It has been found that for optimum design there is a limit to the pull-in performance obtainable with this loop. For these limit designs the following performance is obtained:

(a) The static phase error $\Delta\phi$ may be as small as possible and in fact must be smaller than some specified number in order that pull-in time be minimized.
(b) The pull-in range is equal to $\pm(f_H/2)$.
(c) Except near the limit of pull-in range, the pull-in time and noise bandwidth are very nearly related to the frequency detuning, Δf, by (3)

$$T_F f_{NN} \approx 4\left(\frac{\Delta f}{f_{NN}}\right)^2. \qquad (3)$$

This has been used in Fig. 8 to plot the limit of pull-in performance for optimum design standard APC loops. Fig. 8 represents the pull-in time T_F in seconds as a function of the noise bandwidth f_{NN} in cycles per second. The range of f_{NN} in this log-log plot is from 10 to 1,000 cps with the approximate normal required bandwidth of 100 cps in the center of the graph. Pull-in times ranging from less than one-tenth to approximately one second appear instantaneous and may be characterized as "good." Pull-in times between 1 and 10 seconds are acceptable but probably close to the limit of adequate performance and have been designated "fair." Pull-in times in excess of 10 seconds are definitely "poor."

The relationship between f_{NN} and T_F is shown for several values of Δf. For example an optimum design unit having a noise bandwidth of 100 cycles will require 4 seconds to pull in from 1,000 cycles detuning. This indicates that such a sync system should be adequate for completely automatic phase control but that it apparently does not have an excess of available performance; for example, if the noise bandwidth needed to be reduced to 50 cycles, then 32 seconds would be required to pull in 1 kc.

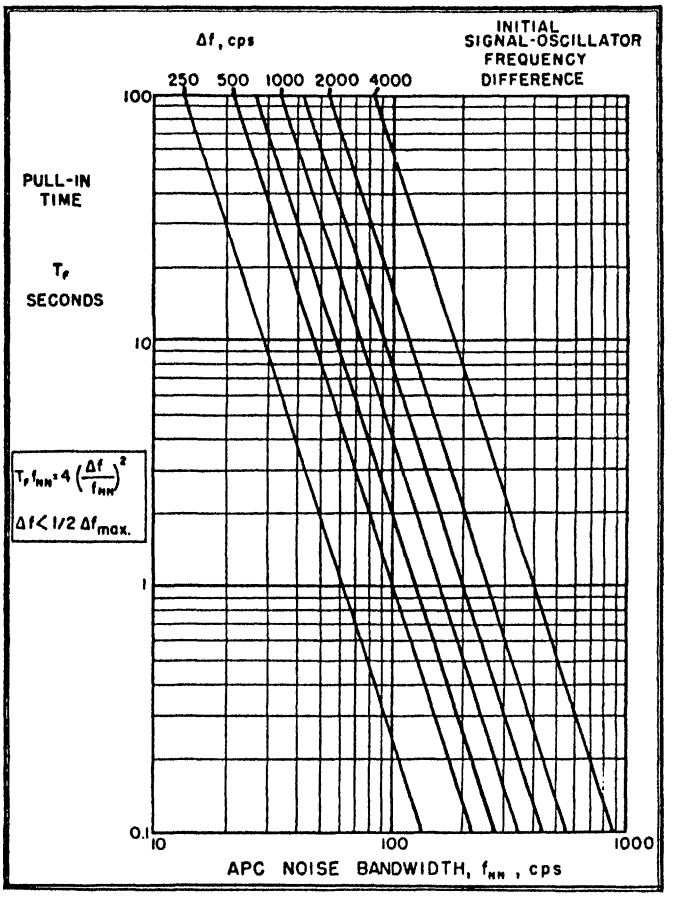


Fig. 8—Standard APC optimum pull-in performance.

Pull-In Performance Attainable with a Standard APC System

Not all designs of APC circuits will achieve the limits of performance discussed with respect to Fig. 8. In fact, partly due to economic limitations, the majority of past designs have fallen short of the limit. Accordingly, Figs. 9 and 10 are presented as a basis for demonstrating the pull-in limitations of the Standard APC System. The curves are expressed in terms of what are believed to be the parameters of interest to the user, specifically the noise bandwidth f_{NN}, the initial frequency difference Δf, and the frequency stabilization time T_F. The dimensionless parameters $T_F f_{NN}$, and $\Delta f/f_{NN}$, are used as ordinate

[10] K. R. Wendt and G. L. Fredendall, "Automatic frequency and phase control of synchronization in television receivers," Proc. I.R.E., vol. 31, pp. 7–15; Jan., 1943.

and abscissa. Two different parameters, designated m and K, which are discussed in Appendix C, appear. The parameter m varies inversely as the dc loop gain for fixed noise bandwidth. The figure shows that increased dc loop gain (smaller m) and hence tighter static phase control permit wider pull-in range and a closer approximation to the minimum pull-in time curve. The parameter K which is a damping coefficient (discussed in

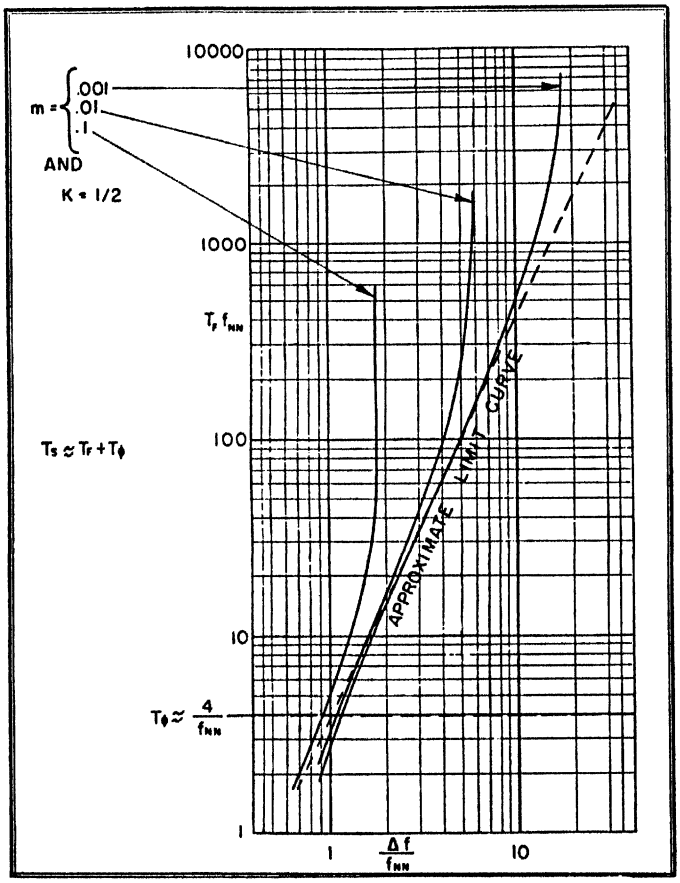


Fig. 9—Pull-in characteristics of standard APC loop.

Appendix C) determines the level of the limit curve as indicated in Fig. 10. Over part of its range of variation the parameter K permits an exchange of minimum pull-in time for pull-in range. The maximum increase, however, is limited to a 50% increase in frequency pull-in range, over designs which approach the optimum pull-in time limit curve.

The mathematics upon which these curves are based is presented in Appendexes C and D. Appendix C introduces and presents the relevant relations between the parameters of the Standard APC System. Derivation of the pull-in time equation and discussion of the pull-in phenomenon is presented in Appendix D.

Theory of Synchronization

Improved Sync Systems

The systems described thus far permit a level of performance which appears to satisfactorily meet the requirements for burst synchronization but do not appear to have a large excess of performance. The signal itself permits substantially better performance.[11] This will be shown below by considering the limitations of the systems presented thus far and introducing the factors which lead to full utilization of the signal information. This leads to a sync system which appears capable of efficiently using all of the timing and synchronizing information in the signal. Then an implementation of this system is described which appears applicable to NTSC color television receivers to produce what may be ideal performance at no substantial cost increase.

Finally, the approximate upper limit of performance capability for the signal is evaluated numerically. The limitations on the previous system relate to the severe restrictions interrelating noise bandwidth and pull-in time. There appear to be a variety of new sync systems which can overcome this limitation. Several varieties have been instrumented and found practical. However, the potentialities of the NTSC burst sync system are perhaps most clearly demonstrated by examining what may be the upper limit of performance.

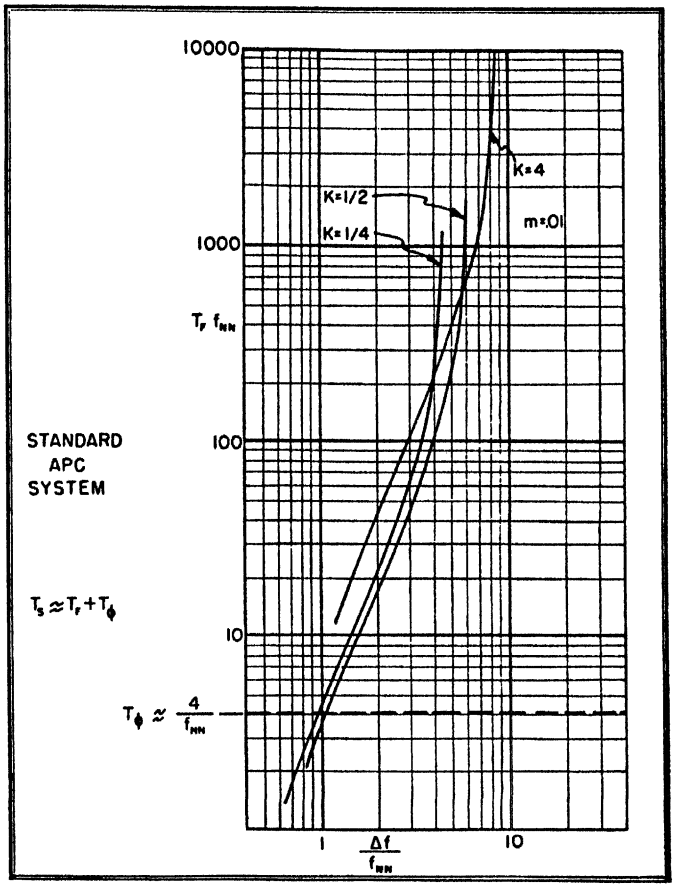


Fig. 10—Effect of variations in the parameter K.

Two Mode Systems

There are two separate and distinct modes of performance of sync systems. These relate to (a) the phase stability attainable after the system has achieved a stable synchronized operating condition, which has been

[11] D. Richman, "Theory of synchronization applied to NTSC color television," IRE Convention Record, Part 4; 1953.

discussed in some detail earlier in this paper, and (b) the performance associated with the system achieving that final state. Each of these modes has fundamental physical restrictions and characteristics associated with it. The full measure of performance permitted by the signal can be achieved by a system which makes these two modes of operation as independent as possible of each other and of each other's limitations.

Some systems use the same mechanism for hold-in and pull-in. The Standard APC System falls into this category. It is inefficient in its use of signal information. Other types of systems use a multiplicity of mechanisms, usually two.[12] One mechanism is designed for stable performance after synchronization, the second mechanism is designed to produce synchronization. *Such a device must have within it the inherent ability to extract from the signal the necessary information with regard to the mode of performance which is required.* For example, it should not confuse noise which may be present when the system is synchronized with a beatnote indicative of a lack of synchronism.

Factors Relating to Frequency Pull-In

There are two basic factors which relate to frequency pull-in. The first problem is concerned with the mechanism whereby a frequency difference is recognized in the presence of strong signals and a control voltage generated which can be utilized for pull-in. The second problem relates to the ability of the mechanism associated with pull-in to discriminate against noise interference.

Frequency Recognition

This separation of the requirements of the system leads to the following principle. *The real limitation of a synchronization system with respect to frequency pull-in is the ability of the system when out of sync to recognize a frequency difference and distinguish it from noise.*

This sets the *real upper limit of performance.* If the frequency determination is effectively linear, then after a time delay which permits the frequency difference to be measured to within a suitable measure of reliability, the reference oscillator may be switched instantaneously by the proper amount to insure synchronization. A system for accomplishing this may be called an *ideal sync system.* Just as with phase measurements this reliability is obtained by integrating the frequency difference information for an adequately long period of time. The shortest stabilization time consistent with reliable performance is therefore determined by the integration time necessary to measure a frequency difference with a suitable measure of reliability.

The Pull-In Control Effect

Fig. 11 represents the generated control effect for pull-in for two important synchronization systems. Fig. 11(a) relates to the frequency pull-in characteristic of a standard APC loop. The generated control voltage for pull-in is shown as a function of instantaneous applied frequency difference Δf. If the frequency is within a range roughly two-thirds that of the noise bandwidth, pull-in (as explained in Appendix D) is effectively instantaneous. The system never slips a cycle; a dc voltage for frequency control is generated which is proportional to the frequency difference. For larger values of Δf the

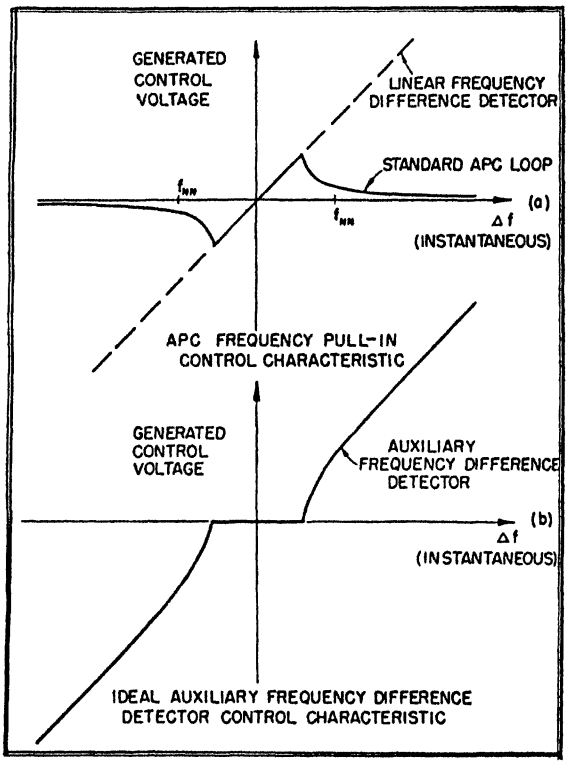


Fig. 11—Synchronization control characteristics.

system slips cycles but by virtue of the feedback in the APC loop generates a dc component of control voltage which varies in the inverse fashion with frequency difference indicated in Fig. 11(a). This inefficient control effect may be compensated for in this system by very high ratios of dc to ac loop gain ($1 \gg m$) but at the expense of the long pull-in times indicated by Fig. 9 and 10. An automatic frequency control system[13] containing a linear frequency difference detector[14] which generates a control voltage proportional to the frequency difference for all frequency differences of interest as indicated in Fig. 11(a) provides a more efficient indication of large frequency differences.

Improved performance may be achieved by supplementing the APC system with an "Ideal Auxiliary Frequency Difference Detector," the control characteristic of which is shown in Fig. 11(b). Such an auxiliary detector can provide a suitable control effect for nearly optimum pull-in performance and as indicated by the flat

[12] Fundamentals relating to systems analyzed here have been applied to automatic gain control circuits as well as to sync systems.

[13] C. Travis, "Automatic frequency control," PROC. I.R.E., vol. 23, pp. 1125–1141; Oct., 1935.

[14] C. F. Shaeffer, "The zero-beat method of frequency discrimination," PROC. I.R.E., vol. 30, pp. 365–367; Aug., 1942.

portion of the curve will *automatically turn itself off when synchronization has been achieved;* this occurs when the frequency difference is reduced to within the linear sloping portion of the curve of Fig. 11(a), within which range the standard APC loop can produce effectively instantaneous pull-in.

A Sync System which Efficiently Uses the Signal Information

Fig. 12 represents the block diagram for a sync system having the auxiliary frequency detection control characteristic described with regard to Fig. 11(b). It includes a Standard APC System such as was shown earlier in Fig. 7 and in addition an auxiliary frequency difference detector which supplements the pull-in performance of the APC system.

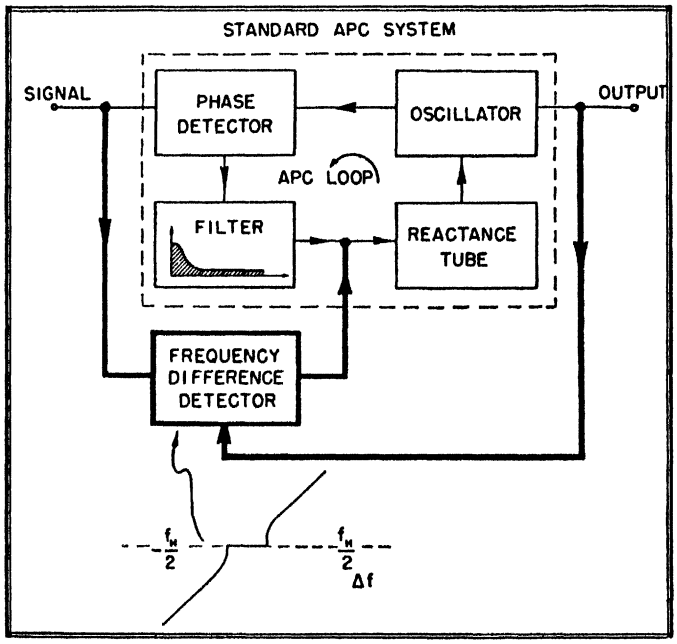


Fig. 12—A synchronization system capable of using total signal information at maximum efficiency.

The idealized upper limit performance described earlier under "Frequency Recognition" may be achieved by means of a suitable interconnection circuit. However, with the stepped characteristic of Fig. 11(b) an essentially direct connection is feasible. The composite system functions as a form of automatic frequency control system when out of sync and as an automatic phase control system when in sync; the auxiliary frequency difference detector turns itself off automatically by virtue of the shape of its control characteristic.

The ideal switched system has pull-in time equal for all frequency differences.

AFC systems normally require high loop gain and are characterized by a pull-in time constant. In some instrumentations of the system of Fig. 12 a loop gain of approximately unity (or a little more for tolerance purposes) may be adequate if the frequency difference detector includes a small amount of delay in its output. As soon as the oscillator is brought near the frequency of the sync signal, the high-gain APC system becomes operative, and the frequency difference detector is automatically inactivated.

The Quadricorrelator: A Frequency Difference Detector

In order to illustrate in more detail the problems and characteristics associated with the achievement of effectively upper limit performance, a form of circuit arrangement is introduced here which appears capable of using elements already present in color television receivers operable on NTSC standards to achieve the ideal frequency difference detection described above. This form of circuit will be called a quadricorrelator in this paper. Analysis of the performance characteristics of the quadricorrelator presented in Appendix E shows that when preceded by a limiter it comes within a few db in signal-to-noise ratio of using all of the signal information for signal-to-noise ratios of interest here. When the limiter is omitted from the system, the quadricorrelator is an efficient frequency detector; the extra noise due to amplitude modulation disappears after pull-in.[15] It is a true frequency difference detector since it is not subject to tuning errors. The excess of available over required noise discrimination suggests that the limiter can be omitted.

There is no real purpose to accomplishing pull-in much more rapidly than perhaps a few tenths of a second. The simple quadricorrelator instrumentations appear (on this basis) to give effectively optimum performance.

A block diagram of a basic form of a quadricorrelator is shown in Fig. 13. Its elements are a pair of synchronous detectors which are fed with reference signals in quadrature with each other so that the phase detector outputs represent "in phase" and "quadrature" components of the applied sync signal. These output signals are then limited in maximum frequency to (for example) $f_H/2$ by filters as indicated in Fig. 13(a). The output of one of the synchronous detectors goes through a differentiating circuit which provides a 90° phase shift through the passband. The two signals are then heterodyned in another synchronous detector, and the output is integrated in a narrow band filter; a low-pass filter is shown. This filter exchanges brevity of integration time for reliability of frequency measurement. The resulting output signal is proportional to the frequency difference (as explained below) and is applied through an interconnection circuit to the controlled oscillator of the APC system of the receiver.

The mechanism by which the frequency difference is determined may be explained as follows: Assume that a frequency difference Δf exists between the sync signal and the local reference oscillator. The input noise may

[15] J. G. Chaffee, "The application of negative feedback to frequency modulation systems," Proc. I.R.E., vol. 27, pp. 317–331; May, 1939.

be considered for simplicity as the sum of two noise-modulated signals in quadrature with each other at the *oscillator* frequency. The output from one synchronous detector will contain a sine beatnote (see Fig. 13(b)) and the noise along one reference axis. The output of the other synchronous detector will contain a cosine beatnote (see Fig. 13(c)) and the noise along an axis in quadrature with the first reference axis. These two *noise* voltages are completely independent of each other.

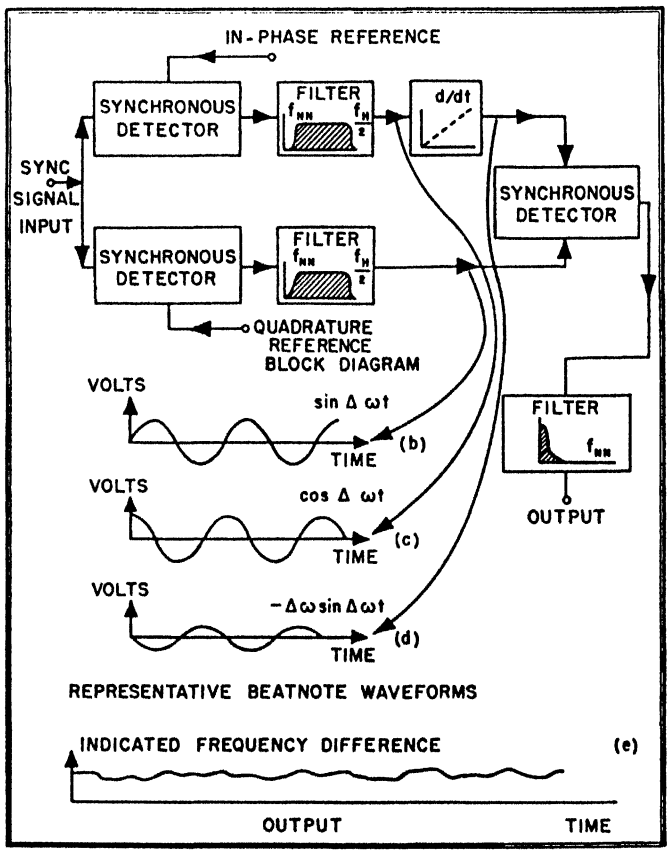


Fig. 13—Basic quadricorrelator.

The cosine beatnote is converted by differentiation to a sine beatnote having an amplitude which is proportional to its frequency, as indicated in Fig. 13(d); its associated noise is differentiated but the two noise voltages are still independent of each other. The output of the cross-multiplying synchronous detector will contain *a dc term proportional to and polarized according to the frequency difference.* In addition the output contains random noise; this noise output is discussed in Appendix E.

The quadricorrelator provides a convenient means for measuring a frequency difference with any selected measure of reliability in the presence of noise by integration of the frequency difference information.

An Illustrative Receiver Operable on NTSC Standards

Fig. 14 shows a partial block diagram of an NTSC color television receiver which includes color difference synchronous detectors, a Standard APC System, and a quadricorrelator for frequency difference detection. The composite video signal is fed to a pair of synchronous detectors for deriving the color difference video signal. The $R-Y$ synchronous detector output may be fed through an amplifier gated during line retrace to a filter, a reactance tube, and an oscillator, the output of which is fed back in the normal fashion to both synchronous detectors. These elements comprise a Standard APC System as described earlier. The gated outputs of both synchronous detectors are fed to a pair of filters as indicated. These may be bandpass filters having low-frequency cutoffs near the noise bandwidth of the APC system and having high frequency cutoffs not higher than half line frequency, as shown in Fig. 14. The differentiating circuit may be included in either beatnote translation path. The third synchronous detector and filter as indicated complete the elements of the quadricorrelator, the output of which is fed to the reactance tube. *The low-frequency attenuation characteristics of the filters in the two channels make the quadricorrelator have an essentially zero transmission characteristic for small beatnote frequency differences.*

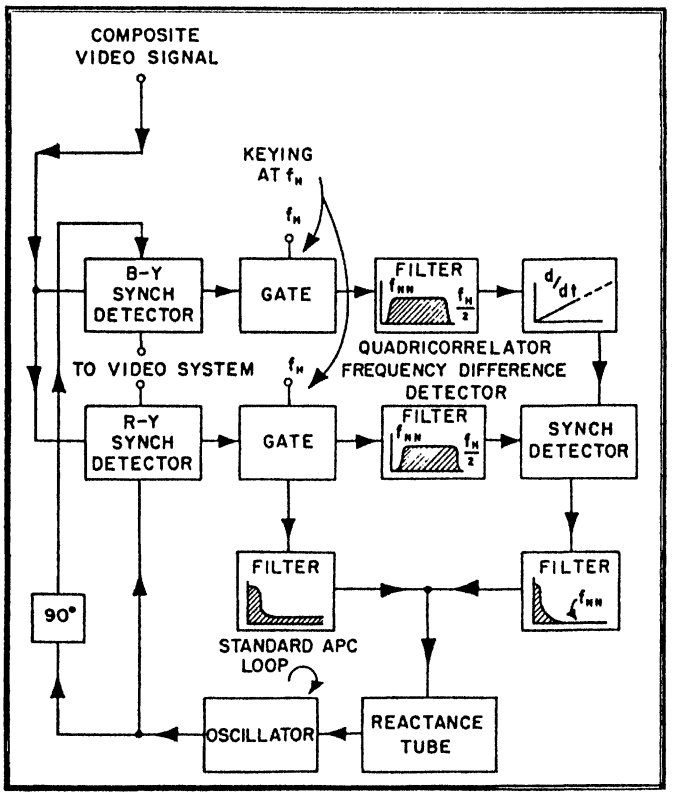


Fig. 14—Application of Fig. 12 sync system to NTSC color television receiver.

The circuit arrangement presented in block form in Fig. 14 provides one means for realizing the sync system of Fig. 12. Since this composite frequency- and phase-control system is essentially free of the previous limitations between noise bandwidth and pull-in time, the over-all system can be more readily designed for desired performance.

There appears to be a variety of frequency difference detector circuits and of linear and nonlinear interconnection

arrangements which can be used to approach upper limit performance in the burst system.

The excess of performance inherent to these arrangements appears exchangeable for receiver economy and long term reliability.

The Approximate Limit of Performance Permitted by Signal Information

There are three requirements on the sync system.

(1) The static phase error shall not exceed some selected value, say 5°. It is shown in Appendexes B and C that for both passive and locked integrators this may be accomplished by use of adequately high loop gain.

(2) The rms phase error shall not exceed some selected value, say 5°, for signal-to-noise ratios at least as high as the approximate lowest level for which monochrome video picture information is acceptable; this is approximately $S_0/N_W = 1$.

The required noise bandwidth for the APC system is

$$f_{NN} = \frac{1}{2T_M} \approx \frac{1}{2}\left(3\phi_{\text{rms}}\frac{S_0}{N_W}\right)^2$$

from (2). (The effect of excess gate width is small and is neglected here for simplicity.)

(3) The stabilization time shall not be annoyingly long. For example, pull-in times shorter than 1 second are acceptable.

The minimum integration time required for frequency difference detection yielding an rms frequency error f_{rms} is shown in Appendix E to be

$$T_I = \sqrt{\frac{2}{d}}\,\frac{1}{\pi h}\,\frac{N_W}{S_0}\sqrt{\frac{f_H}{f_W}}\,\frac{1}{f_{\text{rms}}} \tag{4}$$

for the signal. This is based on a pull-in range of $\pm(f_H/2)$.

It is shown in Appendexes C and D that the linear portion of the curve of Fig. 11(a) extends to a value of Δf approaching $2f_{NN}/\pi$; the control effect is strong to near f_{NN}. Then, if for frequency differences between approximately $(2/\pi)f_{NN}$ and $f_H/2$, the error in frequency difference measurement is less than $(2/\pi)f_{NN}$, pull-in will occur in time T_I. The more severe of the following two requirements then determines the frequency pull-in time T_F.

$$\begin{cases} T_I \geqq \dfrac{\pi}{2f_{NN}} \text{ Approximately} & (5) \\ f_{\text{rms}} \leqq \dfrac{1}{4}\left(\dfrac{2}{\pi}f_{NN}\right) = \dfrac{f_{NN}}{2\pi}. & (6) \end{cases}$$

Combining (4) and (6),

$$T_I \geqq 2\sqrt{\frac{2}{d}}\,\frac{N_W}{S_0}\sqrt{\frac{f_H}{f_W}}\,\frac{1}{f_{NN}}\,\frac{1}{h}. \tag{7}$$

The same adverse tolerances used in obtaining (2) may be used here. If $d = .0352$, $h = .9$, $S_0/N_W = 1$, $f_W = 4.3$ mc, and $f_H = 15734+$cps, then (7) becomes

$$T_I \geqq \frac{1}{f_{NN}}.$$

Thus, the required frequency pull-in time is of the order of magnitude of $1/f_{NN}$ or $(\pi/2)(1/f_{NN})$. After frequency pull-in, phase pull-in occurs. (Both occur effectively simultaneously in the continuous feedback system.) The time for phase pull-in is normally less than

$$T_\phi \approx \frac{4}{f_{NN}}. \tag{8}$$

The constant in (8) depends on the shape of the passband determining f_{NN}.

Then, the stabilization time, T_S is given by

$$T_S \approx T_F + T_\phi. \tag{9}$$

Since the required value of f_{NN} was found earlier to be 100 cps, pull-in times of the order of .05 second are possible. This is considerably shorter than is required, indicating that the information inherently contained in the signal is substantially in excess of what is required.

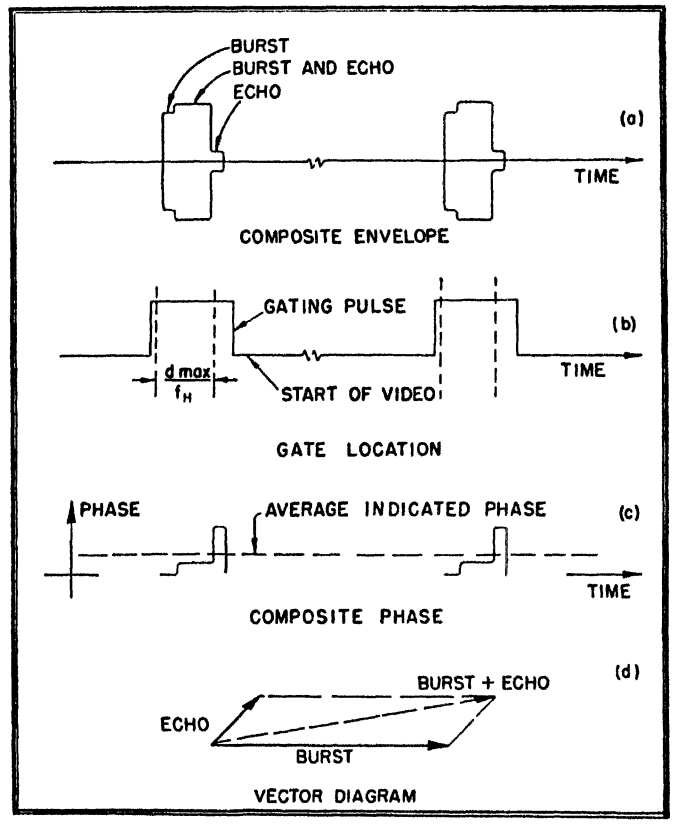


Fig. 15—Effect of echoes on the NTSC color burst.

Other Topics

Effects of Echoes

Some sketches relative to a discussion of the effect of echoes on burst sync are presented in Fig. 15. Fig. 15(a) shows one possible representation of a burst to which an echo has been added. Parameters of interest are the relative delay, the relative amplitude, and the relative phase. If the time-gate exceeds the burst width on the

lagging end as indicated in Fig. 15(b) combined signals may be used to operate the burst sync system. In this case the indicated phase as a function of time is as shown in Fig. 15(c) while Fig. 15(d) is a vector diagram representing the signals of interest. Phase angles of interest are indicated for the burst phase, for the phase of the sum of the burst and echo, and for the phase of the echo. The average phase is not necessarily equal to any one of these but may often be near the phase of burst plus echo. The phase of burst plus echo is the correct reference phase for low detail large area colors. For this reason it appears possible that some extra gate width as indicated in Fig. 15(b) may give a useful and efficient exchange of noise immunity for performance in the presence of echoes. However, the existence of high order correlation between widely separated picture elements[16] may be uncommon enough to make this effect relatively unimportant.

A complete discussion of the effect of echoes in the NTSC system is beyond the scope of the present paper.

Effect of Stability of the Gate

The gate is conveniently obtained from horizontal flyback. The effect of gate stability depends on two factors: the stability of the horizontal sync system which produces the gate; and the relative widths of the gating pulse and the burst, which determines the extent to which noise jitter of the gate can be cross-modulated into the burst channel.

The fundamental physical considerations which have been presented and discussed above with regard to burst synchronization are also true of horizontal synchronization although the shape of the spectral distribution for horizontal sync introduces some additional complications. The static phase may be controlled as closely as desired, limited ultimately by transmission tolerances. The stability may be held to any desired level still permitting effectively instantaneous pull-in.

The effect of cross-modulation when it occurs is to increase the noise power for those low-frequency components to which the horizontal sync system is responsive. The horizontal sync system appears to contain more information than it needs. Stability of the gate is a design consideration but it is not a real limitation of the burst sync system.

Conclusions

The discussion above has shown that standard sync systems appear capable of completely automatic synchronization for NTSC burst sync (although without a large *excess* of performance). In the presence of strong signals the burst sync system is capable of yielding a color-carrier reference having a reliability completely determined by the gain in the receiver sync system, while noise is rejected by integrating the timing information for a suitably long period. An *effective* integration time of the order of 1/200th of a second appears appropriate. Passive integrators using controlled crystal filters, appear capable of meeting the requirements on Q, frequency stability, and rapidity of stabilization. The Standard APC System, when designed for near limit performance, appears capable of providing adequate and usable performance. This means that for reasonable operating tolerances, synchronization will always occur, and with adequate synchronization accuracy.

Improved sync systems which overcome the ultimate limitations of the standard APC sync system have been presented along with a discussion of factors leading to improvement and of the upper limit of performance permitted by the signal. These indicate that the requirement of a high order of noise immunity does not limit synchronization performance in the manner and to the degree experience with previous circuits had indicated. A large excess of attainable as compared to apparently necessary performance appears to exist.

The NTSC color-carrier reference phase synchronization signal contains adequate information for reliable performance down to levels of signal-to-noise ratio where the signals are no longer usable in picture content. A variety of circuits can provide satisfactory performance.

Appendix A

Phase of a Sine Wave Plus Random Noise

Derivation of the Equation

The analysis of the theoretical limits to phasing accuracy may be based on the properties of a signal composed of a sine wave plus random noise.[17] The information of each frequency component may be determined separately and then all of the information may be combined.

The problem is solved here first for a continuous (ungated) sine wave.

The probability density distribution of amplitude coefficients for a sinusoidal signal plus two-dimensional Gaussian noise is shown in Fig. 16. The signal is

$$S(t) = S \cos \omega_{SC} t. \tag{A-1}$$

The noise may be written as

$$N(t) = a(t) \cos \omega_{SC} t + b(t) \sin \omega_{SC} t \tag{A-2}$$

where $a(t)$ and $b(t)$ are time-varying parameters, each having a Gaussian distribution, and defined by the mean square values shown below.

$$\overline{a^2} = \overline{b^2} = \overline{N^2}. \tag{A-3}$$

This equality results from the fact that by symmetry, $\overline{a^2} = \overline{b^2}$ while the total noise power

$$\overline{N^2} = \overline{(a \cos \omega_{SC} t + b \sin \omega_{SC} t)^2}$$

[16] E. R. Kretzmer, "Statistics of television signals," *Bell Sys. Tech. Jour.*, vol. 30, pp. 751–767; July, 1952.

[17] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. Jour.*, vol. 23, p. 282–332, July, 1944; vol. 24, pp. 46–156, Jan., 1945.

$$= \tfrac{1}{2}\overline{a^2} + \tfrac{1}{2}\overline{b^2}. \qquad \text{(A-4)}$$

For the above case it is possible to express the probability distribution of phase angles for the combination of signal and noise, relative to the phase of the signal. This, however, leads to a cumbersome expression.[18]

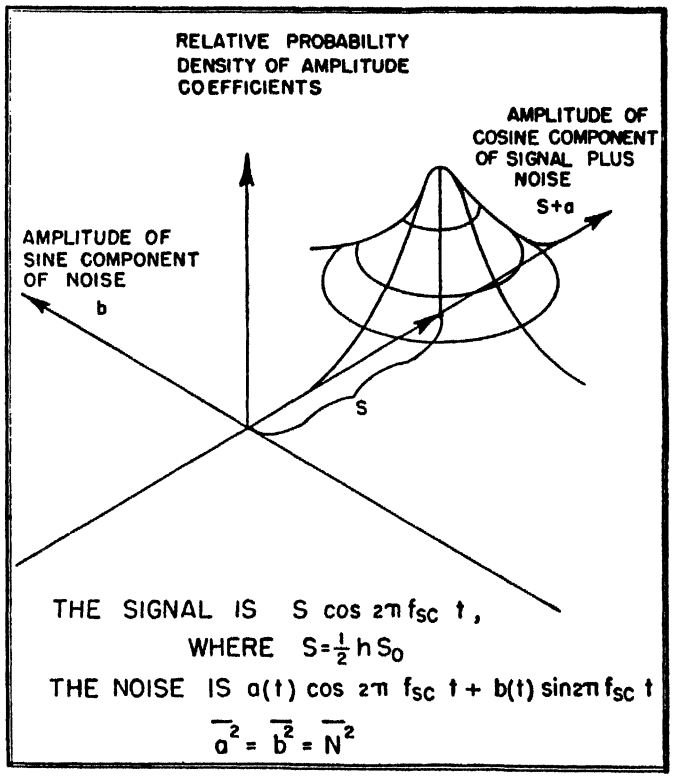


Fig. 16—Probability density distribution of a sine wave and random noise.

It is more convenient to use the simplified vector diagram shown in Fig. 17.[6] Here S represents the signal, and a and b represent the cosine and sine (in-phase and quadrature) components of noise.

Then if ϕ is the phase error,

$$\phi \approx \tan\phi = \frac{b}{S+a} \approx \frac{b}{S} \qquad \text{(A-5)}$$

or, very nearly, since $b_{rms} = N$,

$$\phi_{rms} = \frac{N}{S}. \qquad \text{(A-6)}$$

This equation is a good approximation if N/S is not large; in the case where the sync measuring system is primarily responsive to the noise in quadrature with a reference signal controlled by a long time constant of integration, it is accurate enough.

Then, since

$$N = \frac{N_W}{\sqrt{f_W T_M}} = \text{noise in the noise bandwidth} \qquad \text{(A-7)}$$

$f_N = 1/T_M$, and since $S = \frac{1}{2}hS_0$, we obtain

$$\phi_{rms} = 2\pi f_{sc} t_0' = \frac{\dfrac{N_W}{\sqrt{f_W T_M}}}{\frac{1}{2}hS_0}. \qquad \text{(A-8)}$$

The above equation applies for a continuous sine wave which is not gated. However, because the signal is present only a fraction d of the time, the integration is only $\sqrt{d}$ times as effective, and hence $t_0 = t_0'/\sqrt{d}$. Therefore, by substitution, the following upper limit relationship is obtained.

$$\frac{S_0}{N_W} = \frac{1}{\sqrt{d f_W T_M}} \cdot \frac{1}{f_{sc} t_0} \cdot \frac{1}{\pi h}. \qquad \text{(A-9)}$$

This is (1), presented earlier.

If the signal plus noise is passed through a limiter, the output of the limiter is approximately

$$S \cos \omega_{sc} t + b(t) \sin \omega_{sc} t$$

for signal-to-noise ratios of interest. Thus, the limiter aids in achieving the upper limit, without improving it.

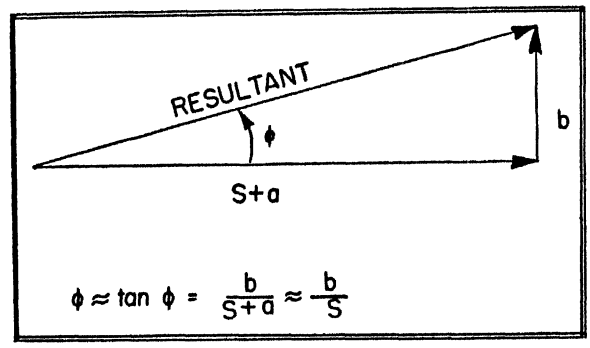


Fig. 17—Simplified vector diagram.

When not all of the signal spectrum is used, the rms error will exceed the limiting value of t_0 computed above.

The burst may be represented by the following Fourier series

$$S(t) = \sum_{k=-k_1}^{k_2} S_k \cos \omega_k t \qquad \text{(A-10)}$$

where

$$S_k = \left(\frac{1}{2} hS_0\right) \cdot d \cdot \left(\frac{\sin dk\pi}{dk\pi}\right) = \frac{hS_0 \sin dk\pi}{2k\pi} \qquad \text{(A-11)}$$

and

$$\omega_k = \omega_{SC} + k\omega_H. \qquad \text{(A-12)}$$

Each of these carries timing information; the error associated with the measurement of any component is very nearly Gaussian. For such a case, the Principle of Least Squares[19] may be applied. Then[20]

$$\frac{1}{t_0^2} = \sum_{k=-k_1}^{k_2} \left(\frac{1}{t_{0k}^2}\right) \equiv \sum_{k=-k_1}^{k_2} \left(\frac{S_k}{N_W}\right)^2 (f_W T_M)(2\pi f_k)^2$$

[18] D. Middleton, "Some general results in the theory of noise through non-linear devices," *Quart. Appl. Math.*, vol. V, p. 471; Jan., 1948.

[19] R. B. Lindsay and H. Margenau, "Foundations of Physics," John Wiley & Sons, Inc., New York, N. Y., chap. IV, pp. 159–187; 1936.

[20] D. Richman, "Frame synchronization for color television," *Electronics*, vol. 25, pp. 146–152; Oct., 1952.

$$= \sum_{k=-k_1}^{k_2} \left(\frac{S_0}{N_W}\right)^2 (f_W T_M)(f_{SC} + kf_H)^2 \left(\frac{h \sin dk\pi}{k}\right)^2 . \quad \text{(A-13)}$$

The factor $(1/t_{0_k}^2)$ has been plotted in Fig. 5(a) as the information per component. The effective accuracy, $1/t_0$ varies as the square root of the area under the curve, for any bandwidth. Although there is an optimum weighting, the weighting is not critical in the vicinity of the correct weighting. This is a general characteristic of integration systems.

Appendix B

Passive Integrators

This appendix presents some equations relevant to the performance of the phase stabilized integrating filter shown in Fig. 6(a).

The basic loop parameters are as follows:

(1) The transfer characteristic of the high Q filter is $F(f)$

$$F(f) \approx \frac{1}{1 + j2\dfrac{f - f_{SC}}{f_{SC}} Q} = \frac{1}{1 + j\dfrac{2\Delta f}{f_3}} = \frac{1}{1 + j\pi \dfrac{\Delta f}{f_N}} = F(\Delta f) . \quad \text{(B-1)}$$

(2) The phase detector sensitivity, for nominal full amplitude input is $\partial E/\partial\phi$

(3) The passband characteristic of the low pass (dc) filter is $Y(f)$, where $Y(0) = 1$. Let

$$Y(f) = \frac{1}{1 + j2\pi fT} . \quad \text{(B-2)}$$

(4) The sensitivity of the phase shifter (assumed broad band) may be represented as

$$\frac{\partial\phi}{\partial E} .$$

(5) The loop gain is G

$$G = \frac{\partial\phi}{\partial E} \cdot \frac{\partial E}{\partial\phi} .$$

(6) The static phase error which would result if there were no feedback is $\Delta\phi_0$

$$\Delta\phi_0 = \arctan\left(-\frac{\pi\Delta f}{f_N}\right) \approx -\frac{\pi\Delta f}{f_N} . \quad \text{(B-3)}$$

The Static Phase Error with Feedback

The static phase error with feedback is $\Delta\phi$

$$|F(\Delta f)|\,\Delta\phi \cdot G \cdot Y_0 = \Delta\phi_0 - \Delta\phi = \Delta\phi_{corr}$$

$$\frac{\Delta\phi}{\Delta\phi_0} = \frac{1}{1 + |F(\Delta f)|\,GY_0} \quad \text{(B-4)}$$

or, since for normal operation $F(\Delta f) \approx 1$ (very nearly) and $Y_0 \equiv Y(0) = 1$

$$\frac{\Delta\phi}{\Delta\phi_0} = \frac{1}{1 + G} . \quad \text{(B-5)}$$

Since $\Delta\phi_0 < 90°$, a loop gain of $G > 17$ makes $\Delta\phi < 5°$ always.

The Effect of the Feedback Loop Upon Noise Performance

When noise is present, the phase detector output produces a noise output, which, after filtering by $Y(f)$ produces extra phase modulation noise.

The equation written earlier can be rewritten in terms of the *phase correction*, $\Delta\phi_{corr}$, since

$$\Delta\phi_{0\ effective} = \Delta\phi_{corr} + \Delta\phi . \quad \text{(B-6)}$$

$\Delta\phi_{0_{effective}}$ is the equivalent phase modulation to produce the actual phase detector noise output.

Then

$$[\Delta\phi_0(p) - \Delta\phi_{corr}(p)] \cdot G \cdot Y(p) = \Delta\phi_{corr}(p)$$

or

$$\frac{\Delta\phi_{corr}(p)}{\Delta\phi_0(p)} = \frac{GY(p)}{1 + GY(p)} = \frac{G}{1 + G}\left[\frac{1}{1 + p\dfrac{T}{1 + G}}\right] . \quad \text{(B-7)}$$

The signals to the phase detector are

(1) The original composite signal+noise, unfiltered.

(2) The filtered signal, with a narrow band of noise having a very small rms value.

Cross beats of signal upon noise produce considerably larger output than the beatnote between noise components, which are therefore negligible.

The output noise may be expressed as a phase:

$$\frac{b(t)}{S} \approx \Delta\phi_{01}(t) \quad \text{or} \quad \frac{b(p)}{S} \approx \Delta\phi_{01}(p)$$

$$\frac{b(p) \cdot F(\Delta p)}{S} \approx \Delta\phi_{02}(p) . \quad \text{(B-8)}$$

The total phase noise is $\Delta\phi_{0_{effective}} = \Delta\phi^0{}_1 - \Delta\phi_{02}$ since, if the filter $F(f)$ were removed, the phase detector output would be identically zero.

$$\Delta\phi_0(p) = \frac{b(p)}{S}\,[1 - F(\Delta p)] . \quad \text{(B-9)}$$

There is little noise energy below approximately $f_N/2$ appearing at the phase detector output.

Since the transfer characteristic for this noise is

$$\frac{\Delta\phi_{corr}}{\Delta\phi_0} = \frac{G}{1 + G}\left[\frac{1}{1 + p\dfrac{T}{1 + G}}\right] \quad \text{(B-10)}$$

(which corresponds to a low pass filter), the following design condition may be employed to insure that the effective Q of the crystal filter will not be degraded by the feedback

$$\frac{1 + G}{2\pi T} < \frac{f_N}{2}$$

or
$$T > \frac{1+G}{\pi f_N} \cdot \tag{B-11}$$

Transient Analysis

The response to a step in differential phase $\Delta\phi_0$, is $\Delta\phi(p)$ or $\Delta\phi(t)$

$$\frac{\Delta\phi(p)}{\Delta\phi_0} = \frac{1}{p}\left[\frac{1+pT}{1+FG+pT}\right] = \frac{1}{pT}\left[\frac{1}{\frac{1+FG}{T}+p}\right] + \frac{1}{\frac{1+FG}{T}+p} \tag{B-12}$$

$$\begin{aligned}\frac{\Delta\phi(t)}{\Delta\phi_0} &= \int_0^t \left[\frac{1}{T}\,\epsilon^{-[1+FG/T]t}\right] dt + \epsilon^{-[1+FG/T]t} \\ &= \frac{1}{1+FG}\left[1-\epsilon^{-[1+FG/T)]t}\right] + \epsilon^{-[1+FG/T]t} \\ &= \frac{1}{1+FG} + \frac{FG}{1+FG}\,\epsilon^{-[1+FG/T]t} \\ &= \text{steady state} + \text{transient response.}\end{aligned} \tag{B-13}$$

For $FG \gg 1$, the transient term is negligible for $t > T$.

The time for the phase error to settle down to twice its final value may be computed, as a measure of stabilization time.

The total transient time consists effectively of an amplitude and phase transient of the high Q filter plus the transient time of the feedback loop. The transient is effectively completed in three times the time constant of the filter. Since the noise bandwidth is f_N, the time is

$$T_A \approx 3\left(\frac{1}{4f_N}\right) \cdot \tag{B-14}$$

This overlaps with the phase loop transient time, which, neglecting amplitude effects, would be

$$T_\phi = \frac{1}{\pi f_N} lnG \approx \frac{1}{\pi f_N} ln\left[\frac{\pi\Delta f_{max}}{f_n\Delta\phi_{max}} - 1\right] \tag{B-15}$$

which is based on

$$\frac{\Delta\phi}{\Delta\phi_0} = \frac{1}{1+G}\left[1+G\epsilon^{-(1+G)(t/T)}\right]$$

$$G = \frac{\Delta\phi_0}{\Delta\phi_\infty} - 1 \approx \frac{\pi\Delta f_{max}}{f_N\Delta\phi_{max}} - 1$$

$$T = \frac{1+G}{\pi f_N} \cdot$$

These two pull-in times overlap.

Appendix C

Performance Characteristics of the Standard APC Loop

This appendix presents a description of the operating characteristics of a standard APC system. The basic parameters of the APC loop are defined. The independence of the primary parameters $\Delta\phi$ (the static phase error) and f_{NN} (the APC loop noise bandwidth) is shown; these parameters characterize the performance after the system has stabilized. The limitations of pull-in are discussed and some formulas which are derived later in Appendix D are introduced. The simple relation presented earlier for pull-in time is then obtained.

The formulas derived may be applied for designs based on any convenient set of assumed criteria.

The Basic APC Loop Parameters

(1) The output voltage ΔE of the phase detector, and the phase difference $\Delta\phi$ between the reference oscillation and the signal are related by the control characteristics. When both signals are sinusoidal,

$$\Delta E = \mu \sin \Delta\phi \tag{C-1}$$

where ΔE is a voltage developed at the phase detector output in response to a phase difference $\Delta\phi$ between signal and reference oscillation. For operation at or very near balance,

$$\frac{\partial E}{\partial \phi} = \mu \qquad \text{volts per radian.}$$

(2) The transfer characteristic of the feedback loop filter is denoted by

$$N(\omega) = \frac{\text{output voltage}}{\text{input voltage}}$$

(3) The sensitivity of the reactance tube is denoted by

$$\beta = \frac{\partial f}{\partial E} \qquad \text{cycles per second per volt.}$$

(4) The factor $|\mu\beta| \equiv f_c$ is a characteristic parameter of the loop; the time constant $t_c \equiv 1/2\pi f_c$ is the transient time constant of the loop when $N(\omega) \equiv 1$. (This may be verified from (C-3) for $Q(\omega)$ presented later.)

(5) The static phase error, $\Delta\phi$, which results from a "free-running" frequency difference, Δf, between signal and local oscillator may be found from the preceding relations:

$$-\sin\Delta\phi = 2\pi\cdot\Delta f\cdot t_c = \frac{\Delta f}{f_c} \cdot \tag{C-2}$$

Although (C-2) contains the appropriate signs, it is the magnitudes of the above quantities which are of interest in design work.

(6) The phase following ratio for an APC loop is

$$\frac{\text{phase variation of output phase}}{\text{phase variation of input phase}} = Q(\omega) = \frac{N(\omega)}{N(\omega)+pt_c} \cdot \tag{C-3}$$

This is the small signal form of the differential equation which characterizes the APC loop. It is used to determine the response of the APC system to noise, after the system is synchronized.

(7) The noise bandwidth of the APC system is f_{NN}. Consistent with the usual practice, this is defined as

$$f_{NN} = \int_0^\infty |Q(\omega)|^2 df = \int_0^\infty Q(\omega)Q(-\omega)df. \quad \text{(C-4)}$$

Representative network configurations for $N(\omega)$ are shown in Fig. 18(a). For each of these networks

$$N(\omega) = \frac{1 + xpT}{1 + (1 + x)pT} \quad \text{(C-5)}$$

where $T=RC$ and $p=j2\pi f=j\omega$. Then

$$Q(\omega) = \frac{1 + xpT}{1 + p(t_c + xT) + p^2(1 + x)t_cT} \cdot \quad \text{(C-6)}$$

This equation suggests one manner in which the meaning of the phase transfer ratio and noise bandwidth of an APC loop may be readily visualized. Fig. 18(b) represents a network having a voltage transfer characteristic which is identical with $Q(\omega)$ given above. If a voltage

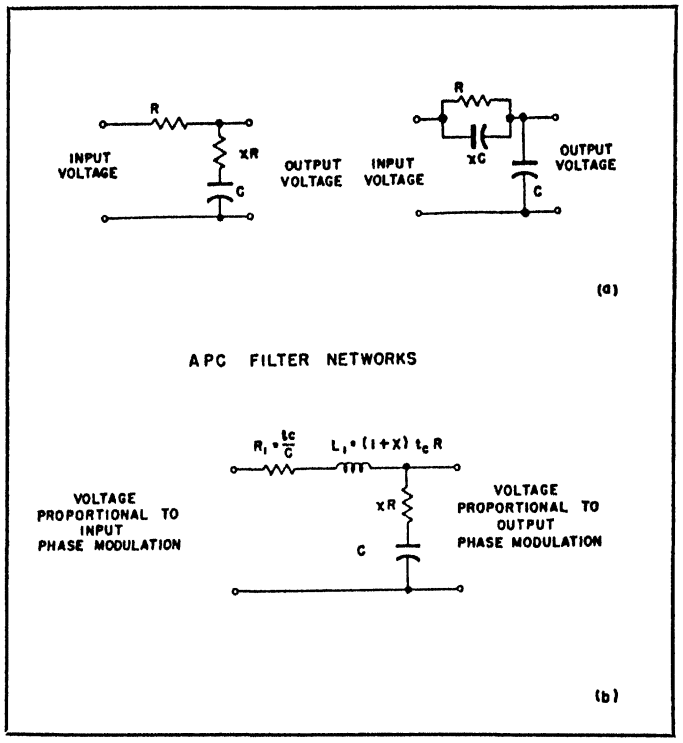


Fig. 18—Equivalent network representing phase following ratio of an APC loop.

proportional to the phase modulation of the synchronizing signal (by noise or any other disturbance) is applied to the input of the network of Fig. 18(b), the output voltage is proportional to the phase modulation of the reference oscillator of the APC loop. The shape of the (low frequency) passband described by $Q(\omega)$ defines the small signal transient response of the loop as well as the noise bandwidth.

(8) The ratio of ac gain/dc gain through the network $N(\omega)$ is

$$m \equiv \frac{x}{1 + x} \quad \text{(C-7)}$$

(from [C-5] when $pT \gg 1$). The parameter m determines the pull-in range of the APC system, when certain other parameters are specified. It is convenient therefore to express the synchronous performance in terms of m.

Also, the term xT/t_c appears often. This is written as

$$y \equiv \frac{xT}{t_c} \cdot \quad \text{(C-8)}$$

Then, rewriting the earlier expressions in terms of these parameters,

$$N(\omega) = \frac{1 + pyt_c}{1 + p\frac{y}{m}t_c} \cdot \quad \text{(C-9)}$$

$$Q(\omega) = \frac{1 + pyt_c}{1 + pt_c(1 + y) + p^2\frac{y}{m}t_c^2} \cdot \quad \text{(C-10)}$$

The noise bandwidth is found by integration (at the end of this Appendix C), using the definition presented earlier, to be

$$f_{NN} = \frac{1}{4t_c} \cdot \frac{1 + my}{1 + y} \cdot \quad \text{(C-11)}$$

(9) In order to prevent resonant ringing on noise impulses, $Q(\omega)$ should have a moderately flat graph. Since the denominator of $Q(\omega)$ contains a quadratic expression, it is convenient to define a damping coefficient, K, which is defined by the following equation:

$$(1 + y)^2 = K \cdot \frac{4y}{m} \cdot \quad \text{(C-12)}$$

Then $K=1$ corresponds to equal roots or critical damping, $K>1$ corresponds to overdamping and makes $Q(\omega)$ approach the shape of the single (RC) low pass filter, and $K<1$ tends to give $Q(\omega)$ a high resonant rise.

Fig. 19 shows the shapes of $|Q(\omega)|$ and $|Q(\omega)|^2$ for several values of K, and subject to the simplifications $y \gg 1$, and $my \approx 4K$, derived below. A value of K close to 1 gives best performance.

The Synchronous Performance of the APC System

The basic equations relating to the synchronous performance of the APC system have been presented above. These are

$$-\sin \Delta\phi = 2\pi \cdot \Delta f \cdot t_c \quad \text{(C-2)}$$

$$f_{NN} = \frac{1}{4t_c} \cdot \frac{1 + my}{1 + y} \quad \text{(C-11)}$$

$$m = \frac{4Ky}{(1 + y)^2} \cdot \quad \text{(C-12)}$$

Since both tight static phase and narrow noise bandwidth are desired, it is possible to define a figure of merit for the system as $|(\sin \Delta\phi)/(\Delta f)| \cdot f_{NN}$; the smaller this product is, the better the over-all performance. However, relations above show that any arbitrarily selected

figure of merit may be obtained by proper design, since, combining the above relations,

$$\left|\frac{\sin \Delta\phi}{\Delta f}\right| \cdot f_{NN} = \frac{\pi}{2}\left[\frac{1+\frac{4Ky^2}{(1+y)^2}}{1+y}\right]. \quad \text{(C-13)}$$

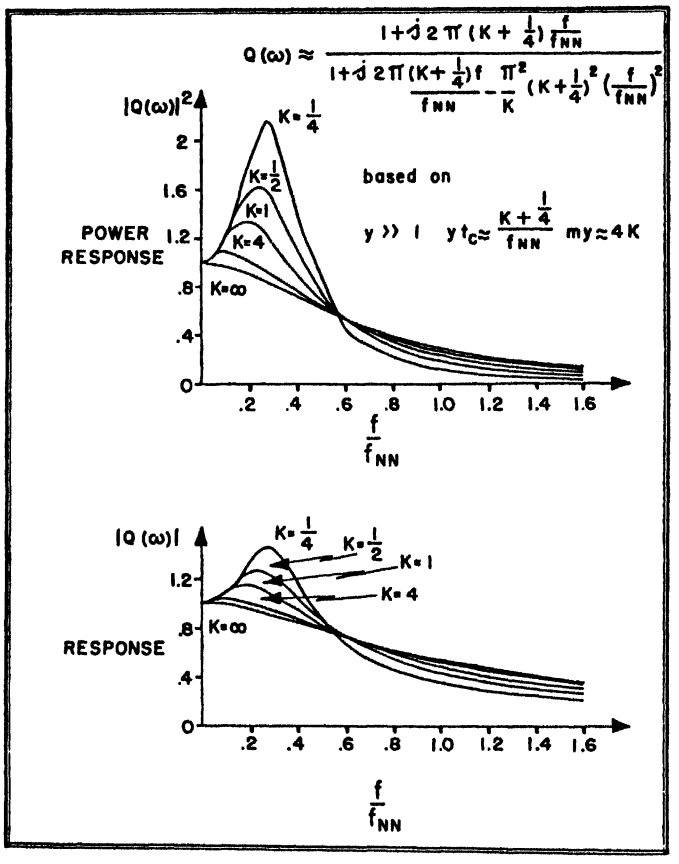


Fig. 19—APC loop small signal modulation response.

For the limiting case of a single time constant filter, $y=0$, and then

$$\left[\left|\frac{\sin \Delta\phi}{\Delta f}\right| \cdot f_{NN}\right]_{y=0} = \frac{\pi}{2}. \quad \text{(C-14)}$$

Thus for the simplified filter the static phase shift and noise bandwidth are interdependent.[21,22] However, for the filters of Fig. 18a, the parameters can be designed for whatever figure of merit is required for synchronous operation.

The above relations may usefully be written in simpler form, since, for the design ranges of interest, $m \ll 1$ and $y \gg 1$; then, very nearly $4K = my$ and hence

$$4 f_{NN} t_c = m\frac{1+my}{m+my} \approx m\left(\frac{1+4K}{4K}\right). \quad \text{(C-15)}$$

This equation will be used in expressing the pull-in performance of the system conveniently.

[21] T. S. George, "Analysis of synchronizing systems for dot-interlaced color television," PROC. I.R.E., vol. 39, pp. 124–131; Feb., 1951.
[22] K. Schlesinger, "Locked oscillator for television synchronization," *Electronics*, vol. 22, pp. 112–118; Jan., 1949.

The figure of merit may be written as

$$\left|\frac{\sin \Delta\phi}{\Delta f}\right| f_{NN} = \frac{f_{NN}}{f_c} \approx \frac{\pi}{2} \cdot m\left(\frac{1+4K}{4K}\right). \quad \text{(C-16)}$$

The Transient (Pull-In) Performance

The pull-in behavior of the APC system is investigated in detail in Appendix D. The significant conclusions are as follows: The pull-in performance is expressible in terms of the relations between the parameters

$$\left(\frac{T_F}{xT}\right) \equiv \left(\frac{T_F}{yt_c}\right)$$

and

$$\left|\frac{\Delta f}{mf_c}\right|.$$

Fig. 20 shows the relation between these parameters.

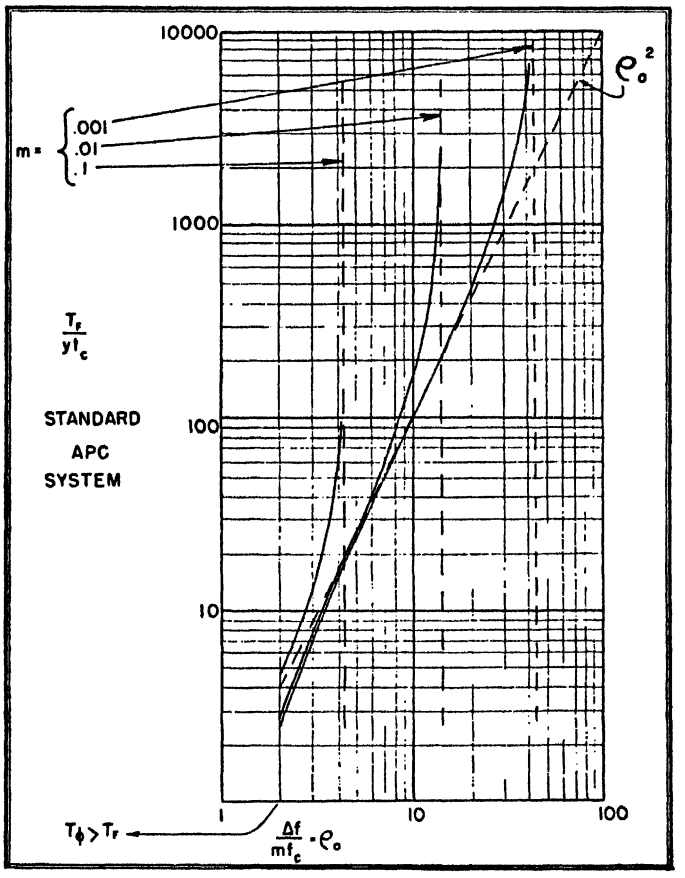


Fig. 20—Universal frequency pull-in characteristics.

The following approximation to the data represented by Fig. 20, based on (D-29), has been found useful in design work, it can also be solved for Δf:

$$T_F \approx xT\frac{\left(\frac{\Delta f}{mf_c}\right)^2}{1-\frac{\Delta f^2}{2f_c \cdot mf_c}}.$$

If $|\Delta f/mf_c| \leqq 1$ the frequency pull-in is effectively instantaneous ($T_F = 0$) but a short period is required for the phase to approach closely its stable value. If $|\Delta f/mf_c| > 1$, the system can slip cycles; often the slip is a great many cycles as this pull-in mechanism is fairly inefficient. The pull-in range is limited to the region

$$\left|\frac{f}{mf_c}\right| < \sqrt{\frac{2}{m} - 1} \qquad \text{(C-17)}$$
$$\text{(D-25)}$$

or

$$\Delta f_{max} = f_c\sqrt{2m - m^2} \approx \sqrt{2f_c \cdot mf_c}$$

Then,

$$|\sin \Delta\phi| \approx |\Delta\phi| \leqq \left|\frac{\Delta f_{max}}{f_c}\right| = m\sqrt{\frac{2}{m} - 1} \approx \sqrt{2m}. \qquad \text{(C-18)}$$

If $m < (1/250)$ the phase angle after pull-in will always be less than 5°. However, not all of the pull-in range is normally used. If

$$|\Delta\phi| < \frac{1}{2}\Delta f_{max}, \; m < \frac{1}{62} \text{ makes } |\Delta\phi| < 5°.$$

When operation is *well within* the pull-in range the *frequency pull-in* time, T_F, which is defined as the time for the oscillator to be pulled from $|\Delta f|$ to within mf_c of the frequency of the color burst, approaches very nearly the relation

$$\frac{T_F}{yt_c} = \left(\frac{\Delta f}{mf_c}\right)^2. \qquad \text{(C-19)}$$
$$\text{(D-28)}$$

By making m smaller and f_c larger it is possible to extend the pull-in range far enough so that the gated nature of the signal provides the only real limitation on pull-in; the range is $|\Delta f| < (f_H/2)$. The pull-in time is then expressed by the square law relation above, except near the limit of the pull-in range. Furthermore, making m smaller improves the synchronous figure of merit.

The pull-in relations may be expressed in terms of f_{NN}, since

$$yt_c = my\frac{t_c}{m} \approx 4K \cdot \frac{1}{4f_{NN}}\left(\frac{1 + 4K}{4K}\right) = \frac{K + 1/4}{f_{NN}} \qquad \text{(C-20)}$$

and

$$mf_c = \frac{m}{2\pi tc} \approx \frac{1}{2\pi} \cdot 4f_{NN}\,\frac{4K}{1 + 4K} = \frac{2}{\pi}\left(\frac{K}{K + 1/4}\right) \cdot f_{NN} \qquad \text{(C-21)}$$

the following equation results

$$T_F f_{NN} = \lambda^2\left(\frac{\Delta f}{f_{NN}}\right)^2 \qquad \text{(C-22)}$$

where, when f_c is large enough so that $\Delta f_{max} \gg \Delta f$,

$$\lambda^2 \equiv \left(\frac{\pi}{2}\right)^2 \frac{(K + 1/4)^3}{K^2} \geqq 4.2. \qquad \text{(C-23)}$$

The approximate value 4 has been used in Figs. 8 and 9. Fig. 21 shows graphically the relation between K and λ^2. The curve has a minimum at $K = 1/2$.

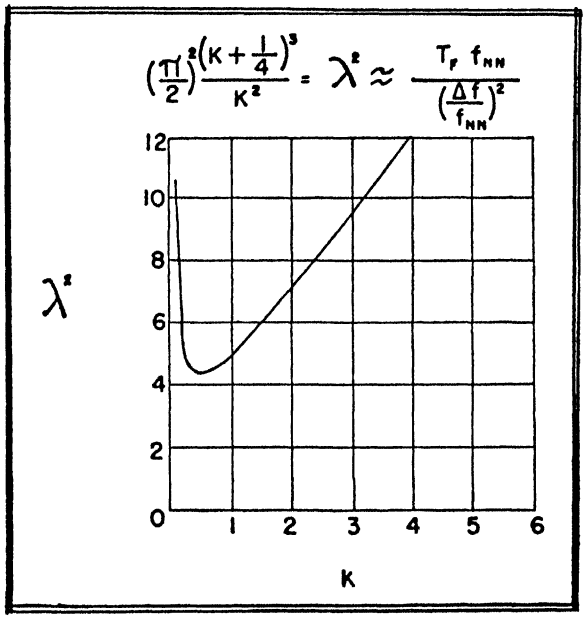


Fig. 21—Graph showing the relation between the damping coefficient, K, and the constant in the APC limit curve equation.

In view of the shape of the curve, and the normal tolerance variations of practical circuits, a value of K near 1 seems desirable. This gives good small signal transient response also. The problem of optimum design is discussed in more detail in a reference.[23]

Derivation of the Noise Bandwidth

The integration is performed as follows. Since

$$Q(p) = \left[\frac{1 + pyt_c}{\frac{m}{y} + pt_c(1 + y)\frac{m}{y} + p^2t_c^2}\right]\left(\frac{m}{y}\right), \qquad \text{(C-10)}$$

then

$$|Q^2| = Q \cdot Q^* = \left(\frac{m}{y}\right)^2 \frac{1 + y^2\theta^2}{(\theta^2 + \theta_\alpha^2)(\theta^2 + \theta_\beta^2)} \qquad \text{(C-24)}$$

where

$$\theta = \omega t_c \qquad \text{(C-25}$$

and

$$\theta_{\alpha,\beta} = \frac{1}{2}\left\{(1+y)\frac{m}{y} \pm \sqrt{\left[(1+y)\frac{m}{y}\right]^2 - 4\frac{m}{y}}\right\} \qquad \text{(C-26)}$$

but

[23] D. Richman, "APC color sync for NTSC color television," IRE Convention Record, part 4; presented March 23, 1953.

$$\frac{1+y^2\theta^2}{(\theta^2+\theta_\alpha^2)(\theta^2+\theta_\beta^2)}=\left[\frac{1-y^2\theta_\alpha^2}{\theta^2+\theta_\alpha^2}-\frac{1-y^2\theta_\beta^2}{\theta^2+\theta_\beta^2}\right]\cdot\frac{1}{\theta_\beta^2-\theta_\alpha^2}. \quad \text{(C-27)}$$

The above is substituted in (C-4) to give

$$\int_0^\infty Q^2(ft_c)d(ft_c) = t_c f_{NN}$$

$$=\frac{\left(\frac{m}{y}\right)^2}{\theta_\beta^2-\theta_\alpha^2}\int_0^\infty\left[\frac{1-y^2\theta_\alpha^2}{\theta_\alpha^2+(2\pi ft_c)^2}-\frac{1-y^2\theta_\beta^2}{\theta_\beta^2+(2\pi ft_c)^2}\right]d(ft_c). \quad \text{(C-28)}$$

Then, since d/dx arctan $x=1/(1+x^2)$ and arctan $0=0$, and arctan $\infty=\pi/2$.

$$t_c f_{NN}=\left(\frac{m}{y}\right)^2\left(\frac{\frac{1-y^2\theta_\alpha^2}{4\theta_\alpha}-\frac{1-y^2\theta_\beta^2}{4\theta_\beta}}{\theta_\beta^2-\theta_\alpha^2}\right). \quad \text{(C-29)}$$

This is simplified as follows. Since

$$\theta_\alpha+\theta_\beta=\frac{m}{y}(1+y) \quad \text{(C-30)}$$

and

$$\theta_\alpha\cdot\theta_\beta=\frac{m}{y},$$

then

$$t_c f_{NN}=\left(\frac{m}{y}\right)^2\cdot\frac{1}{4}\left[\frac{\frac{\theta_\beta-\theta_\alpha}{\theta_\beta\theta_\alpha}+y^2(\theta_\beta-\theta_\alpha)}{(\theta_\beta-\theta_\alpha)(\theta_\beta+\theta_\alpha)}\right] \quad \text{(C-11)}$$

$$=\frac{1}{4}\frac{m^2}{y^2}\left[\frac{\frac{y}{m}+y^2}{\frac{m}{y}(1+y)}\right]=\frac{1}{4}\left(\frac{1+my}{1+y}\right).$$

This is the desired result.

Appendix D

Transient Performance of the APC Loop

This appendix provides a description and derivation of formulas relating to pull-in characteristics and pull-in time of APC loops. Exact analysis of a simplified APC loop provides useful formulas and a basis for understanding some of the phenomena relating to pull-in. This then suggests a simple approximate method for reducing the differential equation of the loop to a form which is readily solvable for the pull-in time. The results are plotted and discussed.

The Simplified Loop

The simplest form of APC network is the one for which $N(\omega)=$ a constant. See Fig. 22(a).

The basic equations are:

$$N(\omega) = m$$

$$Q(\omega) = \frac{m}{m+pt_c}$$

$$f_{NN} = \frac{\pi}{2}mf_c \quad \text{(D-1)}$$

$$|\sin\Delta\phi| = \left|\frac{\Delta f}{mf_c}\right| \leqq 1.$$

The differential equation of the loop is

$$m\cdot\omega_c\sin\phi=\frac{d\phi}{dt}-\Delta\omega. \quad \text{(D-2)}$$

The same equation has been shown applicable for directly synchronized oscillators.[24]

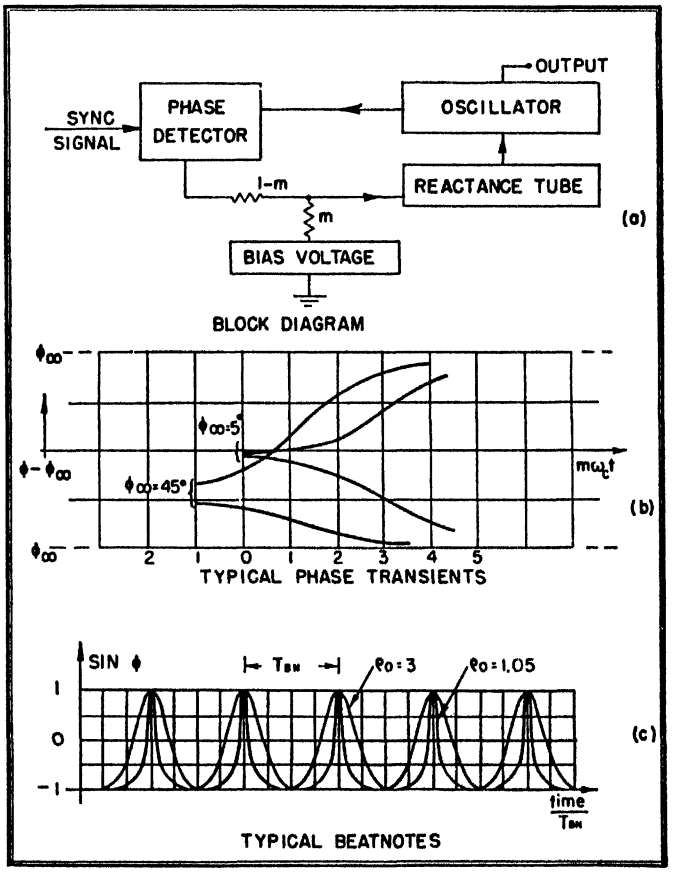


Fig. 22—Basic APC system.

This equation is equivalent to

(Filter transfer characteristic) · (Phase detector output)
= (Rate of change of phase difference)
− (Initial angular frequency difference).

The equation may be rewritten as

$$dt=\frac{d\phi}{\Delta\omega+m\omega_c\sin\phi}. \quad \text{(D-3)}$$

It has two solutions, depending on whether $\Delta\omega/m\omega_c$ is greater than or less than 1. Boundary conditions are

[24] R. Adler, "A study of locking phenomena in oscillators," Proc. I.R.E., vol. 34, pp. 351–357; June, 1946.

$$t = 0 \qquad \frac{d\phi}{dt} = \Delta\omega \qquad \phi = \phi_0$$

$$t = \infty \qquad \frac{d\phi}{dt} = 0 \qquad \phi = \phi_\infty = \arcsin\left(\frac{-\Delta f}{mf_c}\right). \tag{D-4}$$

Equation (D-3) is directly integrable.[25]

The pull-in range is $\Delta f \leqq mf_c$. Within the pull-in range the phase stabilizes according to the following equation, which is the integral of (D-3) under this condition:

$$m\omega_c t \cos\phi_\infty = \ln \left| \frac{\tan\frac{\phi}{2} - \cot\frac{\phi_\infty}{2}}{\tan\frac{\phi}{2} - \tan\frac{\phi_\infty}{2}} \right| \left| \frac{\tan\frac{\phi_0}{2} - \tan\frac{\phi_\infty}{2}}{\tan\frac{\phi_0}{2} - \cot\frac{\phi_\infty}{2}} \right| \tag{D-5}$$

where

$$\frac{\Delta\omega}{m\omega_c} = \rho = -\sin\phi_\infty \qquad (|\rho| < 1) \tag{D-6}$$

and

$$-\sqrt{1-\rho^2} = \cos\phi_\infty. \tag{D-7}$$

Typical phase transients are shown in Fig. 22(b). Phase is plotted relative to ϕ_∞ with a scale calibrated in units of $m\omega_c t$. The starting point on any curve is determined by $\phi_0 - \phi_\infty$.

An approximate time constant of stabilization is

$$\frac{-1}{m\omega_c \cos\phi_\infty} = \frac{1}{\sqrt{(m\omega_c)^2 - (\Delta\omega)^2}},$$

however, *the actual stabilization time is a function of the initial phase.*

Outside the pull-in range $\rho > 1$, and the phase as a function of time is defined by the following equation, which is the integral of (D-3) for this condition:

$$\frac{m\omega_c t\sqrt{\rho^2-1}}{2} = \arctan\left\{\frac{\tan\frac{\phi}{2}+1}{\sqrt{\rho^2-1}}\right\}\Bigg|_{\phi_0}^{\phi}. \tag{D-8}$$

This represents a cyclic variation characterized by its wave form and its fundamental frequency, f_{BN}.

Fig. 22(c) shows examples of the cyclic relationship between $\sin\phi$ and t, $\rho_0 = \Delta f/mf_c$ being specified as 1.05 or 3. The time scale is normalized to the beatnote period $T_{BN} = 1/f_{BN}$. The period T_{BN} is such that t increases by T_{BN} when ϕ increases by 2π, and is found from the following relation:

$$\frac{m\omega_c T_{BN}\sqrt{\rho^2-1}}{2} = \pi \quad \text{when} \quad \Delta\phi = 2\pi. \tag{D-9}$$

Then

$$T_{BN} = \frac{2\pi}{m\omega_c\sqrt{\rho^2-1}} = \frac{1}{\sqrt{(\Delta f)^2 - (mf_c)^2}}. \tag{D-10}$$

This is an important relationship. It states for example, that if in the APC loop block diagram presented above the bias is adjusted so that the effective open loop frequency difference is $\Delta f (> mf_c)$, the operating beatnote frequency difference is $\sqrt{(\Delta f)^2 - (mf_c)^2}$. If the bias is a slowly varying function of time (as compared to f_{BN}), the above relationship accurately describes the variation of f_{BN} with time.

The dc bias or average dc potential developed at the reactance tube input may be determined from the above relationships. It may be expressed in terms of its effect on frequency.

Integrating the differential equation over a cycle, and dividing by the period

$$\frac{1}{T_{BN}}\oint m\omega_c \sin\phi dt = \frac{1}{T_{BN}}\oint \frac{d\phi}{dt}dt - \frac{1}{T_{BN}}\oint \Delta\omega dt \tag{D-11}$$

or

$$m\omega_c \overline{\sin\phi} = \frac{2\pi}{T_{BN}} - \Delta\omega \tag{D-12}$$

and therefore, dividing by 2π,

$$mf_c \overline{\sin\phi} = \sqrt{(\Delta f)^2 - (mf_c)^2} - \Delta f. \tag{D-13}$$

This is plotted in Figs. 11(a) and 23(a) which represents magnitude of the developed bias as a function of Δf.[26] In the standard loop shown later in which the bias battery is replaced by a capacitor it is proportional to the control effect which causes pull-in.

Fig. 23(a) shows that $m\omega_c \overline{\sin\phi}$, the average angular frequency shift, is a maximum when $\Delta\omega/m\omega_c = 1$ and decreases beyond that point, approaching zero asymptotically. When $(\Delta\omega/m\omega_c) < 1$, the phase does not shift 2π radians in a finite time. Enough bias is produced however, to shift the angular frequency by $\Delta\omega$. This bias is represented by the straight line portion, as discussed with regard to Fig. 11(a).

The Standard APC Loop

The standard APC loop is shown in Fig. 23(b). For the network shown,

$$N(p) = \frac{1 + pyt_c}{1 + p\frac{y}{m}t_c} = m\frac{1 + pyt_c}{m + pyt_c} = m + \frac{1-m}{1 + p\frac{y}{m}t_c}$$

[25] H. B. Dwight, "Tables of Integrals and Other Mathematical Data," The Macmillan Co., New York, N. Y., Integral 436.00; 1947.

[26] In experimental work this characteristic may be measured in terms of f_{BN}. From (D-10), above, $f_{BN}^2 + (mf_c)^2 = (\Delta f)^2$.

= wideband direct transfer component

+ long time-constant integration component

= resistive component

+ capacitive component. (D-14)

The differential equation in operational form is

$$N(p)\omega_c \sin\phi = p\phi - \Delta\omega \tag{D-15}$$

which may be written as

$$m\omega_c \sin\phi = p\phi - \Delta\omega - \frac{1-m}{1+p\dfrac{y}{m}t_c}\omega_c \sin\phi. \tag{D-16}$$

The term

$$\Delta\omega + \frac{1-m}{1+p\dfrac{y}{m}t_c}\omega_c \sin\phi \equiv \omega_I \tag{D-17}$$

is the Fourier transform of a time function representing effective instantaneous impressed frequency difference.

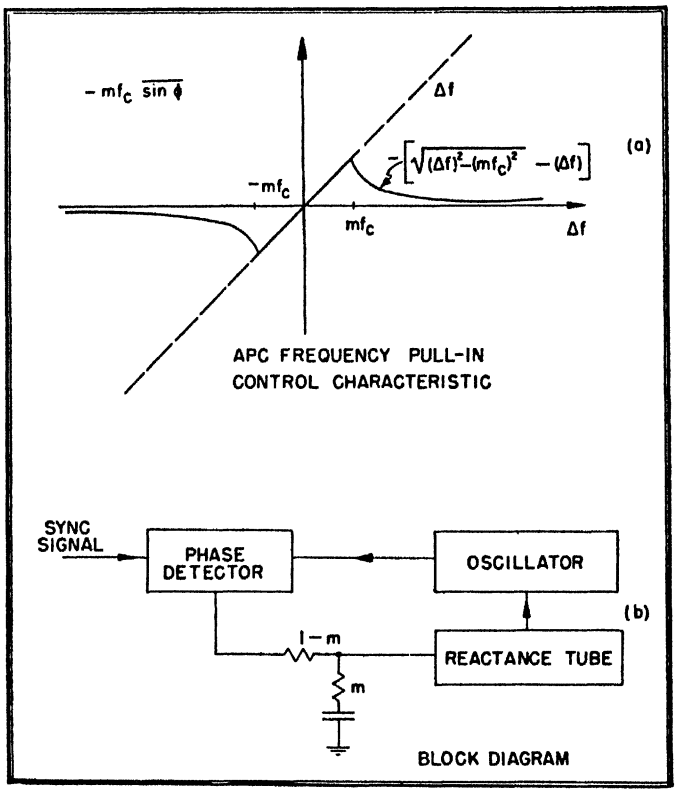


Fig. 23—Standard APC loop.

When this loop is turned on, or has a signal applied to it, the transient of stabilization lasts for a period of time which depends on both the initial phase and the frequency difference. However, the initial phase has only a small effect on the pull-in time and may be neglected for simplicity; the phase transient time, T_ϕ is rapid compared to the frequency pull-in time, T_F. Fig. 22(b) substantiates that for high dc loop gain ($\phi_\infty \ll 90°$) normally $m\omega_c T_\phi < 10$. Then, using (C-15),

$$10 > m\omega_c T_\phi = 4f_{NN}T_\phi \frac{K}{K+\frac{1}{4}}. \tag{D-18}$$

If $K=\frac{1}{2}$, $T_\phi < (15/4f_{NN})$, while if $K=4$, $T_\phi < (2.7/f_{NN})$.

If the frequency difference is such that $\Delta f < mf_c$ the resistive component of loop feedback is adequate to ensure pull-in. The analysis presented above for simplified loop shows the system never slips a complete cycle.

A definition of frequency pull-in time T_F, and phase pull-in time T_ϕ is desirable; the following are useful.

If the system never slips a cycle, then the transient is defined as phase pull-in and measured in terms of the phase pull-in time, T_ϕ.

If the system slips cycles, then the period of time from the instant of switching or excitation until a definable point is reached from which the phase slip does not exceed a cycle is T_F, the frequency pull-in time.

When the initial frequency difference is such that $\Delta f > mf_c$, the long time integration component of feedback must be relied upon for pull-in.

The time constant $(y/m)t_c = y/(2\pi mf_c)$ is long compared to the loop time constant, t_c/m, since $y \gg 1$. Because of this long time constant, the average bias across the capacitor which may result from an unsymmetrical beatnote wave form from the phase detector will not change rapidly with time. It is not unreasonable therefore to integrate the differential equation for this APC loop over a cycle of beatnote.

Then

$$m\omega_c\,\overline{\sin\phi} = \frac{2\pi}{T_{BN}} - \overline{\omega_I} = \sqrt{\overline{(\omega_I)^2} - \overline{(m\omega_c)^2}} - \overline{\omega_I} \tag{D-19}$$

$$\overline{\omega_I} = \Delta\omega + \overline{\frac{1-m}{1+p\dfrac{y}{m}t_c}\omega_c \sin\phi}. \tag{D-20}$$

At this point it is necessary to recognize clearly the nature of the signal circulating in the APC loop. There are two components; there is a *cyclic* component produced as a result of the average frequency difference, and having a harmonic composition which is a function of the frequency difference and hence of time during pull-in; there is a low frequency *drift* component which represents the slow change in frequency difference which constitutes pull-in. It has been shown earlier that the generated frequency shift, $\omega_c\,\overline{\sin\phi}$, varies in an inverse manner with $\Delta\omega$ or $\overline{\omega_I}$; thus, frequency changes slowly except when $\overline{\omega_I}$ is very near $m\omega_c$; stated another way, almost all of the pull-in time is accrued under the condition that the rate of change of the beatnote frequency is not comparable to the beatnote frequency. Therefore, very nearly

$$\oint \frac{1-m}{1+p\dfrac{y}{m}t_c}\omega_c \sin\phi\,dt \approx \frac{1-m}{1+p\dfrac{y}{m}t_c}\oint \omega_c \sin\phi\,dt \tag{D-21}$$

and

$$\frac{1-m}{1+p\frac{y}{m}t_c}\,\omega_c \sin\phi \approx \frac{1-m}{1+p\frac{y}{m}t_c}\,\omega_c \overline{\sin\phi}. \qquad \text{(D-22)}$$

The term $\omega_c \overline{\sin\phi}$ may be eliminated from the above equations, giving a first order differential equation in $\overline{\omega_I}$, the average angular frequency difference.

$$\overline{\omega_I} - \Delta\omega = \frac{1-m}{1+p\left(\frac{y}{m}t_c\right)} \cdot \frac{1}{m}\left[\sqrt{\overline{(\omega_I)^2} - (m\omega_c)^2} - \overline{\omega_I}\right]. \qquad \text{(D-23)}$$

This may be written more conveniently, dividing through by $m\omega_c$, writing $\rho = \omega_I/m\omega_c$ and $\rho_0 = \Delta\omega/m\omega_c$ and by operating on both sides with the differential operator, $1+p(y/m)t_c$.

Then

$$\left(1 + p\,\frac{y}{m}t_c\right)(\rho - \rho_0) = \frac{1-m}{m}\left(\sqrt{\rho^2-1} - \rho\right)$$

or

$$\rho - \rho_0 + \frac{y}{m}t_c\frac{d\rho}{dt} = \frac{1-m}{m}\left(\sqrt{\rho^2-1} - \rho\right).$$

Transposing $\rho-\rho_0$ and separating the variables

$$\frac{dt}{\frac{y}{m}t_c} = \frac{d\rho}{\rho_0 - \rho + \frac{1-m}{m}\left(\sqrt{\rho^2-1} - \rho\right)}. \qquad \text{(D-24)}$$

This equation may be directly integrated (between the limits $\rho=\rho_0$ and $\rho=1$) to yield T_F.

The integration is accomplished with the aid of a change of variable which permits the application of some tabulated integrals. The equations obtained are cumbersome; they are presented at the end of this appendix; they were used for the computations on which the several graphs presented are based. Fig. 20 presents the universal pull-in curves for the standard APC system. The following simplified analysis obtains the significant conclusions, in simpler form.

The limiting pull-in range may be determined as the condition which makes the required pull-in time become infinite. This occurs when the denominator of the above integrand has a real root. It will only occur when

$$\rho_0\left(\equiv\frac{\Delta f}{mf_c}\right) \geqq \sqrt{\frac{2}{m}-1}. \qquad \text{(D-25)} \quad \text{(C-17)}$$

A simple approximate solution for the "limit-curve" may be obtained by eliminating from the equation the factor which produces the above limitation. (Specifically, the small term $(\rho_0-\rho)$ in the denominator is omitted.)

Then, if $m \ll 1$, approximately

$$\frac{dt}{yt_c} \approx \frac{d\rho}{\sqrt{\rho^2-1}-\rho} = -\left(\sqrt{\rho^2-1}+\rho\right)d\rho \qquad \text{(D-26)}$$

and hence, integrating from $\rho=\rho_0$ to $\rho=1$,

$$\frac{T_F}{yt_c} = \frac{\rho_0^2-1}{2} + \frac{\rho_0\sqrt{\rho_0^2-1}}{2} - \frac{1}{2}\ln\left|\frac{\rho_0+\sqrt{\rho_0^2-1}}{1}\right|. \qquad \text{(D-27)}$$

Except for ρ_0 near 1, this is closely equal to

$$\frac{T_F}{yt_c} \approx \rho_0^2 = \left(\frac{\Delta f}{mf_c}\right)^2 \qquad \text{(D-28)} \quad \text{(C-19)}$$

which is the equation presented earlier.

The pole at $\rho_0^2=(2/m)-1$ can be included, writing the simplified equation as

$$\frac{T_F}{yt_c} = \frac{\rho_0^2}{1-\frac{m}{2-m}\rho_0^2} \qquad 1 < \rho_0^2 < \left(\frac{2}{m}-1\right). \qquad \text{(D-29)}$$

The exact integration of (D-24) is accomplished with the aid of the following substitution:

$$z = \sqrt{\rho^2-1} - \rho$$

whence,

$$-\rho = \frac{1+z^2}{2z} \quad \text{and} \quad \frac{d\rho}{dz} = -\frac{1}{2}\left(\frac{z^2-1}{z^2}\right).$$

Then

$$\frac{T_F}{yt_c} = \int_{z_0}^{z_1} \frac{-\left(z-\frac{1}{z}\right)dz}{(2-m)z^2 + 2m\rho_0 z + m}. \qquad \text{(D-30)}$$

The limits are

$$\left|\begin{array}{l}\rho = 1\\ \rho = \rho_0\end{array}\right. \quad \text{and} \quad \left|\begin{array}{l}z_1 = -1\\ z_0 = \sqrt{\rho_0^2-1} - \rho_0.\end{array}\right.$$

Referring to H. B. Dwight, "Tables of Integrals and Other Mathematical Data,"[25] Integrals #160.01, #160.11 and #161.11 are used.

Then

$$\frac{T_F}{yt_c} = -\left\{\frac{1}{2(2-m)}\ln\left|(2-m)z^2 + 2m\rho_0 z + m\right| - \frac{1}{2m}\ln\frac{z^2}{(2-m)z^2+2m\rho_0 z+m} + \left[\frac{2m\rho_0}{2}\left(\frac{1}{m}-\frac{1}{2-m}\right)\frac{2}{\sqrt{4(2-m)m-(2m\rho_0)^2}} \cdot \arctan\frac{2(2-m)z+2m\rho_0}{\sqrt{4(2-m)m-(2m\rho_0)^2}}\right]\right\}\Bigg|_{z_0=\sqrt{\rho_0^2-1}-\rho_0}^{z_1=-1}$$

$$= \frac{-1}{2(2-m)} \ln \left| \frac{2(1-m\rho_0)}{(2-m)z_0^2 + 2m\rho_0 z_0 + m} \right|$$
$$+ \frac{1}{2m} \ln \left(\frac{1}{z_0^2}\right) \left[\frac{(2-m)z_0^2 + 2m\rho_0 z_0 + m}{2(1-m\rho_0)}\right]$$
$$- \left\{ 2\rho_0 \left(\frac{1-m}{2-m}\right) \frac{1}{\sqrt{m(2-m) - (m\rho_0)^2}} \right. \quad \text{(D-31)}$$
$$\cdot \left[\arctan \frac{m\rho_0 - 2 + m}{\sqrt{m(2-m) - (m\rho_0)^2}} \right.$$
$$\left. \left. - \arctan \frac{m\rho_0 + (2-m)z_0}{\sqrt{m(2-m) - (m\rho_0)^2}} \right] \right\}.$$

Appendix E

Reliability of Frequency Difference Detection

This Appendix presents some mathematical derivations relating to the reliability of frequency difference detection.

The relations between rms frequency error and integration time are derived for

(a) the signal

(b) quadricorrelator frequency difference detector preceded by limiter

(c) quadricorrelator frequency difference detector alone.

Basic Signal Characteristics

The combination of signal and noise may be expressed in the following alternate forms (omitting for the moment the time gate factor)

$$S \cos \omega_{SC} + a(t) \cos \omega_{SC} t + b(t) \sin \omega_{SC} t \quad \text{(E-1)}$$

in which the noise is related to the color subcarrier frequency, or

$$S \cos \omega_{SC} t + a_0(t) \cos \omega_0 t + b_0(t) \sin \omega_0 t \quad \text{(E-2)}$$

in which the noise is expressed relative to the local oscillator frequency.

After limiting, the signal can be expressed as

$$S \cos (\omega_{SC} t + \phi(t)). \quad \text{(E-3)}$$

The phase modulation due to noise is $\phi(t)$.

$$\phi(t) = \arctan \frac{b(t)}{S + a(t)} \approx \frac{b(t)}{S} \quad \text{(E-4)}$$

as a first order approximation.

Then

$$\frac{d\phi}{dt} = \frac{S \dfrac{db}{dt} + \dfrac{d}{dt}(ab)}{(S+a)^2 + b^2}. \quad \text{(E-5)}$$

As a second order approximation, the relationship

$$\phi(t) \approx \int \frac{d\phi}{dt} dt \approx \frac{b(t)}{S} + \frac{a(t)b(t)}{S^2} \quad \text{(E-6)}$$

can be used, since $(S+a)^2 + b^2 \approx S^2$ for signals of interest.

The instantaneous frequency of the amplitude limited signal is

$$f(t) = \frac{1}{2\pi}\left[\omega_{SC} + \frac{d\phi(t)}{dt}\right]. \quad \text{(E-7)}$$

The rms frequency error due to noise is f_{rms}.

$$f_{rms} = \frac{1}{2\pi}\left[\frac{d\phi(t)}{dt}\right]_{rms}. \quad \text{(E-8)}$$

The signal amplitudes will also be useful in this analysis.

Then

$S = \frac{1}{2}hS_0$ = amplitude of a burst

$Sd = \frac{1}{2}hS_0 d$ = average amplitude of the component at the burst frequency with gate duty cycle d.

The rms value of $b(t)$ is the square root of the noise power. If *effectively* passed through a filter of bandwidth f_H, and gated with a duty cycle d, the noise power per unit time is $d(N_W^2/f_W)f_H$ and hence, the first order approximation for ϕ_{rms} is

$$\phi_{rms} \approx \frac{b_{rms}}{S} = \frac{N_W}{\frac{1}{2}hS_0}\sqrt{\frac{f_H}{df_W}}. \quad \text{(E-9)}$$

These relations are useful in evaluating the relation between integration time and reliability of the best possible frequency difference detector which might be used for the signal.

To relate reliability to time, the signal information may be averaged over a period T_I, and the rms value of the average then has improved reliability by virtue of integration. As in the case of phase information, it is convenient to use a rectangular time aperture for a standard of comparison for integrators.

Then

$$f_{rms} = \frac{1}{T_I}\left[\int_0^{T_I} [f(t) - f_{SC}]dt\right]_{rms} \quad \text{(E-10)}$$
$$= \frac{1}{T_I}\left[\int_0^{T_I} \frac{1}{2\pi}\frac{d\phi}{dt} dt\right]_{rms}$$
$$= \frac{1}{2\pi T_I}[\phi(T_I) - \phi(0)]_{rms}$$
$$= \frac{\sqrt{2}}{2\pi T_I}\phi_{rms}$$

and therefore, using the first order approximation above,

$$f_{rms} \approx \sqrt{\frac{2}{d}} \cdot \frac{1}{h\pi T_I} \cdot \frac{N_W}{S_0} \cdot \sqrt{\frac{f_H}{f_W}}. \quad \text{(E-11)}$$

The term $(S_0/N_W)\sqrt{f_W}$ is the signal-to-*noise-density* ratio.

The factor $(1/T_I)\sqrt{f_H}$ has the dimensions of (frequency)$^{3/2}$; such terms normally result in frequency

modulation noise analysis due to the triangular spectrum of the noise.[27]

The second order approximation is

$$f_{\text{rms}} \approx \frac{1}{\sqrt{2}\,\pi T_I}\left[\left(\frac{b_{\text{rms}}}{S}\right)^2+\left(\frac{(ab)_{\text{rms}}}{S^2}\right)^2\right]^{1/2} \qquad \text{(E-12)}$$

Note here that rms values add in quadrature.

The second term varies as

$$\left(\frac{N_W}{S_0}\sqrt{\frac{f_H}{f_W}}\right)^2$$

and for signal-to-noise ratios which at present give satisfactory monochrome video signals is small compared to the first term.

Quadricorrelator with Limiter

The quadricorrelator is shown in Figs. 13 and 14. The quadrature reference is R_Q.

$$R_Q = \sin \omega_0 t. \qquad \text{(E-13)}$$

The in-phase reference is R_I.

$$R_I = \cos \omega_0 t. \qquad \text{(E-14)}$$

The cosine beatnote is the beatnote between the input signal and R_I. This is conveniently expressed as

$$\frac{2}{S}\overline{[S\cos(\omega_{SC}t+\phi)]\cos\omega_0 t} = \cos\left[[\omega_0-\omega_{SC}]t-\phi(t)\right]. \qquad \text{(E-15)}$$

The sine beatnote is then

$$\frac{2}{S}\overline{[S\cos(\omega_{SC}t+\phi)]\sin\omega_0 t} = \sin\left[[\omega_0-\omega_{SC}]t-\phi(t)\right]. \qquad \text{(E-16)}$$

The derivative of the cosine beatnote is

$$-\left[\omega_0-\omega_{SC}-\frac{d\phi}{dt}\right]\sin\left[[\omega_0-\omega_{SC}]t-\phi(t)\right]. \qquad \text{(E-17)}$$

Let

$$\omega_0-\omega_{SC}\equiv\Delta\omega\equiv 2\pi\Delta f. \qquad \text{(E-18)}$$

Then, the indicated frequency, which is the integrated output from the product of the signals expressed in (E-16) and (E-17), as multiplied in the output synchronous detector of the quadricorrelator, is, with due regard to signs,

$$f(t)=\frac{1}{\pi T_I}\int_0^{T_I}\left[\Delta\omega-\frac{d\phi}{dt}\right]\sin^2\left[\Delta\omega t-\phi(t)\right]dt. \qquad \text{(E-19)}$$

The polarity of the indicated frequency may be reversed (when so required) by transferring the differentiating circuit from the cosine channel to the sine channel.

Then, since $\sin^2 x=\frac{1}{2}-\frac{1}{2}\cos 2x$, and since

$$\left[\Delta\omega-\frac{d\phi}{dt}\right]\cos 2[\Delta\omega t-\phi]dt = \cos 2[\Delta\omega t-\phi]\cdot d[\Delta\omega t-\phi]$$

we obtain

$$f(t)=\Delta f+\frac{1}{2\pi T_I}\int_{t=0}^{t=T_I}d\phi-\frac{1}{2\pi T_I}\int_{t=0}^{t=T_I}\cos 2\eta\,d\eta \qquad \text{(E-20)}$$

where

$$\eta\equiv\Delta\omega t-\phi(t).$$

Thus the output noise consists of two components: the first represents the frequency noise of the signal; it could be measured as output noise if ω_{SC} were *known*. The second represents extra noise introduced by the measurement of an unknown frequency in this circuit. Then

$$f_{\text{rms}}=\frac{1}{\sqrt{2}\,\pi T_I}\left[\phi_{\text{rms}}{}^2+\text{extra noise}^2\right]^{1/2}. \qquad \text{(E-21)}$$

The extra noise is evaluated as follows:

$$\frac{1}{2\pi T_I}\int_{t=0}^{t=T_I}\cos 2\eta\,d\eta = \frac{1}{2\pi T_I}\left[\tfrac{1}{2}\sin 2\eta(T_I)-\tfrac{1}{2}\sin 2\eta(0)\right]. \qquad \text{(E-22)}$$

Two effects are indicated:

(a) Due to the use of a rectangular time aperture, an extraneous "sampling distortion" term appears unless $\Delta\omega T_I$ = a multiple of 2π, which is therefore assumed for simplicity.

(b) The output noise has the character of random noise which is passed through a nonlinear amplifier having a gain proportional to the sine of the input. This crushes the noise peaks and reduces the rms value.

Then, since $\sin^2 x<x^2$

$$\left[\frac{1}{2\pi T_I}\int_{t=0}^{t=T_I}\cos 2\eta\,d\eta\right]_{\text{rms}}<\frac{\sqrt{2}}{2\pi T_I}\phi_{\text{rms}}. \qquad \text{(E-23)}$$

If there is substantial integration ($f_H T_I \gg 1$), the two noise components approach complete independence and add in quadrature, hence at worst,

$$f_{\text{rms}}\approx\frac{1}{\pi T_I}\phi_{\text{rms}}. \qquad \text{(E-24)}$$

Thus, the quadricorrelator, with a limiter, measures a frequency difference to within a few db of the ultimate reliability permitted by signal information. It has no "detuning" error. The stepped characteristic may be introduced to give

[27] M. G. Crosby, "Frequency modulation noise characteristics," PROC. I.R.E., vol. 25, pp. 472–514; April, 1937.

$$f_{rms} \approx \frac{1}{\pi T_I}\,\phi_{rms}\sqrt{\frac{f_H - 2f_{NN}}{f_H}}. \qquad \text{(E-25)}$$

Since $f_H \gg 2f_{NN}$, the simpler equations above are adequate.

Quadricorrelator without Limiter

The input signal is

$$S\cos\omega_{SC}t + a_0(t)\cos\omega_0 t + b_0(t)\sin\omega_0 t. \qquad \text{(E-2)}$$

The cosine beatnote is proportional to

$$\cos\Delta\omega t + \frac{a_0(t)}{S}. \qquad \text{(E-26)}$$

The sine beatnote is

$$\sin\Delta\omega t + \frac{b_0(t)}{S}. \qquad \text{(E-27)}$$

The derivative of the cosine beatnote is

$$-\Delta\omega\sin\Delta\omega t + \frac{1}{S}\frac{da_0(t)}{dt}. \qquad \text{(E-28)}$$

The quadricorrelator output is

$$\begin{aligned} f(t) &= \frac{1}{\pi T_I}\int_0^{T_I}\left[\Delta\omega\sin\Delta\omega t - \frac{1}{S}\frac{da_0}{dt}\right] \\ &\quad\cdot\left[\sin\Delta\omega t + \frac{b_0}{S}\right]dt \\ &= \Delta f - \frac{1}{\pi T_I}\int_0^{T_I}\frac{1}{S}\frac{da_0}{dt}\sin\Delta\omega t\,dt \\ &\quad + \frac{1}{\pi T_I}\int_0^{T_I}\frac{\Delta\omega}{S}b_0\sin\Delta\omega t\,dt \\ &\quad - \frac{1}{\pi T_I}\int_0^{T_I}\frac{1}{S^2}b_0\frac{da_0}{dt}\,dt. \end{aligned} \qquad \text{(E-29)}$$

The evaluation of the several terms is aided by integration by parts:

$$\begin{aligned} &-\frac{1}{\pi T_I}\int_0^{T_I}\frac{1}{S}\frac{da_0}{dt}\sin\Delta\omega t\,dt \\ &= -\frac{1}{\pi T_I}\sin\Delta\omega t\,\frac{a_0}{S}\Big|_0^{T_I} \\ &\quad + \frac{1}{\pi T_I}\int_0^{T_I}\frac{\Delta\omega}{S}a_0\cos\Delta\omega t\,dt. \end{aligned} \qquad \text{(E-30)}$$

Then

$$\left[-\frac{1}{\pi T_I}\sin\Delta\omega t\cdot\frac{a_0}{S}\Big|_0^{T_I}\right]_{rms} = \frac{1}{\pi T_I}\cdot\frac{a_{0rms}}{S}. \qquad \text{(E-31)}$$

The term

$$\frac{1}{\pi T_I}\int_0^{T_I}\frac{\Delta\omega}{S}\left[b_0(t)\sin\Delta\omega t + a_0(t)\cos\Delta\omega t\right]dt \qquad \text{(E-32)}$$

now appears. This is a two-dimensionally noise modulated sine wave, of the type shown in Fig. 16. The bandwidth of the noise, however, is such that it heterodynes with the carrier to produce a dc component. Then, the integral of this term has the rms value

$$\frac{\Delta\omega}{\pi S}\frac{N_{rms}}{\sqrt{f_H T_I}} = \frac{\Delta\omega}{\pi S}\frac{a_{0rms}}{\sqrt{f_H T_I}} \qquad \text{(E-33)}$$

since there are $f_H T_I$ effective harmonic components. The remaining term is evaluated as follows, integrating by parts:

$$\begin{aligned} &\left[\frac{1}{\pi T_I}\int_0^{T_I}\frac{a_0}{S^2}\frac{db_0}{dt}\,dt\right]_{rms} \\ &= \left[\frac{1}{\pi T_I}\int_0^{T_I}\frac{b_0}{S^2}\frac{da_0}{dt}\,dt\right]_{rms} \\ &= \frac{1}{\sqrt{2}}\left[\frac{1}{\pi T_I}\int_0^{T_I}\frac{1}{S^2}\frac{d}{dt}(a_0 b_0)\,dt\right]_{rms} \\ &= \frac{1}{\pi T_I}\frac{(a_0 b_0)_{rms}}{S^2}. \end{aligned} \qquad \text{(E-34)}$$

Then

$$\begin{aligned} [f(t) - \Delta f]_{rms} &= f_{rms} = \frac{1}{\pi T_I}\left[\left[\left(\frac{a_{0rms}}{S}\right)^2 + \left(\frac{(a_0 b_0)_{rms}}{S^2}\right)^2\right]\right. \\ &\quad \left. + \left[\frac{2\Delta f}{\sqrt{f_H T_I}}\cdot\frac{a_{0rms}}{S}\right]^2\right]^{1/2}. \end{aligned} \qquad \text{(E-35)}$$

These terms add in quadrature as they represent independent random variables. The first bracketed term is of similar form as, but 3 db larger than, the second order signal approximation presented in (E-12), and is nearly equal to

$$\frac{1}{\pi T_I}\frac{a_{0rms}}{S}.$$

The extra noise due to amplitude modulation appears in the last term of (E-35). The ratio of the AM component of noise to the FM component of noise is near

$$\frac{2\pi\Delta f T_I}{\sqrt{f_H T_I}}. \qquad \text{(E-36)}$$

When Δf is small, the quadricorrelator without a limiter approaches the limit of performance permitted by the signal. When Δf approaches $\frac{1}{2}f_H$, a poorer signal-to-noise ratio is obtained. The time T_I must be selected so that f_{rms} does not exceed some selected value, when Δf is the nominally maximum design value for pull-in range.

Equation (E-35) shows that the operation of pulling in results in a large reduction of output noise from the quadricorrelator.

TABLE OF SYMBOLS

Symbol	*Description*
a, $a(t)$	cosine component of noise at the frequency of the sync signal
a_0, $a_0(t)$	same at frequency of oscillator
b, $b(t)$	sine component of noise at the frequency of the sync signal
b_0, $b_0(t)$	same at frequency of oscillator
C	capacitor, Fig. 18
d	the duty cycle of the burst
f_c	the dc loop gain of an APC system, equal to the peak frequency holding range
f_{BN}	beatnote frequency, appearing in Appendix D
f_H, $f_{\text{HORIZONTAL}}$	the line-scanning frequency, 15,750 cps
f_N	effective noise bandwidth of a phase-detection system
$f_{N\text{LIMIT}}$	the value of f_N if all of the phase information is used
f_0	frequency of the local reference oscillator
f_{NN}	the noise bandwidth of an APC loop
f_{rms}	root-mean-square frequency error of a frequency-difference detector
f_{SC}, $f_{\text{SUBCARRIER}}$	the subcarrier (color carrier) frequency
$f(t)$	indicated frequency difference
f_W	video bandwidth occupied by signal and noise
f_3	the 3 db pass-band width of a high Q filter
$F(f)$	transfer characteristic of high Q filter
$F(\Delta f)$	transfer characteristic of high Q filter measured in terms of frequency difference
G	loop gain of phase-control system of Appendix B
h	the ratio of the peak-to-peak amplitude of the burst and the line and field sync pulses
k	index number used in Appendix A
K	a damping coefficient relating to the pass band of an APC loop
m	the resistive divider ratio of a standard APC filter, the ratio of ac gain over dc gain through the network $N(\omega)$
$N(t)$, N	noise signal as a function of time
N_W	the root-mean-square noise in the entire video pass band, assumed flat over the band
$N(\omega)$, $N(p)$	transfer characteristic of the filter of an APC loop
p	$j\omega$, $j2\pi f$, or d/dt as appropriate
$p(\tau)$	the relative probability density distribution function for timing data
$p(\phi)$	the relative probability density distribution function for phasing data
$Q(f)$, $Q(\omega)$	the effective modulation pass band transfer characteristic of an APC system after synchronization
r	ratio of actual gate width to minimum burst width
R	resistor, Fig. 18
R_I	in-phase reference signal
R_Q	quadrature reference signal
S	amplitude of color burst
S_k	amplitude of kth frequency component
S_0	the amplitude of the line and field sync pulses
$S(t)$	the synchronizing signal as a function of time
T	time constant of low-pass filter $Y(f)$
T	time constant RC of Fig. 18
T_A	transient response time of high Q filter in Appendix B
T_{BN}	beatnote period
t_c	characteristic time constant of an APC loop
T_F	frequency pull-in time
T_I	integration time of a frequency-difference detector
T_M	effective integration time
$T_{M\text{LIMIT}}$	the required value of T_M if all of the phase information is used
t_0	root-mean-square time error
t_0	root-mean-square error if the color synchronizing signal were present all of the time
t_{0k}	rms timing error of the k'th frequency component
T_S	stabilization time of a synchronizing system
T_ϕ	phase pull-in time
x	constant relating to Fig. 18(a)
y	parameter defined by (C-8)
$Y(f)$	pass-band characteristic of the low-pass filter in the phase-control system of Fig. 6, Appendix B
z	variable of integration used in Appendix D
β	sensitivity of the reactance tube of an APC loop
ΔE	output voltage of the phase detector of an APC loop
Δf	frequency difference between oscillator and sync signal
Δf_{max}	maximum frequency difference from which pull-in will occur
$\Delta\phi$	the static phase error of an APC loop
$\Delta\phi$	as used in Appendix B—phase error of phase feedback system
$\Delta\phi_{\text{corr}}$	the phase correction produced by the phase feedback system of Appendix B
$\Delta\phi_{\text{corr}}(p)$	frequency spectrum of $\Delta\phi_{\text{corr}}$
$\Delta\phi_0$	static phase error of high Q filter
$\Delta\phi_{0\text{effective}}$	equivalent phase modulation (re. B-6)
$\Delta\phi_0(p)$	frequency spectrum of $\Delta\phi_{0\text{effective}}$
$\Delta\phi_{0I}(p)$	frequency spectrum of beat between filtered signal and direct noise at the phase detector of Fig. 6(a) (re. B-8)
$\Delta\phi_{0Q}(p)$	frequency spectrum of beat between the direct signal and filtered noise at the phase detector of Fig. 6(a) (re. B-8)
$\Delta\omega$	angular frequency difference
η	parameter used in Appendix E
θ, θ_α, θ_β	parameters defined and used in Appendix C
χ^2	see (C-23) and Fig. 21
μ	transfer gain of phase detector of an APC loop
$\rho = \frac{\Delta f}{mf_c}$	normalized frequency difference defined in Appendix C
ρ_0	initial value of ρ
τ	time scale for a probability density
ϕ	phase angle
ϕ_1	a phase variable relating to Fig. 2(c)
ϕ_{rms}	root-mean-square phase error
ϕ_0	initial phase
ϕ_∞	static phase difference due to frequency detuning
ω_c	$2\pi f_c$
ω_H	$2\pi f_H$
ω_I	instantaneous angular frequency difference (re. D-17)
ω_k	kth angular frequency
ω_0	$2\pi f_0$
ω_{SC}	$2\pi f_{SC}$

Charge-Pump Phase-Lock Loops

FLOYD M. GARDNER, FELLOW, IEEE

Abstract—**Phase/frequency detectors deliver output in the form of three-state, digital logic. *Charge pumps* are utilized to convert the timed logic levels into analog quantities for controlling the locked oscillators. This paper analyzes typical charge-pump circuits, identifies salient features, and provides equations and graphs for the design engineer.**

I. INTRODUCTION

PHASE-LOCK loops (PLL's) incorporating sequential-logic, *phase/frequency detectors* (PFD's) have been widely used in recent years [1]-[5], [6, ch. 6]. Reasons for their popularity include extended tracking range, frequency-aided acquisition, and low cost. A *charge pump* usually accompanies the PFD, as illustrated in Fig. 1. The purpose of the charge pump is to convert the logic states of the PFD into analog signals suitable for controlling the voltage-controlled oscillator (VCO).

Good understanding of the PFD itself has been attained but very little has been published on the operation of charge pumps. In consequence, design of PLL's containing charge pumps has often proceeded as an intuitive extension of conventional PLL's. That approach obscures the special benefits and the special problems of a charge-pump PLL.

The intent of this paper is to place the design analysis of a charge-pump PLL on a sound basis so that its special features are recognized and can be either utilized or avoided, as necessary. In Section II we introduce the basic charge-pump model and derive the loop transfer function based on assumptions of small error (linearized loop) and narrow bandwidth as compared to the input frequency (continuous-time approximation).

Section III is devoted to second-order PLL's wherein it is shown that Type-II operation is obtainable even with a passive loop-filter. This behavior is contrary to that obtained in conventional PLL's and is a particular benefit associated with charge pumps.

A continuous-time approximation is not valid if the loop bandwidth approaches the input frequency. In that case, the discrete-time—or sampled—nature of the loop must be recognized. In particular, sampling introduces stability problems that do not exist in continuous time networks; the stability limit for the second-order loop is presented.

Furthermore, the control voltage (v_c in Fig. 1) has large, rectangular excursions (*ripple*) on each cycle of operation. Ripple magnitude is shown to be proportional to loop bandwidth; ripple can easily be so large as to overload the VCO. The existence of ripple places limits on the application of the simple second-order loop.

Paper approved by the Editor for Communication Electronics of the IEEE Communications Society for publication without oral presentation. Manuscript received Janaury 4, 1980; revised May 27, 1980.

The author is at 1755 University Avenue, Palo Alto, CA 94301.

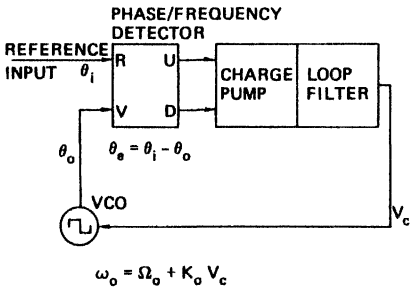


Fig. 1. Phase-lock loop with three-state phase detector and charge pump.

Filters are frequently added after the charge pump to reduce the ripple. Section IV describes the loop performance obtaining from addition of a single capacitor—the simplest-possible ripple filter. The loop is now third order (although still Type II) so analysis is more complicated. Root-locus plots are given for the continuous-time approximation. For wider bandwidths, a discrete-time, linearized analysis yields a z-plane characteristic-function from which pole locations and stability limits may be obtained. Ripple reduction factor is also set forth.

Results of a nonlinear, discrete state-variable analysis of the second-order loop are described in Part V. It turns out that transient settling times of wide-band loops obtained by discrete-time analysis are very similar to the scaled settling times of narrow loops analyzed on the ordinary continuous-time basis. Similar analysis is possible for the third-order loop, but has not been pursued.

II. MODEL

The states of a sequential-logic PFD are initiated by edges of the input waveform. In Fig. 1, if the R-input phase leads the V-input phase, then an edge of the R input sets the U (denoting "up") terminal true. The next V edge resets the U terminal false. As long as R leads V, the D (for "down") terminal remains false. Conversely, if V leads, R, a V edge sets D true and the next R edge resets D false.

Both U and D can be false simultaneously, or either one alone can be true, but both can never be true simultaneously. Therefore, a PFD has three allowable states at its two output terminals. The states will be denoted as U, D, and N, where the last connotes "null" or "neutral."

It is also possible to have *combinatorial* (or *multiplier*; see [6, ch. 6] for terminology) phase detectors with three-state logic outputs as in [7] and [8]. A combinatorial PD does not have the frequency-detector properties of the sequential PFD, but the charge-pump analyses given here apply to either type of circuit in the phase-locked condition. Matters of frequency acquisition are not treated in this paper.

Reprinted from *IEEE Trans. Comm.*, vol. COM-28, pp. 1849-1858, November 1980.

Assume that the PLL is locked and denote the frequency of the input signal as ω_i, radians/second. Let the phase error be $\theta_i - \theta_0 = \theta_e$ radians. The ON time[1] of either U or D, as appropriate, is

$$t_p = |\theta_e|/\omega_i \tag{1}$$

for each period $2\pi/\omega_i$ of the input signal. (The subscript "p" connotes "pump.")

These two features—the three-state description and the ON-time equation—completely characterize the PFD or PD for purposes of this paper.

A charge pump is nothing but a three-position, electronic switch that is controlled by the three states of the PFD. When the switch is set in the U or D position, it delivers a *pump voltage* $\pm V_p$ [Fig. 2(a) and (c)] or a *pump current* $\pm I_p$ [Fig. 2(b) and (d)] to the loop filter. In the N position, the switch is open, thereby isolating the loop filter from the charge pump and the phase detector. This open condition is not encountered in the conventional, analog PLL's and it engenders important, novel characteristics, as will be seen presently.

The loop filter can be either passive, as in Fig. 2(a) and (b), or active, as in Fig. 2(c) and (d). The significant features of the filters to be studied here are contained in the impedance $Z_F(s)$ of Fig. 2, where s is the Laplace-transform complex variable.

Most attention will be given to the current-pump, passive-filter configuration of Fig. 2(b). This choice is made partly because analysis is simplified but also because the configuration is eminently practical under many real-life conditions. It will be shown that performance of the other three configurations is readily obtained, at least approximately, from the analysis of 2(b).

Because of the switching, the charge-pump PLL is a time-varying network; an exact analysis must take account of the time variations of the circuit topology and that is a more-involved procedure than usually found in the common time-invariant networks. In particular, simple transfer-function analysis is not directly applicable to time-varying networks.

In many applications, the state of the PLL changes by only a very small amount on each cycle of the input signal. That is, the loop bandwidth is small compared to the signal frequency. In these cases we may not care about the detailed behavior within a single cycle and may be interested only in the average behavior over many cycles. By applying an averaged analysis, the time-varying operation can be bypassed and the powerful tool of transfer functions retained for our usage. The remainder of this section is devoted to the derivation of average-operation transfer functions. Be aware, though, that the per-cycle behavior can be important even for quite narrow bandwidths, as will be shown later.

Using Fig. 2(b), a pump current I_p sgn θ_e is delivered to the

[1] Equation (1) is exact if the VCO leads the input signal but is an approximation, valid for small bandwidth (compared to input frequency), if the signal leads the VCO. The approximation arises because the VCO phase can change during the ON interval. The effect is usually negligible and is considered further in Section V.

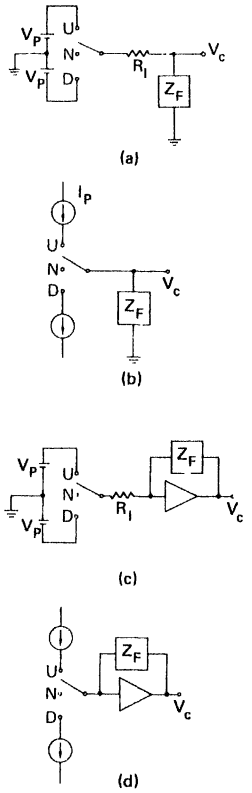


Fig. 2. Charge pumps and loop filters.

filter impedance Z_F for the time t_p on each cycle. Each cycle has a duration $2\pi/\omega_i$ seconds so, utilizing (1), the average *error current* over a cycle is

$$i_d = I_p\theta_e/2\pi \text{ amps.} \tag{2}$$

Equation (2) is also the error current averaged over many cycles, provided that both inputs are periodic—that no input cycles are missing. However, in some applications—notably, in recovery of clock from digital bit streams—edges, or pulses, of the R input will be missing at random. To avoid imposing an erroneously large error current upon the PLL, it is necessary to arrange the logic circuits to recognize the absence of R and to force the circuit into the N state upon those occasions. If the average error current on a single cycle is i_d, as in (2), and if the probability of occurrence of R (the so-called *transition density*) is denoted d, then the average error current over many cycles is $i_d d$. Where applicable, d must be factored into each of the following expressions that treats average behavior.

Oscillator control voltage is given by

$$V_c(s) = I_d(s)Z_F(s) = I_p Z_F(s)\theta_e(s)/2\pi \tag{3}$$

where $I_d(s)$ is the Laplace transform of $i_d(t)$, and similarly for the other symbols. For a locked loop (the only condition for which transfer functions are applicable, because of out-of-lock nonlinearities) the VCO phase is given by

$$\theta_0(s) = K_0 V_c(s)/s \tag{4}$$

where K_0 is the *VCO gain*[2] in radians/second/volt. These expressions, plus $\theta_e(s) = \theta_i(s) - \theta_0(s)$, lead to the loop transfer functions

$$\frac{\theta_0(s)}{\theta_i(s)} = \frac{K_0 I_p Z_F(s)}{2\pi s + K_0 I_p Z_F(s)} = H(s) \tag{5}$$

$$\frac{\theta_e(s)}{\theta_i(s)} = \frac{2\pi s}{2\pi s + K_0 I_p Z_F(s)} = 1 - H(s). \tag{6}$$

These functions apply for any Z_F.

An important property of any PLL is the *static phase error* [6, ch. 4] or *loop stress* that arises from a frequency offset $\Delta\omega$ between the input signal and the free-running frequency of the VCO. Applying the final-value theorem, as in [6], the static phase error is found to be

$$\theta_v = \frac{2\pi\Delta\omega}{K_0 I_p Z_F(0)} \text{ rad.} \tag{7}$$

The foregoing results were all obtained for the configuration of Fig. 2(b): the current switch with passive filter. Much the same expressions arise for each of the other three configurations. For an active filter, it is necessary to take the polarity reversal of the operational amplifier into account. For a voltage switch, the same equations as above occur if we let $I_p \approx V_p/R_1$. For Fig. 2(a)—voltage switch with passive filter—the resulting equations are approximate with the approximation being valid only if $|v_c| \ll V_p$.

III. SECOND-ORDER LOOP

Continuous-Time Approximation

A large preponderance of applications utilize second-order PLL's. To obtain a zero-stabilized, second-order loop, consider a loop filter function

$$Z_{F2}(s) = R_2 + 1/sC \tag{8}$$

which is produced by a series connection of a resistor and a capacitor.

To systematize the notation, define

$$\tau_2 = R_2 C s$$

$$\omega_n = (K_0 I_p / 2\pi C)^{1/2} \text{ rad/s}$$

$$\zeta = \frac{\tau_2}{2}\left[\frac{K_0 I_p}{2\pi C}\right]^{1/2}$$

$$K = \frac{K_0 I_p R_2}{2\pi} \text{ rad/s.} \tag{9}$$

[2] Notation throughout corresponds to that established in [6].

These quantities are interrelated by

$$K = 2\zeta\omega_n$$

$$K\tau_2 = 4\zeta^2$$

$$K/\tau_2 = \omega_n^2 \tag{10}$$

where K is the *loop gain*, ω_n is the *natural frequency* and ζ is the *damping factor*. Any two of the three parameters completely define the linearized, time-averaged behavior of the PLL. Substituting (8) and (9) into (5) and (6) gives the transfer functions for the second-order, charge-pump PLL. They turn out to have exactly the same form as obtained for a conventional second-order PLL [6, ch. 2]. Therefore, to the extent that the various approximations are valid, the charge-pump PLL has exactly the same small-scale behavior as conventional PLL's with the same values for the loop parameters.

To explore further, we note that $Z_F(0) = \infty$ so that the static phase error, from (7), is zero. This desirable performance is achieved with a passive filter. To approach zero static phase error in a conventional PLL requires an active filter with large dc gain. Therefore, the charge pump permits zero static phase error (Type-II response) without the need for dc amplification. This effect arises because of the input open circuit during the N state and does not necessarily depend upon use of an active current switch. The same behavior is found in any of the four configurations of Fig. 2.

Practical circuits will impose some shunt loading across the passive filter impedance. Denote the load as a resistor R_s. The actual static phase error, from (7), will be

$$\theta_v = \frac{2\pi\Delta\omega}{K_0 I_p R_s} \text{ rad.} \tag{11}$$

Shunt loading is most likely to come from input impedance of the VCO control terminal or from the switch itself. Both impedances can be made extremely large. The VCO may be varactor-tuned, which implies near-infinite resistance, and the switch is typically a reverse-biased semiconductor. Some other variety of VCO could utilize a high-impedance buffer, if necessary to isolate a small-input impedance.

When R_s is very large, then leakage current may be more significant in producing phase error. The phase error θ_b resulting from a bias current I_b injected continuously into the filter node can be calculated as

$$\theta_b = 2\pi I_b / I_p \text{ rad.} \tag{12}$$

An active filter, incorporating an ideal op amp, will, of course, obviate any static phase error from VCO control-terminal loading effects.

Although the various results above were obtained specifically for the configuration of Fig. 2(b), they also apply for the other three configurations, as noted at the end of Section II. However, the voltage switch with passive filter [Fig. 2(a)] exhibits a curious nonlinearity that may disqualify it from serious consideration in many applications. Denote the voltage

on the capacitor as v_x. Pump current is $(v_p - v_x)/(R_1 + R_2)$, where $v_p = \pm V_p$ with the sign determined by the phase error direction.

For small v_x, the pump current is influenced little by the capacitor charge so v_x could be neglected in determining the approximate behavior of the circuit. If v_x should become large then it cannot be neglected. (A large v_x would be required if the VCO needed a large v_c to tune it to the proper frequency. Capacitor voltage v_x is "large" if its magnitude reaches a significant fraction of V_p; it can never exceed V_p.)

Let v_x be some positive voltage. When $v_p = +V_p$, then the pump current is $i_{p+} = (V_p - v_x)/(R_1 + R_2)$ while a negative v_p drives a pump current of $i_{p-} = -(V_p + v_x)/(R_1 + R_2)$. These currents are unequal–substantially so if v_x is large enough–so loop gain about this v_x operating point will be larger for negative phase errors than for positive. It is unlikely that any significant asymmetry can be tolerated in most applications.

Granularity Problems

All of the foregoing is based on averaged-response, time-continuous, constant-element operation of the loop. There are features arising from the actual discontinuous operation that need attention, even for narrow bandwidths. The primary features are loop stability and phase-detector ripple.

In some sense, the loop operates on a sampled basis and not as a straightforward continuous-time circuit. A sampled system almost always has more stability problems than arise in continuous-time systems. In particular, an analog, second-order PLL is unconditionally stable for any value of loop gain, but the sampled equivalent will go unstable if the gain is made too large. Prudent design requires that the stability limit be known.[3]

A linearized, sampled analysis is presented in Appendix A. The end result is the characteristic equation (denominator of the transfer function) of the sampled PLL in the z-plane, which has the form

$$D(z) = (z-1)^2 + (z-1)\frac{2\pi K'}{\omega_i \tau_2}\left[1 + \frac{2\pi}{\omega_i \tau_2}\right] + \frac{4\pi^2 K'}{\omega_i^2 \tau_2^2} \tag{13}$$

where $K' = K\tau_2$ may be regarded as a *normalized loop gain*, ω_i is the input frequency, and $\tau_2 = R_2 C$ is the time constant of the filter zero.

Transient response for small phase errors and loop stability are studied by examining the locations of the zeros of $D(z)$–the poles of the z-domain transfer function. The root locus shows pole locations in the z plane for varying K'; an example is sketched in Fig. 3. The shape of the locus is very similar to that of a conventional second-order loop in the s-plane [6, ch. 2].

The two poles start at $z = 1$ for $K' = 0$ and move on a

[3] Tal [9] has investigated sampled-stability of a phase-locked speed-control servo that uses a PFD and a simple lag filter. His problem differs somewhat from that considered here and his method provides an alternate approach.

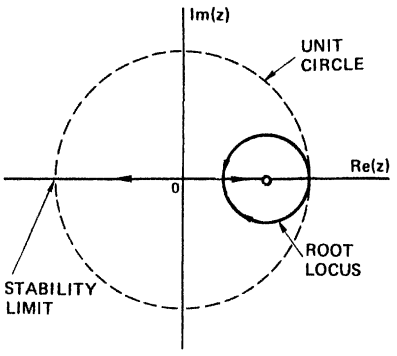


Fig. 3. Root locus plot of second-order loop in z-plane.

circle with center at $z = (1 + 2\pi/\omega_i\tau_2)^{-1}$ for values of

$$K' < \frac{4}{(1 + 2\pi/\omega_i\tau_2)^2}.$$

For larger K', the poles lie on the real axis; one pole migrates towards the center of the locus circle and the other migrates towards $-\infty$.

The loop is stable only if the poles lie inside the unit circle. Instability results where the outbound pole crosses the unit circle at $z = -1$, as noted in Fig. 2. Normalized gain at the crossing point is

$$K' = \frac{1}{\frac{\pi}{\omega_i\tau_2}\left(1 + \frac{\pi}{\omega_i\tau_2}\right)}. \tag{14}$$

This value of K' is the *stability limit* and is plotted in Fig. 4.

Ripple is another granularity effect that demands attention. Upon each cycle of the PFD, the pump current I_p is driven into the filter impedance Z_F, which responds with an instantaneous voltage jump of $\Delta v_c = I_p R_2$. At the end of the charging interval ($t = t_p$), the pump current switches off and a voltage jump of equal magnitude occurs in the opposite direction.

Frequency of the VCO follows the voltage steps so there will be frequency excursions of $\Delta\omega_0 = K_0 I_p R_2 = 2\pi K$ radians/second for each pump pulse. The phase excursion during the pump interval t_p will be $\Delta\theta_0 = 2\pi K |\theta_e|/\omega_i$ [using (1)], so the phase jitter vanishes for $\theta_e = 0$. (A not-unexpected happening since the pump pulses are supposed to vanish for $\theta_e = 0$.)

Some applications (e.g., bit synchronizers) may be able to tolerate such frequency jitter, but others (e.g., frequency synthesizers) may require much better spectral purity.

A possibly more serious consequence of the jumps is the potential for overload of the VCO, even if the indicated ripple is allowable from a spectral-purity standpoint. Any real VCO has only a finite frequency range over which it can be tuned. If control voltages outside of this range are applied, the VCO frequency is unable to follow. (In fact, oscillations may cease or the circuit might even sustain damage.) We require that the frequency jumps remain within the allowable tuning range of the VCO under all conditions.

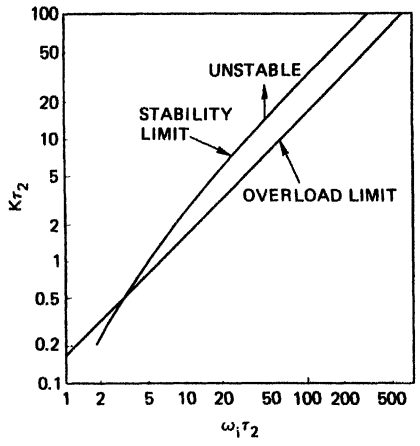


Fig. 4. Stability and overload limits for second-order loop.

As an extreme instance, the frequency jump must not exceed the input frequency. A larger jump would imply that the VCO frequency was driven negative—a meaningless status. For this extreme condition, the overload bound is $2\pi K < \omega_i$ or, in normalized form

$$K' < \omega_i\tau_2/2\pi. \tag{15}$$

A multivibrator, operated near the center of its tuning range, might be able to approach the excursions implied in (15). Most other oscillators will have a much smaller tuning range and therefore will be restricted to use in PLL's with much smaller values of K.

As a comparison to the stability limit, the overload limit of (15) has also been plotted in Fig. 4. It is apparent that overload is the actual restriction on loop gain; overload sets in at a lower value of gain than does instability for any practical circuit.

In all discussion of granularity effects it has been assumed tacitly that all transitions are present. If transitions can be missing at random, as in bit-clock recovery applications, then there may be a data-pattern-dependent jitter induced into the VCO phase. That problem is not treated in this paper.

IV. THIRD-ORDER LOOP

Origination

The frequency jumps inherent to the second-order loop usually cannot be accepted and additional filtering is often included within the PLL in order to mitigate the ripple. The simplest ripple filter is an additional capacitor C_3 in parallel with the earlier RC impedance, as shown in Fig. 5. Defining $b = 1 + C/C_3$, we obtain

$$Z_{F3}(s) = \left(\frac{b-1}{b}\right)\frac{s\tau_2 + 1}{sC\left(\frac{s\tau_2}{b} + 1\right)}. \tag{16}$$

Retaining the previous definition (9) for K, the closed-loop transfer function from (5) is found to be

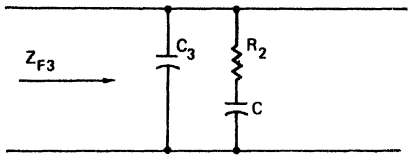


Fig. 5. Filter for third-order loop.

$$H(s) = \frac{K\left(\frac{b-1}{b}\right)\left(s + \frac{1}{\tau_2}\right)}{\frac{s^3\tau_2}{b} + s^2 + K\left(\frac{b-1}{b}\right)s + \frac{K(b-1)}{b\tau_2}} \tag{17}$$

where the continuous-operation, time-averaging assumption has been made.

Simple addition of C_3 across $R_2 + 1/sC$ ought to serve very well for the passive filter but is not likely to be satisfactory for the active filter. The operational amplifiers would be required to deliver step currents of I_p on each cycle, which is likely to be beyond the slew capabilities of most amplifiers. Rather than attempting to accommodate the current by brute force amplifiers, it is more conservative to prevent the current step from ever reaching the op amp, as in the circuits of Fig. 6. The general characteristics of these circuits ought to be much the same as those for the simple passive circuit, but some fine details will differ. This paper treats only the passive-filter, current-switch circuit.

Properties

The transfer function (17) has a denominator of third degree, so the system is a third-order PLL. In the open-loop transfer function, the additional pole is located at $s = -b/\tau_2$, which is far away from the dominant, low-frequency poles for large b. If C_3 is small compared to C ($b \gg 1$), then we should expect only high-frequency effects from the additional filtering. Low-frequency properties should be essentially the same as for the second-order loop.

In particular, the steady-state responses will be the same as for the second-order loop. The static phase error caused by a frequency offset will be zero and the phase lag caused by a frequency ramp will be $\theta_a = \dot{\omega}/\omega_n^2$, [6, ch. 4] where $\dot{\omega}$ is the slope of the ramp. Although the loop is third order, it is only Type II.

The s-plane root locus of (17) has been studied in [6, ch. 8] for another application. Root loci are shown in Fig. 7 for various selections of b. For large b and small-enough K' (the normalized loop gain) the dominant poles are virtually unchanged from the locations expected for the second-order loop. As K' becomes very large, the outward-bound real pole meets the extra pole coming in from $-b/\tau_2$ and the pair go complex asymptotic to a vertical line at $s = -0.5(b-1)\tau_2$. The loop could become seriously underdamped for large gain.

As b is reduced, the breakaway point for the vertical asymptote approaches closer to the low frequency portion of the locus; if $b < 9$, the locus never returns to the real axis and is underdamped for all values of K'.

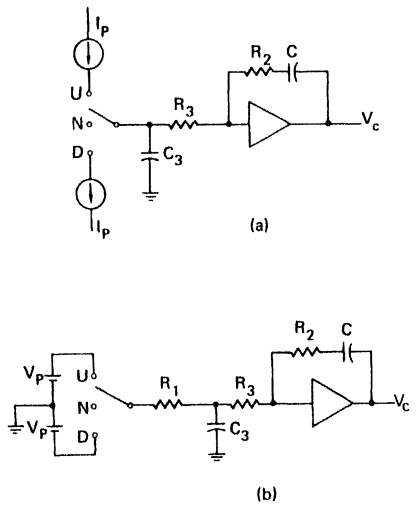


Fig. 6. Jump suppression for active filters.

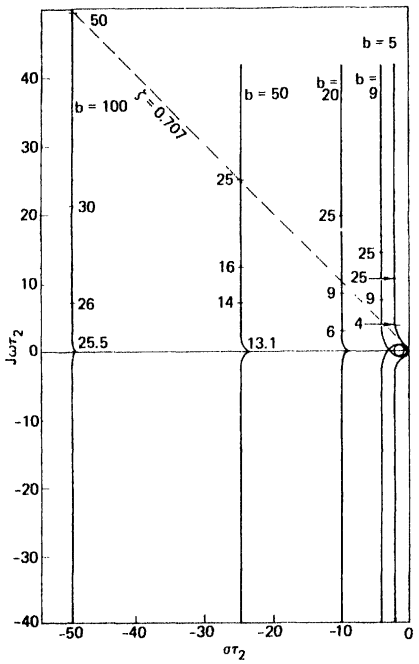


Fig. 7. Root locus plots for third-order, type II PLL. (a) Large scale. (b) Expanded scale. Tick marks show values for $K' = K\tau_2$. (Taken from [6]. Reproduced by permission of publisher.)

If $b < 1$, then the loop is unstable for all K'. In the configuration of Fig. 5 it is impossible to have $b < 1$, but it is entirely possible in Fig. 6. (In terms of the component values, b is defined differently in Fig. 6 than in Fig. 5, but the transfer function and root locus plots have the same form for both configurations.)

This stability impairment caused by the third pole is calculated on the basis of assumed continuous-time operation. Time-discrete operation can be expected to cause even more impairment. To investigate stability of the third-order loop, an analysis similar to that given in the Appendix was performed. (The analysis is omitted here because of space constraints.)

The criterion for stability—all poles inside the unit circle—is satisfied if

$$K\tau_2 < \frac{4(1+a)}{\frac{2\pi(b-1)}{b\omega_i\tau_2}\left[\frac{2\pi(1+a)}{\omega_i\tau_2} + \frac{2(1-a)(b-1)}{b}\right]} \tag{18}$$

where

$$a = \exp\left(-\frac{2\pi b}{\omega_i\tau_2}\right).$$

Knowledge of the stability limit alone is not sufficient for good design; some insight into the transient response is also needed. To that end, the z-plane characteristic equation is

$$D(z) = z^3 + z^2\left[-a - 2 + G\left(\frac{2\pi}{\omega_i\tau_2} + \frac{(1-a)(b-1)}{b}\right)\right] + z\left[2a + 1 - G\left(\frac{2\pi a}{\omega_i\tau_2} + \frac{(1-a)(b-1)}{b}\right)\right] - a \tag{19}$$

where

$$G \triangleq \frac{2\pi K\tau_2(b-1)}{b\omega_i\tau_2}.$$

The zeros of $D(z)$ are the z-plane poles; their location defines the response to transients.

Fig. 8 shows the stability limits for several values of b. Given values for b and $\omega_i\tau_2$, any value of $K\tau_2$ below the curve yields a stable loop while any value above the curve is unstable.

Because of the extra capacitor, control voltage v_c describes a continuous, ramp-like, exponential function for each pump pulse, instead of the rectangular jump that was found for the second-order loop. The same analysis that provided the stability limit gives the ramp amplitude as

$$|\Delta\omega_0|_3 = 2\pi K\left(\frac{b-1}{b}\right)\left[\frac{b-1}{b}\left(1 - e^{-\frac{b|\theta_e|}{\omega_i\tau_2}}\right) + \frac{|\theta_e|}{\omega_i\tau_2}\right] \tag{20}$$

as compared to $|\Delta\omega_0|_2 = 2\pi K$ for the second-order loop. Define $\beta = |\Delta\omega_0|_3 / |\Delta\omega_0|_2$, and assume $b|\theta_e|/\omega_i\tau_2 \ll 1$ (not necessarily true, but a common condition). Then the suppression of unwanted frequency excursion provided by the extra

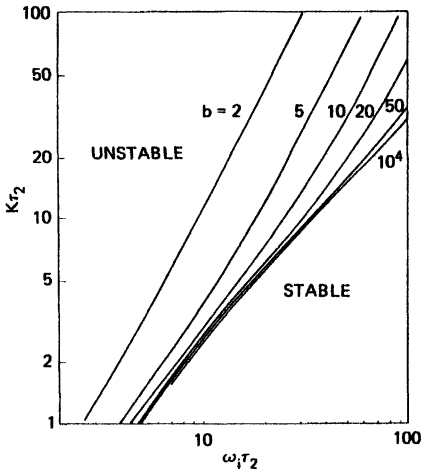


Fig. 8. Stability limits for third-order loop.

capacitor is

$$\beta \cong \frac{(b-1)\,|\,\theta_e\,|}{\omega_i \tau_2}. \tag{21}$$

When the loop is tracking near equilibrium, $|\,\theta_e\,|$ is very small so the suppression afforded by C_3 can be substantial.

V. TRANSIENT RESPONSE

The results in all of the preceding sections—the continuous-time approximation, the z-plane characteristic function, and the stability limits—were all based upon an assumption of small phase error. That assumption fails for large phase errors such as occur during acquisition of lock. An analysis was performed for large phase errors for the second-order loop; the analysis is outlined in the Appendix.

In essence, the method is to consider the loop state variables of phase error and frequency at the instant immediately before each pump pulse. These state variables are related by difference equations which were iterated numerically on a programmable calculator. The resulting printout is a sequence of the state variables along with the times of occurrence.

Two different displays are possible: frequency- or phase-error versus time to show the familiar transient response, or frequency error versus phase error to produce a phase-plane portrait [10]. Examples of both are shown below.

The question to be addressed is the following. If bandwidth (or gain, K) is very small compared to the switching frequency ω_i, then we know that the continuous analysis provides an excellent approximation to the behavior of the charge-pump loop and we can utilize the extensive information available from the study of conventional, analog PLL's. In many applications we want to be able to use a large bandwidth. Therefore we ask, "How small can ω_i/K be made before behavior departs significantly from that predicted by the continuous-time analysis?"

Several example calculations were performed in order to explore the question. A value of $K' = 2$, corresponding to $\zeta = 0.707$ in the continuous PLL, was chosen as representative of many applications. Referring to the overload curve of Fig. 4, it can be seen that the VCO is certain to overload unless $\omega_i\tau_2 > 12.6$, or, in other words, $\omega_i/K > 12.6/2 = 6.3$. To allow some margin on the limit, a value of $\omega_i/K = 10$ was chosen for the example calculations. For practical VCO's that is probably still too small a ratio but it will illustrate the results very well.

Transient phase error in response to a phase step of ±6 rad and of a frequency step of $\pm 2K$ radians per second were calculated for the conditions of $\omega_i/K = 10$ and ∞. The latter corresponds to the continuous-time PLL. Transient curves are plotted in Figs. 9 and 10, respectively.

It is apparent that, even for such a low frequency as compared to bandwidth, the response of the charge-pump loop is very close to that of the classical, continuous loop.

Asymmetry between positive and negative phase errors is evident for the charge-pump loop. (The classical loop, of course, has symmetric response with respect to error polarity.) This asymmetry arises from the polarity-asymmetric dependence of pulse duration upon phase error (see the Appendix) and dwindles as ω_i/K is made larger.

The same program yields a phase-plane solution of the PLL. Example trajectories are shown in Fig. 11. Asymmetry is also apparent in this display. Each marked point represents the state of the system at the starting instant of consecutive charge pulses. The points have been connected by straight lines, to aid in following the individual trajectories, but the actual trajectory between two calculated points has not been determined and there is no reason to suppose that it would be linear.

The shape of the trajectories may seem rather peculiar; the vertical sections do not occur on the $\Delta\omega = 0$ axis as is expected from previous phase-plane plots [10]. The discrepancy arises not from the charge-pump action, but from the choice of state variables. Here the variables are θ_e and $\Delta\omega = \omega_i - \Omega_0 - K_0 v_x$, where Ω_0 is the free-running frequency of the VCO and v_x is the voltage stored on the capacitor C in the loop filter. The usual phase-plane plot uses θ_e and $\dot{\theta}_e$ as the state variables. Proportional and integral elements enter into $\dot{\theta}_e$ whereas only integral elements contribute to the frequency variable in Fig. 11. If equivalent state variables were defined, then similar, skewed trajectories would also be obtained for the classical loop.

It is clear that the loop converges towards equilibrium without difficulty, at least for the trajectories examined. Attempts were made to examine trajectories with larger initial frequency errors. An overload phenomenon intervened: a phenomenon that the program was not designed to accommodate, so an error message was produced instead of a trajectory. Inasmuch as the loop was running very close to the nominal overload as deduced from Fig. 4, the program breakdown for small ω_i/K was not pursued further, on the supposition that a practical loop would break down under even more restrictive conditions.

Trajectories were also obtained for $\omega_i/K = 100$. Much larger $\Delta\omega/K$ values could be accommodated for that condition and the program breakdown was not encountered again. With large enough initial frequency error, the loop does not converge within the phase interval $(-2\pi, 2\pi)$ but slips one or more cycles before settling. The program was not designed to accommodate phase excursions beyond $\pm 2\pi$, so no results are provided.

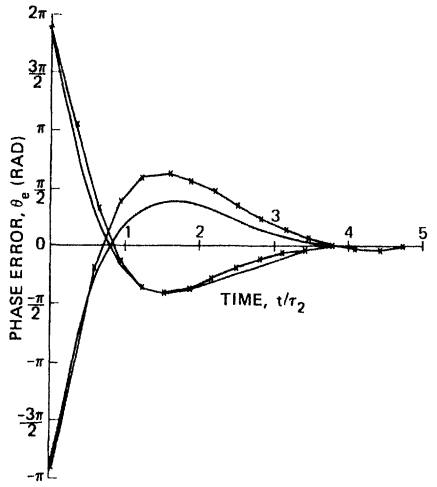


Fig. 9. Response to phase step. $\Delta\theta = \pm 6$ rad; $K\tau_2 = 2$; second-order loop. —— $\omega_i/K = \infty$; ⋊—⋊ $\omega_i/K = 10$.

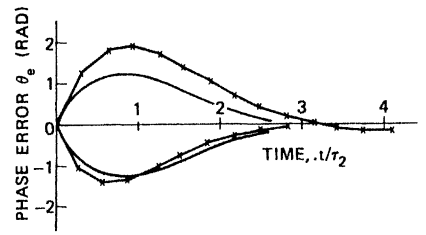


Fig. 10. Response to frequency step. $\Delta\omega/K = \pm 2; K\tau 2 = 2$; second-order loop. —— $\omega_i/K = \infty$, ⋊—⋊ $\omega_i/K = 10$.

(The phase-plane portrait for an ordinary phase detector is periodic in 2π, but the PFD portrait is more complicated. One can consider that Fig. 11 is the central region of a phase portrait but the outer regions to either side of center each extend over only -2π to 0 or 0 to $+2\pi$. The PFD portrait is not strictly periodic.)

Calculations were performed only for the second-order loop, but the third-order loop resulting from the filter of Fig. 5 is more likely to be employed in real applications. It is possible to calculate the transient response of the third-order loop in much the same way employed for second-order by taking account of three state variables. That has not yet been accomplished. Moreover, state trajectories for a third-order loop are three-dimensional and cannot be displayed readily on a two-dimensional sheet.

If the extra capacitor C_3 is effective, the third-order loop ought not suffer from VCO overload. Instead, stability limits the allowable gain for a given switch frequency (as in Fig. 8). From Fig. 7, we see that is it improbable that we would ever take b significantly less than about 10. If $K' = 2$, then the stability limit for $b = 10$ from Fig. 8 is $\omega_i/K \approx 7.5$. To obtain some stability margin, a value of ω_i/K in excess of 15 to 20 might be considered reasonable. In light of the results obtained with the second-order loop it seems fair to predict that response of the third-order charge-pump loop will be very much the same as that of the equivalent continuous-time loop.

VI. CONCLUSIONS

The conventional-wisdom rule-of-thumb has been that switching granularity effects can be neglected if the switching frequency exceeds 10 times the loop bandwidth. This paper

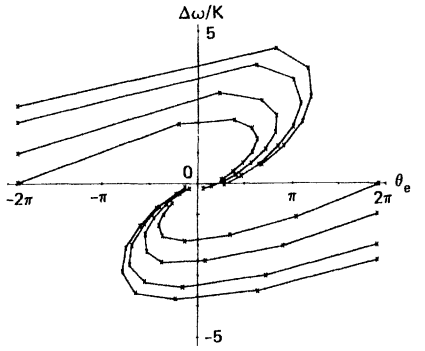


Fig. 11. Phase-plane portrait; second-order loop. $K\tau_2 = 2$; $\omega_i/K = 10$.

has shown that the rule-of-thumb is not far wrong if considered as an approximate outer limit beyond which troubles begin to appear. Somewhat more conservative design would be prudent in most circumstances.

The passive filter with current switching has been shown to have attractive properties. A fast, balanced, current-switch integrated circuit would be very helpful to the hardware designer.

The second-order loop has switching-rate frequency-excursions that are excessive for most applications. Any smoothing results in at least a third-order loop, although still Type II. Root loci for the third-order loop are presented to aid design efforts.

Transient response of practical charge-pump PLL's can be expected to be nearly the same as the response of the equivalent classical PLL.

APPENDIX

DIFFERENCE-EQUATION ANALYSIS OF CHARGE-PUMP PLL

Analysis of the charge-pump circuit is impeded by the switching of the pump current between the values $-I_p$, 0, and I_p. Moreover, the switching times are complicated functions of the relative, time-varying phases $\theta_i(t)$ of the signal input and $\theta_0(t)$ of the VCO. However, during any one switch condition, the circuit is a linear, time-invariant network and is described by linear differential equations with constant coefficients. Given the initial conditions at the start of a switching interval, it is straightforward to calculate the state variables at any time within the interval. The final state variables at the end of one interval become the initial conditions for the next interval.

If we define the phase and frequency errors at the start of a current pulse as the discrete-time state variables, then it is possible to write difference equations that describe a recursive sequence of the state. The exact difference equations were iterated to obtain the transient responses of Section V while discrete-time stability was examined for Sections III and IV by means of linearized difference equations.

This Appendix derives the linearized difference equations for the second-order PLL and shows an outline of the derivation of the exact difference equations of the second-order PLL. The process for the exact equations is shown as comments in parentheses following the corresponding portion of the linear-equations derivation.

Linearized equations were also obtained for the third-order

loop, but the derivation is not shown here because of space limitations. Exact difference equations could be found for the third-order loop, but that problem has not yet been attacked.

The notation and circuit configuration for the analysis are shown in Fig. 12. It is convenient to set the time origin to coincide with an instant of turn-on of the current switch. Observing that convention, we obtain the following equations that are valid for the entire first cycle (i.e., until the next turn-on instant of the current switch).

$$\theta_i(t) = \theta_i(0) + \omega_i t \tag{A1}$$

$$\omega_0(t) = \Omega_0 + K_0 v_c(t) \tag{A2}$$

$$\theta_0(t) = \theta_0(0) + \Omega_0 t + K_0 \int_0^t v_c(\tau)\, d\tau \tag{A3}$$

$$i_p = I_p \operatorname{sgn} \theta_e(0); \quad 0 < t < t_p \tag{A4}$$

$$= 0; \quad t_p < t < 2\pi/\omega_i$$

$$\theta_e = \theta_i - \theta_0 \tag{A5}$$

$$t_p \cong |\theta_e| / \omega_i \tag{A6}$$

$$v_c(t) = i_p R_2 + v_x \tag{A7}$$

$$v_x(t) = \frac{1}{C} \int_0^t i_p(\tau)\, d\tau. \tag{A8}$$

Suppose that the switching is initiated by an edge of the VCO waveform. Then $-I_p$ is switched on so as to retard the VCO phase; also $\theta_0(0) = 0$, and $\theta_i(0) = \theta_e(0)$, which is a negative number. The input-signal edge that shuts off the current switch occurs when $\theta_i(t) = \theta_i(t_p) = 0$; in other words, when the input phase has advanced by $\theta_e(0)$ radians at a rate of ω_i radians/second. In this case, that time is exactly $t_{p-} = |\theta_e|/\omega_i$ s.

Now suppose that the switch-on is initiated by an edge of input signal. The pump current is $+I_p$; $\theta_i(0) = 0$, $\theta_0(0) = -\theta_e(0)$, $\theta_e(0)$ is positive, and the pump current remains on until the next edge of the VCO. For the linearized analysis, that time is approximated by $t_{p+} = \theta_e(0)/\omega_i$, the same expression as for the opposite polarity of phase error.

(In actual fact, since v_c is not constant during the pump interval, the frequency of the VCO changes during $(0, t_{p+})$ so a linear equation for t_p is incorrect for positive phase error. In the nonlinear analysis it was found that the correct charging interval is a solution of a quadratic equation involving the initial conditions at $t = 0$ and the loop parameters. The quadratic solution is carried through in the numerical iteration of the nonlinear difference equations.)

Define $v_{x0} = v_x(0)$; $v_{xp} = v_x(t_p)$. Ordinary linear-network analysis methods yield

$$\theta_0(t_p) = \theta_0(0) + \Omega_0 t_p + K_0(v_{x0} t_p + i_p R_2 t_p + i_p t_p^2/2C). \tag{A9}$$

(For positive $\theta_e(0)$, setting $\theta_0(t_p) = 0$ in (A9) gives the quadratic equation for the exact value of t_{p+}, as described above.) Furthermore,

$$v_x(t_p) = v_{x0} + i_p t_p / C. \tag{A10}$$

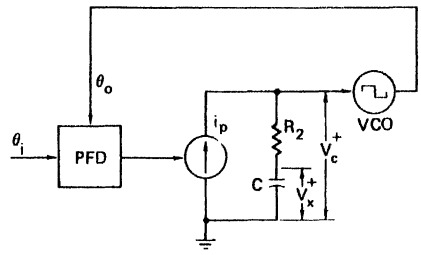


Fig. 12. Equivalent circuit of charge-pump PLL.

These equations, (A9) and (A10), are exact. (By substituting the exact solution for t_p into (A9) and (A10), the calculator program carries the state variables numerically up to t_p.)

Let t^* be the time following t_p at which the next edge—from signal or VCO, as the case may be—activates the PFD and starts a new pump pulse. Charge on the capacitor remains constant from t_p to t^*. Therefore, the VCO phase at t^* is

$$\theta_0(t^*) = \theta_0(t_p) + \Omega_0(t^* - t_p) + K_0(t^* - t_p) v_{xp}. \tag{A11}$$

Substituting (A9) and (A10) into (A11) gives

$$\theta_0(t^*) = \theta_0(0) + \Omega_0 t^* + K_0 \left[i_p R_2 t_p - \frac{i_p t_p^2}{2C} + v_{x0} t^* + \frac{i_p t_p t^*}{C} \right]. \tag{A12}$$

This last equation is also exact. To pursue the linearized analysis, substitute $i_p t_p \cong I_p \theta_e / \omega_i$ from (A4) and (A6), and approximate t^* by $2\pi/\omega_i$ to obtain

$$\theta_0(t^*) \cong \theta_0(0) + \frac{2\pi}{\omega_i}(\Omega_0 + K_0 v_{x0}) + \frac{K_0 I_p \theta_e}{\omega_i} \left[R_2 + \frac{2\pi}{\omega_i C} - \frac{|\theta_e|}{2\omega_i C} \right]. \tag{A13}$$

Except for the very last term, (A13) is linear in $\theta_e(0)$. By dropping the last term—a valid approximation for small θ_e—we obtain a linear equation for the VCO phase at time t^* in terms of the initial phase and frequency and the loop parameters.

Define $\Delta\Omega = \omega_i - \Omega_0$ and recall that $\theta_e = \theta_i - \theta_0$. By the previous approximations, θ_i advances by 2π in the time interval $(0, t^*)$. With these substitutions, we obtain the linear difference equations

$$\theta_e(t^*) = \theta_e(0) + 2\pi\Delta\Omega/\omega_i - \frac{K_0 I_p \theta_e(0)}{\omega_i}(R_2 + 2\pi/\omega_i C) - \frac{2\pi K_0 v_{x0}}{\omega_i} \tag{A14}$$

and

$$v_x(t^*) = v_{x0} + \frac{\theta_e(0) I_p}{\omega_i C}. \tag{A15}$$

(All approximations are avoided in the calculator program. The quantities $\theta_0(t_1) = 2\pi$, from (A12), and $\theta_i(t_2) = \theta_i(0) +$

$\omega_i t_2 = 2\pi$ are solved for t_1 and t_2. The smaller of these is taken as the value for t^*. Having obtained the correct values of t_p and t^*, the values for $\theta_e(t^*)$ and $\Delta\omega(t^*)$ are calculated and the process repeats with these state variables as the new initial conditions. The program starts at specified initial conditions and stops after arriving within a specified tolerance band about the zero state.)

The linearized analysis continues by taking z-transforms [11] of (A14) and (A15). Treating the initial frequency error as a frequency step gives the z-transformed equations

$$z\theta_e(z) = \theta_e(z) + \frac{2\pi\Delta\Omega z}{\omega_i(z-1)} - \frac{K_0 I_p \theta_e(z)}{\omega_i}(R_2 + 2\pi/\omega_i C) - 2\pi K_0 V_x(z)/\omega_i \tag{A16}$$

$$zV_x(z) = V_x(z) + \frac{\theta_e(z) I_p}{\omega_i C}. \tag{A17}$$

Solving for $\theta_e(z)$ gives

$$\theta_e(z) = \frac{2\pi z\Delta\Omega/\omega_i}{(z-1)^2 + (z-1)\dfrac{K_0 I_p}{\omega_i{}^2 C}(2\pi + \omega_i C R_2) + \dfrac{2\pi K_0 I_p}{\omega_i{}^2 C}}. \tag{A18}$$

Applying notation definitions from (9), the denominator of (A18) becomes $D(z)$, as shown in (13). Analysis of pole locations follows by standard methods [11].

Linearized difference equations (three of them) were derived in the same manner for the third-order loop. The volume of algebra is substantially greater than for the second-order loop so only the results are given in Section IV. Only linear approximations have been performed for the third-order loop; the exact equations have not been attempted.

REFERENCES

[1] R. C. E. Thomas, "Frequency comparator performs double duty," *EDN*, pp. 29–32, Nov. 1, 1970.

[2] J. I. Brown, "A digital phase and frequency-sensitive detector," *Proc. IEEE*, vol. 59, p. 717, Apr. 1971.

[3] *Phase-Locked Loop Data Book*, 2nd ed. Motorola, Inc., Aug. 1973.

[4] D. K. Morgan and G. Steudel, "The RCA COS/MOS phase-locked-loop," RCA, Somerville, NJ, Application Note ICAN-6101, Oct. 1972.

[5] C. A. Sharpe, "A 3-state phase detector can improve your next PLL design," *EDN*, pp. 55–59, Sept. 1976.

[6] F. M. Gardner, *Phaselock Techniques*, 2nd ed. New York: Wiley, 1979.

[7] P. Lue, "A multispeed digital regenerative repeater for digital data transmission," in *Conf. Rec., 1979 Nat. Telecommun Conf.*, paper 14.1.

[8] J. A. Afonso, A. J. Quiterio, and D. S. Arantes, "A phase-locked loop with digital frequency comparator for timing signal recovery," in *Conf. Rec., 1979 Nat. Telecommun. Conf.*, paper 14.4.

[9] J. Tal, "Speed control by phase-locked servo systems—New possibilities and limitations," *IEEE Trans. Ind. Electron. Contr. Instrum.*, vol. IECI-24, pp. 118–125, Feb. 1977.

[10] A. J. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966, ch. 3.

[11] J. T. Tou, *Digital and Sampled-Data Control Systems*. New York: McGraw-Hill, 1959.

z-Domain Model for Discrete-Time PLL's

JERRELL P. HEIN, MEMBER, IEEE, AND JEFFREY W. SCOTT

Abstract—**The well-known s-domain model for continuous-time phase-locked loops (PLL's) is a fundamental tool for the linearized analysis of these systems. For PLL's with digital inputs and outputs, however, a discrete-time z-domain model more accurately describes loop behavior. In this paper, a methodology is described for obtaining an accurate z-domain description of a discrete-time PLL. This method is an alternative approach to the analysis presented in [4]. The modeling technique transforms portions of the s-domain PLL model directly into the z-domain, requiring only straightforward algebraic manipulations even for complex loop filters. This methodology is demonstrated for a simple loop filter, and measurements from the digital signaling interface (DSI) integrated circuit are used to compare s-domain, z-domain, and time step analysis results for a more complicated loop filter. The z-domain model, although only incrementally more complicated than the s-domain model, is shown to be more accurate, especially at higher jitter frequencies.**

I. Introduction

THE basic s-domain PLL model presented in numerous texts [1]–[3] treats a loop in the locked condition as a linear continuous-time system. The input and output waveforms are assumed to be sinusoidal and the phase detector is modeled as a linear analog multiplier with an inherent ideal low-pass filter. Although the linear continuous-time model is useful within these constraints, many PLL's operate under conditions not accurately represented by these assumptions. In particular, a large class of PLL's used most notably in data communications have digital waveforms as both inputs and outputs. For these PLL's, the phase information is contained in the digital waveform transitions and should be viewed as a discrete-time sequence. The linear, continuous-time model can approximate the operations of these loops only if the jitter frequencies of interest are much less than the incoming data transition rate.

To obtain an accurate discrete-time model of a PLL, one can write the complete set of differential equations describing the system, convert these into difference equations, linearize the equations along the way, and finally z-transform the result to obtain $H(z)$, the z-domain jitter transfer function. This procedure quickly becomes cumbersome for all but the simplest loop filters as noted by Gardner [4].

In this paper, a z-domain description for a PLL is presented which is only incrementally more complicated than the linear continuous-time model. The model uses the impulse invariant transformation to convert the s-domain description of a portion of the loop directly into the z-domain. In addition to the model derivation and implementation for a simple loop filter, results will be shown comparing the continuous- and discrete-time models with a PLL timestep simulator, and actual measured devices. In Appendix A, the assumptions needed to linearize the network analysis will be discussed and it will be shown that these assumptions are identical to those required to analyze the PLL with the use of linearized difference equations. Finally, in Appendix B, the analysis methodology will be extended to PLL's with switched capacitor loop filters.

Manuscript received October 16, 1987; revised March 18, 1988. This paper was recommended by Associate Editor C. A. T. Salama.
The authors are with AT&T Bell Laboratories, Reading, PA 19612.
IEEE Log Number 8823450.

II. Discrete-Time Model Development

A. Continuous-Time PLL's and the s-Domain Model

A functional block diagram for a continuous-time PLL is shown in Fig. 1. The block diagram of the associated s-domain PLL model is shown in Fig. 2. This model assumes that the input waveform is a phase modulated sine wave, i.e., it has the form,

$$s_a(t) = a \cdot \sin\left(w_c t + \phi_i(t)\right).$$

The input phase modulation $\phi_i(t)$, and oscillator output $\phi_o(t)$ are continuous functions of time.

In Fig. 2, the summing node and K_p gain block represent the operation of the phase detector in the frequency domain. The phase detector is assumed to be a linear analog multiplier which multiplies the input and output waveforms. The result is a multifrequency signal which contains the phase difference information desired $(\phi_i(s) - \phi_o(s))$ in the low frequency portion of the phase detector output spectrum. The higher frequency multiplicative products are ignored in the analysis. The effect of this last assumption can be included by assuming that an ideal low-pass filter sits behind the multiplier. In practice, the loop filter following the phase detector approximates this ideal filter.

Finally, the loop filter $(F(s))$ and voltage-controlled oscillator (K_o/s) are included in the block diagram. In general, the loop filter is modeled accurately as a linear continuous-time element, especially if it is implemented with passive components. Relaxation and current ramping oscillators exhibit linear and wideband voltage-to-frequency relationships and are also accurately modeled as linear continuous-time elements.

The jitter transfer characteristic for the s-domain model is also shown in Fig. 2. $H(f)$ will be used to compare the accuracy of the s- and z-domain models in a later section.

The assumption inherent in the application of the s-domain model to a PLL are generally valid for loops operating with sinusoidal inputs and outputs (continuous-time PLL's). It will be shown, however, that for PLL's

Reprinted from *IEEE Trans. Circuits and Systems*, vol. 35, pp. 1393-1400, November 1988.

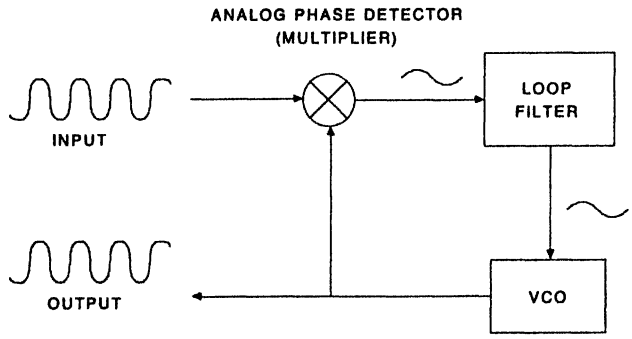


Fig. 1. Functional block diagram for continuous-time PLL.

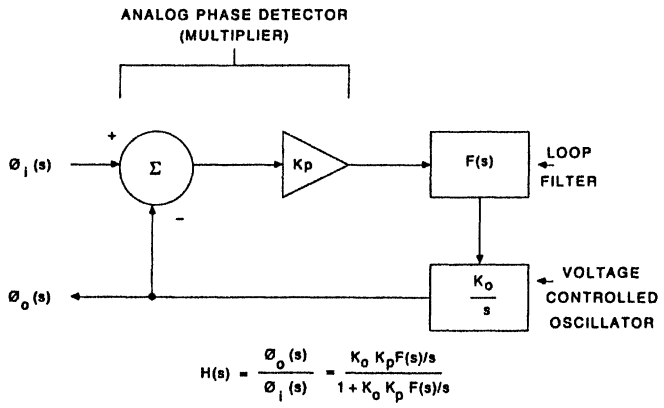


Fig. 2. S-domain model block diagram.

operating with digital inputs and outputs (discrete-time PLL's), some of the assumptions may result in significant inaccuracies in the analysis. In these cases, it is necessary to develop a discrete-time model which more accurately reflects the actual operation of the loop.

B. Discrete-Time PLL's and the z-Domain Model

The functional block diagram for a discrete-time PLL is shown in Fig. 3. The input waveform to the PLL can be described as

$$s_d(t) = a \cdot \mathrm{sgn}\left[\sin\left(w_c t + \phi_i(t)\right)\right]$$

where sgn is the signum function. The difference between the discrete- and the continuous-time PLL structures lies in the implementation of the phase detector. The phase detector for a discrete-time PLL is a digital circuit which drives a pulse width modulated digital pulse into the loop filter. The width of the pulse is determined by the time difference between the input data reference edge and the recovered clock edge. The digital phase detector operation is depicted in Fig. 4. Phase difference information arrives at the phase detector input only when data reference edges occur. Therefore, the phase error between the data and clock is properly viewed as a discrete-time sequence with values spaced at intervals approximately equal to the time between input data reference edges. For a repetitive data pattern, the time interval between data pulses is given by: $T = 1/$(recovered clock frequency·input data one's density). For example, a repetitive 1,0 return-to-zero (RZ) data pattern (as shown in Fig. 4) would have $T = 1/$(recovered clock frequency·$\frac{1}{2}$).

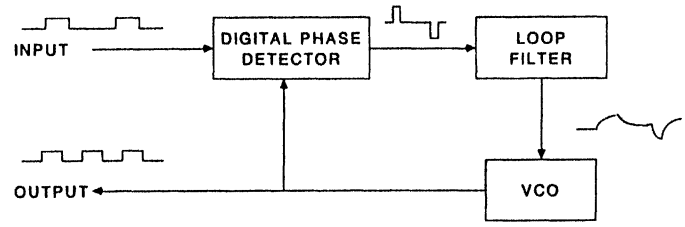


Fig. 3. Functional block diagram for discrete-time PLL.

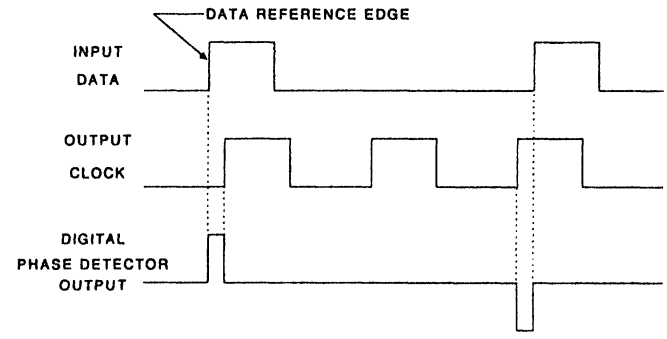


Fig. 4. Digital phase detector operation.

For small phase errors, the pulses driving the loop filter can be modeled accurately as weighted impulses. The accuracy of this approximation is calculated in Appendix A for a simple RC loop filter. The phase detector samples the difference between the data and clock phases at intervals of T seconds and drives the loop filter with weighted impulses. The gain factor K_p simply represents the conversion factor between input phase error and output impulse area. Therefore, the digital phase detector in a discrete-time PLL can be modeled as a summing node and gain block in the z-domain. Note that the type of blocks required to model the digital phase detector in the z-domain are the same type of blocks used to model the analog phase detector in the s-domain.

With the digital phase detector now properly modeled in the z-domain, the problem remaining is the accurate modeling of the continuous-time loop filter and VCO elements in the z-domain. For arbitrary signals driving the loop filter and VCO, it would be impossible to accurately map the elements' entire s-domain response into the z-domain. But the loop filter and VCO are not being driven by arbitrary signals, they are being driven by a series of weighted impulses from the phase detector. Therefore, it is only necessary to preserve the loop filter and VCO's impulse response in transforming from the s- to z-domain. In other words, the essential characteristics of the loop filter and VCO will be preserved if the derived discrete-time network has a unit impulse response with values equal to T spaced samples of the continuous-time impulse response. A transformation exists which guarantees exactly this type of relationship—the impulse invariant transformation.

Fig. 5 illustrates the relationship guaranteed by the impulse invariant transformation. Given a continuous-time network with impulse response $h_a(t)$, the impulse invari-

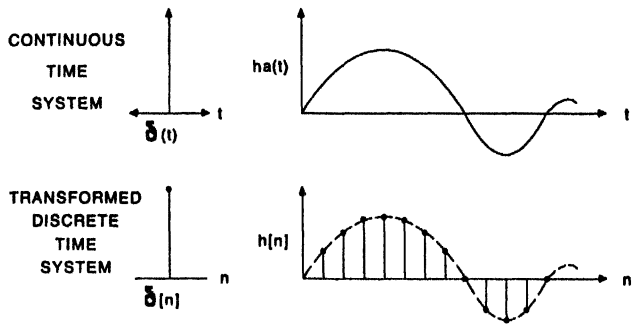


Fig. 5. Impulse invariant transformation.

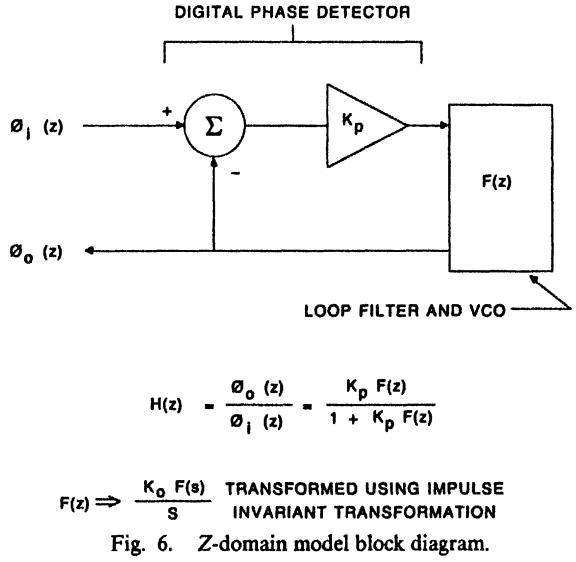


Fig. 6. Z-domain model block diagram.

ant transformed discrete-time network will have a unit sample response $h[n]$ where the $h[n]$ values are equal to the $h_a(t)$ values sampled at intervals of T seconds [5]. The actual transformation of some function $F(s)$ to $F(z)$ is a straightforward process which will be demonstrated in a later section. Basically, the function $F(s)$ must be divided using partial fraction methods into simple $a/(s-s_k)$ and $1/s^2$ terms. Each of these terms then transforms directly into the z-domain. Poles at s_k in the s-domain map to poles at $z_k = e^{s_k T}$ in the z-domain. Zeros in the s-domain move to points in the z-plane which depend on the pole locations. Because the impulse-invariant transformation is not an algebraic mathematical mapping, the combined loop filter and VCO s-domain terms $(K_o F(s)/s)$ must be transformed together.

Once the loop filter and VCO are transformed into $F(z)$, the entire loop can be evaluated in the z-domain. The z-domain model is shown in Fig. 6 along with the jitter transfer characteristic $H(z)$. The development of the complete discrete-time PLL model involves only two steps: (1) calculation of the phase detector gain constant K_p, and (2) transformation of the loop filter and VCO s-domain descriptions into the z-domain using the impulse invariant transformation. The model accurately represents the operation of the digital phase detector with digital data and clock inputs, and correctly transforms the continuous-time networks $F(s)$ and K_o/s into the z-domain. The application of this technique to a second-order PLL will be demonstrated in the next section.

C. Application of z-Domain Model to a Second-Order PLL

In this section, the z-domain analysis will be performed on a discrete-time PLL with a first-order loop filter. For comparison purposes, the classical s-domain model of this PLL will be described first.

The s-domain PLL model is shown in Fig. 2. A first-order transimpedance loop filter transfer function $F(s)$, (see Fig. 11) is given by

$$F(s) = R\left(\left(s + \frac{1}{RC}\right)\Big/ s\right)$$

where R and C are the series resistance and capacitance of the loop filter. Using this filter transfer function in the expression for $H(s)$ shown in Fig. 2 yields the following jitter transfer characteristic:

$$H_c(s) = RK_oK_p \frac{s + \dfrac{1}{RC}}{s^2 + RK_oK_p s + \dfrac{K_oK_p}{C}}$$

where the subscript "c" denotes the continuous-time nature of the jitter transfer expression; K_o is the voltage-to-frequency conversion gain of the VCO; and K_p is the phase detector conversion gain given by

$$K_p = dI_p/2\pi$$

where d is the one's density of the PLL input data and I_p is the magnitude of the phase detector pump current.

The loopgain expression required for root locus construction is given by

$$\text{loopgain} = K_p \cdot \frac{K_o}{s} \cdot F(s) = K \cdot \frac{s + \dfrac{1}{RC}}{s^2}$$

where K is defined as

$$K = K_oK_pR.$$

The root locus for the PLL modeled in the s-domain is shown in Fig. 7. There are two open loop poles at the origin of the s-plane and one open loop zero at $s = 1/RC$. Note that, since the loopgain parameter K contains the one's density information, the closed-loop pole locations of the s-domain jitter transfer model depend on the input data pattern. However, it is of further interest to note that the s-domain model never predicts an unstable loop for any combination of PLL parameters.

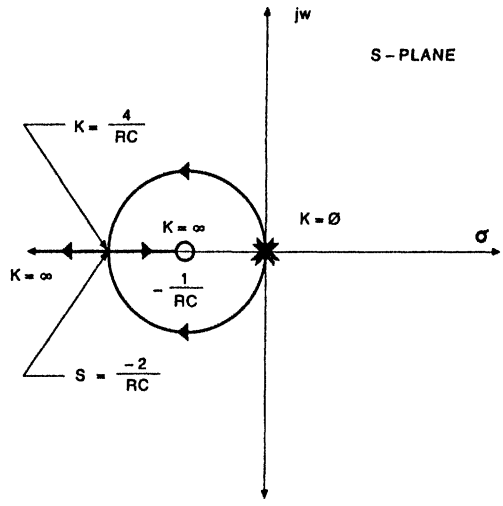


Fig. 7. *S*-domain root locus.

The discrete-time or *z*-domain analysis of the second-order PLL begins by applying the impulse invariant transformation to the term $(K_o F(s)/s)$. For the first order loop filter under investigation, this term is given by

$$\frac{K_o}{s}\cdot F(s) = RK_o\frac{s+\frac{1}{RC}}{s^2} = \frac{K_o}{C}\left[\frac{1}{s^2}+\frac{RC}{s}\right].$$

The impulse response of this section of the PLL is found by taking the inverse Laplace transform, which yields

$$h_a(t) = \frac{K_o}{C}[t+RC]u(t).$$

The impulse-invariant property requires that

$$h[n] = h_a(nT) = \frac{K_o}{C}[nT+RC]u[n]$$

where T is the effective sampling or impulse arrival rate given by $T = 1/df_{vco}$ where d is the one's density of the incoming data and f_{vco} is the VCO recovered clock frequency. The desired *z*-domain description of the loop filter and VCO is found by taking the *z*-transform of $h[n]$ (the $nT\cdot u[n]$ term is most easily transformed by invoking the differentiation property), which results in the following expression for $F(z)$:

$$F(z) = \frac{K_o}{C}\left[\frac{Tz^{-1}}{(1-z^{-1})^2}+\frac{RC}{1-z^{-1}}\right] = RK_o\frac{z\left[z-\left(1-\frac{T}{RC}\right)\right]}{(z-1)^2}.$$

For simplification, let $\alpha = 1-(T/RC)$. Then

$$F(z) = RK_o\frac{z(z-\alpha)}{(z-1)^2}.$$

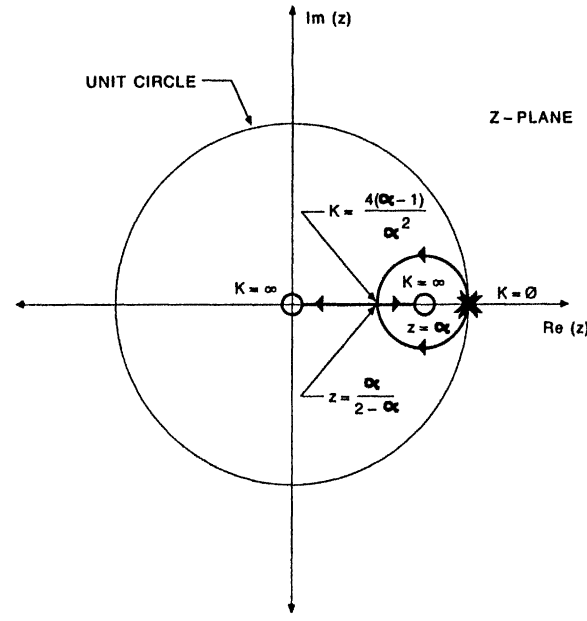


Fig. 8. *Z*-domain root locus.

Using this transformed filter/VCO transfer function in the expression for $H(z)$ shown in Fig. 6 yields the following jitter transfer characteristic:

$$H_d(z) = \frac{K_oK_pR}{1+K_oK_pR}\cdot\frac{z(z-\alpha)}{z^2-\frac{2-\alpha K_oK_pR}{1+K_oK_pR}z+\frac{1}{1+K_oK_pR}}$$

where the subscript "d" denotes the discrete-time nature of the jitter transfer expression, and K_p is the phase detector conversion gain given by

$$K_p = \frac{I_p}{2\pi f_{VCO}}$$

where I_p is the magnitude of the phase detector pump current.

The loopgain expression required for *z*-domain root locus construction is given by

$$\text{loopgain} = K_p\cdot F(z) = K\frac{z(z-\alpha)}{(z-1)^2}$$

where the loopgain parameter K is again defined as K_oK_pR.

The root locus for the PLL modeled in the *z*-domain is shown in Fig. 8. There are two open loop poles at $z=1$ and open loop zeros at $z=0$ and $z=\alpha$. It is interesting to note that, since the open-loop zero location (α) is a function of the input data one's density, the *z*-domain model can predict unstable loop performance for the condition of $T > 2RC$ (in which case an open-loop zero resides on the negative real axis outside the unit circle).

For $T \ll RC$, the quantity α is close to unity. In this case, the *z*- and *s*-domain models predict similar jitter

transfer characteristics for jitter frequencies within the loop's bandwidth. Mathematically, the reason is that minimal aliasing occurs as the loop filter/VCO s-domain transfer characteristic is transformed to the z-domain; physically, the results predicted by the two models converge in this case because the s-domain assumption concerning the ability of the loop filter to reject phase detector high frequency output energy is valid. The two models diverge for jitter frequencies outside the loop's bandwidth where the z-domain analysis correctly models the repetitive nature of the input/output jitter spectra. For wider bandwidth loops, the z- and s-domain models will diverge even for jitter frequencies within the loop's bandwidth. The reason is that the relatively wideband loop filter does not reject high frequency phase detector components as the s-domain approximation would suggest. In the next section, it will be shown that the s- and z-domain models begin to diverge for jitter frequencies within the loop's bandwidth for practical PLL's. The two models will also be compared with experimental results taken from the DSI (Digital Signaling Interface) integrated circuit.

III. Measured Results

In the design of the wideband PLL used for clock recovery in the DSI device, three modeling techniques were used to estimate loop performance. The jitter transfer characteristic was derived using both the linear s- and z-domain models presented in this paper. In addition, a timestep simulator was written in Fortran which performed a transient analysis on the loop in the locked condition using time domain models of the loop elements. The timestep simulation was thought to be the most accurate of the three methods because it included a number of nonlinear effects in the element models. For example, in the timestep simulator, the output of the phase detector was treated as a finite width pulse, instead of a weighted impulse and the finite pull range of the VCO was taken into account.

The nominal design parameters were entered into each of three models and the jitter response was evaluated. For the timestep simulator, a 0.1 Unit Interval (1 U.I. = 1 clock period) input jitter magnitude was used as the "small signal" input. Plots of the magnitude of $H(f)$ are shown in Fig. 9. It is seen that the timestep simulator results agree with the z-domain model results within ± 0.3 dB over all frequencies. With an input data transition rate of 193 kHz and VCO clock rate of 1.544 MHz the linear models agree within ± 0.3 dB up to about 1/20th of the input transition rate. For the z-domain model and the timestep simulator, the loop response is periodic with a period of 96.5 kHz, whereas the s-domain response continues to roll off at higher frequencies. As a result of these simulations, the z-domain model and timestep simulator were chosen as the design tools for determining loop parameters in the actual PLL.

Upon receipt of silicon, the individual loop parameters were measured for a specific device. These parameters

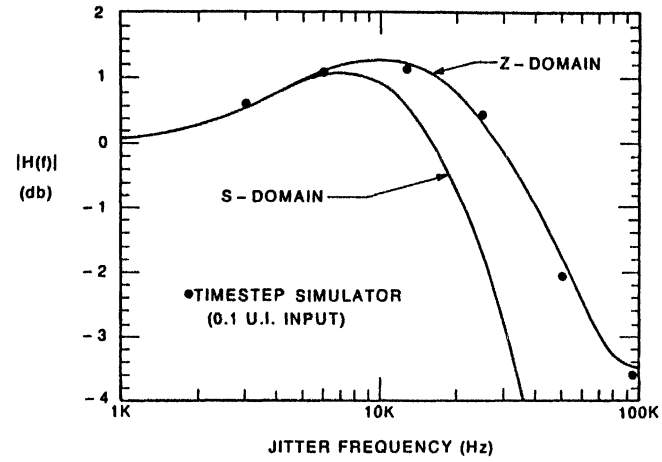


Fig. 9. $|H(f)|$ comparisons using simulators.

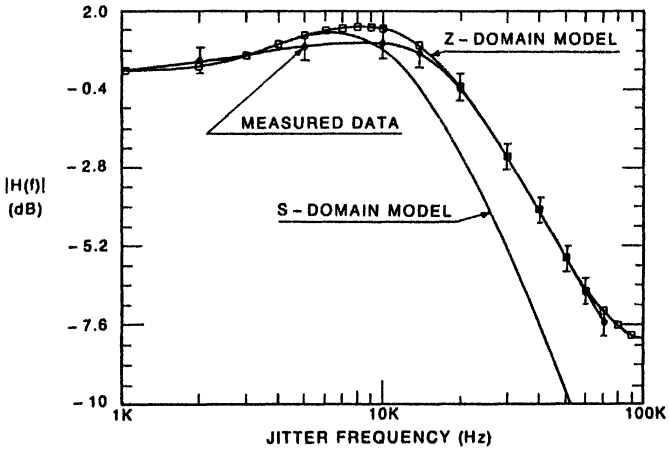


Fig. 10. $|H(f)|$ comparisons using measured data.

were then entered into both the s- and z-domain models and $H(f)$ was calculated in both cases. The results were compared with the measured $H(f)$ for the device under test. The measured $H(f)$ is defined to be the output jitter amplitude (peak-to-peak) at the input jitter frequency divided by the input jitter amplitude (pk–pk). Fig. 10 shows the three curves. It is seen that the z-domain results agree with the measured results within the ± 0.5-dB measurement error bars. Again, the s-domain $H(f)$ tracks the z-domain $H(f)$ for low jitter frequencies, but diverges at about 1/20th the input data transition rate. The agreement of the timestep simulator, fabricated PLL, and z-domain $H(f)$ verifies the accuracy of the discrete-time linear model.

IV. Summary and Conclusions

In this paper, two classes of PLL's were described: continuous-time and discrete-time PLL's. Classic s-domain analysis, although valid for continuous-time PLL's, cannot always accurately predict the behavior of discrete-time loops. The z-domain model, presented here, takes into account the sampled data nature of the digital phase detector and accurately predicts overall loop performance. The interface between discrete- and continuous-time ele-

ments is handled using the impulse invariant s-to-z-domain transformation. In this way, the entire PLL is analyzed in the discrete-time domain. Measured results from the DSI phase-locked loop support the z-domain model.

The modeling philosophy presented in this paper can be generalized to many systems in which discrete- and continuous-time elements coexist. One such system, a PLL with a switched capacitor loop filter driving a continuous-time VCO, is described in Appendix B. In this case, the interface waveform is staircase in nature; therefore, the step invariant transformation is appropriate. In general, if a well-defined interface waveshape exists, an appropriate transformation can be chosen to convert all s-domain elements into the z-domain. In this way, a complete linear model for the mixed continuous-time/discrete-time system can be developed.

Appendix A
Network Linearization Approximations

In the derivation of the z-domain model for discrete-time PLL's presented in this paper, a number of approximations were made to linearize the network. In [4], Gardner uses the difference equation approach to derive a discrete-time PLL's response to an input phase ramp (frequency step). In linearizing the difference equations, a number of approximations are also required. In this appendix, each of the linearizing assumptions will be discussed. It will be shown that the assumptions required to derive the z-domain model through the impulse invariant transformation are the same as the assumptions required to linearize the difference equations. Therefore, both approaches lead to the same results, although the impulse invariant approach significantly reduces the amount of computation required.

Linearizing Assumptions

1) In any linear PLL model, it must be assumed that the loop filter and VCO operate in a linear fashion. As stated previously, many practical loop filters and VCO's are well modeled as linear elements.

2) In both discrete-time models, it must be assumed that phase samples occur at constant intervals and that the loop responds symmetrically to leading and lagging phase information. This implies that:

a) the phase tracking errors are small,
b) the data pattern is constant,
c) the bit-to-bit data jitter is small,
d) no significant amount of high frequency energy can propagate directly around the loop, generating asymmetric behavior for leading and lagging phase errors.

Typically, the finite high frequency response of the VCO and of ripple filters built into the loop filter remove this high frequency energy sufficiently to allow the loop to operate in a linear fashion.

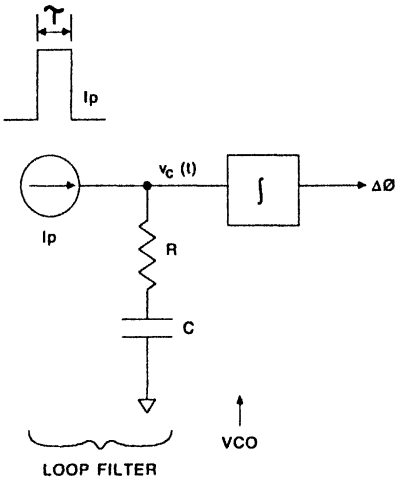


Fig. 11. PLL response to finite width pulse input.

3) Finally, in the model using the impulse invariant transformation, it was assumed that the pulse width modulated output of the digital phase detector could be approximated as a series of weighted impulses. The validity of this approximation can be investigated by analyzing the loop filter and VCO response to a finite width pulse for a simple RC loop filter. In Fig. 11, a simple RC charge pump loop filter is shown. The loop filter is driven with a current pulse of magnitude I_p and duration τ.

The output of the loop filter from time 0 to T (the time at which the next reference data edge arrives) is given by

$$v_c(t) = I_p R + \frac{I_p}{C} t, \qquad 0 \leqslant t \leqslant \tau$$

$$= \frac{I_p}{C}\tau, \qquad \tau \leqslant t \leqslant T.$$

The VCO integrates its control voltage to produce an output phase shift. Therefore, the total phase shift at time T caused by the finite width pulse starting at time 0 is given by

$$\Delta\phi = K_o \int_0^T v_c(t)\, dt$$

$$= K_o \left[\int_0^{\tau} \left(I_p R + \frac{I_p}{C} t \right) dt + \int_{\tau}^{T} \frac{I_p}{C} \tau\, dt \right]$$

$$= (I_p \tau) \frac{K_o}{2C} [2(RC + T) - \tau]. \tag{1}$$

From this equation, one can see that the phase shift is a linear function of pulse area $(I_p \cdot \tau)$ if $2 \cdot (RC + T) \gg \tau$. For reasonable loop filter components and small tracking errors, this is a very good approximation. For the wideband loop used to verify the z-domain model in this paper, the RC time constant of the loop filter was $\sim 40\ \mu$s, much larger than the maximum phase tracking error (< 50 ns). Therefore, approximating the finite width pulse with a weighted impulse of area $I_p \cdot \tau$ introduces negligible error in the analysis.

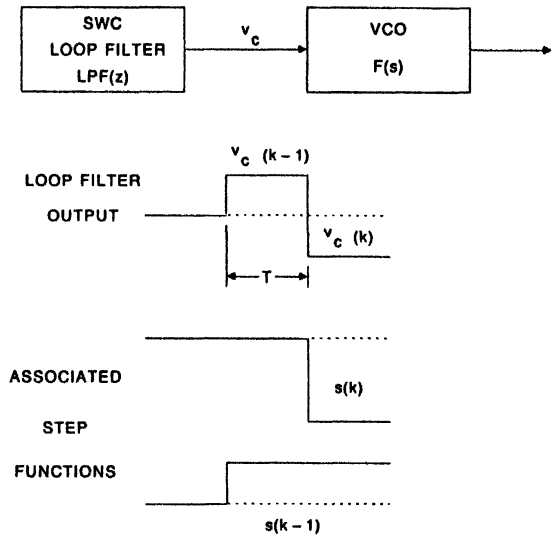


Fig. 12. Switched capacitor loop filter/continuous-time VCO interface.

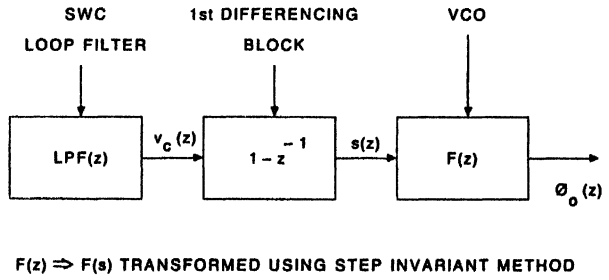


Fig. 13. Z-domain model for SWC loop filter/VCO interface using step invariant transformation.

In the difference equation analysis from [4], the $|\phi_c|/2w_iC$ term was dropped from (A13) in order to linearize the equation. This is equivalent to dropping the nonlinear τ term in (1) above. Thus, the difference equation approach also approximates the finite width pulses coming from the digital phase detector as a series of weighted impulses.

From the above analysis, it is shown that the linearizing approximations required for the impulse invariant transformation method and difference equation method are the same. Also, for most practical PLL's, the linearizing assumptions are valid and the loop operates in a linear fashion for small phase tracking errors.

Appendix B
Extension of Model to Switched-Capacitor Loop Filter/Continuous Time VCO Interface

The previous sections of this paper have dealt with a PLL architecture which consisted of a digital phase detector which was well modeled directly in the z-domain and a continuous-time loop filter and VCO which were described by s-domain transfer functions. The modeling technique presented addressed the problem of converting the continuous-time components of the PLL into the discrete-time domain when the loop filter is being driven by a weighted series of impulses. In other PLL architectures, switched capacitor loop filters may be used to drive a continuous-time VCO. In these cases, an extension of the modeling philosophy presented in this paper can be used to develop a discrete-time linear model for the PLL.

The functional block diagram for such a switched-capacitor loop filter/continuous-time VCO interface is shown in fig. 12. The output of the loop filter is a staircase waveform which changes values at intervals of T seconds. This staircase waveform can be viewed as a superposition of step functions whose value for the kth time interval is given by the difference between the output of the SWC loop filter in the k and $(k-1)$th intervals. Quantitatively, if $s(k)$ is the output step function in the kth interval, then:

$$v_c(k) = v_c(k-1) + s(k)$$

$$= \sum_{j=0}^{k} s(j)$$

or

$$s(k) = v_c(k) - v_c(k-1).$$

Transforming this relationship into the z-domain yields

$$s(z)/v_c(z) = 1 - z^{-1}.$$

Therefore, the input to the VCO can be modeled as a summation of step functions $s(k)$ if the output of the loop filter $v_c(k)$ is passed through a first-differencing block with transfer characteristic $1 - z^{-1}$. The first-differencing block, of course, does not exist in the real system, it is only required in the model in order to represent the loop filter output as a sum of step functions.

The treatment of a continuous-time system being driven by a series of weighted step functions is analogous to the treatment of a continuous-time network being driven by a weighted sum of impulses. For a step function input to the VCO, it is necessary to use the step invariant transformation to model the VCO in the z-domain. A reconstruction filter or VCO finite frequency response can be handled by including these s-domain singularities with the basic K_o/s VCO term and step transforming the entire expression into the z-domain. The resulting z-domain model for the SWC loop filter/continuous-time VCO interface is shown in Fig. 13.

Once the continuous-time elements have been transformed into the z-domain using the step invariant method, the discrete-time analysis of the PLL can proceed completely in the discrete-time domain.

References

[1] F. Gardner, *Phaselock Techniques*. New York: Wiley, 1979.
[2] P. Gray and R. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: Wiley, 1984.
[3] R. Best, *Phase-Locked Loops*. New York: McGraw-Hill, Inc. 1984.
[4] F. Gardner, "Charge pump phase-lock loops," *IEEE Trans. Commun.*, vol. COM-28, Nov. 1980.
[5] A. Oppenheim and R. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

ANALYZE PLLs WITH DISCRETE TIME MODELING

This multi-rate, discrete time-domain model provides an accurate description of phase-lock-loop dynamics.

DESIGN and optimization of phase-locked loops (PLLs) for frequency-synthesizer applications often require analysis of jitter and loop behavior in the time domain. Since these circuits operate on digital signals, where phase information is represented by signal edges, classical s-domain analysis is a first-degree approximation at best.[1]

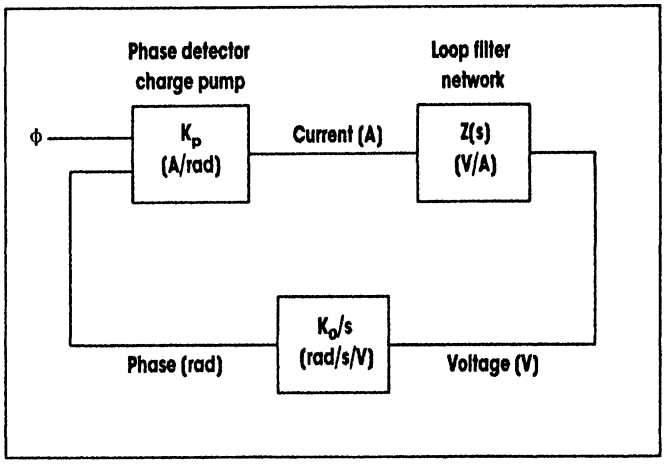
Classical PLL analysis models are based on assumptions regarding the behavior of the loop and its component hardware.[2] These assumptions are that the loop is frequency locked, the loop operates in continuous time, the voltage-controlled-oscillator (VCO) signal source is an ideal integrator, and that the phase detector is a linear adder.

A PLL in the s-domain (Fig. 1) can be calculated as:

$$G(s) = K_p Z(s) \frac{K_o}{s} \quad (1)$$

With a typical series resistance-capacitance (RC) filter configuration

JANOS KOVACS, Senior Design Engineer, Semiconductor Div., Analog Devices, Inc., 1 Technology Way, Norwood, MA 02062; (617) 937-1328.



1. Individual function blocks can be tied together to form an s-domain diagram of a PLL in the time domain.

(R_1, C_1) with a smaller capacitor in parallel (C_2), the filter impedance is:

$$Z(s) = \left(R_1 + \frac{1}{sC_1}\right) \times \frac{\frac{1}{sC_2}}{R_1 + \frac{1}{sC_1} + \frac{1}{sC_2}} \quad (2)$$

assuming $C_1 >> C_2$, leads to:

$$Z(s) = \frac{1}{sC_1} \frac{1 + sR_1C_1}{1 + sR_1C_2} \quad (3)$$

The open-loop transfer function becomes:

$$G(s) = K_p \frac{1}{sC_1} \frac{1 + sR_1C_1}{(1 + sR_1C_2)} \frac{K_o}{s} \quad (4)$$

This third-order system has two poles at $\omega=0$, one zero at $\omega_1 = 1/R_1C_1$, and another pole at $\omega_2 = R_1C_2$. The closed-loop system will be stable if the open-loop crossover frequency falls between ω_1 and ω_2, where the rolloff is 20 dB/decade. At the crossover frequency, the loop

Table 1: VCO jitter versus input frequency

Input frequency	ΔT (peak-to-peak)	Sigma
0 to 10 MHz	10 ns	3 ns
11.8 MHz	7.1 ns	1.5 ns
15.5 MHz	5 ns	0.91 ns

Note: Measurements are for the AD897 with VCO frequency set at 30 MHz.

Reprinted with permission from *Microwaves & RF*, J. Kovacs, "Analyze PLLs with Discrete Time Modeling," pp. 224-229, May 1991.

gain is one, therefore:

$$\omega_o = (K_pK_o)R_1 = KR_1 \quad (5)$$

The loop response can be approximated as a second-order system, Assuming $\omega_2 >> \omega_1$:

$$H(s) = \frac{G(s)}{1 + G(s)}$$

$$= \frac{K}{C_1} \frac{\left(1 + \frac{s}{\omega_1}\right)}{\left(s^2 + s(KR_1) + \frac{K}{C_1}\right)} \quad (6)$$

The closed-loop transfer function has its own 0-dB point, ω_0. By analyzing the PLL's amplitude characteristics (Fig. 2), the natural frequency of the system can be derived:

$$\omega_n = \sqrt{\frac{K}{C_1}} \quad (7)$$

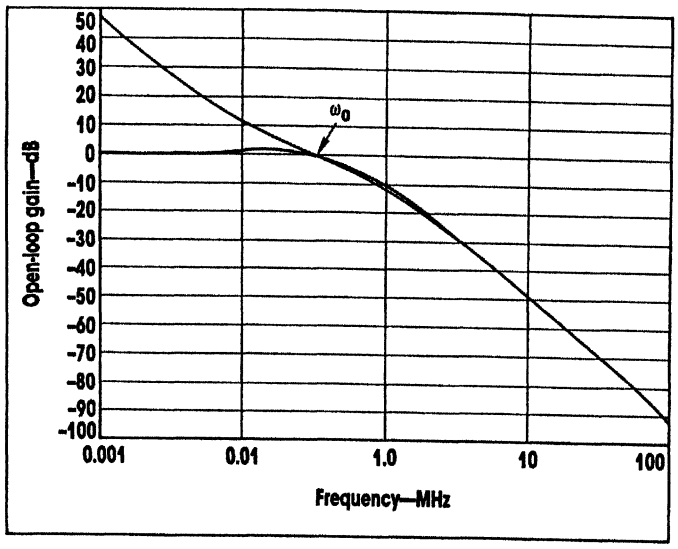
Using the above expressions for ω_0 and ω_1, several useful formulas can be gained involving the damping factor:

$$\zeta = \frac{\omega_o}{2\omega_n} \quad \text{or} \quad \omega_1 = \frac{\omega_o}{(2\zeta)^2} \quad (8)$$

By placing ω_1 relative to ω_0, it is possible to adjust the damping factor at will. For example, selecting critical damping would fix ω_1 at half of ω_0. A typical choice for ω_2 is about 4 to 8 times the value of ω_1. (For designers willing to experiment with the AD897 PLL, source code for a short BASIC program, "PLL filter calculations," is available from the author.)

In PLLs based on digital signals, the phase detector/charge pump presents a new current value to the loop filter only when it receives a data pulse. Between these updates, the charge pump is usually tristated (i.e., not supplying any current to the filter). PLL performance is greatly degraded if this hold mode in not efficient. Any current leakage produces a voltage change on the VCO control node. Resulting frequency drift causes pattern-dependent phase errors at the next update cycle, decreasing the error margin.

To more closely model real PLL operation, this complex frequency description must be translated into the z-domain. For practical circuits, an impulse-invariant transformation is justified.[3] Performing this transformation for a fourth-order loop, where the finite bandwidth of the voltage to current converter found in the VCO is considered with an additional pole at ω_3, the open-loop gain is:



2. Open-loop gain of a third-order phase-locked loop is closely followed by the closed-loop-gain characteristics below 0 dB (ω_0).

$$G(s) = \frac{K}{sC_1} \frac{1 + \frac{s}{\omega_1}}{1 + \frac{s}{\omega_2}} \frac{1}{1 + \frac{s}{\omega_3}} \frac{1}{s} \quad (9)$$

The next step in the PLL model development is to derive the partial fraction form of the above expression:

$$G(s) = KR_1 \left(\frac{\alpha}{s} + \frac{\omega_1}{s^2} - \frac{1}{\beta} \frac{1}{s + \omega_2} + \frac{\alpha\omega_2}{\beta\omega_3} \frac{1}{s + \omega_3}\right) \quad (10)$$

where:

$$\alpha = 1 - \frac{\omega_1}{\omega_3} \quad \beta = 1 - \frac{\omega_2}{\omega_3}$$

$$\gamma = \omega_1 T_s \quad (11)$$

The impulse-invariant z-transform with T_s sampling period becomes:

$$G(z) = \frac{\omega_0}{f_{vco}} \left(\frac{\alpha z}{z - 1} + \frac{\gamma z}{(z - 1)^2} - \frac{1}{\beta} \frac{z}{z - a} + \frac{\alpha\omega_2}{\beta\omega_3} \frac{z}{z - b}\right) \quad (12)$$

with the new constant defined as:

$$a = \exp(-\omega_2 T_s) \quad (13)$$

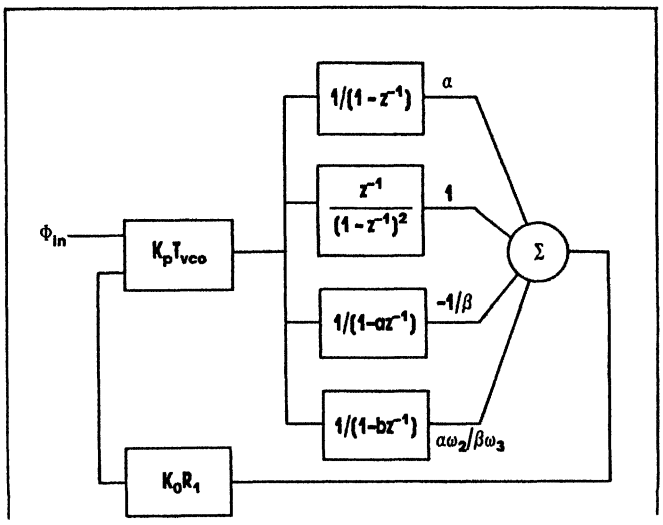
$$b = \exp(-\omega_3 T_s) \quad (14)$$

When ω_3 approaches ω_2, the third and fourth terms in the partial-fraction form derived for G(s) cancel. In this case, the original open-loop gain becomes:

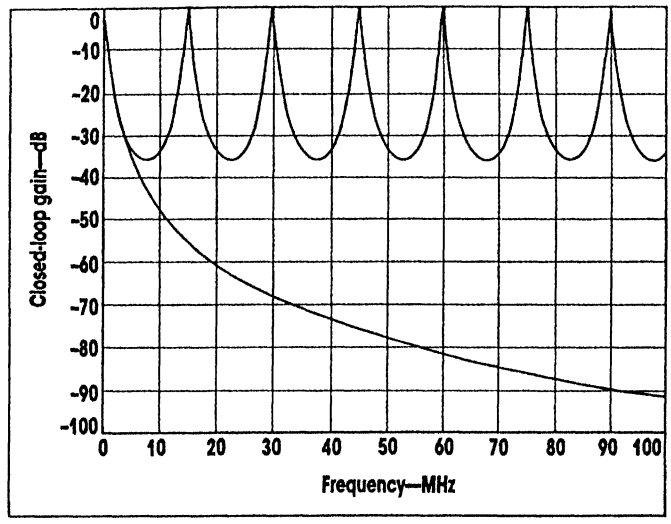
$$G(s) = \frac{K}{sC_1} \frac{1 + \frac{s}{\omega_1}}{\left(1 + \frac{s}{\omega_2}\right)^2} \frac{1}{s} \quad (15)$$

and the partial-fraction form changes to:

$$G(s) = \omega_o \left(\frac{1}{s} + \frac{\omega_1}{s^2} - \frac{1}{s + \omega_2} - \frac{\omega_2}{(s + \omega_2)^2}\right) \quad (16)$$



3. A PLL can be represented in the z-domain, with discrete phase-error updates.



4. Closed-loop gain is plotted in the z- and s-domains, as a function of frequency, for a data rate of 15 MHz.

Assuming that $\omega_1 < \omega_2 < \omega_3$ normally holds for practical systems, the frequency-domain response can be simulated. The PLL can be represented by a weighted sum of parallel and series combinations (Fig. 3) of blocks that implement the various terms of G(z), such as:

$$g(z)_i = \frac{1}{1 - b_i z^{-1}} \qquad (17)$$

During AC analysis, the transfer functions are evaluated by setting $z = e^{j2\pi f/f_s}$. It should be noted that $f_s = f_{data}$; that is, the sampling rate is determined by the reference frequency supplied by the input data.[4] The consequences of switching to discrete time representation can be measured by comparing the amplitude transfer characteristics gained in the s- and z-domains.

The difference between the continuous and sampled time-domain results becomes even more obvious when examining a PLL's closed-loop response (Fig. 4). The act of sampling continuous signals gives rise to aliasing.[5] In the case of a lowpass transfer characteristic, spectral components of signals (including noise) at frequencies higher than one-half the sampling rate might be mirrored to the passband. The repetitive peaks of the PLL closed-loop z-domain response of $f_{data} \pm f_{3dB} \pm 2f_{data} \pm f_{3dB}$, $3f_{data} \pm f_{3dB}$, and so on, will be aliased to the passband unless they are removed by filtering.

Although the input signal spectrum is limited by lowpass filtering in the read channel, the bandwidth is typically set to $f_{max} = f_{vco}/2$, which overlaps quite a few peaks at higher f_{vco}/f_{data} ratios. Aliasing should be a concern when evaluating the noise performance of these PLLs. The periodic characteristic of the transfer function also adversely effects high-frequency rolloff.

Until now, the integrator that represents the VCO in this PLL model has been included with the remainder of the loop's frequency-dependent element. However, the VCO must also be closely examined within this model to ensure accurate loop simulations.

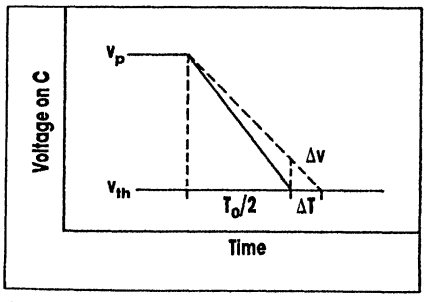
In a monolithic PLL circuit (the Analog Devices AD897) with two cross-coupled voltage-controlled ramp generators that provide 50 percent duty-cycle output voltages, the voltage ramp is generated by a voltage-controlled current source with a transconductance of g_m discharging a capacitor C. When the voltage on the capacitor becomes less than a preset threshold value, v_{th}, a flip-flop is tripped and resets the voltage on the capacitor to a maximum value v_p, thus energizing the other ramp generator. Neglecting the propagation delay in the comparator and flip-flop, the capacitor is discharged over a period of $T_0/2$, determined by the allowed voltage swing:

$$v_p - v_{th} = \frac{v_{ino} g_m}{C} \frac{T_o}{2} \qquad (18)$$

The oscillator frequency is:

$$\omega_{osc} = 2\pi \frac{g_m}{C} \frac{v_{ino}}{2(v_p - v_{th})} = K_o v_{ino} \qquad (19)$$

where v_{ino} is the DC value of the voltage applied to the voltage-controlled current source or voltage-to-current (V/I) converter. The VCO period changes by ΔT when v_{in} changes only slightly relative to v_{ino}, due to a Δv voltage change on the capacitor (Fig. 5):



5. A VCO's period changes with control voltage; for small changes, $\Delta T << T_0$.

$$\Delta v = \frac{g_m}{sC} v_{in} \qquad (20)$$

with

$$\frac{\Delta v}{v_p - v_{th}} = 2\frac{\Delta T}{T_o} \qquad (21)$$

The change in phase represented in Eq. 21 can be written as:

$$\Delta\phi = \frac{2\pi g_m}{2(v_p - v_{th})C} \frac{v_{in}}{s} = K_o \frac{v_{in}}{s} \qquad (22)$$

With a time-interval analyzer (TIA), it is possible to check how the VCO period changes with the frequency of excitation by forcing a sinusoidal voltage from a signal generator onto the VCO input and measuring the period of the oscillator output. The TIA will give a distribution with the RMS value of ΔT. Based on the above result for $v_{in} = V_{in}\sin(\omega_{in}t)$,

$$\Delta T = \frac{1}{v_{ino}} \int v_{in} dt = -\frac{1}{\omega_{in}} \frac{V_{in}}{v_{ino}} \cos\omega_{in} t \qquad (23)$$

that is, RMS(ΔT) should decrease with ω_{in}.

Measurements of the AD897 (Table 1) indicate flat response to $f_3 =$ 12 MHz where the V/I converter's finite bandwidth affects the response. If ΔT (or $\Delta\phi$ since they differ only in a fixed multiplier at a given VCO frequency) is independent of the excitation frequency, the VCO obviously can not be considered as an integrator.

The problem lies in the crudeness

of the PLL model. Although the oscillator's operation is based on the interaction of the two ramp generators, the model does not account for the synchronous resetting action. It can be included by adding a parallel feed-forward path to the output where Δv can be measured (Fig. 6). The z^{-1} block samples the voltage on the capacitor at a rate of $f_s = f_{vco}$, which means that the voltage at the time of sampling is subtracted from the output node, resetting it to zero. Now, $\Delta\phi$ becomes:

$$\Delta\phi = K_o \frac{v_{in}}{s}(1 - z^{-1}) \qquad (24)$$

Since the term $1\text{-}z^{-1}$ adds a differentiating characteristic to the overall gain at lower frequencies, the VCO gain is flat at lower frequencies, then starts rolling off. In the AD897, the V/I converter has an exponential transfer function:

$$\omega_{vco} = \frac{2\pi}{T_o} e^{-k_{exp} v_{in}} \qquad (25)$$

and the VCO gain becomes $K_o = k_{exp}\omega_{vco}$ (i.e., it increases with frequency). The consequence is a constant fractional loop bandwidth, since $G(s)/\omega_{vco} = K_p k_{exp} R_1$. By rewriting ΔT as

$$\Delta T = \Delta\phi \frac{T_o}{2\pi}$$

$$= K_{exp} \frac{v_{in}}{s}(1 - z^{-1}) \qquad (26)$$

ΔT can be plotted as a function of frequency. Besides shifting the flat portion of the ΔT curve with the VCO frequency, the periodic nature of $1\text{-}z^{-1}$ and the fact that it peaks at multiples of one-half the sampling frequency is reflected in the oscillator's noise transfer function. This can be characterized by injecting Δv directly into the loop circuit.[6] TIA measurements will show that jitter introduced by this method changes with frequency and is periodic with the trigger or sampling frequency.

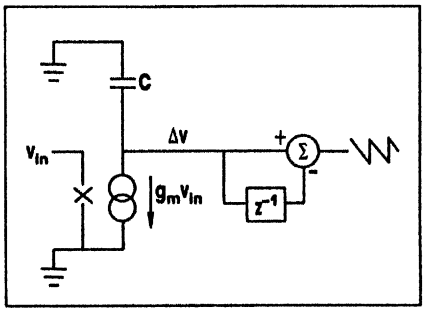
The PLL model also must include the phase detection process. Digital phase detectors apply a control signal to the charge pump, commanding it to either source or sink current. The time of pumping charge into or out of the loop filter is equal the phase error t_ϕ measured between the input data edge and the center of the window. The window size is defined by the clock signal period present at the phase detector input; without any divider in the feedback path, it is equal to the VCO period T_{vco}.

The phase-error measurement reference center of the window is marked by one of the edges of the clock. The phase measurement operation is triggered by the arrival of a data edge. Let $t_\phi = t_{\phi i}$ be the phase error at the ith data pulse. If the data pulses are N clock cycles apart, the phase error $t_{\phi(t=1)}$ can be calculated as:

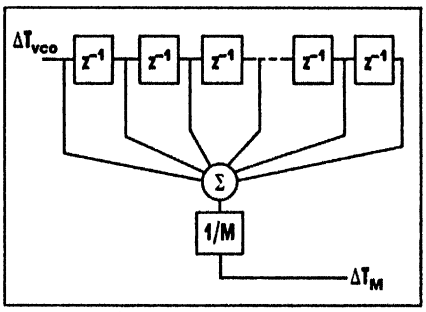
$$t_{\phi(i+1)} = \sum_{j=1}^{N} (T_o + \Delta T_j) - NT_o + t_{\phi i}$$

$$= \sum_{j=1}^{N} \Delta T_j + t_{\phi i} \qquad (27)$$

In the discrete time domain, this accumulation of ΔT_j can be implemented with an adder where one input is receiving ΔT_j while the other is the value of the adder's output delayed by one clock cycle. The block just described implements an integrator in the z-domain and its output is the position of the clock edge relative to some starting point. Phase error $t_{\phi i}$ is the difference between this reference $t_{\phi 0}$ and the accumulated clock position, and is sampled at the update rate.



6. This VCO representation incorporates a parallel feed-forward path to measure Δv.



7. In a divide-by-M model, the counter averages VCO jitter over M clock periods.

The phase detection process, described in this manner, brings the integrator back into the loop. By separating the loop's elements in this way, it is possible to consider the effect of introducing dividers into the PLL's feed-forward and feed-back paths.

Classical theory suggests that adding a frequency divider between the VCO and phase detector will result in increasing VCO jitter as the division ratio M increases. Measured data contradict this prediction, however.

A divide-by-M counter can be modeled by an M-bit-long shift register, which in the z-domain is equivalent to cascading M-1 unit delay blocks (Fig. 7). Since integration of the VCO period change ΔT is performed on a cycle basis during the phase detection process, the jitter at the output of the M divider can be written as:

$$\Delta T_M = \frac{\Delta T_{vco}}{M} \sum_{m=0}^{M-1} z^{-m} \qquad (28)$$

It is clear by now that the divider implements a moving average filter, with no averaging at the lower frequencies.

The full PLL can now be built by assembling z-domain building blocks (Fig. 8). The phase detector is sampled at the data rate while other components run at the VCO frequency. The VCO frequency is actually an integer multiple of the data frequency, i.e.:

$$f_{vco} = df_{data} = [2,8]f_{data} \qquad (29)$$

The different sampling rates automatically take care of the question of how the phase detector/charge pump gain changes with frequency. In frequency synthesizer applications, there is a divider in the feed-forward path so that the VCO output frequency becomes:

$$f_{vco} = \frac{M}{N} f_{ref} \quad (30)$$

and the data frequency becomes:

$$f_{data} = \frac{f_{ref}}{N} \quad (31)$$

which should be the data sampling rate for the phase detector.

How does noise in the VCO effect overall jitter when the oscillator is followed by a divider stage? Since ΔT_M is given by a geometric series:

$$\Delta T_M = \frac{\Delta T_{vco}}{M} \sum_{m=0}^{M-1} z^{-m}$$

$$= \frac{\Delta T_{vco}}{M} \frac{1 - z^{-M}}{1 - z^{-1}} \quad (32)$$

When forcing Δv in the VCO, the transfer function becomes:

$$\frac{\Delta T_M}{\Delta v} = \frac{1}{M} \frac{1 - z^{-M}}{1 - z^{-1}} (1 - z^{-1}) \quad (33)$$

or

$$\frac{\Delta T_M}{\Delta v} = \frac{1}{M} (1 - z^{-M}) \quad (34)$$

The resulting amplitude curve is very much like the original noise transfer function in the VCO, but with z^{-1} replaced by z^{-M} and the gain reduced by a factor M. For $\Delta v=1$, the peak amplitude in dB-s is:

$$\Delta T_{M,\,dB} = 20\log\frac{2}{M} \quad (35)$$

Varying data frequency can also affect open-loop gain. The average charge pumped into the loop filter depends on $d=f_{vco}/f_{data}$, the number of VCO cycles between data updates. In traditional s-domain models, this is considered by including d in the formula for the charge pump gain:

$$K_p = \frac{d}{d\phi} i = \frac{I_p}{2\pi} \frac{1}{d}$$

$$= \frac{I_p}{2\pi} \frac{f_{data}}{f_{vco}} \quad (36)$$

After replacing the filter components with their discrete time equivalent, the update rate is reflected in the sampling frequency of the z-domain network, $f_s=f_{data}$. The charge pump gain is

$$K_p = \frac{d}{d\phi} i = \frac{I_p}{2\pi} \frac{1}{f_{vco}} \quad (37)$$

In the final multi-rate sampled data system, the same thing happens to f_{vco}. The charge pump gain definition is changed to reflect time-domain operation, since the phase detector that measures the phase error in seconds is:

$$K_p = \frac{d}{dt} i = I_p \quad (38)$$

Variations in the open-loop gain with the update rate should be carefully considered as a potential source of closed-loop instability. Under normal circumstances, the open-loop crossover frequency ω_0 should fall above the loop-filter zero ω_1 since the damping factor is:

$$\zeta = \frac{1}{2} \sqrt{\frac{\omega_0}{\omega_1}} \quad (39)$$

The loop-filter component values fix the position of ω_1 (i.e. it is independent of the data frequency, whereas ω_0 is not). Increasing the spacing between updates will decrease the open-loop gain, moving ω_0 closer to ω_1 and resulting in less and less damping. This leads to oscillatory transient behavior under closed-loop conditions.

Introducing a feedback divider with division ratio M is equivalent to adding a moving average filter between the VCO and phase detector. The transfer function of an M-point filter is:

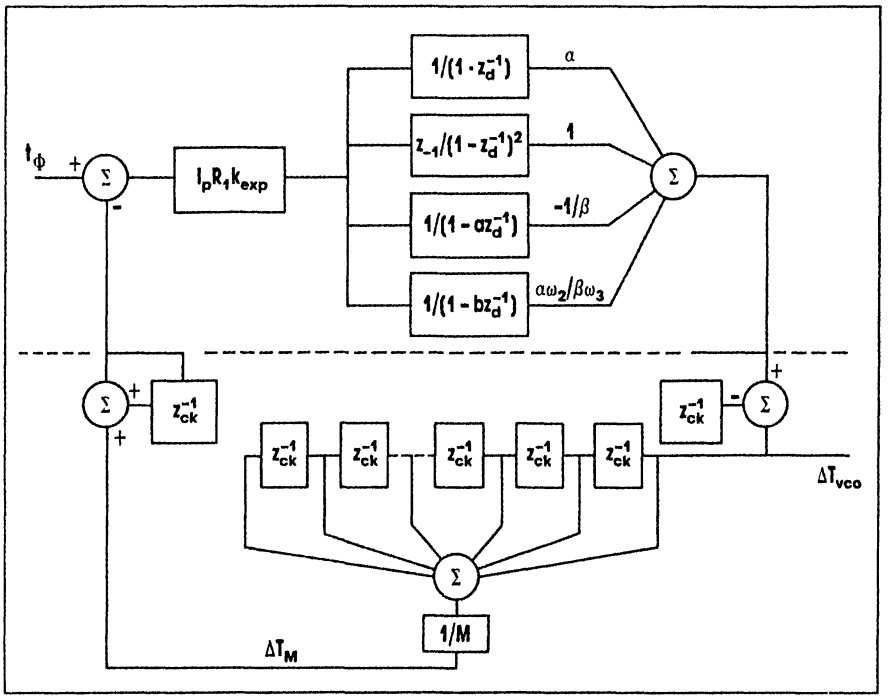
$$H(f) = \frac{\sin\left(\pi M \frac{f}{f_s}\right)}{M \sin\left(\pi \frac{f}{f_s}\right)} \quad (40)$$

Besides the additional low-frequency rolloff in the open-loop transfer function, the filter creates an extra delay in the loop since its group delay is:

$$t_{delay} = \frac{M - 1}{2} T_o \quad (41)$$

where T_0 is the VCO period.

High values of division factor M create the potential for unstable closed-loop behavior as the rate of change around the crossover point in the open-loop transfer function gets close to 40 dB/decade. These problems can be avoided by choosing ω_1 as low as possible, with the tradeoff being poor transient response. ••



8. A complete z-domain PLL model includes two domains with different sampling rates (separated by the dotted line).

■

References

1. F. Gardner, *Phase Lock Techniques, 2nd ed*, John Wiley and Sons, NY, 1981.
2. U. Rohde, *Digital PLL Frequency Synthesizers, Theory and Design*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1983.
3. J. Hein and J. Scott, "z-Domain Model for Discrete Time PLLs," *IEEE Transactions on Circuits and Systems*, Vol. 35, No. 11.
4. G. Gutierrez and D. DeSimone, "An Integrated PLL Clock Generator for 275 MHz Graphic Displays," IEEE 1990 Custom Integrated Conference.
5. R. Higgins, "Digital Signal Processing in VLSI," *Analog Devices Technical Reference Books*, Prentice Hall, Englewood Cliffs, NJ, 1990.
6. "Programmable Time Delay Generation Techniques with the AD9500 Programmable Delay Generator," Analog Devices Application Notes.

Properties of Frequency Difference Detectors

FLOYD M. GARDNER, FELLOW, IEEE

Abstract—**Among other applications, frequency-tracking loops are employed in digital-data receivers, either as a frequency-acquisition aid for phase-locked coherent reception, or as the sole carrier-frequency control for noncoherent reception. This article provides details of design and performance of the frequency-difference detector that lies at the heart of the loop.**

I. Introduction

RECEIVERS of modulated-carrier data signals sometimes contain frequency tracking loops. Reasons for employing such loops include the following.

- Frequency must be adjusted accurately for differentially coherent or incoherent reception.
- A Costas or other phase-locked loop for coherent reception may require frequency aiding for acquisition purposes.

A frequency-tracking loop, in its simplest form, is arranged as shown in Fig. 1. The loop consists of a frequency-difference detector (FDD), a loop filter, and a voltage-controlled oscillator (VCO).

Such a loop strongly resembles automatic frequency control (AFC) loops, which have been known for many years (e.g., [1], [2]). A conventional AFC loop employs a frequency discriminator, which relies upon a passive tuned circuit to furnish the frequency reference. By contrast, the FDD uses a local oscillator to furnish the frequency reference.

This paper examines the characteristics of a particular type of FDD in some detail. Features covered include: circuit principles; response to signals, with emphasis upon data signals; and effects of noise. Several properties important to satisfactory designs are brought to light.

Only analog implementations are treated here. Digital implementations are also of considerable interest and may be found in [3] and [4].

II. Operating Principles

Simple Quadricorrelator

The best known FDD is the quadricorrelator of Fig. 2. It was first presented by Shaeffer [5] but given its name and described in some detail by Richman [6]. The quadricorrelator (and some modification thereof) is the only FDD considered in this paper.

A pair of mixers are used to convert the input passband signal into the corresponding in-phase and quadrature baseband components. For analysis purposes, the mixers are represented as ideal multipliers, but physical circuits could be switching devices, with no alteration in the results [7, p. 108]. Transfer gain of the mixer is the dimensionless factor K_m.

Outputs from the mixers consist of sum and difference frequency products between the input signal and the local oscillator. Low-pass arm filters following the mixers suppress the sum frequency and pass the difference frequency. For purposes of the immediate analysis, it will be assumed that the difference-frequency component is passed without attenuation or phase shift. That assumption is removed later.

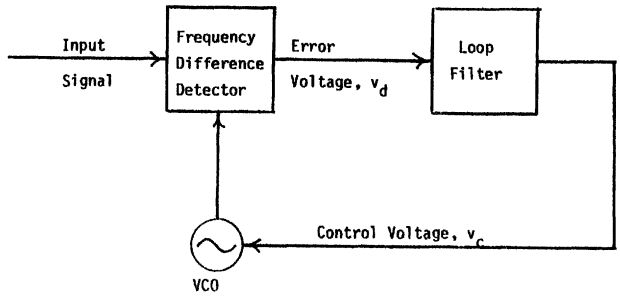


Fig. 1. Frequency-tracking loop, simplified block diagram.

In order for the quadricorrelator to operate, the difference frequency between input signal and local oscillator must fall within the passband of the arm filters. A signal that is well outside of passband will be attenuated by the filter selectivity. Bandwidth of the arm filter therefore provides a rough estimate of the capture range of a frequency tracking loop.

Output of one arm filter (the I-arm in Fig. 2) is differentiated. Time constant (or "gain") of the differentiator is T_d seconds.

A perfect differentiator is assumed for ease of analysis. One can object that a perfect differentiator cannot be built and that the analysis is therefore unrealistic. In fact, perfect relative differentiation is readily obtainable and so the analysis is quite realistic. The term "relative differentiation" is explained as follows.

Let one arm filter have a low-pass filter transfer function denoted as $H_a(s)$. Let the other arm filter have a transfer function $sH_a(s)$, which is always physically realizable, provided only that $H_a(s)$ is low-pass. Then if the same signal should be applied to both filters, the output from the $sH_a(s)$ filter is exactly the derivative of the output of the $H_a(s)$ filter. Therefore, perfect relative differentiation can be achieved. Outputs of the arm filters are designated $v_I(t)$ and $v_Q(t)$.

Differentiation of the I-channel produces $T_d\dot{v}_I(t)$. Arm outputs are multiplied in the third multiplier to produce

$$v_d(t) = K_3 T_d v_Q(t)\dot{v}_I(t) \tag{1}$$

where K_3 is the gain of the third multiplier and has dimensions of $(\text{volts})^{-1}$.

To gain insight into the behavior of a quadricorrelator, let the input signal be a simple sinusoid

$$v_{in}(t) = V_s \cos(\omega_i t + \theta_i) \tag{2}$$

where θ_i is an arbitrary, time-invariant phase angle. Defining frequency error as $\Delta\omega = \omega_i - \omega_0$, where ω_0 is the radian frequency of the reference signals, the arm-filter outputs are calculated to be

$$\begin{aligned} v_I(t) &= K_m V_s \cos(\Delta\omega t + \theta_i) \\ v_Q(t) &= K_m V_s \sin(\Delta\omega t + \theta_i). \end{aligned} \tag{3}$$

Paper approved by the Editor for Communication Theory of the IEEE Communications Society for publication without oral presentation. Manuscript received April 3, 1984. This work was supported by a contract from the European Space Agency, Noordwijk, The Netherlands.

The author is a Consulting Engineer at 1755 University Avenue, Palo Alto, CA 94301.

Reprinted from *IEEE Trans. Comm.*, vol. COM-33, pp. 131-138, February 1985.

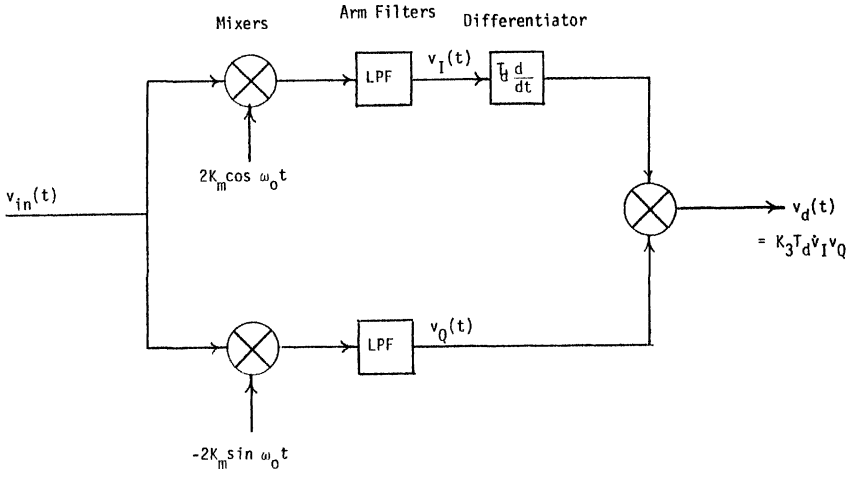


Fig. 2. Quadricorrelator.

Differentiating and multiplying, the output of the third multiplier is found as

$$v_d(t) = -\tfrac{1}{2}\Delta\omega T_d K_m{}^2 K_3 V_s{}^2 [1 - \cos(2\Delta\omega t + 2\theta_i)]. \quad (4)$$

There are two components in this product: a dc component proportional to the frequency difference (including its sign) and a ripple component at double the difference frequency. Notice that the phase θ_i appears only in the ripple and not in the dc component.

The dc component is the useful output; it could be used for FM demodulation or as the error signal in a frequency-tracking loop. The ripple, which has the same peak amplitude as the dc component, can be a major nuisance if the circuit is used to recover modulation or if the frequency tracker is the only frequency-control element. A method of cancelling ripple will be presented below.

There are numerous variations on the basic quadricorrelator. One of them is to recognize that the differentiator need not be perfect; indeed, it is not even necessary that the box labeled "differentiator" be a high-pass network.

To see how this can be, let v_I and v_Q be as shown in (3) and let the differentiator be replaced by a selective network with transfer function $H_d(f) = A(f) \exp(j\phi(f))$, where A is the amplitude and ϕ is the phase shift of the frequency response of the network. Denoting $\Delta f = \Delta\omega/2\pi$, the output of this network will be

$$\tilde{v}_I(t) = K_m V_s A(\Delta f) \cos(\Delta\omega t + \theta_i + \phi(\Delta f)). \quad (5)$$

(The tilde denotes a filtered signal.)

The output of the quadricorrelator is

$$\begin{aligned} v_d(t) &= K_3 v_Q \tilde{v}_I \\ &= -\tfrac{1}{2} K_3 K_m{}^2 V_s{}^2 A(\Delta f)[\sin\phi(\Delta f) \\ &\quad - \sin\{2\Delta\omega t + 2\theta_i - \phi(\Delta f)\}]. \end{aligned} \quad (6)$$

For any physically realizable transfer function $H_d(f)$, the amplitude $A(f)$ is an even function of frequency while the phase $\phi(f)$ is an odd function. Therefore, the dc component, proportional to $\sin\phi(\Delta f)$, reverses polarity as the difference frequency passes through zero, as is required for a frequency detector. This null at zero frequency difference occurs for any filter H_d whatsoever; it is a property of the quadricorrelator and not of the filter.

Frequency-difference information is provided by that portion of v_I that is rotated into phase with v_Q by the network $H_d(f)$. Since v_I and v_Q are generated 90° out of phase, the most effective phase shift is 90°. A perfect differentiator provides 90° phase shift at all frequencies.

Amplitude of the dc component is proportional to $\sin\phi$, whereas amplitude of the ripple is independent of ϕ. To obtain the largest possible dc component relative to ripple requires $\phi = 90°$

There are numerous different networks that might be used for H_d; some examples include:

- a differentiator: $H_d(s) = sT_d$
- a high-pass filter: $H_d(s) = s^n H_L(s)$, where $H_L(s)$ is a low-pass filter
- a low-pass filter: $H_d(s) = H_L(s)$ (for example, [3], [8])
- a delay line: $H_d(s) = \exp(-sT_d)$ (for example, [3])
- a delay-differencing network (approximating a differentiator): $H_d(s) = 1 - \exp(-sT_d)$.

As pointed out in [8], if the phase shift of H_d becomes excessive, then $\sin\phi$ reverses sign and the FDD output has the wrong polarity for controlling frequency of the VCO. The frequency error would be increased instead of decreased by action of the tracking loop.

If phase shift exceeds 270°, there will be one or more points of false lock where the loop comes to equilibrium at a frequency error other than zero. This phenomenon is similar to the false lock sometimes encountered in phase-locked loops [7, ch. 8].

A perfect differentiator will be assumed for H_d in the remainder of this analysis.

Balanced Quadricorrelator

Ripple may be cancelled by the balanced quadricorrelator of Fig. 3. This is a single-sideband cancellation scheme where the double-frequency component (the ripple) is cancelled and the zero-frequency components (the desired error signal) add together.

The balanced circuit, or variations thereon, have appeared previously in [8] (using low-pass filters instead of differentiators), [3], and [9], and has been mentioned but not pursued in [10].

Straightforward analysis shows that if the sinusoidal signal of (2) is applied to the balanced quadricorrelator of Fig. 3, the output voltage will be

$$v_d(t) = -\Delta\omega K_3 T_d K_m{}^2 V_s{}^2. \quad (7)$$

Ripple is gone. Moreover, the phase θ_i of the input does not appear in the output expression. Therefore, time-invariant input phases will be omitted from further consideration.

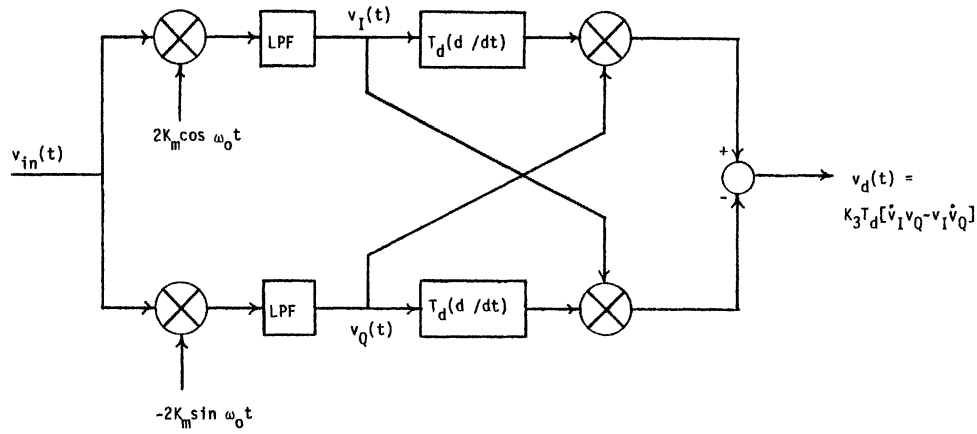


Fig. 3. Balanced quadricorrelator.

The remainder of this paper deals exclusively with balanced quadricorrelators.

III. Response to Signals and Noise

Bandpass Inputs

Let the input be

$$v_{in}(t) = x(t) \cos \omega_i t - y(t) \sin \omega_i t. \tag{8}$$

All bandpass signals can be reduced to this format. The coefficients $x(t)$ and $y(t)$ are real variables but no constraints are placed, yet, on their stationarity, cross correlation, or spectra.

To simplify the discussion, it will be assumed that the arm filters merely suppress sum-frequency components of the mixer outputs and do not affect the modulation components $x(t)$ and $y(t)$. Arm filter functions can be translated to a fictitious passband filter located before the mixers. Then $v_{in}(t)$ can be regarded as the output of the passband filter. This expedient permits filter effects to be included without cluttering the descriptions to follow.

If this bandpass signal is applied to the balanced quadricorrelator of Fig. 3, the output is found to be

$$v_d(t) = -K_3 T_d K_m{}^2 [\Delta\omega(x^2 + y^2) + x\dot{y} - y\dot{x}] \tag{9}$$

where the time argument has been suppressed for notational convenience. Some special examples give insight into the implications of this expression.

First, as a shorthand convenience, define

$$K_q = K_3 T_d K_m{}^2. \tag{10}$$

DSB-AM: Let $y(t) = 0$ for all t. Binary PSK is one possible example of such a signal. Then (9) reduces to

$$v_d(t) = -K_q x^2(t) \Delta\omega. \tag{11}$$

We recognize $x^2(t)$ as the squared envelope of the input signal (More generally, in (9) the squared envelope is $x^2 + y^2$.)

Mean value of the output for DSB-AM is

$$\text{Avg } [v_d] \triangleq V_d = -K_q \Delta\omega \sigma_x{}^2 \tag{12}$$

where $\sigma_x{}^2 \triangleq$ Avg $[x^2(t)]$. (We write Avg [ ] instead of statistical expectation $E[\ ]$ to allow for time averaging as well as ensemble averaging.)

Two features are apparent from (11).

- There is no additive self-noise generated for any DSB-AM signal, no matter what the form of $x(t)$. (Self-noise is explained in [11].)
- Error voltage $v_d(t)$ is modulated by the signal envelope. When the FDD is used in a feedback frequency-tracking loop, the loop gain will fluctuate with fluctuations in the envelope.

Uncorrelated Channels: Let $E[x(t_1)y(t_2)] = 0$ for all t_1 and t_2. A band-limited signal with this property has a spectrum that is symmetric about ω_i in the passband. It can be shown that Avg $[x\dot{y}] = 0 =$ Avg $[y\dot{x}]$ and so the FDD mean output is

$$V_d = -K_q \Delta\omega(\sigma_x{}^2 + \sigma_y{}^2). \tag{13}$$

This equation has been derived previously in [10] under slightly more restrictive conditions.

PAM-QAM Signal: Let x and y be synchronous PAM data streams of the form

$$\begin{aligned} x(t) &= \Sigma a_n g(t - nT) \\ y(t) &= \Sigma b_n g(t - nT) \end{aligned} \tag{14}$$

where $g(t)$ is a standard signaling pulse and the $\{a_n\}$ and $\{b_n\}$ are multilevel data sequences. Impose the special, but often-encountered conditions that

$$\begin{aligned} E[a_n] &= 0 = E[b_n] \\ E[a_m a_n] &= \sigma_a{}^2 \delta_{mn} \\ E[b_m b_n] &= \sigma_b{}^2 \delta_{mn} \\ E[a_n b_m] &= 0. \end{aligned} \tag{15}$$

Output of the balanced quadricorrelator for such an input is

$$\begin{aligned} v_d(t) = -K_q \Bigg[\Delta\omega \Bigg\{ \sum_n \sum_m (a_n a_m + b_n b_m) \\ \cdot g(t - nT) g(t - mT) \Bigg\} \Bigg] \\ - K_q \Bigg[\sum_n \sum_m a_n b_m \{ g(t - nT) \dot{g}(t - mT) \\ - g(t - mT) \dot{g}(t - nT) \} \Bigg]. \end{aligned} \tag{16}$$

Taking statistical expectation and applying the conditions of (15) yields

$$E[v_d(t)] = -K_q \Delta\omega(\sigma_a{}^2 + \sigma_b{}^2) \sum_n g^2(t - nT). \tag{17}$$

Only the first bracketed term of (16) contributes useful average output; the second bracketed term has zero mean and only contributes pattern (or self) noise [11].

Expectation of the first term, as given by (17), is nonstationary—in fact, periodic in T. That is not surprising inasmuch as the input signal (8) and (14) is cyclostationary [12], [13]. Useful output of the circuit is the dc value, which is found by time-averaging (17) over one period, to obtain

$$V_d = -K_q\Delta\omega(\sigma_a^2 + \sigma_b^2)(1/T)\sum_n \int_0^T g^2(t - nT)\,dt$$

$$= -K_q\Delta\omega(\sigma_a^2 + \sigma_b^2)(1/T)\int_{-\infty}^{\infty} g^2(t)\,dt. \tag{18}$$

But the integral is just the energy of the pulse $g(t)$ (dissipated in a 1 Ω resistor):

$$E_g = \int_{-\infty}^{\infty} g^2(t)\,dt \tag{19}$$

so the useful dc output of the balanced quadricorrelator is

$$V_d = -K_q\Delta\omega(\sigma_a^2 + \sigma_b^2)E_g/T. \tag{20}$$

Consider pattern noise more closely: in particular, the difference of products contained in braces in the second term of (16). For $n = m$, that difference is zero; the double summation can contribute pattern noise only for $m \neq n$.

If the signaling pulse $g(t)$ is time limited to a single interval T (that is, $g(t) = 0$, $t < 0$, and $t > T$) then there is no pattern noise whatever. Pattern noise can arise only if pulses overlap.

The cancellation that is noted arises because of the balanced circuit; an unbalanced quadricorrelator (e.g., Fig. 2) not only does not afford the pattern-noise cancellation, but also contains terms that are products of pattern noise and ripple. This is just one more reason to employ a balanced circuit.

Gaussian Noise

Let the input $v_{in}(t)$ be bandpass Gaussian noise with two-sided spectral density $S_n(f)$. Bandpass Gaussian noise can be expanded in quadrature components about any arbitrary frequency ω_i to obtain

$$v_{in}(t) = n_c(t)\cos\omega_i t - n_s(t)\sin\omega_i t. \tag{21}$$

There is no implication that the spectrum is centered on ω_i or is in any way symmetric about it. Furthermore, there is no requirement that the spectrum $S_n(f)$ be symmetric about any frequency other than zero.

Adapting (8) and (9), the output of the balanced quadricorrelator, for noise input, is found to be

$$v_d(t) = -K_q[\Delta\omega(n_c^2 + n_s^2) + \dot{n}_s n_c - \dot{n}_c n_s]. \tag{22}$$

Taking statistical expectations, the average output voltage is

$$V_d = E[v_d(t)] = -K_q[\Delta\omega E(n_c^2 + n_s^2) + E(\dot{n}_s n_c - \dot{n}_c n_s)]$$
$$= -K_q[2\Delta\omega\sigma_n^2 + E(\dot{n}_s n_c - \dot{n}_c n_s)] \tag{22a}$$

where σ_n^2 is the variance of the input noise $v_{in}(t)$. To proceed further we must evaluate $E(\dot{n}_s n_c - \dot{n}_c n_s)$.

To that end we follow the approach described in [14, sect. 6-4, 8-5]. Consider a finite segment of the noise input of duration T_0. Expand $n_c(t)$ and $n_s(t)$ in Fourier series over this interval to obtain expressions of the form

$$n_{c0}(t) = \sum_{k=1}^{\infty}[x_{ck}\cos(2\pi k/T_0 - \omega_i)t + x_{sk}\sin(2\pi k/T_0 - \omega_i)t]$$

$$n_{s0}(t) = \sum_{k=1}^{\infty}[x_{ck}\sin(2\pi k/T_0 - \omega_i)t - x_{sk}\cos(2\pi k/T_0 - \omega_i)t] \tag{23}$$

where the subscripts "0" in n_{c0} and n_{s0} indicate that the series are valid only in the finite interval T_0.

These expressions are differentiated and substituted into (22). The statistical-expectation operator is next formally applied and the interval T_0 is caused to grow towards infinity while k is constrained so that $k/T_0 = f_k$ holds constant. In the limit, the cross correlations all vanish and the autocorrelations become delta functions, viz.,

$$E[x_{ck}x_{sn}] = 0$$
$$E[x_{ck}x_{cn}] = \delta_{kn}E[x_{ck}^2]$$
$$E[x_{sk}x_{sn}] = \delta_{kn}E[x_{sk}^2]. \tag{24}$$

Therefore, following [14], the expectations of the cross products are

$$E[n_c\dot{n}_s] = -E[\dot{n}_c n_s] = 2\int_0^{\infty}(\omega - \omega_i)S_n(f)\,df \tag{25}$$

wherefore the dc output is

$$V_d = -K_q\left[2\Delta\omega\sigma_n^2 + 4\int_0^{\infty}(\omega - \omega_i)S_n(f)\,df\right]. \tag{22b}$$

But

$$(\omega - \omega_i) = (\omega - \omega_0) + (\omega_0 - \omega_i) = (\omega - \omega_0) - \Delta\omega$$

so

$$4\int_0^{\infty}(\omega - \omega_i)S_n(f)\,df = 4\int_0^{\infty}(\omega - \omega_0)S_n(f)\,df - 2\Delta\omega\sigma_n^2$$

since

$$\sigma_n^2 = 2\int_0^{\infty} S_n(f)\,df.$$

Therefore, the dc output of the balanced quadricorrelator is

$$V_d = -8\pi K_q\int_0^{\infty}(f - f_0)S_n(f)\,df \tag{26}$$

irrespective of the arbitrary input frequency ω_i.

In other words, the average error signal generated by the FDD is proportional to the first moment about the local reference frequency f_0 of the spectrum $S_n(f)$ of the input $v_{in}(t)$. The error signal will be zero only if the center-of-gravity (c.g.) of the input spectrum coincides with the local reference frequency. In that sense the frequency-tracking loop tracks the center-of-gravity of the input spectrum.

Conjecture: Inasmuch as the FDD is incoherent, it seems plausible that it tracks the c.g. of *any* input spectrum, not just that of bandpass Gaussian noise.

Noise Bias

This result has important consequences for circuit design. In a coherent phase-locked loop, additive noise causes phase jitter in the loop, but no bias is generated. However, (26) shows clearly that noise can generate a bias in a frequency-tracking loop, in addition to the fluctuations in tracking. An equipment designer must be able to predict the bias and often wants to be able to avoid it entirely.

Bias will be generated if the c.g. of the noise spectrum does not coincide with the local reference frequency f_0. In many receivers, the noise spectrum is shaped by passing white noise through bandpass filters in intermediate-frequency portions of the receiver. To avoid noise bias, it is sufficient to make those filters symmetric and center them on the reference frequency f_0.

There are two ways to close a tracking loop: a short-loop connection or a long loop. These options are illustrated in Fig. 4.

The bandpass filter (BPF) in the IF portion of the receiver has a fixed characteristic. If the frequency f_0 is allowed to vary, as is necessary in a short loop, then the filter center—and therefore the c.g. of the noise spectrum—cannot coincide with f_0 except by rare accident. In general, a noise bias in tracking must be anticipated whenever a short loop is employed.

By contrast, in a long loop, both f_0 and the filter center-frequency are fixed. Tracking is accomplished by controlling the frequency of an oscillator that precedes the reference source. If the filter and the reference source are properly aligned and stable, then the c.g. of the noise spectrum will always coincide with the reference frequency and no noise bias will be generated.

Signal Plus Noise

In this section, analysis of the response of the balanced quadricorrelator to signal-plus-noise is presented. The results are in the form of the spectrum of the noise output of the FDD. These results can then be applied to a frequency-tracking loop to calculate the fluctuations of tracking error.

Previous workers have analyzed signal-plus-noise from differing standpoints. Pickard [15] dealt with a simple quadricorrelator that had hard limiters in each arm. He performed his analyses entirely in the time domain and did not derive output spectra. Pawula [16] extended Pickard's work, with the same circuit; his paper has references to other predecessors.

Park [9] treated the balanced quadricorrelator and derived output spectra. He was most interested in the use of the FDD as an FM demodulator, whereas the emphasis here is on frequency tracking. A portion of Park's article deals with the balanced quadricorrelator without any limiters—the circuit of greatest interest here. He states that the FM clicks that trouble a conventional FM discriminator below threshold do not arise if there is no limiting. Also, he concludes that the limiterless circuit has better output signal-to-noise ratio if input SNR is very small. (At large input SNR's, the circuit with limiter will be superior because the limiter suppresses AM noise. It is not possible to achieve the FM noise advantage unless the AM noise is somehow removed.)

Cahn [3] performed an analysis of the extra phase fluctuation introduced by the presence of the frequency tracker in a combined phase-frequency tracking loop.

In this analysis we shall be concerned only with fluctuations in a frequency tracking loop. To that end, the input to the FDD will be assumed to be a pure sine wave plus bandpass Gaussian noise.

$$v_{\text{in}}(t) = V_s \cos \omega_i t + n(t). \tag{27}$$

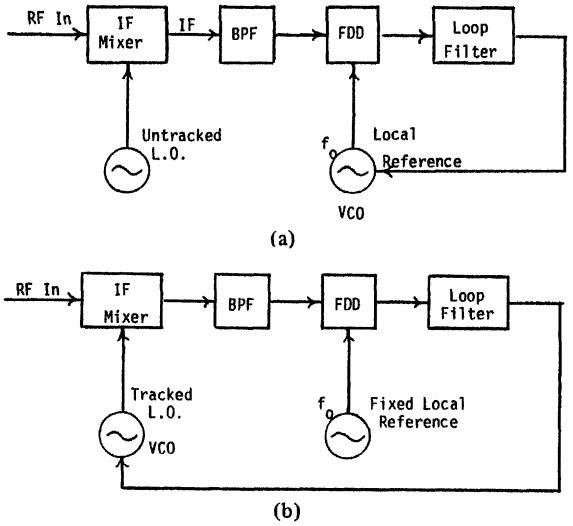


Fig. 4. Frequency-tracker loop connections. (a) Short loop. (b) Long loop.

Obviously, a data signal is not a pure sine wave so the assumed signal is a simplification of reality. The simplification affords better visibility into the operation of the circuit while, it is hoped, providing a useful approximation of performance with real signals.

The noise has a spectrum of $S_n(f)$ (two-sided), which is unrestricted other than being required to be bandpass. (Strictly speaking, the spectrum should be sufficiently band limited that foldover problems do not arise in the mixers of the quadricorrelator.)

The input may be resolved into

$$v_{\text{in}}(t) = (V_s + n_c) \cos \omega_i t - n_s \sin \omega_i t \tag{28}$$

where the time dependence of $n_c(t)$ and $n_s(t)$ has been suppressed for compactness of notation.

Assume that the arm filters remove the double-frequency mixer products, but that the filters are broad enough not to affect n_c or n_s. That is contrary to standard practice but convenient for analysis. The assumption will be removed later.

Under these conditions the filtered outputs of the mixers are

$$\begin{aligned} v_I(t) &= K_m [(V_s + n_c) \cos \Delta\omega t - n_s \sin \Delta\omega t] \\ v_Q(t) &= K_m [(V_s + n_c) \sin \Delta\omega t + n_s \cos \Delta\omega t]. \end{aligned} \tag{29}$$

Performing the balanced-quadricorrelator operations of differentiation, multiplication, and subtraction on these arm voltages gives

$$\begin{aligned} v_d(t) = -K_q [\Delta\omega(V_s^2 + 2V_s n_c + n_c^2 + n_s^2) \\ + V_s \dot{n}_s + n_c \dot{n}_s - \dot{n}_c n_s]. \end{aligned} \tag{30}$$

(This result is the remnant after combining 30 terms of the function $\dot{v}_I v_Q - v_I \dot{v}_Q$. Output of the simple unbalanced quadricorrelator is half of the above, plus 13 distinct ripple terms.)

Average output is

$$\begin{aligned} V_d = E[v_d] = -K_q \Delta\omega V_s^2 \\ -K_q [2\sigma_n^2 \Delta\omega + E(n_c \dot{n}_s - \dot{n}_c n_s)]. \end{aligned} \tag{31}$$

Ordinary statistical expectations (instead of double-averaging as for a QAM input) suffice because the noise is assumed to be stationary. The terms n_c, n_s, $\dot{n}_c$, $\dot{n}_s$ are all zero-mean by the bandpass assumption. Variance of the noise input is σ_n^2.

The terms containing n_c^2 and n_s^2 would superficially appear to contribute to the useful dc output of the FDD. That conclusion is contrary to common sense; the terms should be regarded as an artifact of the particular selection of noise representation. The last term—$E(n_c\dot{n}_s - \dot{n}_c n_s)$—generates a component to cancel the anomaly, as has been demonstrated in (21)-(26).

As is characteristic of quadratic devices, the output in (30) is composed of $S \times S$, $S \times N$, and $N \times N$ terms [14, ch. 12]. Desired frequency-difference information is contained in the $S \times S$ term, which is independent of the noise. There is no signal suppression effect at low SNR, as would arise in a circuit containing a limiter.

Some portion of the noise in (30) is proportional to the frequency difference $\Delta\omega$. Total noise decreases as the frequency error is reduced to zero. This dependence of output noise on frequency deviation has been noted earlier [6], [9].

Frequency error in a successful frequency tracker should be very small so the error-proportional noise should also be small. In the sequel we neglect the error-proportional noise and concentrate on the last three terms of (30). Our objective is to determine their spectrum; to that end we first obtain their autocorrelation

$$\begin{aligned} &E[v_d(t_1)v_d(t_2)] \\ &= K_q^2(V_s^2 E[\dot{n}_{s1}\dot{n}_{s2}] + V_s(E[n_{s1}n_{s2}n_{c2}] - E[\dot{n}_{s1}\dot{n}_{c2}n_{s2}] \\ &\quad + E[n_{c1}\dot{n}_{s1}\dot{n}_{s2}] - E[\dot{n}_{c1}n_{c2}\dot{n}_{s2}]) + E[n_{c1}\dot{n}_{s1}n_{c2}\dot{n}_{s2}] \\ &\quad - E[n_{c1}\dot{n}_{s1}\dot{n}_{c2}n_{s2}] - E[\dot{n}_{c1}n_{s1}n_{c2}\dot{n}_{s2}] \\ &\quad + E[\dot{n}_{c1}n_{s1}\dot{n}_{c2}n_{s2}]). \end{aligned} \tag{32}$$

The subscripts 1 and 2 refer to times t_1 and t_2.

Next, the noise functions are truncated to finite time segments and expanded in Fourier series, as in (23). These functions are substituted into (32), like terms are combined, and the time segment is allowed to approach infinite extent. The details are extremely space consuming, so they are omitted. Underlying principles are identical to those of [14, sect. 6-4, 8-5].

Ultimately, it is found that the expectations of the triple products [second line of (32)] are all zero for Gaussian noise and that the autocorrelation of the remaining terms is given by

$$\begin{aligned} &E[v_{d1}v_{d2}]/K_q^2 \\ &= 2V_s^2 \int_0^\infty (\omega_k - \omega_0)^2 S_n(f_k)\cos(\omega_k - \omega_0)\tau\, df_k \\ &\quad + 8\iint_0^\infty (\omega_p - \omega_0)^2 S_n(f_k)S_n(f_p) \\ &\quad \cdot \cos(\omega_k - \omega_p)\tau\, df_k\, df_p \\ &\quad + 8\iint_0^\infty (\omega_p - \omega_0)(\omega_k - \omega_0)S_n(f_k)S_n(f_p) \\ &\quad \cdot \cos(\omega_k - \omega_p)\tau\, df_k\, df_p + 4\iint_0^\infty (\omega_m - \omega_0) \\ &\quad \cdot (\omega_k - \omega_0)S_n(f_m)S_n(f_k)\, df_k\, df_m \end{aligned} \tag{33}$$

where f_k, f_m, f_p are dummy frequency variables and $\tau = t_2 - t_1$.

The last step is to take the Fourier transform of (33) to find the spectrum of the noise on $v_d(t)$. After additional labor, the two-sided spectrum is found to be

$$\begin{aligned} &S_{vd}(f)/K_q^2 \\ &= 2V_s^2(2\pi f)^2[S_n(f_0 - f) + S_n(f_0 + f)] \\ &\quad + 8\int_0^\infty (\omega_p - \omega_0)^2 S_n(f_p) \\ &\quad \cdot [S_n(f_p - f) + S_n(f_p + f)]\, df_p \\ &\quad + 8\int_0^\infty (\omega_p - \omega_0)S_n(f_p)[(\omega_p - \omega_0 - 2\pi f) \\ &\quad \cdot S_n(f_p - f) + (\omega_p - \omega_0 + 2\pi f)S_n(f_p + f)]\, df_p \\ &\quad + 8\delta(f)\left[\int_0^\infty (\omega_p - \omega_0)S_n(f_p)\, df_p\right]^2. \end{aligned} \tag{34}$$

The first term in the spectrum arises from $S \times N$; the other terms are all $N \times N$. The last term is a dc component that is zero only if the center of gravity of the noise spectrum coincides with f_0.

The broad-band assumption on the arm filters can now be removed. Low-pass arm filters are equivalent to a bandpass filter placed in front of the mixers; this equivalent bandpass filter is always symmetric and always centered on f_0. To take account of an arm filter, simply translate it into the equivalent bandpass IF filter. The noise spectrum $S_n(f)$ then becomes the actual input noise spectrum (often effectively white), filtered by any actual bandpass filters in the receiver and by the equivalent bandpass filter corresponding to the arm filters.

Spectrum Examples

Two particular spectra were investigated in further detail. They are given by

$$\begin{aligned} S_n(f_a) &= N_0/2, \qquad |f_a - f_0| < W/2 \\ &= 0, \qquad |f_a - f_0| > W/2 \end{aligned} \tag{35}$$

and

$$S_n(f_b) = \frac{N_0}{2}\,\frac{\sin^2 \pi(f_b - f_0)/B}{[\pi(f_b - f_0)/B]^2}. \tag{36}$$

The first is a rectangular spectrum of width W, centered at f_0. The second is the spectrum that would be imposed by an integrate-and-dump arm filter on white noise. These spectra are related in the sense that if $W = B$, then the two spectra contain equal noise powers.

Noise spectra at the FDD output are found by substituting into (34) and evaluating the integrals. The results are

$$\begin{aligned} &S_{vd}(f) \\ &= \begin{cases} 2V_s^2(2\pi f)^2 K_q^2 N_0 & (|f| < W/2) \\ 0 & (|f| > W/2) \end{cases} \\ &\quad + \begin{cases} \frac{8}{3}\pi^2 N_0^2 W^3(1 - |f|/W)^3 K_q^2 & (|f| < W) \\ 0 & (|f| > W) \end{cases} \end{aligned} \tag{37}$$

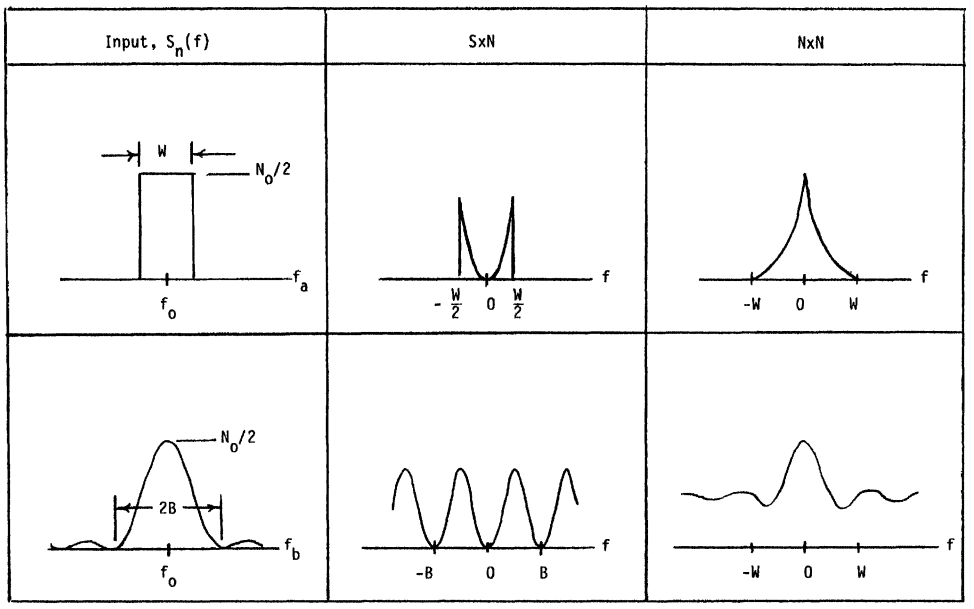


Fig. 5. Spectrum examples.

for the rectangular spectrum and

$$S_{vd}(f) = 8K_q^2 V_s^2 B^2 N_0 \sin^2(\pi f/B) + 8K_q^2 B^3 N_0^2 \left(1 + \frac{\sin 2\pi f/B}{2\pi f/B}\right) \quad (38)$$

for the integrate and dump spectrum. Since both input spectra are symmetric about f_0 there is zero dc component present.

These are two-sided spectra, valid for both positive and negative values of f. Each spectrum consists of two portions: one arises from $S \times N$ and the other from $N \times N$. The $N \times N$ portions are quite different for the two different input spectra, but the $S \times N$ portions are the same for small magnitudes of f. To see that they are the same, approximate $\sin^2(\pi f/B)$ by $(\pi f/B)^2$, which makes the first term of (38) equal to the first term of (37).

At low frequencies at least, the $S \times N$ term is the familiar quadratic-spectrum noise that appears at the output of all FM discriminators.

The various spectra are sketched in Fig. 5. They have been drawn with $B = W$ for ready comparison between the two input shapes. Vertical scales of the output spectral components are not significant.

It is apparent that the $N \times N$ noise has its spectrum concentrated near zero frequency. A narrow-band tracking loop is likely to be more troubled by $N \times N$ noise than by $S \times N$, even for fairly large signal-to-noise ratios.

Modulated Signals

These results have all been obtained for a pure sinusoidal signal. A data signal is not a sinusoid so the noise spectra resulting from a data signal plus noise will be altered from the results shown here. The more difficult analysis needed to accommodate the modulation sidebands of the data signal will not be pursued here. Nonetheless, it is useful to speculate on the noise spectrum that would arise if a data-modulated signal replaced the sinusoid. Some possibilities include the following.

- Pattern-noise components might appear, as discussed earlier.
- The $N \times N$ output components are likely to remain the same inasmuch as they do not depend upon the signal.
- The $S \times S$ components will be those already discussed for QAM signals.
- The $S \times N$ components will be modified substantially. There are two effects that might appear.

i) The $S \times N$ components derived for the sinusoid signal are likely to be spread in frequency by convolving noise against the spread spectrum of the data signal.

ii) Additional $S \times N$ components may be generated (e.g., [9]).

In the absence of better analysis of the noisy data-signal input, an engineering approximation should provide useful guidance to the equipment designer. In two parts, the approximation is as follows.

1) Calculate useful dc error-signal output (thus obtaining the gain of the FDD) and pattern noise from the QAM analysis presented in a previous section.

2) Calculate output noise caused by the additive input noise according to the sinusoidal-signal analysis of the immediately preceding sections.

This approximation is valid only if the $N \times N$ noise is indeed dominant.

IV. Conclusions

This article has examined quadricorrelators used as frequency-difference detectors (FDD). The FDD develops an error signal that is proportional to the frequency difference between an incoming signal and a local reference oscillator.

The simplest quadricorrelator consists of a differentiator (or other, less effective network), an I-Q demodulator, and a baseband multiplier. It suffers from ripple in its output.

A balanced quadricorrelator requires an additional multiplier and differentiator; it cancels ripple and is easier to analyze.

It has been demonstrated that the loop will track the center-of-gravity of the spectrum of certain classes of inputs. It is conjectured that the loop will track the c.g. of any bandpass input spectrum.

A quadricorrelator can be used for frequency tracking of a PAM-QAM data signal. Pattern noise may arise if the data pulses overlap and the modulation is two-dimensional.

If the input noise spectrum is not symmetric on the center frequency of the data signal, the FDD (or any other frequency detector) will develop a noise bias. To avoid bias, use a long loop for tracking and a bandpass filter that is symmetric about the final demodulation frequency f_0.

Noise spectra of the FDD output have been ascertained for a sinusoidal signal plus bandpass, Gaussian noise. These output spectra depend upon filter shaping in the receiver. Two important examples of shaping are given; most practical data-filter shapes will lie between the two examples.

Both $S \times N$ and $N \times N$ noise components are generated. The $N \times N$ component is likely to dominate in a frequency tracker because the $S \times N$ spectrum goes to zero at zero frequency.

References

[1] C. Travis, "Automatic frequency control," *Proc. IRE,* vol. 23, p. 1125, Oct. 1935.

[2] G. H. Nibbe, F. E. Towsley, and E. Durand, "AFC systems and circuits," in *Microwave Receivers,* S. N. van Voorhis, Ed. M.I.T. Radiation Lab. Series, vol. 23, ch. 3.

[3] C. R. Cahn, "Improving frequency acquisition of a Costas loop," *IEEE Trans. Commun.,* vol. COM-25, pp. 1453-1459, Dec. 1977.

[4] F. D. Natali, "AFC tracking algorithms for satellite links," in *Conf. Rec., Int. Conf. Commun.,* Boston, MA, 1983.

[5] C. F. Shaeffer, "The zero-beat method of frequency discrimination," *Proc. IRE,* vol. 30, pp. 365-367, Aug. 1942.

[6] D. Richman, "Color carrier reference phase synchronization accuracy in NTSC color television," *Proc. IRE,* vol. 42, pp. 106-133, Jan. 1954.

[7] F. M. Gardner, *Phaselock Techniques,* 2nd ed. New York: Wiley, 1979.

[8] R. W. D. Booth, "A note on the design of baseband AFC discriminators," in *Conf. Rec. Nat. Telecommun. Conf.,* Houston, TX, 1980, vol. 2, paper 24.2.

[9] J. H. Park, Jr., "An FM detector for low *S/N*," *IEEE Trans. Commun.,* vol. COM-18, pp. 110-118, Apr. 1970.

[10] D. G. Messerschmitt, "Frequency detectors for PLL acquisition in timing and carrier recovery," *IEEE Trans. Commun.,* vol. COM-27, pp. 1288-1295, Sept. 1979.

[11] F. M. Gardner, "Self-noise in synchronizers," *IEEE Trans. Commun.,* vol. COM-28, pp. 1159-1163, Aug. 1980.

[12] W. R. Bennett, "Statistics of regenerative digital transmission," *Bell Syst. Tech. J.,* vol. 37, pp. 1501-1542, Nov. 1958.

[13] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Trans. Inform. Theory,* vol. IT-21, pp. 4-14, Jan. 1975.

[14] W. B. Davenport and W. L. Root, *Random Signals and Noise.* New York: McGraw-Hill, 1958.

[15] T. B. Pickard, "The effect of noise on a method of frequency measurement," *IRE Trans. Inform Theory,* vol. IT-4, pp. 83-88, June 1958.

[16] R. B. Pawula, "Analysis of an estimator of the center frequency of a power spectrum," *IEEE Trans. Inform. Theory,* vol. IT-14, pp. 669-676, Sept. 1978.

Frequency Detectors for PLL Acquisition in Timing and Carrier Recovery

DAVID G. MESSERSCHMITT, SENIOR MEMBER, IEEE

Abstract—A significant problem in phase-locked loop (PLL) timing and carrier extraction is the initial acquisition. Very narrow loop bandwidths are generally required to control phase jitter, and acquisition may depend on an extremely accurate initial VCO frequency (VCXO) or sweeping. We describe two simply implemented frequency detectors which, when added to the traditional phase detector, can effect acquisition even with very small loop bandwidths and large initial frequency offsets.

The first is the quadricorrelator, previously applied to timing recovery by Bellisio, while the second is new, and called a rotational frequency detector. The latter, while limited to lower frequencies and higher signal-to-noise ratios, is suitable for many applications and can be implemented with simpler circuitry.

1.0. INTRODUCTION

THE initial acquisition of a phase-locked loop (PLL) when used for timing or carrier extraction is a significant practical problem, since the narrow loop bandwidth generally required for jitter requirements severely restricts the pull-in range. Methods widely employed to effect acquisition include [1]

a) compromises in loop filter design,
b) highly accurate initial VCO frequency (VCXO),
c) sweeping of the VCO, and
d) in-lock detection with switching of loop filter.

In many instances, as in carrier recovery, several of these methods may be simultaneously employed.

There is a fifth method of effecting acquisition [1], which seems to have been first suggested by Richman [2], and that is to add a frequency detector (FD) to the traditional PLL phase detector (PD) in the manner of Figure 1. With a large initial VCO frequency offset, the PD output has essentially a zero d.c. output, and the FD generates a voltage proportional to the frequency difference between input and VCO, driving that difference to zero. The PD takes over when the frequency difference is small, completing the acquisition. When the PLL is in-lock, the FD output will have at the least zero mean, and optimistically will be identically zero, automatically allowing the PD and its loop filter to govern the loop dynamics. The beauty of this approach is that a crystal controlled VCO (VCXO) can often be exchanged for the additional FD circuitry in timing recovery applications, an advantageous tradeoff in this age of integrated circuitry. In carrier recover, a VCXO is often still required because of the problem of false locking to a data sideband, but the sometimes troublesome in-lock detector and/or sweeping circuitry can be eliminated and the PD loop filter can be designed virtually independently of acquisition considerations, removing a significant burden from the designer.

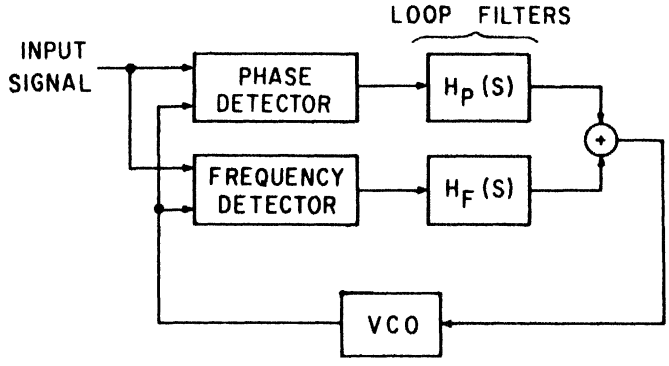


Fig. 1. PLL with Phase and Frequency Detector.

This paper will discuss two specific FD's, each of which is applicable to both timing and carrier recovery. The first is the quadricorrelator described by Richman [2], which was more recently rediscovered by Pickard [3-5], Bellisio [6], and in modified form by Park [7], Cahn [8] and Citta [14]. These authors have discussed its applicability to sinusoid [2, 3] and narrowband Gaussian process [2-4] input signals, to timing recovery [6], and to Costas loop carrier recovery for biphase modulation [8]. We will show here that the quadricorrelator is more generally applicable to carrier recovery for any modulation method which has a power spectrum symmetrical about the carrier frequency. This includes most data modulation methods, with the notable exceptions of single and vestigal sideband modulation.

The second FD, called a rotational FD, is new, and unlike the quadricorrelator is implemented with predominately digital circuitry. As a consequence, its operation is limited to lower frequencies, but where applicable it is more amenable to integrated circuitry realization because of the elimination of multipliers and filtering functions. Its operation depends on detecting, with simple circuitry, the direction of rotation of the signal constellation.

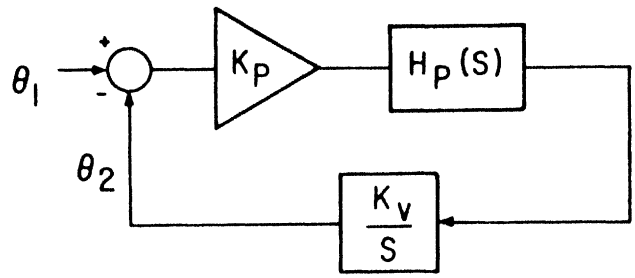
For completeness we mention the papers by Oberst [9], describing an FD for two square waves (useful in frequency synthesis *), and Runge [10], describing an unrelated FD for timing and carrier recovery applications.

Paper approved by the Editor for Data Communication Systems of the IEEE Communications Society for publication without oral presentation. Manuscript received October 23, 1978; revised May 4, 1979. This research was performed for the VIDAR Division of TRW, Mountain View, CA.

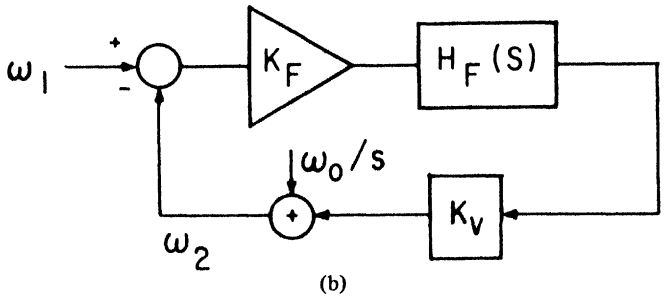
The author is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

* The FD's described here can be used for two sinusoids or square waves, but appear to have greater complexity than Oberst's circuits.

Reprinted from *IEEE Trans. Comm.*, vol. COM-27, pp. 1288-1295, September 1979.



(a)



(b)

Fig. 2. Linearized Loop Models.

While the primary purpose of this paper is to describe and analyze the FD techniques, we first discuss in Section 2.0 the choice of loop filters. Then in Section 3.0 we focus on the quadricorrelator and rotational FD, and describe experimental results in Section 4.0.

2.0. LOOP FILTERS

Assuming the input signal to the PLL of Figure 1 is of the form $\sin(\omega_1 t + \theta_1)$, the VCO output is $\cos(\omega_2 t + \theta_2)$, and the PD and FD are both linear, the linearized models of Figure 2 result. The phase-locked loop of Figure 2(a) governs after lock has been achieved, while the frequency locked loop of Figure 2(b) governs the acquisition behavior. The design parameters are the PD and FD constant K_p and K_f, the VCO constant K_v, and the VCO free-running frequency ω_0.

The loop dynamics are governed by the standard closed loop phase transfer function

$$\frac{\theta_2(s)}{\theta_1(s)} = \frac{K_p K_v H_p(s)}{s + K_p K_v H_p(s)} \tag{2.1}$$

plus a transfer function governing acquistion

$$\omega_1(s) - \omega_2(s) = \frac{\omega_1(s) - \omega_0/s}{1 + K_f K_v H_f(s)}. \tag{2.2}$$

Bellisio [6] recommends a proportional plus integral PD loop filter,

$$H_p(s) = \mu_1 + \frac{\mu_2}{s} \tag{2.3}$$

which is a good choice since the static phase error is small [1] and the usual concern with the integrator being initially saturated is alleviated due to the action of the FD. He also recommends that the FD use the same loop filter (that is, the summer in Figure 1 be placed in front of a single loop filter of type (2.3)). This latter choice is shown to be disadvantageous when we calculate the time response due to a step frequency change $\omega_1(s) = \omega_1/s$ from (2.2),

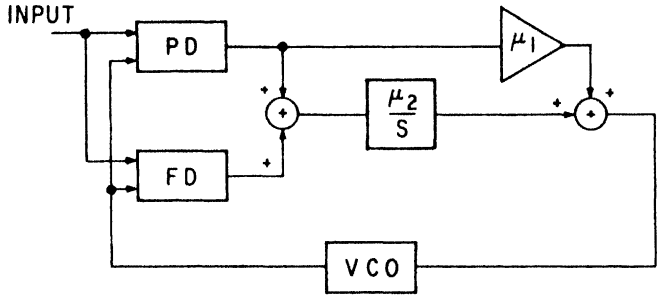


Fig. 3. Choice of Loop Filters.

$$\omega_1(t) - \omega_2(t) = \frac{\omega_1 - \omega_0}{1 + \mu_1 K_f K_v} e^{-t/\tau} \tag{2.4}$$

where the time constant is

$$\tau = \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2 K_f K_v}. \tag{2.5}$$

Thus, we see that, as expected, fastest acquisition occurs for K_f large, but the time constant is limited to $\tau = \mu_1/\mu_2$. Physically, this limitation on speed of acquisition is due to the proportional part of the filter, which initially reduces the frequency error and slows the charging of the integrator. The solution is to eliminate the proportional filter

$$H_f(s) = \frac{\mu_2}{s} \tag{2.6}$$

resulting in the configuration of Figure 3. The FD charges the integrator capacitor to the correct voltage to reduce the frequency error to zero (in spite of any initial saturation), and in-lock the PD maintains that charge. While (2.4) predicts that increasing K_f can result in arbitrarily fast acquisition, in practice the fact that the FD output will have a randomly fluctuating voltage on its output in-lock places a practical limit on the size of K_f.

3.0. SPECIFIC FD DESIGNS

3.1. Quadricorrelator Frequency Detector

The quadricorrelator, as shown in Figure 4, consists of two quadrature mixers, a differentiator in the in-phase channel, and a cross-correlator.** The mean value of $p(t)$ is proportional to the difference between the center frequency of the power spectrum of $r(t)$ and ω_2. While this property has been demonstrated for sinusoidal [2, 3] and Gaussian [3-5] inputs $r(t)$, it can be easily established in general. In particular, if

** The similarity of the quadricorrelator to the PD of a Costas loop [11] is striking.

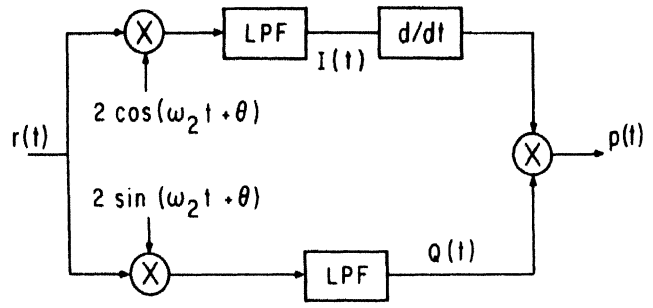


Fig. 4. Quadricorrelator.

$r(t)$ has a power spectrum symmetric about the radian frequency ω_1, it can be expanded in the form

$$r(t) = x_c(t) \cos \omega_1 t - x_s(t) \sin \omega_1 t \tag{3.1}$$

where $x_c(t)$ and $x_s(t)$ are uncorrelated. It is shown in Appendix A, using (3.1), that

$$Ep(t) = \Delta\omega \sigma_r^2 \tag{3.2}$$

where $\Delta\omega$ is the radian frequency difference,

$$\Delta\omega = \omega_1 - \omega_2 \tag{3.3}$$

and σ_r^2 is the variance of $r(t)$. Thus, $p(t)$ is an unbiased estimate of $\Delta\omega$ when properly scaled by σ_r^{-2}.

Many data transmission modulation methods have a signal power spectrum symmetric about the carrier frequency, the most important examples being PSK and QAM [11]. The quadricorrelator is thus a suitable FD for carrier recovery with these modulation methods. For the particular case of biphase modulation, Cahn [8] has suggested a FD structure similar to the quadricorrelator, except that it includes an additional $I(dQ/dt)$ term. From the foregoing, it is evident that the simpler quadricorrelator would suffice. Bellisio [6] applied the quadricorrelator to baseband PAM timing recovery, exploiting the symmetry about the baud frequency of the pulse waveform spectrum generated by a NRZ data transition detector (differentiator followed by dead-zone quantizer).

Finally, we mention that many authors include limiters in both I and Q channels. This simplifies implementation of the correlation multiplier, which must have a very small offset to control static phase error, as well as insures a zero FD output after acquisition and eliminates the σ_r^2 dependence in (3.2).

3.2. Rotational Frequency Detector

The rotational FD, in contrast to the quadricorrelator, is constructed of predominately digital circuitry and includes no filtering functions. Consequently, it is particularly well suited to integrated circuit implementation, but is also inherently limited to lower frequency operation than the quadricorrelator.

The rotational FD is simplest to describe for measurement of the frequency difference between two square waves, although it offers no particular advantage for that application over circuits described by Oberst [9]. That description is given in Section 3.3, and the simple generalizations to timing and carrier recovery are described in Section 3.4 and 3.5. The effect of noise and phase jitter is analyzed in Section 3.6.

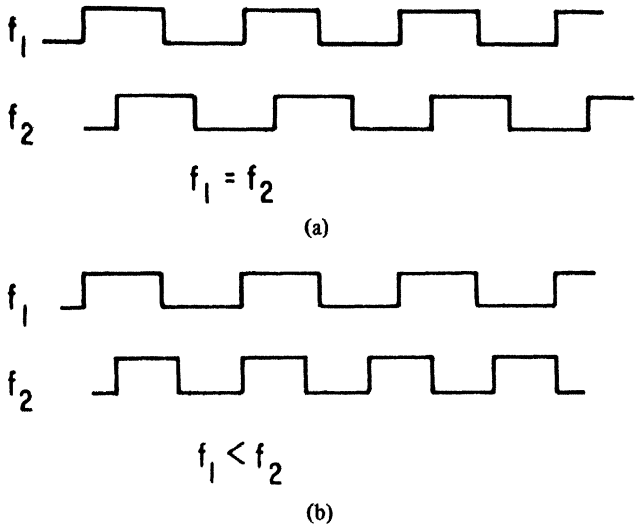


Fig. 5. Situations to be Distinguished by FD.

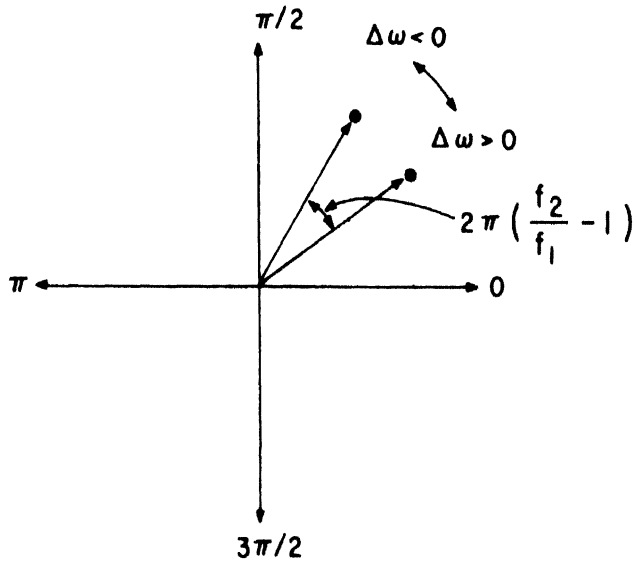


Fig. 6. Phasor Diagram of Two Successive Transitions of f_1 Relative to Phase of f_2.

3.3. Two Square Waves

Two of the three cases to be distinguished by the FD are shown in Figure 5. These cases would easily be recognized by a human observer watching the waveforms on an oscilloscope. When $f_1 = f_2$, the transitions of f_1 maintain a fixed relationship to those of f_2. When $f_1 < f_2$, the transitions of f_1 advance in phase relative to those of f_2, and vice versa when $f_1 > f_2$. An excellent way to view the situation is to draw a phasor diagram as in Figure 6. One cycle (2π radians) of f_2 is shown and the two phasors represent the relative phase of two successive transitions of f_1. The angle of rotation is readily shown to be $2\pi\,((f_2/f_1) - 1)$, which is counterclockwise if $f_1 < f_2$ and clockwise if $f_1 > f_2$. Hence detecting the sign of the frequency difference is equivalent to determining the direction of rotation in Figure 6, while the magnitude of the frequency difference is related to the angle of rotation.

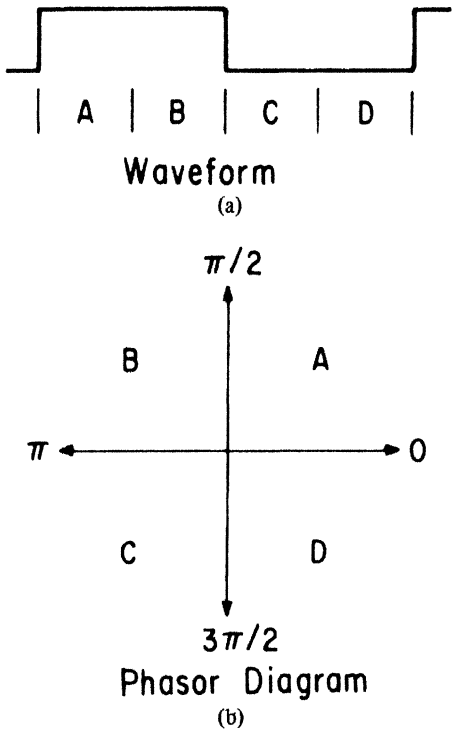


Fig. 7. Division of VCO Cycle into Four Quadrants.

A circuit which detects the direction of rotation can be built as follows: Assume f_2 is the VCO frequency, and divide each cycle into four quadrants labeled A, B, C, and D as in Figures 7(a) and (b). This can be accomplished by actually running the VCO at four times frequency f_2, and dividing by four to obtain f_2 itself. Further assume that the PD is designed so that in-lock the PLL will maintain the positive transition of f_1 in the vicinity of the positive transitions of f_2 (other PD designs can be handled in like manner). Therefore, in-lock we would expect to observe positive transitions of f_1 predominately or exclusively in quadrants A and D. To ensure that the FD will produce an output rarely if ever in-lock, it will operate only upon the observation of positive transitions of f_1 in quadrants B and C.

Let the kth cycle of f_2 be denoted by a k-subscript. The situation $f_1 > f_2$ can be recognized by observation of C_k followed by B_{k+1}, in which case the FD generates a positive pulse. Similarly, if a B_k is followed by C_{k+1}, the FD generates a negative pulse, in recognition that $f_1 < f_2$.

The FD does not generate a pulse for every pair of f_1 transitions, since the rather special conditions of the last paragraph must be met. In particular, they will hopefully seldom be met in-lock, when no FD output is desired, since the PLL should serve to maintain the transition of f_1 in quadrants A and D.

The FD is characterized by the mean value of the pulses at its output, since that mean value serves to charge or discharge the integrating capacitor in the loop filter. That mean value is, assuming FD positive and negative output pulses have equal area,

$$\mu_{FD} = \Pr\{\text{positive pulse}\} - \Pr\{\text{negative pulse}\}. \tag{3.4}$$

If the frequency difference is $\Delta\omega$, the angle of rotation is $-\Delta\omega/f_1$ radians, and μ_{FD} can readily be calculated by assuming that the phasor in cycle k is uniformly distributed from 0 to 2π radians in Figure 7(b). For example, if the angle of rotation of two successive phasors is $\phi < \pi/2$, then an FD output is generated only when the first phasor is within an angle ϕ of the π-axis, an event which has probability $\phi/2\pi$. By a simple extension of this argument, the plot of μ_{FD} of Figure 8 can be generated. The characteristic is periodic for $f_1 > f_2$, since multiple cycles of f_1 in a period of f_2 cannot be distinguished from a single cycle by the FD circuit as described. It is not periodic for $f_1 < f_2$ since, if the period of f_1 is too great, successive positive transition of f_1 will not occur within two periods of f_2 and the FD will generate no output.

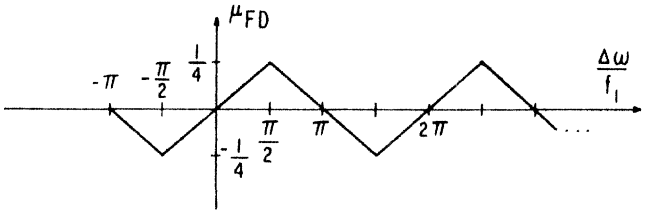


Fig. 8. FD Characteristic for Two Square Waves.

As seen from Figure 8, the useful range of the FD is

$$|\Delta\omega| < \pi f_1 = \frac{\omega_1}{2}. \tag{3.5}$$

That is, a 50% offset in the initial VCO frequency f_2 can be tolerated. The range of linearity of the FD, that is, the range over which the model of Figure 2(b) is accurate, is $|\Delta\omega| \leq \omega_1/4$. The largest FD output is at $\Delta\omega = \pm\omega_1/4$, where the probability of an output pulse is 0.25.

3.4. Timing Recovery

In timing recovery it is standard to generate a sequence of timing pulses from the data waveform. For example, Bellisio [6] describes a circuit consisting of a differentiator and dead-zone rectifier which generates data transition pulses from an NRZ data waveform. The nominal spacing between two successive pulses can be any multiple of the baud interval $T = 1/f_1$ since a pulse is only generated by a data transition. As in Section 3.3 we let f_2 be the VCO output frequency.

The FD described in Section 3.3 works for this case, where the quadrants of f_2 in which the data transition pulses occur are observed. The calculation of the FD output mean is similar, except that in addition to the requirement for two successive phasors to span the π-axis, there must be two data transitions in a row in order for an FD pulse to be generated. Thus, the FD characteristic of Figure 8 remains valid, except that μ_{FD} must be multiplied by the probability of two data transitions in a row (0.25 for equally likely independent data). The FD range is 50% of the baud rate, which is comparable to that reported by Bellisio [6] and more than adequate for the elimination of a VCXO.

3.5. Carrier Recovery

As mentioned in the introduction, the motivation for using an FD to aid acquisition in carrier recovery is somewhat dif-

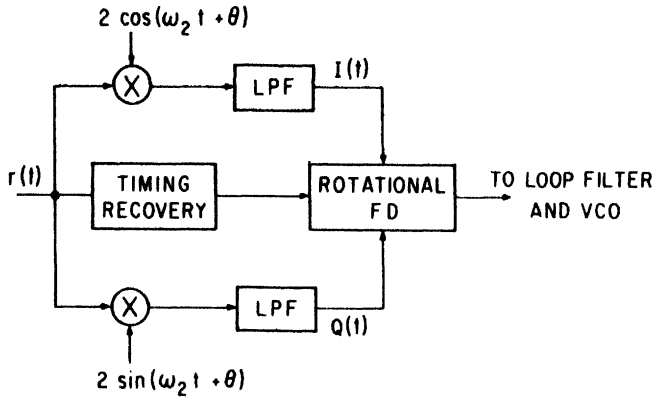


Fig. 9. Configuration of the Rotational FD in Carrier Recovery.

ferent than in timing recovery, for it is usually not practical to eliminate the VCXO due to the problem of false lock to a data sideband to be described shortly. Acquisition remains a problem, however, since worst-case frequency offsets can still exceed the desired loop bandwidth by many orders of magnitude. The problem is particularly acute in microwave radio transmission, where even very accurate RF oscillator frequencies can result in absolute frequency offsets of 50 to 100 kHz, while carrier recovery loop bandwidths are more typically in the range of 0.1 to 1.0 kHz.

The signal used by the FD to extract carrier frequency offset is assumed to be of the form of (3.1), where ω_1 would typically be the carrier frequency at IF. The object of the FD is to estimate $\Delta\omega$ given by (3.3), where ω_2 is the IF local oscillator frequency generated by the VCO in our carrier recovery PLL. The configuration of the FD is shown in Figure 9. The first step is to demodulate to baseband with quadrature carriers (at frequency ω_2) just as for the Costas loop and quadricorrelator; the resulting quadrature baseband signals $I(t)$ and $Q(t)$ are given by (A.1). The FD operates on $I(t)$ and $Q(t)$, while timing recovery is performed on $r(t)$. The major requirement of the rotational FD is that timing recovery acquisition occur *before* the FD output is valid and carrier recovery acquisition is initiated. The interesting property that timing recovery can be achieved independent of carrier recovery can be seen from squaring $r(t)$ in (3.1) and eliminating the double frequency term; the result is $(x_c^2(t) + x_s^2(t))/2$, which will have a baud frequency component suitable for extraction.

The first operation of the FD is to sample I and Q at the baud interval kT; from (A.1) the result is

$$I(kT) = x_c(kT)\cos(k\Delta\omega T - \theta) - x_s(kT)\sin(k\Delta\omega T - \theta)$$
$$Q(kT) = -x_c(kT)\sin(k\Delta\omega T - \theta) - x_s(kT)\cos(k\Delta\omega T - \theta). \quad (3.6)$$

Consider first the case $\Delta\omega = \theta = 0$ following acquisition. The point $(I(kT), Q(kT))$ when plotted in a two-dimensional plane is, in fact, one of the data points in the two-dimensional signal constellation corresponding to the modulation method. We show two examples in Figure 10(a) and (b), two-level (biphase) and four-level (QPSK) phase-shift keying. Biphase serves as a basis of comparison to the work of Cahn [8], while

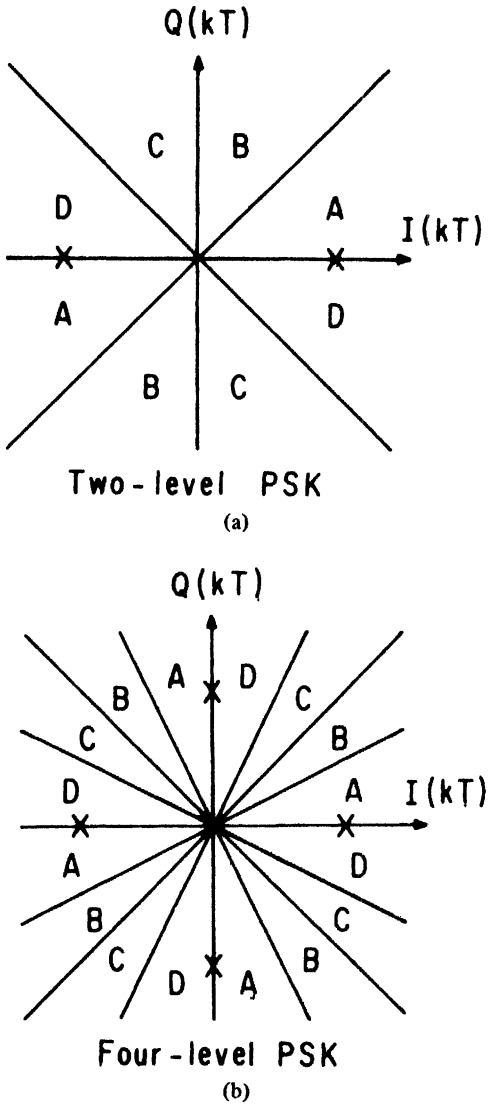


Fig. 10. Two-Dimensional Signal Constellations and Rotational FD Thresholds (Constellation shown for $\Delta\omega = \omega = 0$).

QPSK demonstrates how the technique generalizes to more complicated constellations.***

When there is a frequency difference, we recognize (3.6) as the parametric equations of a circle; that is, the signal constellation is rotated by an angle $k\Delta\omega T$-θ. Rotation is clockwise if $\Delta\omega > 0$. The nature of the problem of false lock to a data sideband is now clearly evident; in biphase modulation rotation by $\Delta\omega T$ equal to multiples of π radians can clearly not be distinguished from $\Delta\omega T = 0$. Thus, any FD characteristic must be periodic in $\Delta\omega = \pi/T$, and the maximum useful range of any FD is

$$|\Delta\omega| \leq \pi/2T, \text{ biphase.} \quad (3.7)$$

Thus, the initial VCO frequency must not deviate from the carrier frequency in magnitude by more than one-quarter the

*** It is also possible to restrict operation of the FD to a subset of the data points, if that subset can be unambiguously identified in the face of rotation. For example, in 16-level QAM, restriction to the four inner data points results in operation identical to QPSK.

baud rate. For QPSK, (3.7) is replaced by

$$|\Delta\omega| \leq \pi/4T, \text{ QPSK} \tag{3.8}$$

since false lock occurs when rotation is by $\pi/2$ radians, and the maximum VCO offset is one-eighth the baud rate.

Shown on Figure 10, in addition to the signal constellation, are the radial thresholds required for the rotational FD. Some of these thresholds can also serve as slicers for data decisions as well as for implementing a bang-bang type of PD. The radial thresholds divide the angle between each pair of adjacent data points into four quadrants A, B, C, D. In each baud interval the quadrant actually observed is independent of the data, and depends only on the angle of rotation ($k\Delta\omega T$-θ). As before, to insure infrequent FD pulses in lock when $\theta \cong 0$, operation is restricted to quadrants B and C. The actual circumstances in which an FD pulse is generated are identical to the square wave and timing recovery cases.

The FD characteristic is plotted in Figure 11. The highest probability of a pulse output is 0.25 and occurs at $|\Delta\omega T| = \alpha/2$, and the range of useful operation is $|\Delta\omega T| \leq \alpha$, where $\alpha = \pi/2$ for biphase and $\alpha = \pi/4$ for QPSK. Since these figures are consistent with (3.7) and (3.8) it follows that the rotational FD has as large a range of operation for carrier recovery as any FD.

3.6. Effect of Phase Jitter

The plots of μ_{FD} presented thus far have not taken into account the effects of noise and intersymbol interference. Since the rotational FD is sensitive to the angle of rotation, which in turn is influenced by these factors, there is concern that they might significantly affect FD operation.

The situation is considered in Appendix B, where it is shown that if the phase jitter is small relative to $\pi/2$, $\pi/4$, or $\pi/8$ for the timing recovery, biphase, or QPSK situations, respectively, the effect of phase jitter is virtually absent, this in spite of any statistical dependencies which may exist between successive samples of phase jitter. For phase jitter with amplitude less than twice the previously mentioned values, the effect is to change the shape of the FD characteristic (basically, round the corners), but not otherwise adversely affect its operation. Even larger phase jitter will have a significant adverse effect on FD operation, but is not likely to be encountered in practice, since the effect of this large jitter on error rate would also be substantial.

4.0. EXPERIMENTAL RESULTS

Experimental results on the use of the quadricorrelator in timing recovery were reported by Bellisio [6]. We report here on experimental results obtained in the implementation of a rotational FD in a carrier recovery application. The terrestrial microwave system to which it was applied employed a 16-point signal constellation and 10 Mbit/s data rate. References [12-13] describe timing and carrier recovery techniques which are typical for this type of system.

This particular system protection switches at an error rate of 10^{-6}, which corresponds to a baseband SNR of about 22 dB. Reliable acquisition was experimentally observed for an SNR of 15 dB or less, which is substantially lower than necessary for this type of system. Figure 12 shows the VCO control line during several acquisitions at SNR's of 22, 20, and 15 dB starting from a worst-case carrier frequency offset (133 kHz). The initial flat portion of the curves corresponds to the period of timing recovery acquisition, which must precede carrier acquisition for the rotational FD. Total acquisition time is about 15, 25, and 45 ms for the three cases.

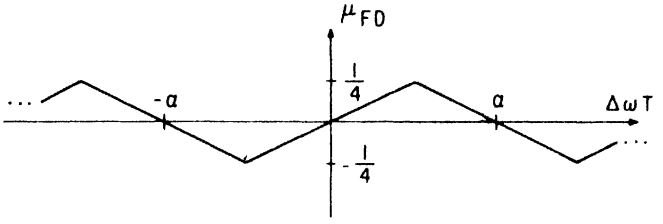


Fig. 11. FD Characteristic for Carrier Recovery.

5.0. CONCLUSIONS

The use of a FD to aid PLL timing and carrier acquisition is a very advantageous technique; the major impediment to its use appears to have been the lack of suitable FD circuits. We have described two such circuits, the quadricorrelator and rotational FD, both of which have a broad applicability. For timing recovery the rotational FD is somewhat simpler, particularly for integrated circuit implementation, since it is digital and requires no multipliers or filters; however, it is also limited to lower data rates. For carrier recovery there appears to be no great difference in the difficulty of implementation, since both circuits require quadrature mixers followed by circuitry which operates at approximately the baud rate. In some instances a substantial portion of the rotational FD circuitry can simultaneously serve other purposes (such as PD and data thresholding), in which case it becomes more attractive.

APPENDIX A
QUADRICORRELATOR OUTPUT MEAN VALUE

Let wide-sense stationary input signal $r(t)$ be written in the form of (3.1) and assume that the power spectrum of $r(t)$ is symmetrical about radian frequency ω_1 so that $x_c(t)$ and $x_s(t)$ are uncorrelated. Assuming that $\omega_1 + \omega_2$ terms are rejected by the low pass filters in Figure 4,

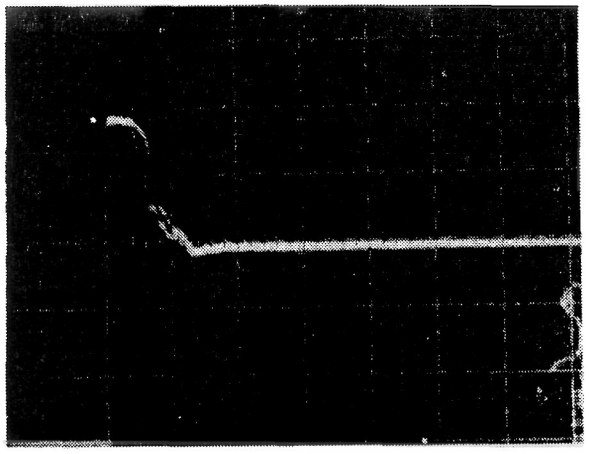
$$I(t) = x_c(t)\cos(\Delta\omega t - \theta) - x_s(t)\sin(\Delta\omega t - \theta)$$

$$Q(t) = -x_c(t)\sin(\Delta\omega t - \theta) - x_s(t)\cos(\Delta\omega t - \theta) \tag{A.1}$$

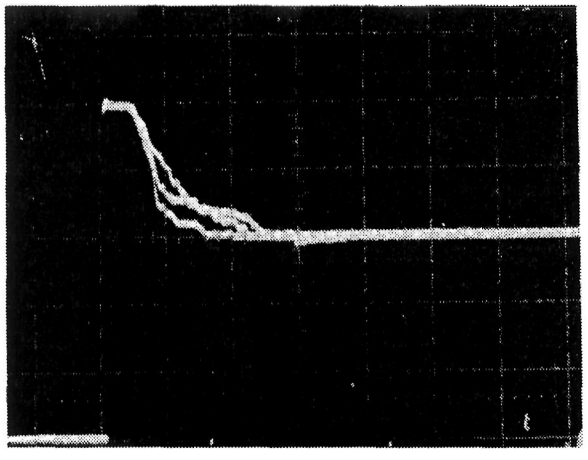
where $\Delta\omega$ is given by (3.3). The crosscorrelation of $I(t)$ and $Q(t)$ is then easily shown to be, using the fact that $x_c(t)$ and $x_s(t)$ are uncorrelated,

$$R_{IQ}(\tau) = E[I(t)Q(t+\tau)]$$

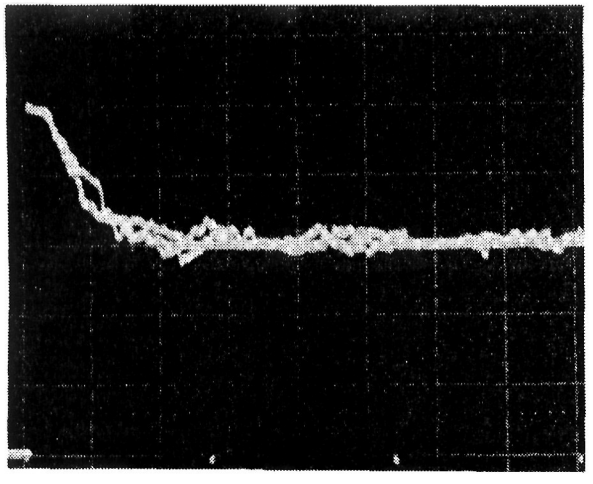
$$= \frac{j}{4}(e^{j\Delta\omega\tau} - e^{-j\Delta\omega\tau})(R_c(\tau) + R_s(\tau)) \tag{A.2}$$



(a)



(b)



(c)

Fig. 12. Rotational FD Carrier Recovery Acquisition. (a) SNR = 22 dB (10 ms/division). (b) SNR = 20 dB (10 ms/division). (c) SNR = 15 dB (20 ms/division).

where $R_c(\tau)$ and $R_s(\tau)$ are the autocorrelation functions of $x_c(t)$ and $x_s(t)$. The cross power spectral density, the Fourier transform of (A.2), is

$$S_{IQ}(\omega) = \frac{j}{4}\{S_c(\omega - \Delta\omega) - S_c(\omega + \Delta\omega) + S_s(\omega - \Delta\omega) - S_s(\omega + \Delta\omega)\}. \quad \text{(A.3)}$$

The mean of $p(t)$ is thus

$$\begin{aligned} Ep(t) &= E\left[Q(t)\frac{dI(t)}{dt}\right] \\ &= \int_{-\infty}^{\infty} (j\omega)^* S_{IQ}(\omega)\frac{d\omega}{2\pi} \\ &= \frac{1}{4}\int_{-\infty}^{\infty} \omega\{S_c(\omega - \Delta\omega) - S_c(\omega + \Delta\omega) \\ &\quad + S_s(\omega - \Delta\omega) - S_s(\omega + \Delta\omega)\}\frac{d\omega}{2\pi}. \end{aligned} \quad \text{(A.4)}$$

Changing variables in this integral, we obtain finally

$$Ep(t) = \tfrac{1}{2}\Delta\omega \int_{-\infty}^{\infty} (S_c(\omega) + S_s(\omega))\frac{d\omega}{2\pi}. \quad \text{(A.5)}$$

Recognizing that the power in $r(t)$ is one-half the sum of the powers in $x_c(t)$ and $x_s(t)$, (3.2) follows.

APPENDIX B

We can model the effect of phase jitter by assuming the angle of rotation at the kth baud interval is $(k\Delta\omega T - \theta + \theta_k)$, where θ_k is a random phase jitter component. It is important that we *not* assume that θ_k and θ_{k+1} are independent, since intuitively dependencies should have a particularly strong influence. In order to recalculate μ_{FD} for this case, we let θ be uniformly distributed on $[0, 2\pi]$ as before. The starting angle is $(k\Delta\omega T - \theta + \theta_k)$, and the angle of rotation is $(\Delta\omega T + \theta_{k+1} - \theta_k)$. The key to simplifying the problem is to first condition on θ_k and θ_{k+1}, and take the expectation over θ, that expectation being the same as previously determined but with $\Delta\omega T$ replaced by $(\Delta\omega T + \theta_{k+1} - \theta_k)$. Thus, completing the expectation over θ_k and θ_{k+1},

$$\mu_{FD} = E[F(\Delta\omega T + \theta_{k+1} - \theta_k)] \quad \text{(B.1)}$$

where $F(\Delta\omega T)$ is the FD characteristic of Figures 8 or 11.

Equation (B.1) is exact, but for the special case where the argument of F is in the linear region with high probability, where $F(\omega) = K\omega$,

$$\begin{aligned} \mu_{FD} &\cong E[K(\Delta\omega T + \theta_{k-1} - \theta_k)] \\ &= K(\Delta\omega T + E\theta_{k+1} - E\theta_k) \\ &= F(\Delta\omega T) \end{aligned} \quad \text{(B.2)}$$

where we have made the further assumption that $E\theta_{k+1} = E\theta_k$ (not necessarily zero). The implication of (B.2) is that the phase jitter has had no effect whatsoever on μ_{FD}. Note that no assumption of independence of θ_k and θ_{k+1} has been made.

When this special case is violated, (B.1) can be used to estimate the effect. If $\theta_{k+1} - \theta_k$ has probability density $f(\cdot)$, then (B.1) becomes

$$\mu_{FD} = \int F(\Delta\omega T + \phi) f(\phi)\, d\Phi. \tag{B.3}$$

Thus, if the argument $(\Delta\omega T + \phi)$ is not confined to the linear region of F, the effect is seen to be a smoothing of the corners of the FD characteristic. If $f(\phi)$ spans a significant portion of the period of F, then there is a significant deterioration of the FD operation.

ACKNOWLEDGMENT

The author is indebted to F. Stevens for the experimental results provided in Section 4.0.

REFERENCES

1. F. M. Gardner, *Phase-Lock Techniques*, New York: Wiley, 1966.
2. D. Richman, "Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," *Proc. IRE*, Vol. 42, p. 106, January 1954.
3. T. B. Pickard, "The Effect of Noise on a Method of Frequency Measurement," *IRE Trans. Information Thy*, Vol. IT-4, p. 83, June 1958.
4. R. F. Pawula, "Analysis of an Estimator of the Center Frequency of a Power Spectrum," *IEEE Trans. Information Thy*, Vol. IT-14, p. 669, September 1968.
5. S. R. J. Axelsson, "Analysis of the Quantizing Error of a Zero-Counting Frequency Estimator," *IEEE Trans. Information Thy*, Vol. IT-22, P. 596, September 1976.
6. J. A. Bellisio, "New Phase-Locked Timing Recovery Method for Digital Regenerators," *Int. Conf. Communications Record*, Philadelphia, pp. 10-17, June 1976.
7. J. H. Park, Jr., "An FM Detector for Low S/N," *IEEE Trans. Comm.*, Vol. COM-18, p. 110, April 1970.
8. C. R. Cahn, "Improving Frequency Acquisition of a Costas Loop," *IEEE Trans. on Comm.*, Vol. COM-25, p. 1453, December 1977.
9. J. F. Oberst, "Generalized Phase Comparators for Improved PLL Acquisition," *IEEE Trans. on Comm. Tech.*, Vol. COM-19, p. 1142, December 1971.
10. P. K. Runge, "Phase Locked Loops with Signal Injection for Increased Pull-In Range and Reduced Output Phase Jitter," *IEEE Trans. Comm.*, Vol. COM-24, p. 636, June 1976.
11. R. W. Lucky, J. Salz, E. J. Weldon, Jr., *Principles of Data Communication*, New York, McGraw-Hill, 1968.
12. C. W. Anderson, S. G. Barber, "Modulation Considerations for a 91 Mb/s Digital Radio," *IEEE Trans. on Communications*, Vol. COM-26, May 1978, p. 523.
13. C. R. Hogge, "Carrier and Clock Recovery for 8 PSK Synchronous Demodulation," *IEEE Trans. on Communications*, Vol. COM-26, May 1978, p. 528.
14. R. Citta, "Frequency and Phase Lock Loop," *IEEE Trans. Consumer Electronics*, Vol. CE-23, Aug. 1977, p. 358.

Analysis of Phase-Locked Timing Extraction Circuits for Pulse Code Transmission

ENGEL ROZA

Abstract—**An analysis is presented of the performance of phase-locked timing extraction circuits for baseband pulse code transmission. The phase error of the extracted timing wave is influenced by the properties of three essential stages in signal processing: prefiltering, nonlinear treatment, and narrow-band filtering. The analysis enables us to calculate quantitatively the quasi-static and the dynamic part of the phase error for arbitrary but specified types of signal processing. This is more than can be done with existing theory in the case of resonant-type timing extraction circuits. Examples are given for practical cases, and conditions for optimum performance are derived.**

Furthermore, the behavior of such phase-locked circuits in a chain of repeaters is investigated, and in particular, the propagation law for jitter. Byrne's model, as used for resonant-type timing extraction circuits, therefore is generalized. It is shown analytically and experimentally that by proper implementation of the timing extraction circuit, considerable improvement can be obtained as compared with resonant-type circuits.

I. INTRODUCTION

IN ORDER to regenerate pulses in a transmission system for pulse coded signals, a clock signal should be available. In self-timing repeaters, this clock signal is derived from the baseband information signal itself. This includes filtering by means of a resonant circuit or by a phase-locked loop (PLL). The properties of a single resonant circuit with respect to timing derivation have been analyzed by Sunde, Bennett, and Manley [1]–[3]. Byrne *et al.* [4] have studied the propagation of jitter of the extracted timing wave in a chain of repeaters with resonant-type clock extraction circuits. In this paper, the PLL is analyzed in these respects. In addition, a more general approach is followed, permitting the influence of undesired signal interference to be taken into account.

II. SPECTRAL DENSITY OF THE INFORMATION SIGNAL

The transmission model used consists of a signal source, sending synchronous digital impulses into an equivalent baseband transmission channel, which also includes all linear processing of the receiving and transmitting end. Generally, the bandwidth of this transmission channel will be restricted such that a well-defined transfer of information is just possible. The first Nyquist criterion states that a bandwidth, which extends from zero to a frequency between the digit frequency and its half, is sufficient.

Manuscript received November 15, 1973.
The author is with the Philips Research Laboratories, Eindhoven, The Netherlands.

The series of synchronous impulses from the signal source can be described as a superposition of a deterministic series of impulses with equal amplitudes and a stochastic series of impulses with discrete amplitudes of a random distribution. The first series is responsible for lines in the power spectrum at the digit frequency and its multiples. The second series produces the continuous part of this spectrum.

Because of the bandwidth restriction of the transmission channel, it is not possible to transfer energy at the digit frequency or its multiples. That means that nonlinear signal processing is required to derive a clock signal at the digit frequency from the received signal.

III. NONLINEAR SIGNAL PROCESSING

Two methods are commonly used which avoid undesired interference of neighboring pulses in the clock path. We shall denote this undesired interference as *interpulse interference,* which has to be distinguished from intersymbol interference in the information path.

1) According to the first method, the input signal of the clock extractor is shaped in such a way that it fulfills Nyquist's second criterion [5]. Such a signal has the property that halfway between the centers of two successive pulses, the signal value of all other pulses is zero. This signal is then processed by a nonlinear circuit generating pulses of a short duration (shorter than one digit interval) when the signal crosses thresholds which correspond to the values of a single pulse halfway between two pulse centers. In this way, a random series of equidistant pulses originates, possessing a power spectrum with lines at the digit frequency and its multiples, so that with narrow-band filtering, a continuous clock signal can be derived. This method will be referred to as the *threshold method.*

2) According to the second method, the transmission characteristic of the clock path is changed such that an even symmetrical transfer characteristic around the Nyquist frequency is obtained. Afterwards, the signal is squared. As a result, a signal component is produced with a frequency equal to the digit frequency, modulated in amplitude, but with a constant phase. All components in the original continuous spectrum around the Nyquist frequency contribute to the power of this signal. Also, this signal has to be filtered afterwards by a narrow bandpass filter centered at the digit frequency. This process, denoted as the *symmetry method,* will be analyzed in Section VII in more detail.

The threshold method is used in systems with a digit

Reprinted from *IEEE Trans. Comm.*, vol. COM-22, pp. 1236-1249, September 1974.

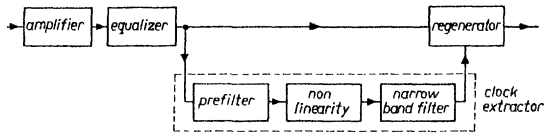


Fig. 1. Block diagram of a regenerative repeater.

frequency that is relatively low with respect to the frequency range of the applied electronic components. The symmetry method is used at higher frequencies, although its fundamental property of avoiding interpulse interference is not, or hardly, recognized. For example, the rectification of the symbols of a band-limited first-order bipolar signal has to be conceived as a nonideal application of the described principle. The bipolar coding of the first order [12] is a linear process which yields a symmetrical power spectrum around the Nyquist frequency between zero and the digit frequency. Although bandwidth limitation may somewhat deteriorate this symmetry and although rectification substitutes the squaring, the principle mentioned can readily be recognized. The resulting signal to be filtered will show phase errors caused by interpulse interference due to this imperfect execution.

The three essential parts of the clock extraction process, i.e., prefiltering, nonlinear treatment, and narrow-band filtering, are shown in the block diagram of a regenerative repeater in Fig. 1. Various possibilities for implementation will be investigated and compared in the following sections.

IV. FILTERING

At first, the description of the filter process after the nonlinear treatment of the signal will be based upon the short pulses as obtained by the threshold method. The symmetry method will be considered later. The obtained signal after the nonlinear process can be approximated by

$$s(t) = \sum_{n=0}^{N} c_n \delta(t - nT - \tau_n). \tag{1}$$

Herein the coefficient c_n is determined by the statistical properties of the signal, and it possesses in its most simple form the binary values 0 or 1. T is the digit interval and N is related to t such that

$$t = NT + \tau, \qquad 0 \leq \tau < T. \tag{2}$$

The quantity τ_n denotes the deviation of the nth pulse from its nominal value.

There are three possible causes for this shift:

1) additive noise in the transmission channel,

2) the input signal to the nonlinear circuit does not obey the second criterion of Nyquist,

3) the incoming symbols contain timing errors, introduced by preceding repeaters.

In a later section, the influences of more general preprocessing will be considered. It will be found that these influences can be taken into account by assigning appropriate values to c_n and τ_n. Statistical mutual dependence of those values may occur in this general outline.

It is useful to make the following remarks about c_n and τ_n *a priori*.

1) c_n and τ_n are assumed to be random variables in a wide sense stationary random process, i.e., their autocorrelation functions $R_c(k)$ and $R_\tau(k)$, defined by (3), are independent of n.

$$R_c(k) = \overline{c_n c_{n+k}}$$
$$R_\tau(k) = \overline{\tau_n \tau_{n+k}}. \tag{3}$$

(The ensemble average of a random variable x will be indicated by $\bar{x}$.)

As a consequence, their discrete Fourier transforms determine their spectral densities.

2) c_n can vary randomly and considerably from digit interval to digit interval. Its spectral density, therefore, usually extends beyond the digit frequency.

3) Because of the filtering of preceding repeaters, the spectral density of τ_n can show strong components in the low-frequency part.

4) Due to the growth of the deviations τ_n along a chain of repeaters, τ_n cannot simply be considered to be small compared with one digit interval T.

A. Bandpass Filtering

Let the signal $s(t)$ of (1), represented by pulses from a current source, excite a narrow bandpass filter with center frequency as close as possible to the digit frequency. As a result, the voltage across the filter will be a harmonic function fluctuating in amplitude and phase with a nominal frequency equal to the digit frequency. The amplitude fluctuation can be eliminated by a hard limiter. The phase fluctuation deteriorates the quality of the derived clock signal.

This process has been investigated by Sunde [1] and Bennett [2] for a resonant circuit.

B. PLL Filtering

An alternative solution for obtaining a clock signal from $s(t)$ is the synchronization of the oscillator of a PLL. In order to analyze the PLL, a model for it, using digital signals, should be available. Such models have been derived in the past by Byrne [6] and Saltzberg [7]. We develop here an alternative model to show the relationship with the common linear model for sinusoidal signals so as to make possible an easy comparison with bandpass filtering. Fig. 2(a) shows the general block diagram for a large class of PLL's. It shows that the loop filter is fed by the product of the input signal $s_i(t)$ and the oscillator signal $s_o(t)$.

In the case that $s_i(t)$ and $s_o(t)$ are sinusoidal signals, the low-frequency part of the output of the multiplier is linearly proportional to the actual difference $\theta_{ea}(t)$ of the input phase $\theta_{ia}(t)$ and the oscillator phase $\theta_o(t)$. The block diagram can then be modified to Fig. 2(b). The phase of the oscillator is determined by the convolution of the phase error $\theta_{ea}(t)$ and $g(t)$. By $g(t)$ we denote the pulse response of the loop filter followed by an ideal integrator

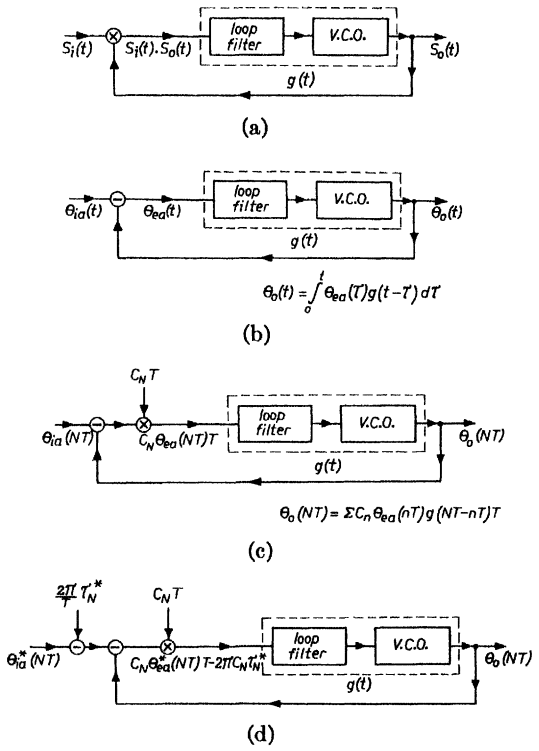


Fig. 2. (a) General block diagram of a PLL. (b) PLL model for sinusoidal signals. (c) PLL model for digital signals. (d) PLL model for digital signals with interpulse interference.

which represents the behavior of the phase of the oscillator due to an applied signal.

So, assuming causality for $g(t)$ and $\theta_{ea}(t)$,

$$\theta_o(t) = \int_0^t \theta_{ea}(\tau)g(t-\tau)\,d\tau. \tag{4}$$

The input signal applied to the multiplier in the case of digital signals consists of a series of equidistant pulses with random amplitudes. These pulses should have a finite width which should be, in principle, less than one digit interval T. The output of the multiplier also consists of a random series of pulses (with another shape). Because of the integrating properties of the loop, the momentary signal value of the pulses as such is not significant, but merely the integrated signal values over nonoverlapping time intervals, short with respect to the duration of $g(t)$. For this interval, the period T is a convenient choice. Consequently, the output of the multiplier can be modeled as a random series of Dirac-like impulses at discrete intervals T. The magnitudes of these impulses are proportional—linearly as a first approximation—to the product of:

1) the relative delay $\theta_{ea}(NT)T/2\pi$ of the oscillator signal with regard to the input pulse positions, and

2) the discrete pulse amplitudes c_N of the input signal.

From these considerations, the evaluation of the general block diagram of Fig. 2(a) to the PLL model for digital signals, as shown in Fig. 2(c), is evident.

For the phase of the oscillator signal, responding to an input signal $s(t)$ of (1), one can write

$$\theta_o(NT) = \sum_{n=0}^{N} c_n\theta_{ea}(nT)g(NT-nT)T. \tag{5}$$

The correspondence with the sinusoidal model is evident if we compare (5) and (4).

$\theta_{ea}(NT)$ represents the phase difference between the *actual* phase of the input signal $\theta_{ia}(NT)$ and the phase of the oscillator $\theta_o(NT)$, so that

$$\theta_{ea}(NT) = \theta_{ia}(NT) - \theta_o(NT). \tag{6}$$

For $\theta_{ia}(NT)$ we may write

$$\theta_{ia}(NT) = \Delta\omega_b NT - 2\pi\frac{\tau_N}{T}. \tag{7}$$

$\Delta\omega_b$ is the detuning of the quiescent frequency of the voltage-controlled oscillator (VCO) from the angular digit frequency ω_b being

$$\omega_b = \frac{2\pi}{T}. \tag{8}$$

τ_N has been defined by (1).

The term $\Delta\omega_b NT$ represents the *nominal* phase of the input signal $\theta_{ia}(NT)$.

Substituting (6) into (5), we find that

$$\theta_{ia}(NT) - \theta_{ea}(NT) = \sum_{n=0}^{N} c_n\theta_{ea}(nT)g(NT-nT)T. \tag{9}$$

We are more interested in the phase error of the oscillator signal with regard to the *nominal* value $\Delta\omega_b NT$ rather than to the actual value $\theta_{ia}(NT)$ because the nominal value represents an ideal imaginary phase reference. So, substituting into (9)

$$\theta_e(nT) = \theta_{ea}(nT) + 2\pi\frac{\tau_n}{T} \tag{10}$$

in which $\theta_e(nT)$ is the phase error with regard to the nominal input phase, we have

$$\theta_e(NT) + \sum_{n=0}^{N} c_n\theta_e(nT)g(NT-nT)T = \Delta\omega_b NT + \frac{2\pi}{T}\sum_{n=0}^{N} c_n\tau_n g(NT-nT)T. \tag{11}$$

We rewrite (11) as

$$\theta_e(NT) + \sum_{n=0}^{N}\frac{c_n}{\bar{c}_n}\theta_e(nT)g_w(NT-nT) = \Delta\omega_b NT + \frac{2\pi}{T}\sum_{n=0}^{N}\frac{c_n\tau_n}{\bar{c}_n}g_w(NT-nT) \tag{12}$$

with

$$g_w(NT-nT) = \bar{c}_n g(NT-nT)T \tag{13}$$

being the normalized weighted open-loop pulse response of the PLL.

Equation (12) will be referred to as the system equation

of the PLL for digital signals. In Appendix I the system equation has been solved. The phase error is found to consist of a static part and a fluctuating part.

For the static part we find

$$\theta_{eo} = \theta_{eo}(1) + \theta_{eo}(2) \tag{14}$$

$$\theta_{eo}(1) = \lim_{N\to\infty} \Delta\omega_b \frac{NT}{\sum_{n=0}^{N} g_w(NT - nT)} \tag{15}$$

$$\theta_{eo}(2) = \frac{2\pi}{T} \frac{\overline{c_N \tau_N}}{\bar{c}_N}. \tag{16}$$

And for the fluctuating part we get

$$\Delta\theta_e(NT) = \sum_{n=0}^{N} b_n h_w(NT - nT) \tag{17}$$

with

$$b_n = -\frac{c_n - \bar{c}_n}{\bar{c}_n}\theta_{eo}(1) - \frac{c_n}{\bar{c}_n}\left(\frac{\overline{c_n\tau_n}}{\bar{c}_n} - \tau_n\right). \tag{18}$$

$h_w(NT)$ is the normalized pulse response of the weighted closed-loop transfer function of the PLL. Its discrete Fourier transform $H_w(l\Omega)$, defined by (94) in Appendix I, will be denoted as the phase-transfer function of the system.

We summarize the conditions which have been assumed in the derivation of the model and the solution of its system equation.

The first two conditions relate to the model. They express the extent to which the model represents the physical reality.

1) Condition of Small Bandwidth: The bandwidth can be characterized most suitably by the closed-loop noise bandwidth, that is,

$$B_L = \frac{1}{2\pi} \lim_{N\to\infty} \sum_{n=0}^{N} |H_w(l\Omega)|^2 \Omega. \tag{19}$$

The condition of small bandwidth is then mathematically expressed by

$$B_L T \ll 1. \tag{20}$$

2) Condition of Linearity: We have assumed that the loop operates in its linear range. This can only be true if the actual phase errors are small. The actual phase errors are identical with the nominal ones if the pulses of the input signal are not deviated from their nominal positions, i.e., if $\tau_n = 0$. The linearity condition is then given by

$$\theta_{eo}(1) \ll 1, \qquad (\overline{\Delta\theta_e^2})^{1/2}|_{\tau_n=0} \ll 1. \tag{21}$$

Two other conditions have been introduced during evaluation of the system equation in Appendix I. These conditions have mainly a theoretical value in order to indicate the bounds of the developed theory. Practical systems operate far from these bounds. The conditions are expressed in terms of spectral densities of the random variables, which are the discrete Fourier transforms of their autocorrelation functions.

3) Condition of Sufficient Clock Content:

$$S_{cc}(0)\,(2\pi B_L T)^{1/2} \ll 1. \tag{22}$$

Here S_{cc} is the spectral density of the random variable $(c_n - \bar{c}_n)/\bar{c}_n$, as defined in Appendix I by (87).

If c_n are statistical independent values of a binary sequence with mark probability Pr, (22) reduces to

$$\frac{1 - \mathrm{Pr}}{\mathrm{Pr}}\,(2\pi B_L T)^{1/2} \ll 1. \tag{23}$$

This condition is apparently strongly related to the small bandwidth condition. Furthermore, the condition of sufficient clock content requires that the spectral density S_{cc} does not vary significantly in the frequency range covered by the phase-transfer function $H_w(l\Omega)$. The name of this condition reflects the importance of the expected value $\bar{c}_n$ ($\bar{c}_n$ must be sufficiently large).

4) Condition of Restricted Interpulse Interference:

$$[R_{cc}(0)]^{1/2} S_{crc}(0)\,(2\pi B_L T)^{1/2} \ll [R_{crcr}(0)]^{1/2}. \tag{24}$$

$R_{cc}(0)$ and $R_{crcr}(0)$ are the variances of the random variables $(c_n - \bar{c}_n)/\bar{c}_n$ and $c_n(\overline{c_n\tau_n}/\bar{c}_n - \tau_n)/\bar{c}_n$. S_{crc} is the cross spectral density of those variables, as defined by (93). It has been assumed, in addition, that S_{crc}, like S_{cc} in the former condition, does not vary significantly in the frequency range covered by the phase-transfer function. Only strong dependence of c_n and τ_n may deteriorate this condition. As will be explained in Section VII, dependence of c_n and τ_n is caused by the mechanism of interpulse interference.

V. JITTER

Equation (17) expresses the fluctuating part of the phase error or jitter. In this form, however, it hardly produces relevant information for the circuit designer. Let us therefore discuss the formula in more detail. In Appendix II an expression for the jitter has been derived in spectral terms. For its mean-square value, which stabilizes for large N, we can write

$$\overline{\Delta\theta_e^2} = \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} S_{bb}(l\Omega)\,|H_w(l\Omega)|^2. \tag{25}$$

$S_{bb}(l\Omega)$ is the spectral density of the random variable b_n or, in other terms, the discrete Fourier transform of its autocorrelation function $R_{bb}(k)$, as defined by (90). $H_w(l\Omega)$ is the discrete Fourier transform of the pulse response of the phase-transfer function of the system, as defined by (94).

Let us consider some particular cases.

1) Assume that c_n and τ_n are mutually statistically independent. This is true, for instance, if the nonlinear preprocessing is ideally performed according to the threshold method stated in Section III. We may then distinguish two contributions to the jitter: first, $\overline{\Delta\theta_{ec}^2}$ due to c_n, and second, $\overline{\Delta\theta_{e\tau}^2}$ due to τ_n:

$$\overline{\Delta\theta_e^2} = \overline{\Delta\theta_{ec}^2} + \overline{\Delta\theta_{er}^2}. \quad (26)$$

Let us first examine $\overline{\Delta\theta_{ec}^2}$. From (25) and the definition of b_n as given in (18), we obtain

$$\overline{\Delta\theta_{ec}^2} = \theta_{eo}(1)^2 \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} S_{cc}(l\Omega) \mid H_w(l\Omega) \mid^2 \quad (27)$$

in which $S_{cc}(l\Omega)$ is the spectral density of $(c_n - \bar{c}_n)/\bar{c}_n$, which is the discrete Fourier transform of its autocorrelation function $R_{cc}(k)$, as defined by (87).

2) Equation (27) can be evaluated further if c_n are statistically independent values. Then

$$\Delta\theta_{ec}^2 = \theta_{eo}(1)^2 R_{cc}(0) \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} \mid H_w(l\Omega) \mid^2. \quad (28)$$

Using (19), we write for (28)

$$\overline{\Delta\theta_{ec}^2} = \theta_{eo}(1)^2 R_{cc}(0) B_L T. \quad (29)$$

By analogy to a resonant circuit, we define for a general transfer function an effective quality factor Q_{eff} as

$$Q_{\text{eff}} = \frac{\pi}{T} \frac{1}{4B_L} \quad (30)$$

so that

$$\Delta\theta_{ec}^2 = \frac{\theta_{eo}(1)^2}{4} \frac{\pi}{Q_{\text{eff}}} R_{cc}(0). \quad (31)$$

A similar result has been found by Bennett [2] for the resonant circuit.

3) In addition, let c_n be a symbol out of a series of marks and spaces, and let Pr denote the probability for a mark; then

$$R_{cc}(0) = \frac{1 - \text{Pr}}{\text{Pr}}. \quad (32)$$

It is evident that the jitter vanishes if Pr $\to$ 1 because the signal is then fully deterministic. For very small values of Pr, the condition of sufficient clock content no longer holds, and an exact calculation of the mean-square jitter requires numerical computation of the system equation (12).

4) Let us now consider $\Delta\theta_{er}^2$.

Mutual independence of c_n and τ_n results in

$$\overline{\Delta\theta_{er}^2} = \frac{4\pi^2}{T^2} \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} S_{crcr}(l\Omega) \mid H_w(l\Omega) \mid^2 \quad (33)$$

in which $S_{crcr}(l\Omega)$ is the spectral density of the random variable $(c_n/\bar{c}_n)(\overline{c_n\tau_n}/\bar{c}_n - \tau_n)$.

The time domain expression can be derived directly from (17):

$$\overline{\Delta\theta_{er}^2} = \frac{4\pi^2}{T^2} \frac{1}{\bar{c}_n^2} \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \sum_{m=0}^{N} \overline{c_n c_m (\tau_n - \bar{\tau}_n)(\tau_m - \bar{\tau}_n)} \cdot h_w(NT - nT) h_w(NT - mT). \quad (34)$$

If, in addition, c_n are statistically independent, (34) reduces to

$$\overline{\Delta\theta_{er}^2} = \frac{4\pi^2}{T^2} \frac{\overline{c_n^2}}{\bar{c}_n^2} \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \sum_{m=0}^{N} (\tau_n - \bar{\tau}_n)(\tau_m - \bar{\tau}_n) \cdot h_w(NT - nT) h_w(NT - mT) \quad (35)$$

or in spectral terms

$$\overline{\Delta\theta_{er}^2} = \frac{4\pi^2}{T^2} \frac{\overline{c_n^2}}{\bar{c}_n^2} \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} S_{rr}(l\Omega) \mid H_w(l\Omega) \mid^2 \quad (36)$$

in which $S_{rr}(l\Omega)$ is the spectral density of the random variable $(\tau_n - \bar{\tau}_n)$.

In the special case of zero-mean independent values for τ_n, we find an analogous expression as found for the resonant circuit [1], [2]:

$$\overline{\Delta\theta_{er}^2} = \frac{4\pi^2}{T^2} \frac{\overline{c_n^2}}{\bar{c}_n^2} \overline{\tau_n^2} \frac{\pi}{2Q_{\text{eff}}}. \quad (37)$$

The causes for the origin of τ_n, as catalogued in Section IV, however, do not generally provide statistical independence for τ_n.

VI. SECOND-ORDER PLL

Let us consider the results obtained for a second-order PLL. The closed- and open-loop transfer functions in general notation are given by [11]

$$H_w(s) = \frac{s\omega_n(2\zeta - \omega_n/K_v) + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (38)$$

$$G_w(s) = \frac{s\omega_n(2\zeta - \omega_n/K_v) + \omega_n^2}{s^2 + s(\omega_n/K_v)\omega_n}. \quad (39)$$

$H_w(l\Omega)$ and $G_w(l\Omega)$ may be regarded as sampled values of $H_w(j\omega)$ and $G_w(j\omega)$. These transfer functions are determined by three parameters: the damping factor ζ, the natural frequency ω_n, and the velocity constant K_v. As follows from (13), these parameters are weighted according to the average energy content of the signal, expressed by the expected value c_n. For K_v this weighting is linearly proportional; for ζ and ω_n it is more complicated. The effective quality factor Q_{eff} for the second-order PLL, calculated by (38), (30), and (19), amounts to

$$Q_{\text{eff}} = \frac{2\pi}{T} \frac{\zeta}{\omega_n\{1 + (2\zeta - \omega_n/K_v)^2\}}. \quad (40)$$

This expression is for high-gain loops, i.e., loops for which

$$2\zeta \gg \frac{\omega_n}{K_v} \quad (41)$$

independent of the velocity constant. As seen from (31) and (37), this Q_{eff} controls the jitter. Evaluation of the first contribution to the static phase error results in

$$\theta_{eo}(1) = \frac{\Delta\omega_b}{K_v}. \quad (42)$$

The detuning $\Delta\omega_b$ is not a real constant, but is influenced by temperature, aging, and moisture, which means that

$\theta_{eo}(1)$ is a *quasi*-static error. The influence of the second part $\theta_{eo}(2)$, however, can be eliminated by an equal shift of the reference position. It may appear odd that the loop exhibits this second contribution because, as is commonly known, the response of the phase error on a phase input step should be zero in the stationary condition. The explanation comes from the fact that in our analysis, we have considered the nominal phase error instead of the actual one.

As may be seen from (42), it is apparently possible to reduce the quasi-static phase error $\theta_{eo}(1)$ to an arbitrarily small amount without influencing Q_{eff}. This is not possible for resonant-type timing extraction circuits. The corresponding expression for the quasi-static phase error, as found by Sunde and Bennett [1], [2], is

$$\theta_{eo}(1) = \beta Q \tag{43}$$

with

$$\beta = \Delta\omega_b T/\pi \tag{44}$$

on the assumption that

$$\beta \ll 1, \qquad \beta Q \ll 1.$$

The independence of static and dynamic phase error is, in fact, the most important advantage of the PLL system if compared with the resonant-type system.

On the other hand, faulty performance for worst case signals is possible, due to the modulation of the loop parameters by $\bar{c}_n$.

VII. INFLUENCE OF INTERPULSE INTERFERENCE

Up to now, τ_n has been regarded as an undesired time shift in an idealized model in which it is assumed that the input signal to the PLL consists of pulses short with respect to the digit interval T. In this section, however, we shall drop this assumption, and we shall show that the same model of Fig. 2(c) applies if the effects of general preprocessing, resulting in pulses of arbitrary shapes and duration, are taken into account by appropriately adapted values for τ_n and c_n.

Let $f_1(t)$ represent the waveform of a single pulse of the signal $s_1(t)$ before the nonlinear processing, so

$$s_1(t) = \sum_{n=0}^{N} a_n f(t - nT) \tag{45}$$

in which a_n is a discrete random variable.

The signal $s_2(t)$, after a general nonlinear treatment, is given by

$$s_2(t) = k_1 s_1(t) + k_2 s_1(t)^2 + k_3 s_1(t)^3 + \cdots. \tag{46}$$

The signal $s_2(t)$ corresponds with signal $s_i(t)$ of Fig. 2(a). Because of the integrating properties of the loop, we may approximate the output signal $s_3(t)$ of the multiplier by a series of Dirac-like impulses at interval T:

$$s_3(t) = \sum_{n=0}^{N} s_3(nT)\delta(t - nT). \tag{47}$$

If T is short with respect to the duration of $g(t)$, which generally is true, the output signal $s_o(t)$ of the oscillator will be unaffected by this approximation. The oscillator signal consists of a sequence of elementary waveforms $f_o(t)$ with maximum duration T, which are shifted in phase by an amount $\theta_{ea}^*(nT)$ from the actual phase $\theta_{ia}^*(nT)$ of the fundamental signal component out of $s_2(t)$ on which $s_o(t)$ locks. Therefore, we may write for the sample values

$$s_3(nT) = \int_{-T/2}^{T/2} s_2(nT + \tau) f_o\left(\tau - \frac{T}{2\pi}\theta_{ea}^*(nT)\right) d\tau. \tag{48}$$

Later we shall argue that the actual phase of the fundamental component of $s_2(t)$ will be, in general, equal to the actual phase of the pulses of $s_1(t)$ of (45).

As long as $\theta_{ea}^*(nT) \ll 1$, according to the linearity condition (21), we may approximate (48) by

$$s_3(nT) = \int_{-T/2}^{T/2} s_2(nT + \tau) f_o(\tau)\, d\tau + \frac{T}{2\pi}\theta_{ea}^*(nT) \cdot \int_{-T/2}^{T/2} s_2(nT + \tau) f_o'(\tau)\, d\tau. \tag{49}$$

Defining now

$$c_n = \frac{T}{2\pi}\int_{-T/2}^{T/2} f_o'(\tau) s_2(nT + \tau)\, d\tau \tag{50}$$

$$c_n \cdot \tau_n^* = \frac{T}{2\pi}\int_{-T/2}^{T/2} f_o(\tau) s_2(nT + \tau)\, d\tau \tag{51}$$

and substituting (49)–(51) into (47), we may write for the output signal $s_3(t)$ of the multiplier

$$s_3(t) = \sum_{n=0}^{N} c_n \theta_{ea}^*(nT) T\delta(t - nT) + \sum_{n=0}^{N} \frac{2\pi}{T} c_n \tau_n^* T\delta(t - nT). \tag{52}$$

The model of the PLL for digital signals with interpulse interference can then be evaluated from Fig. 2(c) and (d).

The actual phase $\theta_{ia}^*(NT)$ of the input signal $s_i(t)$ may be written as

$$\theta_{ia}^*(NT) = \Delta\omega_b NT - \frac{2\pi}{T}\tau_N^{**}. \tag{53}$$

Here $\Delta\omega_b$, as before, is the detuning of the quiescent frequency of the VCO from the angular digit frequency $2\pi/T$ and τ_N^{**} are the deviations of the pulses from their nominal positions.

The block diagram of Fig. 2(d) can be reduced again to that of Fig. 2(c) by defining

$$\theta_{ia}(NT) = \theta_{ia}^*(NT) - \frac{2\pi}{T}\tau_N^* \tag{54}$$

$$\theta_{ea}(NT) = \theta_{ia}(NT) - \theta_o(NT). \tag{55}$$

Combining (53) and (54), we get

$$\theta_{ia}(NT) = \Delta\omega_b NT - \frac{2\pi}{T}\tau_N \qquad (56)$$

with

$$\tau_N = \tau_N^{**} + \tau_N^{*}.$$

Apparently, we may use the same model as we had before for short pulses, provided we interpret c_n and τ_n suitably:

1) c_n is a random variable, defined by (50), which is influenced by interpulse interference;

2) τ_n consists of two parts, τ_n^{**} and τ_n^{*}. The first part represents the phase deviations of the incoming pulses. The second part originates from interpulse interference and may be calculated by (51) and (50).

About the actual phases of the input signal $s_1(t)$, defined by (45), and the fundamental signal component $s_2(t)$, it may be stated that these are equal provided

1) interpulse interference is restricted to a limited number of neighboring pulses,

2) neighboring pulses are equidistant, which is true as long as τ_n^{**} has a low-frequency character. The noise contributions to τ_n^{**} may be neglected in practical systems. Other contributions to τ_n^{**} are introduced by preceding repeaters, and will indeed have a low-frequency character because the power spectrum of τ_n^{**} is shaped by phase-transfer functions of those repeaters.

We are now able to calculate the jitter due to intersymbol interference for a single repeater, i.e., if $\tau_n^{**} = 0$. As we have seen, this jitter is fully determined by the following.

1) The properties of the transmission channel. The pulse response of this channel should be known.

2) The statistics of the digital signal. The transmission code should be known. Conditions 1) and 2) are necessary to calculate (45).

3) The nature of the essential nonlinear process. The nonlinearity should be specified to calculate (46).

4) The properties of the PLL filter. A numerical calculation can now be made by applying successively formulas (45), (46), (50), (51), and (17).

As an example, calculated results are shown in Fig. 3 for transmission channels, fulfilling Nyquist's first criterion at the input of the nonlinearity. The exhibited variables are defined as follows:

$\overline{\Delta\theta_{er}^2}$ is the mean-square jitter due to $(c_n/\bar{c}_n)(\overline{c_n\tau_n}/\bar{c}_n - \tau_n)$,

$\overline{\Delta\theta_{ec}^2}$ is the mean-square jitter due to

$$(c_n - \bar{c}_n)\theta_{eo}(1)/\bar{c}_n,$$

ρ is the correlation coefficient of both contributions.

The total mean-square jitter $\overline{\Delta\theta_e^2}$ can be found from

$$\overline{\Delta\theta_e^2} = \overline{\Delta\theta_{er}^2} + \overline{\Delta\theta_{ec}^2} + 2\rho(\overline{\Delta\theta_{er}^2}\cdot\overline{\Delta\theta_{ec}^2})^{1/2}. \qquad (57)$$

From (18) it is obvious that $(\overline{\Delta\theta_{ec}^2})^{1/2}$ can be expressed per radian quasi-static phase error.

Fig. 3(a) and (b) show the influence of the rolloff factor of the channel (rolloff factor as defined in [5, ch. 5]) for two different processes. Fig. 3(c) shows the influence of the effective quality factor Q_{eff} of the PLL.

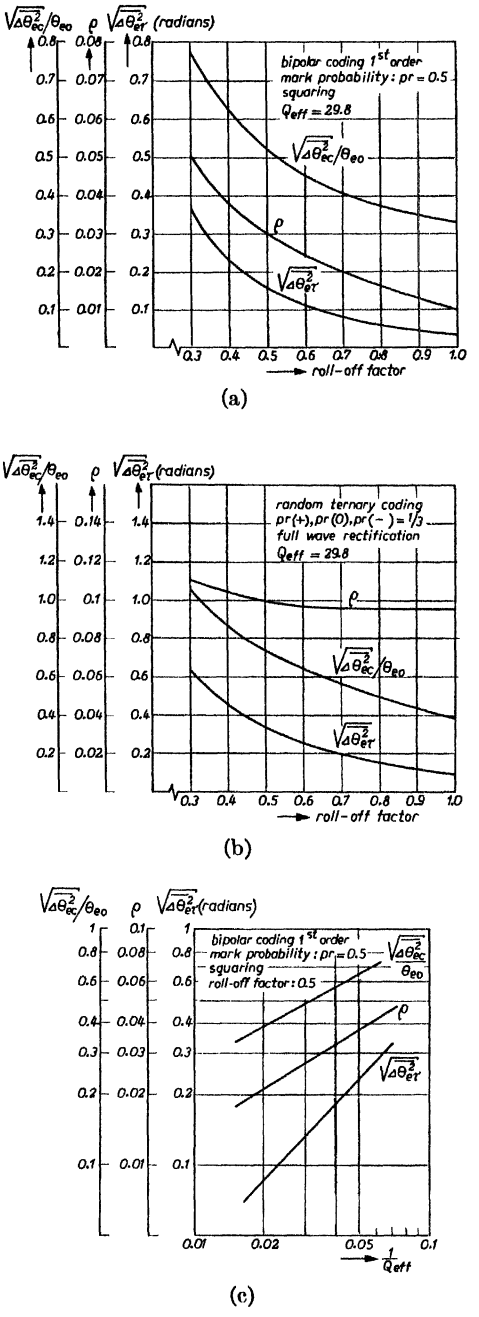


Fig. 3. (a) Jitter due to interpulse interference for Nyquist I channels and squared bipolar signals. (b) Jitter due to interpulse interference for Nyquist I channels and rectified random ternary signals. (c) Influence of effective quality factor on jitter due to interpulse interference.

From the graphs we may conclude the following.

1) Squaring of symmetrical pulses, e.g., Nyquist I pulses, gives no guarantee of jitter-free operation, as is sometimes thought and found in the literature [8].

2) The rms value $(\overline{\Delta\theta_{er}^2})^{1/2}$ is inversely proportional to the effective quality factor. Apparently, this means that the spectral density $S_{c_nc_n}$ of the random variable $(c_n/\bar{c}_n)(\overline{c_n\tau_n}/\bar{c}_n - \tau_n)$ increases linearly with frequency for low frequencies. This can be explained from the fact that

the average interpulse interference does not influence the symmetry of the waveforms. Consequently, the spectral density S_{crcr} is zero for zero frequency.

We may state that the effective quality factor in band-limited systems, where there is a significant influence of interpulse interference, is a more important parameter than suggested by the results of the theory for idealized systems as given by (31) and (37) and the corresponding expressions for idealized resonant-type systems.

3) The rms value $(\overline{\Delta\theta_{ec}^2})^{1/2}$ increases with the square root of the inverse effective quality factor. Apparently, the spectral density S_{cc} of the random variable $(c_n - \bar{c}_n)/\bar{c}_n$ is almost constant for low frequencies, as might be expected.

4) Even for large rolloff factors ($\approx$1), there appears to be a very considerable influence of interpulse interference. Calculating $(\overline{\Delta\theta_{ec}^2})^{1/2}/\theta_{eo}(1)$ for the case of zero interpulse interference by use of (31) yields

$$(\overline{\Delta\theta_{ec}^2})^{1/2}/\theta_{eo}(1) = 0.16, \qquad \text{if } Q_{\text{eff}} = 29.8.$$

Note the difference with the results in Fig. 3(a).

5) The correlation of $\Delta\theta_{ec}$ and $\Delta\theta_{er}$ is neglectably small.

VIII. OPTIMAL PREPROCESSING

In Section III, two methods were mentioned that avoid jitter due to interpulse interference. In this section, we shall prove this statement for the symmetry method.

The equivalent transmission channel up to the non-linearity has a symmetrical transfer function around the Nyquist frequency, yielding a signal

$$s_1(t) = \sum_{n=0}^{N} a_n f(t - nT) \cos \frac{\omega_b}{2} (t - nT). \tag{58}$$

a_n is a discrete random variable. $f(t)$ is the pulse response of the transposed equivalent low-pass filter of the symmetrical transfer function.

$$\omega_b = \frac{2\pi}{T} \tag{59}$$

is the digit frequency.

Squaring the signal $s_1(t)$, we find that

$$\begin{aligned} s_2(t) &= \left\{ \sum_{n=0}^{N} a_n f(t - nT) \cos \frac{\omega_b}{2} (t - nT) \right\}^2 \\ &= \tfrac{1}{2}(\cos \omega_b t + 1) \{ \sum\sum_{n+m=\text{odd}} a_n a_m f(t - nT) f(t - mT) \\ &\quad - \sum\sum_{n+m=\text{even}} a_n a_m f(t - nT) f(t - mT). \end{aligned} \tag{60}$$

Let the spectrum of $f(t)$ be limited in frequency:

$$F(\omega) = 0, \qquad \text{for } \omega > \omega_1. \tag{61}$$

The signal $s_2(t)$ can then be conceived as a superposition of an amplitude-modulated signal with a nominal frequency ω_b whose sidebands reach from $\omega_b - 2\omega_1$ to $\omega_b + 2\omega_1$ and a signal with frequency components below $2\omega_1$. Let this signal be filtered by a symmetrical bandpass filter with angular center frequency ω_b and a bandwidth limited between $\omega_b + \omega_H$ and $\omega_b - \omega_H$. As a result, the zero crossings of the filtered signal have a constant spacing T as long as

$$2\omega_1 < \omega_b - \omega_H. \tag{62}$$

After elimination of the amplitude modulation, a jitterless clock signal is obtained. The proof for the filtering by means of a PLL can be given as follows. In the phase detector, which will be an ideally balanced multiplier, signal $s_2(t)$ of (60) is multiplied by a locked oscillator signal. Suppose, this oscillator signal does not exhibit phase fluctuations, so that it can be represented by $\sin \omega_b t$. As a result, two signal components are produced, one with a spectrum from $2\omega_b - 2\omega_1$ to $2\omega_b + 2\omega_1$, and one with a spectrum from $\omega_b - \omega_1$ to $\omega_b + \omega_1$. As long as the noise bandwidth of the loop is limited to ω_H, fulfilling condition (62), the phase of the VCO will not change.

Summarizing

In order to derive from a band-limited synchronous digital signal a clock signal which does not exhibit jitter due to interpulse interference, the following processes could be performed:

1) a linear process such that the equivalent transmission path has a symmetrical transfer function around half the digit frequency,

2) a nonlinear process such that the signal is squared after process 1),

3) a narrow-band filtering with a noise bandwidth, limited to a lower frequency value than the digit frequency minus the bandwidth of the transmission path of 1).

IX. PROPAGATION OF PHASE ERRORS

The process of jitter propagation in a chain of repeaters has been studied by several authors, including Byrne *et al.* [4], for the clock extraction procedure by a resonant circuit. We shall, in this section, use essentially the same model to derive results for more general processes.

If the phase errors in the input signal have a more or less uniform frequency spectrum in the frequency range, covered by the phase-transfer function $H(j\omega)$ of the narrow-band filter or PLL, and if fluctuations introduced by interpulse interference do not possess pronounced spectral components in this frequency range, it may be stated for the phase of the derived clock signal that

$$\Theta(j\omega) = kH(j\omega) \tag{63}$$

in which k is a constant.

This relation can only be valid for the first repeater in the chain. The second repeater will show in its output signal not only an identical contribution of its own, but also a response on the spectral components, introduced by the first repeater. Because its own contribution is due to the structure of the pattern of digits in the input signal, the fluctuations will add coherently to the fluctuations of the preceding repeater, so that

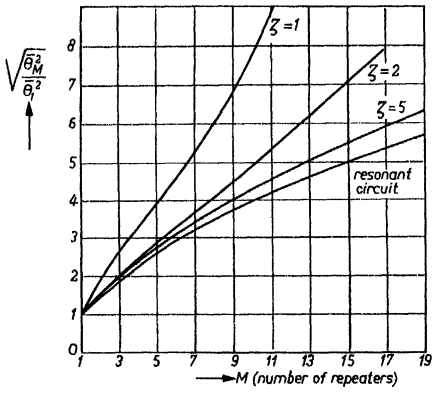


Fig. 4. Normalized absolute jitter as a function of the number of repeaters with high-gain PLL's.

$$\Theta_2(j\omega) = kH(j\omega) + kH(j\omega)\cdot H(j\omega). \qquad (64)$$

The first term represents the second repeater's own contribution, and the second term the response on the spectral components originated from the first repeater. Hence, after M repeaters, we find

$$\Theta_M(j\omega) = kH(j\omega) + kH(j\omega)^2 + \cdots kH(j\omega)^M$$

$$= kH(j\omega)\frac{H(j\omega)^M - 1}{H(j\omega) - 1}. \qquad (65)$$

The mean-square value $\overline{\theta_M^2}$ after M repeaters can then be calculated from

$$\overline{\theta_M^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\Theta_M(j\omega)|^2\, d\omega. \qquad (66)$$

This fluctuation can be denoted as absolute jitter, and has as a reference the nominal position of the digits of the original transmitted signal. This jitter is important because at too large a value, it causes foldover distortion in the decoding process of the digital signal (see [2]).

In Fig. 4 a comparison is given of the jitter propagation for clock extraction by a resonant circuit and for a second-order high-gain PLL. This result has been obtained by a numerical computation of (65) and (66). The phase-transfer function of the second-order PLL is defined by (38). The loop is supposed to be of the high-gain type as long as condition (41) is fulfilled.

It is very clear from the figure that proper performance of a chain of repeaters with high-gain PLL's requires high damping factors.

The exponential growth of the jitter for low damping factors is caused by the fact that in the Bode diagram of the phase-transfer function, the corner frequency of the zero has a lower value than the corner frequencies of the two poles, as is indicated in Fig. 5. Fluctuations with spectral components near the natural frequency are then the more increased the longer the chain is.

This is avoided for a first-order function, like a resonant circuit, a first-order loop, or a degenerated second-order loop with infinite high damping factor [Fig. 5(b)]. If the corner frequency of the zero has a value between those

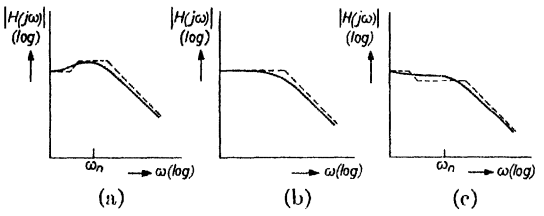


Fig. 5. Phase-transfer function for various types of timing filters. (a) High-gain PLL. (b) Resonant circuit. (c) Special low-gain PLL.

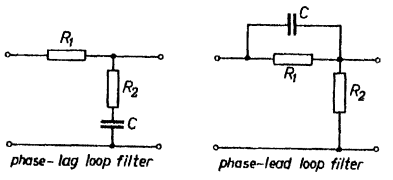


Fig. 6.

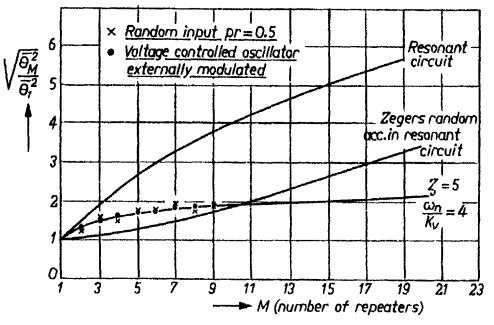


Fig. 7. Normalized absolute jitter as a function of the number of repeaters with well-designed low-gain PLL's.

of the poles, a further improvement in this respect is possible [Fig. 5(c)]. This condition is fulfilled if

$$\zeta - \zeta\left(1 - \frac{1}{\zeta^2}\right)^{1/2} < 2\zeta - \frac{\omega_n}{K_v} < \zeta + \zeta\left(1 - \frac{1}{\zeta^2}\right)^{1/2}. \qquad (67)$$

It can easily be proved that this condition can be obtained by replacing the usual phase-lag loop filter by a phase-lead type (Fig. 6). Of course, the transfer function as a whole remains of the integrating type, due to the influence of the VCO, which is an ideal integrator in the mathematical model. An illustration of possible results is shown in Fig. 7. Even better results are possible than those obtained by Zegers [9], who, in a repeater with a resonant timing extraction circuit, used a scrambler in order to break the coherence of the systematic components.

An experimental chain of ten repeaters has been built to confirm the theoretical expectation. The dots in Fig. 7 represent a test of a chain operating with a fully deterministic signal, while the VCO's in succeeding repeaters are modulated coherently with a sine wave. The crosses are the results of a test with a random digital pattern without external modulation. Whether such an implementation can be used in a practical system depends on the stability and accuracy of the quiescent frequency of the VCO. Normal practice for the design of a second-order phase-

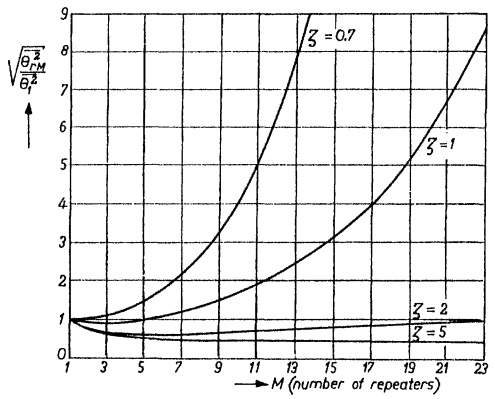


Fig. 8. Normalized relative jitter as a function of the number of repeaters with high-gain PLL's.

locked timing extraction circuit for a chain of repeaters is

1) to fix the required damping factor from propagation properties, as shown in Fig. 7,

2) to fix the natural frequency from a specification of the required noise bandwidth and acquisition time; this is possible by using the figures of Richman [10],

3) to fix the velocity constant such that the static phase error for a given stability of the quiescent VCO frequency is small enough.

An implementation according to the parameters of Fig. 7 eliminates the degree of freedom, as indicated in 3), meaning that the stability and accuracy of the VCO should be high enough to afford the desired velocity constant. Inspection of required parameters has shown that implementation with crystal-stabilized VCO's gives satisfactory results.

Apart from absolute jitter, relative jitter is also important, i.e., fluctuations between the phase of the digits of the input signal of a repeater and the phase of the derived clock signal. This type of jitter influences the error probability in the regeneration process. This jitter may be calculated, as Byrne has done, from the difference of the spectra of phase fluctuations in two succeeding repeaters, i.e.,

$$\Theta_{rM}(j\omega) = \Theta_M(j\omega) - \Theta_{M-1}(j\omega) = kH(j\omega)^M.$$

The propagation of relative jitter for high-gain phase-locked timing extraction circuits is shown in Fig. 8.

X. CONCLUSIONS

Comparing the properties of clock extraction systems, either implemented by a resonant circuit or by a second-order PLL, we conclude the following.

1) Reduction of both quasi-static as well as dynamic phase error is by no means simply possible with resonant-type circuits: the errors are dependent; a compromise is necessary. There is no such dependence in the case of a high-gain second-order PLL.

2) In resonant-type circuits, in contrast to the signal from the PLL, the derived signal is amplitude modulated. This modulation has to be removed by a hard limiter, which may introduce new phase errors. This is the more serious the higher the frequency is and the more the energy density of the pattern varies between wider limits.

3) The phase-transfer function of the resonant circuit is of the first order. The PLL under consideration has a second-order phase-transfer function. To avoid exponential growth of jitter in a chain of repeaters, high damping factors ($\zeta \approx 5$) are necessary for high-gain loops. If low-gain loops can be tolerated, which requires a highly stable VCO, a very small jitter accumulation can be realized by proper dimensioning.

4) The phase-transfer function of the PLL is weighted by the energy density of the signal. Faulty performance in worst case signals is possible.

A method has been given to calculate the influence of interpulse interference on phase errors for the case that the three essential parts of the clock extraction procedure, i.e., prefiltering, nonlinear process, and narrow-band filtering, are properly specified. Although this method has only been derived for the PLL system, it can be proved that it also applies to resonant-type systems. Optimum processing requires a clock path in which prefiltering yields a symmetrical transfer function around the Nyquist frequency in which nonlinearity is a squaring circuit and where narrow-band filtering fulfills certain conditions, as specified in Section VIII.

APPENDIX I

A. Solution of the System Equation

We write for the system equation (12)

$$\theta_e(NT) + \sum_{n=0}^{N} \theta_e(nT) g_w(NT - nT) + \sum_{n=0}^{N} \frac{c_n - \bar{c}_n}{\bar{c}_n} g_w(NT - nT) = \Delta\omega_b NT + \frac{2\pi}{T} \sum_{n=0}^{N} g_w(NT - nT) + \frac{2\pi}{T} \sum_{n=0}^{N} \frac{c_n\tau_n - \overline{c_n\tau_n}}{\bar{c}_n} g_w(NT - nT). \quad (68)$$

We may split $\theta_e(nT)$ into an expected value $\overline{\theta_e(nT)}$ and a zero-mean random part $\Delta\theta_e(nT)$:

$$\theta_e(nT) = \overline{\theta_e(nT)} + \Delta\theta_e(nT). \quad (69)$$

Substitution of (69) into (68) and separation into expected values and zero-mean random contributions yields

$$\overline{\theta_e(NT)} + \sum_{n=0}^{N} \overline{\theta_e(nT)} g_w(NT - nT) = \frac{2\pi}{T} \frac{\overline{c_n\tau_n}}{\bar{c}_n} \sum_{n=0}^{N} g_w(NT - nT) + \Delta\omega_b NT$$

$$- \sum_{n=0}^{N} \overline{\Delta\theta_e(nT) \frac{c_n - \bar{c}_n}{\bar{c}_n}} g_w(NT - nT) \tag{70}$$

and

$$\Delta\theta_e(NT) + \sum_{n=0}^{N} \Delta\theta_e(nT) g_w(NT - nT)$$

$$= \frac{2\pi}{T} \sum_{n=0}^{N} \frac{c_n\tau_n - \overline{c_n\tau_n}}{\bar{c}_n} g_w(NT - nT) - \sum_{n=0}^{N} \overline{\theta_e(nT)}$$

$$\cdot \frac{c_n - \bar{c}_n}{\bar{c}_n} g_w(NT - nT) + \text{term} \tag{71}$$

with

$$\text{term} \equiv - \sum_{n=0}^{N} \left\{ \Delta\theta_e(nT) \frac{c_n - \bar{c}_n}{\bar{c}_n} - \overline{\Delta\theta_e(nT) \frac{c_n - \bar{c}_n}{\bar{c}_n}} \right\}$$

$$\cdot g_w(NT - nT). \tag{72}$$

We shall prove in Section C of this Appendix that the contribution of "term" can be neglected in most cases, and we shall then discuss the conditions for which this is allowed.

Normally, there is a pole in the origin of the open-loop transfer function of a PLL. We may then expect that in the stationary condition the expected error will be a constant, so that

$$\lim_{N\to\infty} \overline{\theta_e(NT)} = \theta_{eo} = \text{constant}. \tag{73}$$

Substituting this in (70), we find

$$\theta_{eo} = \theta_{eo}(1) + \theta_{eo}(2) + \theta_{eo}(3) \tag{74}$$

with

$$\theta_{eo}(1) = \Delta\omega_b NT / [\lim_{N\to\infty} \sum_{n=0}^{N} g_w(NT - nT)] \tag{75}$$

$$\theta_{eo}(2) = \frac{2\pi}{T} \frac{\overline{c_n\tau_n}}{\bar{c}_n} \tag{76}$$

$$\theta_{eo}(3) = -\overline{\Delta\theta_e(NT) \frac{c_N - \bar{c}_N}{\bar{c}_N}} \tag{77}$$

provided

$$\lim_{N\to\infty} \sum_{n=0}^{N} g_w(NT - nT) \gg 1. \tag{78}$$

Condition (78) is fulfilled because of the pole in the origin of the open-loop transfer function. After substitution of the results (73)–(77) into (71), we find for the equation for the random part $\Delta\theta_e(nT)$ neglecting "term," for large N (so that $\overline{\theta_e(nT)}$ has become independent of N), that

$$\Delta\theta_e(NT) + \sum_{n=0}^{N} \Delta\theta_e(nT) g_w(NT - nT)$$

$$= \sum_{n=0}^{N} b_n g_w(NT - nT) \tag{79}$$

in which

$$b_n = - \frac{c_n - \bar{c}_n}{\bar{c}_n} \theta_{eo} + \frac{2\pi}{T} \frac{c_n\tau_n - \overline{c_n\tau_n}}{\bar{c}_n} \tag{80}$$

or, using (74) and (76),

$$b_n = - \frac{c_n - \bar{c}_n}{\bar{c}_n} (\theta_{eo}(1) + \theta_{eo}(3))$$

$$+ \frac{2\pi}{T} \frac{c_n}{\bar{c}_n} \left(\frac{\overline{c_n\tau_n}}{\bar{c}_n} - \tau_n \right). \tag{81}$$

Equation (79) can be conveniently solved by application of the z-transform, giving

$$\Delta\Theta_e(z) + \Delta\Theta_e(z) G_w(z) = \sum_{n=0}^{N} b_n G_w(z) z^{-n} \tag{82}$$

or

$$\Delta\Theta_e(z) = \sum_{n=0}^{N} b_n \frac{G_w(z)}{1 + G_w(z)} z^{-n}. \tag{83}$$

After the inverse transform, the result is

$$\Delta\theta_e(NT) = \sum_{n=0}^{N} b_n h_w(NT - nT), \tag{84}$$

still under the condition of large N.

Herein $h_w(NT)$ is the inverse transform of $H_w(z)$, defined as

$$H_w(z) = \frac{G_w(z)}{1 + G_w(z)} \tag{85}$$

which is, in fact, the weighted closed-loop transfer function.

The interrelation of θ_{eo} and $\Delta\theta_e(NT)$ via $\theta_{eo}(3)$ as expressed by (74), (77), and (81) makes a solution of these equations difficult. In Section C of this Appendix we shall prove that $\theta_{eo}(3)$ gives a negligible contribution in practical systems. The solution of the system equation is then given by (74)–(76), (84), and (81) neglecting $\theta_{eo}(3)$.

B. Definitions

For the next discussions, we shall use discrete Fourier pairs of some variables. The time interval $(N + 1)T$ will be regarded as the fundamental period. The corresponding fundamental frequency is then given by

$$\Omega = \frac{2\pi}{(N + 1)T}. \tag{86}$$

We define discrete Fourier pairs of the following.

1) Autocorrelation Functions:

$$R_{cc}(kT) = \overline{\left(\frac{c_n - \bar{c}_n}{\bar{c}_n} \right) \left(\frac{c_{n+k} - \bar{c}_n}{\bar{c}_n} \right)}$$

$$S_{cc}(l\Omega) = \sum_{k=0}^{N} R_{cc}(kT) \exp(-j\Omega Tlk)$$

$$R_{cc}(kT) = \frac{1}{N+1}\sum_{l=0}^{N} S_{cc}(l\Omega)\exp(j\Omega Tkl). \quad (87)$$

[c_n has been defined first by (1)].

Analogously,

$$R_{\tau\tau}(kT) = \overline{(\tau_n - \bar{\tau}_n)(\tau_{n+k} - \bar{\tau}_n)}, \qquad \tau_n \text{ cf. (1)}$$

$$S_{\tau\tau}(l\Omega) \Leftrightarrow R_{\tau\tau}(kT) \quad (88)$$

$$R_{c\tau c\tau}(kT) = \overline{\frac{c_n}{\bar{c}_n}\left(\frac{\overline{c_n\tau_n}}{\bar{c}_n} - \tau_n\right)\frac{c_{n+k}}{\bar{c}_n}\left(\frac{\overline{c_n\tau_n}}{\bar{c}_n} - \tau_{n+k}\right)}$$

$$S_{c\tau c\tau}(l\Omega) \Leftrightarrow R_{c\tau c\tau}(kT) \quad (89)$$

$$R_{bb}(kT) = \overline{b_n b_{n+k}}, \qquad b_n \text{ cf. (81)}$$

$$S_{bb}(l\Omega) \Leftrightarrow R_{bb}(kT) \quad (90)$$

$$R_{\Delta\theta_e\Delta\theta_e}(kT) = \Delta\theta_e(nT)\,\Delta\theta_e(nT + kT), \quad \Delta\theta_e(nT) \text{ cf. (84)}$$

$$S_{\Delta\theta_e\Delta\theta_e}(l\Omega) \Leftrightarrow R_{\Delta\theta_e\Delta\theta_e}(kT) \quad (91)$$

$$R_{(c\Delta\theta_e)^2}(kT) = \overline{\left(\frac{c_n - \bar{c}_n}{\bar{c}_n}\right)\Delta\theta_e(nT)\left(\frac{c_{n+k} - \bar{c}_n}{\bar{c}_n}\right)\Delta\theta_e(nT + kT)}$$

$$S_{(c\Delta\theta_e)^2}(l\Omega) \Leftrightarrow R_{(c\Delta\theta_e)^2}(kT). \quad (92)$$

2) Cross Correlation Function:

$$R_{c\tau c}(kT) = \overline{\frac{c_n}{\bar{c}_n}\left(\frac{\overline{c_n\tau_n}}{\bar{c}_n} - \tau_n\right)\left(\frac{c_{n+k} - \bar{c}_n}{\bar{c}_n}\right)}$$

$$S_{c\tau c}(l\Omega) \Leftrightarrow R_{c\tau c}(kT). \quad (93)$$

3) Pulse Response:

$$H_w(l\Omega) \Leftrightarrow h_w(nT). \quad (94)$$

C. Constraints

In this section, we shall discuss the conditions for which the solution of the system equation, as derived in Section *A*, is true.

1) First we discuss the irrelevance of $\theta_{eo}(3)$ in (80).

From (84) and (77) it follows that

$$\theta_{eo}(3) = -\overline{\frac{c_N - \bar{c}_N}{\bar{c}_N}}\sum_{n=0}^{N} b_n h_w(NT - nT). \quad (95)$$

Defining a new variable $k = N - n$, we obtain

$$\theta_{eo}(3) = -\sum_{k=0}^{N}\overline{\frac{c_N - \bar{c}_N}{\bar{c}_N}}\, b_{N-k} h_w(kT). \quad (96)$$

Using (81), (87), and (93), this can be written as

$$\theta_{eo}(3) = \sum_{k=0}^{N}\left[R_{cc}(-kT)\{\theta_{eo}(1) + \theta_{eo}(3)\} - \frac{2\pi}{T}R_{c\tau c}(-kT)\right]h_w(kT). \quad (97)$$

By application of (87), (93), and the well-known orthogonality relationships [13]

$$\sum_{k=0}^{N}\exp(j\Omega Tlk) = N + 1, \qquad \text{if } l = 0 \pmod{N+1}$$

$$= 0, \qquad \text{otherwise} \quad (98)$$

we find

$$\theta_{eo}(3) = \theta_{eo}(3a) + \theta_{eo}(3b) \quad (99)$$

with

$$\theta_{eo}(3a) = \{\theta_{eo}(1) + \theta_{eo}(3)\}\frac{1}{N+1}\sum_{l=0}^{N} S_{cc}(l\Omega)H_w(l\Omega) \quad (100)$$

$$\theta_{eo}(3b) = -\frac{2\pi}{T}\frac{1}{N+1}\sum_{l=0}^{N} S_{c\tau c}(l\Omega)H_w(l\Omega). \quad (101)$$

From (100) we conclude that $\theta_{eo}(3a)$ may be neglected with regard to $\{\theta_{eo}(1) + \theta_{eo}(3b)\}$ if

$$\frac{1}{N+1}\sum_{l=0}^{N} S_{cc}(l\Omega)H_w(l\Omega) \ll 1. \quad (102)$$

$\theta_{eo}(3b)$ is neglectable in (81) if the last term of the right-hand part of (81) is much smaller than the contribution of $\theta_{eo}(3b)(c_n - \bar{c}_n)/\bar{c}_n$. Let us compare these quantities in terms of their standard deviations. $\theta_{eo}(3b)$ is neglectable provided

$$\theta_{eo}(3b)[R_{cc}(0)]^{1/2} \ll \frac{2\pi}{T}[R_{c\tau c\tau}(0)]^{1/2}. \quad (103)$$

Substituting (101), we may write

$$[R_{cc}(0)]^{1/2}\frac{1}{N+1}\sum_{l=0}^{N} S_{c\tau c}(l\Omega)H_w(l\Omega) \ll [R_{c\tau c\tau}(0)]^{1/2}. \quad (104)$$

Let us discuss the conditions (102) and (103). In practical cases, $S_{cc}(l\Omega)$ and $S_{c\tau c}(l\Omega)$ do not change significantly in the frequency range covered by $H_w(l\Omega)$. We may write then for (102)

$$S_{cc}(0)\frac{1}{N+1}\sum_{l=0}^{N} H_w(l\Omega) \ll 1. \quad (105)$$

Also

$$\frac{1}{N+1}\sum_{l=0}^{N} H_w(l\Omega) = h_w(0). \quad (106)$$

Therefore (105) reduces to

$$S_{cc}(0)h_w(0) \ll 1. \quad (107)$$

Using result (134) of Appendix III, we may bound (107) by

$$S_{cc}(0)(2\pi B_L T)^{1/2} \ll 1. \quad (108)$$

Analogously, we may evaluate (104) to

$$[R_{cc}(0)]^{1/2}S_{c\tau c}(0)(2\pi B_L T)^{1/2} \ll [R_{c\tau c\tau}(0)]^{1/2}. \quad (109)$$

Very small values of $\bar{c}_n$ deteriorate condition (108). Therefore we denote (108) as the condition of sufficient clock content.

From the definition of $R_{crc}(kT)$ as given by (93), we may conclude that in the case of mutually statistical independence of c_n and τ_n, $R_{crc}(kT)$ is zero. In that case, (109) is always fulfilled. Because of the fact that mutual statistical dependence is caused by the mechanism of interpulse interference as described in Section VII, we denote (109) as the condition of restricted interpulse interference.

2) In this section, we shall motivate the neglect of "term" in (71). The influence of "term" can be estimated by considering it as a second-order effect. Substituting results (84) and (77) into "term" and again solving (71) for the stationary condition ($N \rightarrow \infty$) yields a contribution $\Delta\theta_e(NT)_a$ in addition to $\Delta\theta_e(NT)$ as given by (84). Let us assume that $\Delta\theta_e(NT)_a \ll \Delta\theta_e(NT)$ and investigate whether this assumption can be verified. We may write for the total dynamic phase error $\Delta\theta_e(NT)_t$

$$\Delta\theta_e(NT)_t = \Delta\theta_e(NT) + \Delta\theta_e(NT)_a \tag{110}$$

with

$$\Delta\theta_e(NT) = \sum_{n=0}^{N} b_n h_w(NT - nT) \tag{111}$$

$$\Delta\theta_e(NT)_a = \sum_{n=0}^{N} \left[\frac{c_n - \bar{c}_n}{\bar{c}_n} \Delta\theta_e(nT) - \theta_{eo}(3) \right] \cdot h_w(NT - nT). \tag{112}$$

In the previous section, we already proved that $\theta_{eo}(3)$ is neglectable with regard to b_n, provided conditions (108) and (109) are fulfilled. For (112) there remains a relevant contribution:

$$\Delta\theta_e(NT)_b = \sum_{n=0}^{N} \frac{c_n - \bar{c}_n}{\bar{c}_n} \Delta\theta_e(nT) h_w(NT - nT). \tag{113}$$

Its variance can be written, using Appendix II and (92), as

$$\lim_{N\to\infty} \overline{\Delta\theta_e(NT)_b^2} = \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} S_{(c\Delta\theta_e)^2}(l\Omega) \mid H_w(l\Omega) \mid^2 \tag{114}$$

in which $S_{(c\Delta\theta_e)^2}(l\Omega)$ is the spectral density of the random variable $\Delta\theta_e(nT)(c_n - \bar{c}_n)/\bar{c}_n$.

Analogously, we may write for the variance of $\Delta\theta_e(NT)$ of (111)

$$\lim_{N\to\infty} \overline{\Delta\theta_e(NT)^2} = \lim_{N\to\infty} \frac{1}{N+1} \sum_{l=0}^{N} S_{bb}(l\Omega) \mid H_w(l\Omega) \mid^2 \tag{115}$$

in which $S_{bb}(l\Omega)$ is the spectral density of b_n, as defined by (90). The spectral density of $\Delta\theta_e(NT)$ is limited by $H_w(l\Omega)$. Supposing the condition of sufficient clock content (108) to be fulfilled, we conclude that the spectal density of $(c_N - \bar{c}_N)/\bar{c}_N$ has a very small intersection with the spectral density of $\Delta\theta_e(NT)$, so that these random variables may be regarded to be statistically independent. Then considering (92), (87), and (91),

$$R_{(c\Delta\theta_e)^2}(kT) = R_{cc}(kT) R_{\Delta\theta_e\Delta\theta_e}(kT), \tag{116}$$

and consequently

$$S_{(c\Delta\theta_e)^2}(l\Omega) = S_{cc}(l\Omega) * S_{\Delta\theta_e\Delta\theta_e}(l\Omega) \tag{117}$$

in which * denotes a convolution.

Also, from (115)

$$S_{\Delta\theta_e\Delta\theta_e}(l\Omega) = S_{bb}(l\Omega) \mid H_w(l\Omega) \mid^2. \tag{118}$$

After substitution of (117) and (118) into (114), we find

$$\lim_{N\to\infty} \overline{\Delta\theta_e(NT)_b^2} = \lim_{N\to\infty} \frac{1}{N+1} \cdot \sum_{l=0}^{N} \{ S_{cc}(l\Omega) * S_{bb}(l\Omega) \mid H_w(l\Omega) \mid^2 \} \cdot \mid H_w(l\Omega) \mid^2. \tag{119}$$

Assuming that $S_{cc}(l\Omega)$ is almost constant and equal to $S_{cc}(0)$ in the frequency range covered by $H_w(l\Omega)$, we may write, using (19),

$$\lim_{N\to\infty} \overline{\Delta\theta_e(NT)_b^2} = 2\pi B_L T S_{cc}(0) \lim_{N\to\infty} \frac{1}{N+1} \cdot \sum_{l=0}^{N} S_{bb}(l\Omega) \mid H_w(l\Omega) \mid^2. \tag{120}$$

Comparing now (120) and (115), we may conclude that

$$\lim_{N\to\infty} \overline{\Delta\theta_e(NT)_b^2} \ll \lim_{N\to\infty} \overline{\Delta\theta_e(NT)^2} \tag{121}$$

if the condition (108) of sufficient clock content is fulfilled. Consequently, our initial assumption $\Delta\theta_e(NT)_a \ll \Delta\theta_e(NT)$ is also allowed, and therefore "term" in (71) may be neglected.

APPENDIX II

Jitter in Spectral Terms

For the following evaluation the author is indebted to H. van den Elzen.

From (17) it follows that

$$\Delta\theta_e(NT) = \sum_{n=0}^{N} b_n h_w(NT - nT). \tag{122}$$

Hence

$$\overline{\Delta\theta_e(NT)^2} = \sum_{n=0}^{N} \sum_{m=0}^{N} \overline{b_n b_m} h_w(NT - nT) h_w(NT - mT). \tag{123}$$

Let

$$m = n + k. \tag{124}$$

Then

$$\overline{\Delta\theta_e(NT)^2} = \sum_{n=0}^{N}\sum_{m=0}^{N} \overline{b_n b_{n+k}} h_w(NT - nT) \cdot h_w(NT - nT - kT). \quad (125)$$

Using (90) and (94) we find

$$\overline{\Delta\theta_e(NT)^2} = \frac{1}{(N+1)^2}\sum_{n=0}^{N}\sum_{k=-n}^{N-n} R_{bb}(kT)\sum_{l_1=0}^{N} H_w(l_1\Omega) \cdot \exp\left[j\Omega T(N-n)l_1\right]\sum_{l_2=0}^{N} H_w(l_2\Omega) \cdot \exp\left[j\Omega T(N-n-k)l_2\right]. \quad (126)$$

After a change of the order of summation and some rearranging, this becomes

$$\Delta\theta_e(NT)^2 = \frac{1}{(N+1)^2}\sum_{l_1=0}^{N}\sum_{l_2=0}^{N} H_w(l_1\Omega)H_w(l_2\Omega) \cdot \sum_{n=0}^{N}\exp\left[-j\Omega Tn(l_1+l_2)\right]\sum_{k=-n}^{N-n} R_{bb}(kT) \cdot \exp\left[-j\Omega Tkl_2\right]. \quad (127)$$

Assume for the moment that $R_{bb}(kT)$ is a periodic function with period $(N+1)T$; then

$$\sum_{k=-n}^{N-n} R_{bb}(kT)\exp\left(-j\Omega Tkl_2\right) = \sum_{k=0}^{N} R_{bb}(kT)\exp\left(-j\Omega Tkl_2\right) = S_{bb}(l\Omega) \quad (128)$$

which is the discrete Fourier transform of the autocorrelation function $R_{bb}(kT)$.

Using (128) and the orthogonality relationship of (98), we write for (127)

$$\overline{\Delta\theta_e(NT)^2} = \frac{1}{N+1}\sum_{l=0}^{N} S_{bb}(l\Omega)H_w(l\Omega)H_w(-l\Omega) \quad (129)$$

or

$$\overline{\Delta\theta_e(NT)^2} = \frac{1}{N+1}\sum_{l=0}^{N} S_{bb}(l\Omega)\mid H_w(l\Omega)\mid^2. \quad (130)$$

By extension of the series to $N \to \infty$, the condition of periodicity of $R_{bb}(kT)$ becomes trivial. Therefore

$$\overline{\Delta\theta_e^2} = \lim_{N\to\infty}\overline{\Delta\theta_e(NT)^2} = \lim_{N\to\infty}\frac{1}{N+1}\sum_{l=0}^{N} S_{bb}(l\Omega)\mid H_w(l\Omega)\mid^2. \quad (131)$$

APPENDIX III

Relation Between B_L and $h_w(0)$

The relation between B_L and $h_w(0)$ can be found from Cauchy's inequality, which can be written in the general form

$$\left[\sum_{l=0}^{N} a_l b_l\right]^2 \le \sum_{l=0}^{N} a_l^2 \sum_{l=0}^{N} b_l^2. \quad (132)$$

Taking $a_l = H_w(l\Omega)$ and $b_l = 1$, we find

$$\left[\sum_{l=0}^{N} H_w(l\Omega)\right]^2 \le (N+1)\sum_{l=0}^{N}\mid H_w(l\Omega)\mid^2. \quad (133)$$

Using (106) and (19), it follows that

$$h_w(0)^2 \le 2\pi B_L T. \quad (134)$$

ACKNOWLEDGMENT

The author wishes to thank N. A. M. Verhoeckx for his contributions resulting from many discussions, J. J. Martony for the experimental work, and members of T. F. S. Hargreaves' advanced development team for their stimulating questions and discussions.

REFERENCES

[1] E. D. Sunde, "Self-timing regenerative repeaters," *Bell Syst. Tech. J.*, vol. 36, pp. 891–937, July 1957.
[2] W. R. Bennett, "Statistics of regenerative digital transmission," *Bell Syst. Tech. J.*, vol. 37, pp. 1501–1542, Nov. 1958.
[3] J. M. Manley, "The generation and accumulation of timing noise in PCM systems—An experimental and theoretical study," *Bell Syst. Tech. J.*, vol. 48, pp. 541–613, Mar. 1969.
[4] C. J. Byrne, B. J. Karafin, and D. B. Robinson, "Systematic jitter in a chain of digital regenerators," *Bell Syst. Tech. J.*, vol. 42, pp. 2679–2714, Nov. 1963.
[5] W. R. Bennett and J. R. Davey, *Data Transmission.* New York: McGraw-Hill, 1965.
[6] C. J. Byrne, "Properties and design of the phase controlled oscillator with a sawtooth comparator," *Bell Syst. Tech. J.*, vol. 41, pp. 559–602, Mar. 1962.
[7] B. R. Saltzberg, "Timing recovery for synchronous binary data transmission," *Bell Syst. Tech. J.*, vol. 46, pp. 593–622, Mar. 1967.
[8] Y. Takasaki, "Timing extraction in baseband pulse transmission," *IEEE Trans. Commun.*, vol. COM-20, pp. 877–884, Oct. 1972.
[9] L. E. Zegers, "The reduction of systematic jitter in a transmission chain with digital regenerators," *IEEE Trans. Commun. Technol.*, vol. COM-15, pp. 542–551, Aug. 1967.
[10] D. Richman, "Color-carrier reference phase synchronization accuracy in NTSC color television," *Proc. IRE*, vol. 42, pp. 106–133, Jan. 1954.
[11] T. M. Gardner, *Phase Lock Techniques.* New York: Wiley, 1966.
[12] P. J. van Gerwen, "On the generation and application of pseudo-ternary codes in pulse transmission," *Philips Res. Rep.*, vol. 20, pp. 469–484.
[13] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "The finite Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 77–85, June 1969.

Optimization of Phase-Locked Loop Performance in Data Recovery Systems

Ramon S. Co, *Member, IEEE*, and J. H. Mulligan, Jr., *Life Fellow, IEEE*

Abstract— **Optimized design conditions are presented for a phase-locked loop (PLL) used as a functional block in data recovery systems with the primary function of timing recovery. A mathematical model is presented which takes into account the nonlinear and discrete-time nature of the PLL when used in data recovery applications. Performance attributes for these systems such as acquisition, tracking, and noise are considered. A systematic design procedure is presented which permits quantitative trade-offs among these performance attributes. The validation of the mathematical model and the systematic design procedure on a practical circuit implementation in CMOS technology is described.**

I. Introduction

THERE is an ever-increasing need for digital data transmission and recovery. Digital signals are less susceptible than analog signals to noise and are compatible with the rapid advancements in digital technologies and digital signal processing techniques. Digitally encoded speech, for example, can be transmitted over a long distance with almost no degradation in signal quality. Central to any data transmission and recovery system is the recovery of the timing (or the clock) of the digital information. Data recovery systems have found widespread use in the digital telephone network (Bell T1 carrier system) [1]–[3], local area networks (Ethernet, Token Ring, FDDI) [3]–[5], and in disk drive data storage systems.

A simple mechanization of a digital transmission system is one wherein the digital information is transmitted in the form of rectangular pulses. The presence of a pulse, for instance, signifies binary one, and the absence of a pulse signifies binary zero. At the receiving end, the clock signal is recovered from the received pulses, and the edges of the recovered clock are used to sample the received pulses to determine the values of the binary information. Most often, the incoming pulses would have been propagated through a nonideal channel. The bandwidth limitation of the channel as well as the noise induced on the channel can impose severe restrictions on how well the clock can be recovered from the pulses, and subsequently, on how well the correct information can be detected at the receiving end.

The data and clock recovery process is typically performed through the use of a feedback control system such as a phase-locked loop [6], [7]. Current approaches [8]–[12] in the analysis of data recovery systems have assumed small error signals and small loop bandwidth so that the control loop can be conveniently modeled as a linear (time-invariant) continuous-time system. Such an analysis cannot be used, in general, when the jitter on the incoming data signal is high. With high jitter on the data signal, the error signal in the PLL is large, and the dynamics of the control loop are highly nonlinear. Such an analysis cannot also be used, in general, to optimize the loop parameters (such as widening the loop bandwidth) in order to satisfy a given performance requirement. In addition, the PLL error signal is not a continuous function of time, but a train of aperiodic rectangular pulses. Moreover, by modeling the VCO as K_0/s, a sinusoidal VCO has been implicitly assumed in the phase domain. For data recovery, one is concerned in the *timing error* between the transitions of the data signal and a rectangular VCO (which produces the clock signal), not the phase error of a sinusoidal input and a sinusoidal VCO. In this paper, a mathematical model that exactly describes the dynamical behavior of PLL in data recovery systems is first developed. Given the exact model, optimization techniques are applied to find the optimal solutions according to the desired performance criteria.

In the design of a phase-locked loop for data recovery systems, one is concerned with a number of performance attributes. For instance, rapid acquisition is highly desirable. Faithful tracking of the input signal in the presence of noise or perturbation is very important for the successful recovery of data. If the recovered clock is to be used as timing for retransmission (such as in a repeater chain), the jitter impressed on the recovered clock should preferably be very small. The optimization of these performance attributes is the subject of this paper. As with any engineering design problem, the optimization of one performance attribute usually leads to the deterioration of another. This paper also provides a systematic design procedure which enables the relative importance of these performance attributes to be traded off against one another.

The results are presented in five sections. Section II contains a discussion of the basic PLL model used in a data recovery system. Section III contains a derivation of the mathematical model of the PLL when used in data recovery applications. The mathematical model is compared with the classical linear PLL model. Section IV describes the most significant performance attributes of a data recovery system, i.e., acquisition, tracking, and noise. In Section V, an objective function which can be used to optimize the performance of the data recovery system is presented. A systematic design procedure is outlined and applied to an illustrative example. In Section VI, there is

Manuscript received October 16, 1989; revised May 27, 1994. This work was conducted while R. S. Co was employed by Western Digital Corporation and a doctoral student at the University of California, Irvine, CA 92717 USA.

R. S. Co is with Pericom Semiconductor Corporation, San Jose, CA USA.

J. H. Mulligan, Jr. is with the Department of Electrical and Computer Engineering, University of California, Irvine, CA 92717 USA.

IEEE Log Number 9404027.

Reprinted from *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 1022-1034, September 1994.

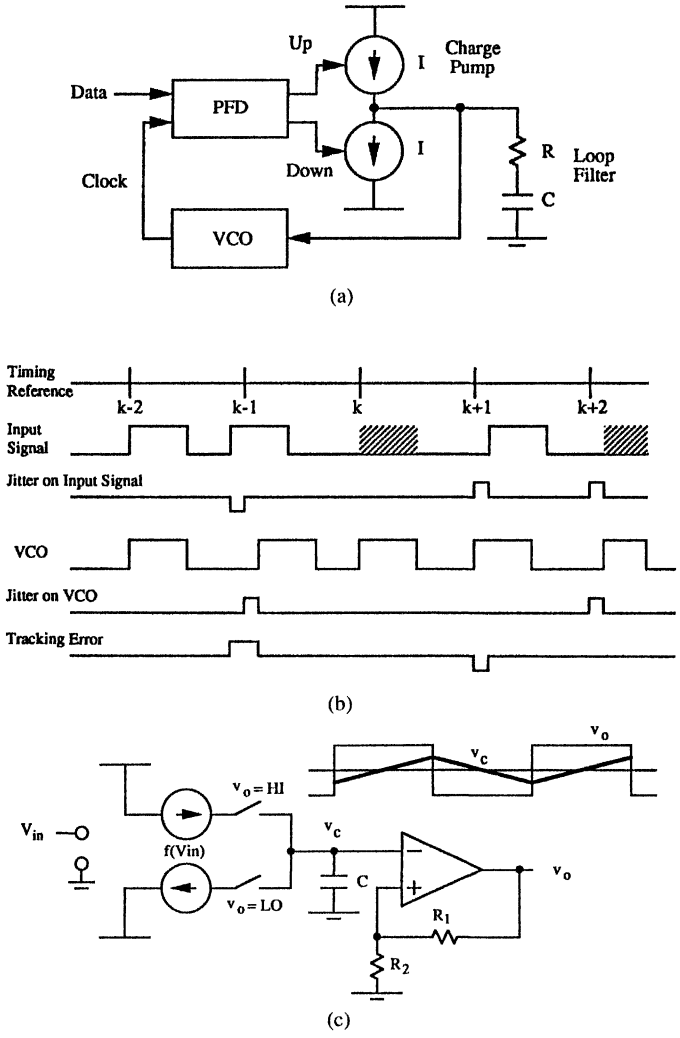


Fig. 1. (a) Phase-locked loop model of data recovery system. (b) Phase-locked loop associated waveforms. (c) Relaxation oscillator circuit diagram.

a description of the validation process for the mathematical model of the data recovery system and the systematic design procedure. The paper concludes with a summary of the principal results, including presentation of the design parameters determined to yield optimum performance.

II. Data Recovery System Model

The PLL model used for the data recovery system is shown in Fig. 1(a), and the associated waveforms are shown in Fig. 1(b). The circuit configuration of the PLL is also known as a *charge-pump* PLL [12]. For the example shown, data one is represented by the presence of a pulse in the first half cycle of the bit period, and data zero is represented by the absence of a pulse (shaded) in the bit period. The positioning of the pulses that represents the data depends on the line coding scheme chosen. In general, the denser the number of pulses, the easier it is to recover the clock since the timing information is imbedded on the occurrence of the pulses. It is these pulses that are tracked by the PLL.

The jitter on the data signal can be described by the time displacement of the positive transitions of the data signal relative to the timing reference k. Each unit of the timing reference is equal to one *bit interval*. The magnitude of the jitter is typically expressed as a fraction of the bit interval. Since a pulse is not present for a data zero, the jitter is measured with respect to the positive transition of a fictitious pulse (indicating the bit boundary). Jitter magnitude can exceed one bit interval. The rate of excursion of the positive transitions of the data signal relative to the timing reference is the *jitter frequency*. The jitter on the VCO signal can be described in the same manner as the data signal jitter. The *tracking error* is the time difference between the positive transitions of the VCO signal and the data signal. It is also equal to the time difference between the VCO signal jitter and the data signal jitter.

A data sampler is generated from the VCO. It is delayed by one quarter of a bit time (for the particular example shown) relative to the VCO output, and its rising edge is used to sample the data signal. It can be seen that if the tracking error exceeds one quarter of a bit time, the data pulse would be sampled incorrectly, and a data error would result. Thus, the probability of error (bit error rate) is determined by the probability in which the tracking error exceeds one quarter of a bit interval. It is to be noted that the data sample point is a function of the line coding scheme used. In NRZ code, for instance, the data pulse is as wide as the bit interval. The optimum sample point is the middle of the bit (which is delayed by half a bit time from the VCO output), and the peak tracking error is one-half of a bit interval.

The Phase/Frequency Detector (PFD) is a *sequential* (as opposed to multiplier) type phase detector as described in [12], [13]. A sequential phase detector can be configured such that a phase comparison is made only whenever there is a data pulse (or a data transition). Whenever there is no data pulse, an error signal is not generated, and the VCO essentially free runs. The PFD in the example shown (see Fig. 1(a) and (b)) is enabled if there is a data one at its input. If the VCO signal lags the data signal, the up signal is activated (meaning the VCO is running too slowly); and if the VCO signal leads the data signal, the down signal is activated (meaning the VCO is running too fast). The duration of the up and down signals is equal to the time difference of the positive transitions of the data signal and the VCO.

The loop filter is a proportional plus integral type, i.e., series RC. The PLL is thus a second order loop. A third order loop is commonly encountered in practice in which a small capacitor is shunted from the filter node to ground. This small capacitor serves to smooth out the voltage developed across the filter resistor R whenever the charge pump (current sources) is turned on and off. The small capacitor together with the filter resistor R form a high frequency pole which is typically placed well beyond the unity gain frequency of the loop so that the loop essentially behaves as a second order loop. It is important to fully understand the behavior of a second order loop. Thus, a simpler loop filter is used for the analysis in this paper.

The VCO is modeled as a relaxation oscillator [14], [15]. A circuit diagram of the relaxation oscillator is shown in Fig. 1(c). The relaxation oscillator belongs to a class of triggered oscillators in which the output switches state when

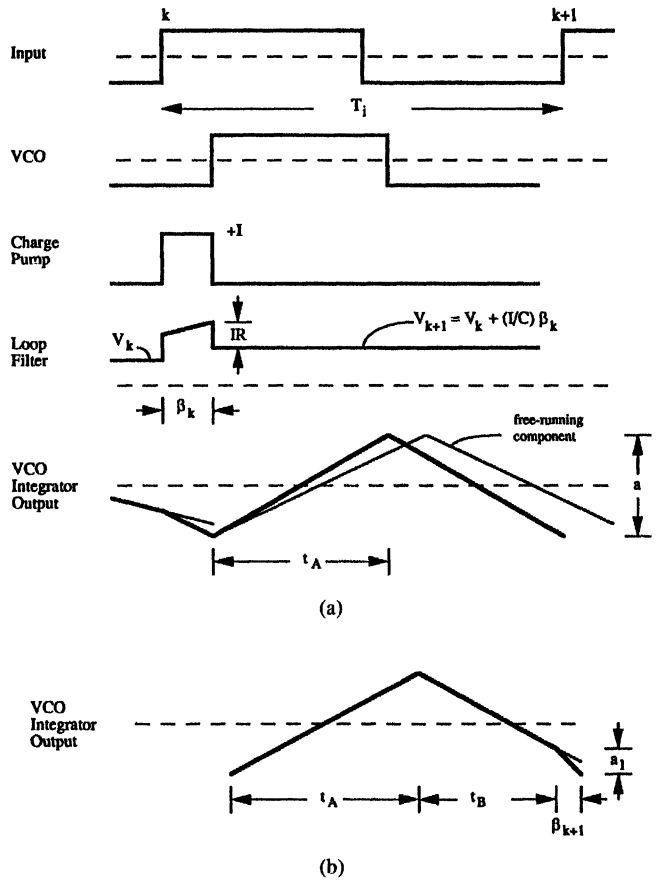


Fig. 2. (a) VCO signal lags input signal at cycle k. (b) VCO signal lags input signal at next cycle $(k+1)$, $T_i < (\beta_k + 2t_A)$.

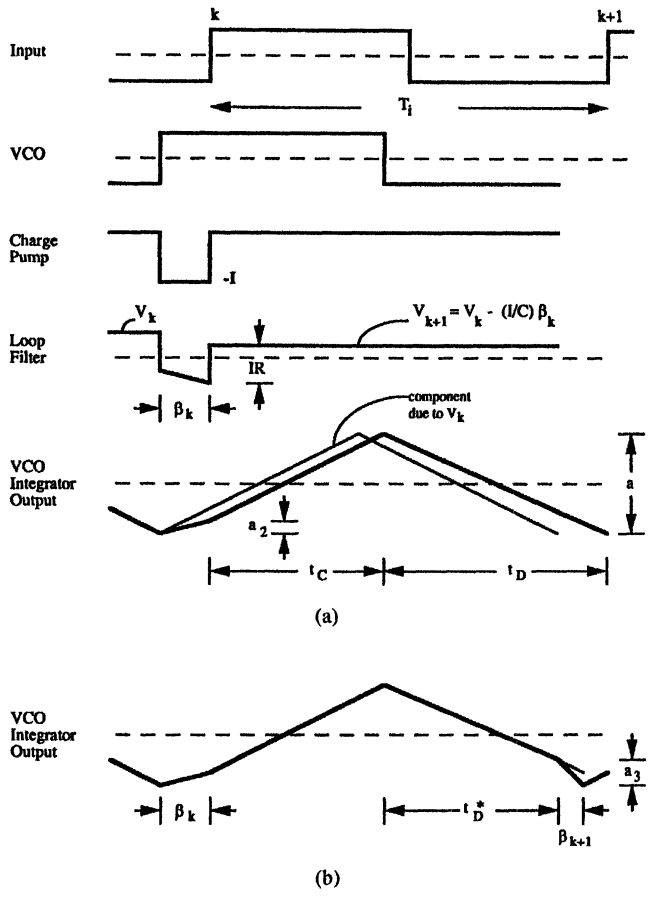


Fig. 3. (a) VCO signal leads input signal at cycle k. (b) VCO signal lags input signal at next cycle $(k+1)$, $T_i < (t_C + t_D)$.

a certain threshold is reached. The output of a relaxation oscillator is naturally a square wave which is required for the clock generation. Its time domain description is also easily formulated.

III. Mathematical Model

A. Formulation of Dynamical Difference Equations

The PLL is a sampled-data system, and its operation can be formulated using a difference equation description. The voltage across the capacitor in the loop filter, and the tracking (zero crossing) error between the data signal and the VCO are taken as the state variables. In the steady-state, the positive transitions of the input signal and the VCO output coincide. Since there are no zero crossing errors, the charge pump is disabled. The output voltage of the loop filter is the voltage across the capacitor C. This is the voltage necessary to hold the VCO frequency equal to the input signal frequency.

To analyze the behavior of the loop in terms of the tracking error sequence, the input to the PLL is assumed to be a train of rectangular periodic pulses. There are two possible conditions at time instant k of the input signal: either the VCO signal lags the input signal or the VCO signal leads the input signal. The waveforms when the VCO signal lags the input signal are shown in Fig. 2(a) and (b), and the waveforms when the VCO signal leads the input signal are shown in Fig. 3(a) and (b).

VCO Signal Lags Input Signal: When the VCO signal lags the input signal, the UP signal of the charge pump is activated by the positive edge of the input signal, and is terminated by the positive edge of the VCO. The duration of the charge pump output is equal to the tracking error at time instant k, and is denoted by β_k. During this time interval, a current of magnitude $+I$ is pumped into the loop filter. If V_k denotes the voltage across the capacitor C in the loop filter prior to the turning on of the charge pump, then the voltage across the capacitor after the charge pump is turned off is equal to

$$V_{k+1} = V_k + \frac{I}{C}\beta_k. \tag{1}$$

Let "a" be the peak-to-peak threshold of the VCO integrator (see capacitor voltage V_c in Fig. 1(c)), then

$$a = (m_0 + k_i V_{k+1}) t_A$$

where m_0 is the slope associated with the free-running frequency of the VCO, and k_i is the integration constant associated with the VCO gain. If we denote the VCO gain K_0 as a fractional change of the VCO free-running frequency per unit (volt) of input, then

$$k_i = K_0 m_0.$$

Also,

$$a = m_0 \frac{T_0}{2}$$

where T_0 is the VCO free-running period. Hence,

$$t_A = \frac{m_0 T_0/2}{m_0 + K_0 m_0 V_{k+1}} = \frac{T_0/2}{1 + K_0 V_{k+1}}. \quad (2)$$

From Fig. 2(a), it can be seen that if $(\beta_k + 2t_A)$ is equal to the period of the input signal T_i, then the positive transition of the input signal signal will coincide with the positive transition of the VCO signal at the next timing reference $k+1$. If $T_i < (\beta_k + 2t_A)$ then the VCO lags the input at the next positive transition of the input signal (time instant $k+1$), and if $T_i > (\beta_k + 2t_A)$ then the VCO leads the input at the next positive transition of the input signal.

a) *Calculation of β_{k+1} for $T_i < (\beta_k + 2t_A)$:* Referring to Fig. 2(b), the charge pump is reactivated at time instant $k+1$, and is held enabled (pumping a current of $+I$) until the VCO integrator reaches the negative threshold level. During the β_{k+1} time interval, the voltage drop across the loop filter consists of additional contributions from the IR drop across the series resistor and the integrated voltage across the series capacitor. Thus,

$$\begin{aligned} a_1 &= (m_0 + k_i V_{k+1})\beta_{k+1} + k_i \int_0^{\beta_{k+1}} \mathrm{IR}\, dt \\ &\quad + k_i \int_0^{\beta_{k+1}} (I/C)t\, dt \\ &= m_0\beta_{k+1} + K_0 m_0 V_{k+1}\beta_{k+1} + K_0 m_0 \mathrm{IR}\beta_{k+1} \\ &\quad + K_0 m_0 (I/C)\beta_{k+1}^2/2. \end{aligned}$$

Also,

$$t_B = \frac{a - a_1}{m_0 + K_0 m_0 V_{k+1}}. \quad (3)$$

Therefore,

$$T_i = \beta_k + t_A + t_B.$$

After some algebraic manipulations, there is obtained

$$\begin{aligned} &(K_0/2)(I/C)[\beta_{k+1}^2] + (1 + K_0 V_{k+1} + K_0 \mathrm{IR})[\beta_{k+1}] \\ &\quad + (T_i - \beta_k)(1 + K_0 V_{k+1}) - T_0 = 0. \end{aligned} \quad (4)$$

β_{k+1} is the positive real solution to the above quadratic equation.

b) *Calculation of β_{k+1} for $T_i > (\beta_k + 2t_A)$:* When the VCO signal leads the input signal at time instant $k+1$, β_{k+1} is simply

$$\begin{aligned} \beta_{k+1} &= T_i - (\beta_k + 2t_A) \\ &= T_i - \beta_k - T_0/(1 + K_0 V_{k+1}) \end{aligned} \quad (5)$$

since the VCO has already made a transition and is just waiting for the input signal to make its transition.

VCO Signal Leads Input Signal: When the VCO signal leads the input signal, the DOWN signal of the charge pump is activated by the positive edge of the VCO, and is terminated by the positive edge of the input signal. The duration of the charge pump output is equal to the tracking error β_k. During this time interval, a current of magnitude $-I$ is pumped into the loop filter. If V_k denotes the voltage across the capacitor in the loop filter prior to the turning on of the charge pump, then the voltage across the capacitor after the charge pump is turned off is equal to

$$V_{k+1} = V_k - \frac{I}{C}\beta_k. \quad (6)$$

Referring to Fig. 3(a) and using the same notations as defined previously,

$$\begin{aligned} a_2 &= (m_0 + k_i V_k)\beta_{k+1} - k_i \int_0^{\beta_{k+1}} , \mathrm{IR}\, dt \\ &\quad - k_i \int_0^{\beta_{k+1}} , (I/C)t\, dt \\ &= m_0\beta_k + K_0 m_0 V_k \beta_k - K_0 m_0 \mathrm{IR}\beta_k \\ &\quad - K_0 m_0 (I/C)\beta_k^2/2. \end{aligned}$$

Also,

$$t_C = \frac{a - a_2}{m_0 + k_i V_{k+1}} \quad (7)$$

$$t_D = \frac{a}{m_0 + k_i V_{k+1}}. \quad (8)$$

It can be seen that if $T_i = (t_C + t_D)$ then the positive transition of the input signal will coincide with the positive transition of the VCO signal at the next timing reference $k+1$. If $T_i < (t_C + t_D)$ then the VCO lags the input at the next positive transition of the input (time instant $k+1$), and if $T_i > (t_C + t_D)$ then the VCO leads the input at the next positive transition of the input.

a) *Calculation of β_{k+1} for $T_i < (t_C + t_D)$:* Referring to Fig. 3(b), the charge pump is reactivated at time instant $k+1$, and is held enabled (pumping a current of $-I$) until the VCO integrator reaches the negative threshold level. During the β_{k+1} time interval, the voltage drop across the loop filter consists of additional contributions from the IR drop across the series resistor and the integrated voltage across the series capacitor. Thus,

$$\begin{aligned} a_3 &= (m_0 + k_i V_{k+1})\beta_{k+1} + k_i \int_0^{\beta_{k+1}} \mathrm{IR}\, dt \\ &\quad + ki \int_0^{\beta_{k+1}} (I/C)t\, dt \\ &= m_0\beta_{k+1} + K_0 m_0 V_{k+1}\beta_{k+1} + K_0 m_0 \mathrm{IR}\beta_{k+1} \\ &\quad + K_0 m_0 (I/C)\beta_{k+1}^2/2. \end{aligned}$$

Also,

$$t_D^* = \frac{a - a_3}{m_0 + k_i V_{k+1}}.$$

Therefore,

$$T_i = t_C + t_D^*.$$

After some algebraic manipulations, there is obtained

$$\begin{aligned} &(K_0/2)(I/C)[\beta_{k+1}^2] + (1 + K_0 V_{k+1} + K_0 \mathrm{IR})[\beta_{k+1}] \\ &\quad + T_i(1 + K_0 V_{k+1}) - T_0 \\ &\quad + \beta_k(1 + K_0 V_k - K_0 \mathrm{IR}) \\ &\quad - (K_0/2)(I/C)\beta_k^2 = 0. \end{aligned} \quad (9)$$

β_{k+1} is the positive real solution to the above quadratic equation.

b) *Calculation of β_{k+1} for $T_i > (t_C + t_D)$:* When the VCO signal leads the input signal at time instant $k+1$, β_{k+1} is simply

$$\begin{aligned}\beta_{k+1} &= T_i - (t_C + t_D) \\ &= T_i - [T_0 - \beta_k(1 + K_0V_k - K_0\mathrm{IR}) \\ &\quad + (K_0/2)(I/C)\beta_k^2]/(1 + K_0V_{k+1}) \end{aligned} \quad (10)$$

since the VCO has already made a transition and is just waiting for the input signal to make its transition.

In summary,

i) when the VCO lags the input signal at time k

$$V_{k+1} = V_k + \frac{I}{C}\beta_k$$

and

$$t_A = (T_0/2)/(1 + K_0V_{k+1}).$$

a) If $Ti < (\beta_k + 2t_A)$ then the VCO lags the input signal at time $k+1$, and β_{k+1} is the positive real solution of:

$$\begin{aligned}&(K_0/2)(I/C)[\beta_{k+1}^2] \\ &\quad + (1 + K_0V_{k+1} + K_0\mathrm{IR})[\beta_{k+1}] \\ &\quad + (T_i - \beta_k)(1 + K_0V_{k+1}) - T_0 = 0.\end{aligned}$$

b) If $T_i > (\beta_k + 2t_A)$ then the VCO leads the input signal at time $k+1$, and β_{k+1} is computed from:

$$\beta_{k+1} = T_i - \beta_k - T_0/(1 + K_0V_{k+1}).$$

ii) When the VCO leads the input signal at time k

$$V_{k+1} = V_k - \frac{I}{C}\beta_k$$

and

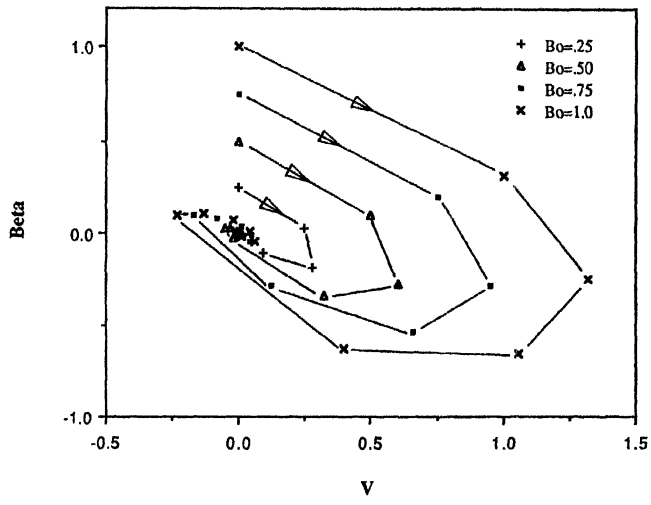
$$\begin{aligned}(t_C + t_D) = [&T_0 - \beta_k(1 + K_0V_k - K_0\mathrm{IR}) \\ &+ (K_0/2)(I/C)\beta_k^2]/(1 + K_0V_{k+1}).\end{aligned}$$

a) If $T_i < (t_C + t_D)$ then the VCO lags the input signal at time $k+1$, and β_{k+1} is the positive real solution of:

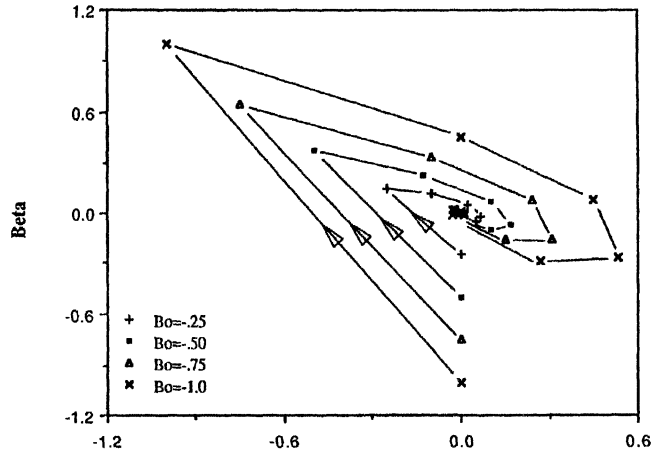
$$\begin{aligned}&(K_0/2)(I/C)[\beta_{k+1}^2] \\ &\quad + (1 + K_0V_{k+1} + K_0\mathrm{IR})[\beta_{k+1}] \\ &\quad + T_i(1 + K_0V_{k+1}) - T_0 \\ &\quad + \beta_k(1 + K_0V_k - K_0\mathrm{IR}) \\ &\quad - (K_0/2)(I/C)\beta_k^2 = 0.\end{aligned}$$

b) If $T_i > (t_C + t_D)$ then the VCO leads the input signal at time $k+1$, and β_{k+1} is computed from

$$\beta_{k+1} = T_i - (t_C + t_D).$$



(a)

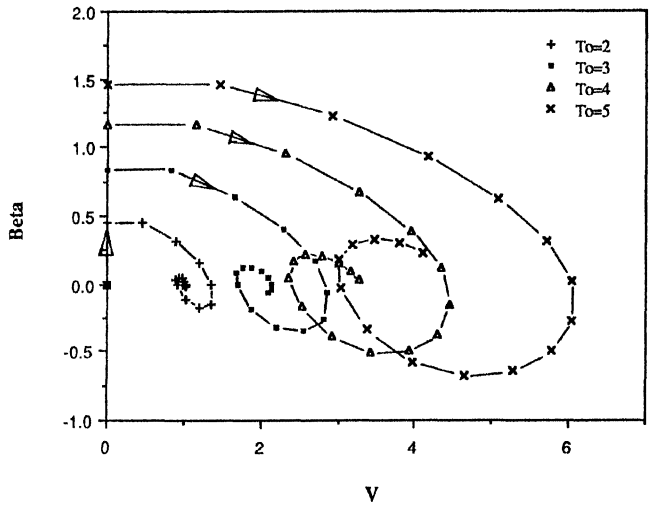


(b)

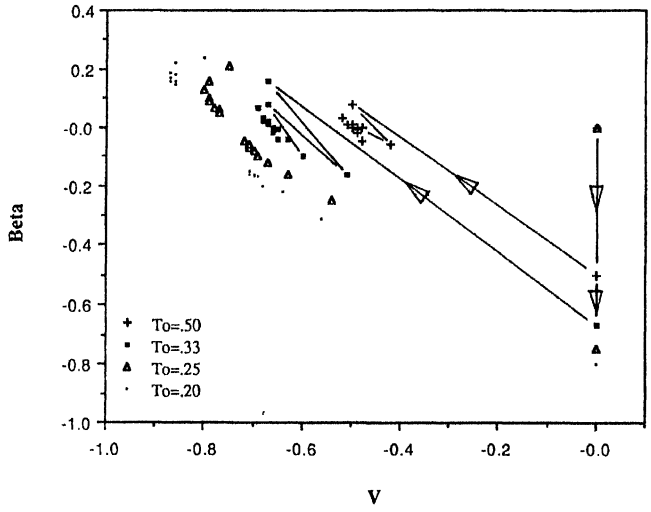
Fig. 4. (a) State trajectories for phase step input ($T_i = T_0 = 1$); direction of trajectory is indicated by arrows. (b) State trajectories for phase step input ($T_i = T_0 = 1$); direction of trajectory is indicated by arrows.

The solutions to the set of difference equations, due to their sequential and recursive nature, can be easily generated with the aid of a computer program. Careful examination of the difference equations (by normalizing the tracking error β and the free-running period T_0 to the input signal period T_i) indicates that the unitless parameters K_0IT_i/C and $K_0\mathrm{IR}$ are sufficient to describe the loop. The two parameters can be further reduced to K_0I/C and $K_0\mathrm{IR}$ by setting T_i equal to unity. This is interesting because K_0I/C is related to the *natural frequency*, and $K_0\mathrm{IR}$ is related to the *unity gain bandwidth* of the loop in linear analysis.

Fig. 4(a) and (b) show the *state trajectories* for phase step input with the tracking error β and the loop filter capacitor voltage V as state variables. Each entry in the trajectory is equivalent to an elapsed time of one clock period or one bit period. The period T_i of the input signal is equal to the free-running period T_0 of the VCO; both are normalized to one. The initial tracking error β_0 is varied (from 0.25 to 1.0 and -0.25 to -1.0) using the normalized loop parameters $K_0I/C = 1$



(a)



(b)

Fig. 5. (a) State trajectories for frequency step input ($T_i = 1$); direction of trajectory is indicated by arrows. (b) State trajectories for frequency step input ($T_i = 1$); direction of trajectory is indicated by arrows.

and $K_0 \mathrm{IR} = 1$. The inherent asymmetry in which the error signals are generated in the loop is evident in the asymmetry of the state trajectories for positive and negative initial tracking errors. The trajectory is the "familiar" spiral shape for positive initial tracking error; whereas, the trajectory is a "pointed" spiral for negative initial tracking errors. Recall that when the input signal leads the VCO signal, the generation of the error signal is causing the VCO transition to occur earlier from its nominal transition time. However, when the VCO signal leads the input signal, the generation of the error signal does not perturb the time of occurrence of the input signal transition (the input signal is the excitation source).

Fig. 5(a) and (b) show the state trajectories for frequency step input. The initial tracking error β_0 is equal to zero. T_i is normalized to one, and T_0 is varied (from 2 to 5 and 0.2 to 0.5) using the same loop parameters as the phase step input.

For either type of input, the steady-state solution (V_{ss}, β_{ss}) is found by setting $V_k = V_{k+1} = V_{ss}$ and $\beta_k = \beta_{k+1} = \beta_{ss}$. Hence,

$$(V_{ss}, \beta_{ss}) = ([T_0/T_i - 1]/K_0, 0). \tag{11}$$

The system is *unstable* if the state trajectories do not converge to the steady-state solution given the initial conditions. In order to determine that the system will be *stable* over the expected range of initial conditions, the state trajectories should be computed once the loop parameters have been established for a tentative design (see Section V for an illustrative example). One can also determine from the state trajectories the *acquisition time* of the system, i.e., the number of clock cycles required for the system to reach and stay within a bounded region of the steady-state equilibrium condition. If so desired, it is also possible to plot the state trajectories for a phase and frequency step input; that is, $\beta_0 \neq 0$ and $T_i \neq T_0$.

B. Comparison With Classical Linear Model

A well-studied model used in the classical analysis of a second order loop is depicted in Fig. 6. The governing equations [13] using Gardner's notations are:

$$\begin{aligned} \omega_n &= (K_0 K_d/\tau_1)^{1/2} = (K_0(K_d/R_1)/C)^{1/2} \\ \zeta &= \frac{\omega_n \tau}{2} = (K_0(K_d/R_1)/C)^{1/2}\frac{RC}{2} \\ K &= 2\zeta\omega_n = K_0\frac{K_d}{R_1}R \end{aligned} \tag{12}$$

where ω_n is the natural frequency, ζ is the damping factor, K is the unity gain bandwidth, K_0 is the VCO gain, and K_d is the phase detector constant. Notice that the quantity $(K_d/R_1)\theta_e$ is the average error current which is driven into the RC feedback elements of the loop filter, where θ_e is the phase error. For a charge pump PLL, the average error current over a bit interval which is pumped into a similar RC loop filter is equal to $(I/2\pi)\theta_e$. Therefore, a linear continuous-time approximation can be made for the charge pump PLL by equating (K_d/R_1) equal to $(I/2\pi)$. Using this transformation, the following equivalent relations are obtained for the charge-pump PLL:

$$\begin{aligned} \omega_n &= (K_0(I/2\pi)/C)^{1/2} \\ \zeta &= (K_0(I/2\pi)/C)^{1/2}\frac{RC}{2} \\ K &= 2\zeta\omega_n = K_0(I/2\pi)R. \end{aligned} \tag{13}$$

These equations are in agreement with those defined by Gardner [12]. The transient response plots comparing the classical model and the mathematical model for a phase step input and a frequency step input are shown in Fig. 7(a) and (b) respectively. Note that the discrepancy in the response of the classical model from the mathematical model is particularly severe for the frequency step input.

C. Dynamical Difference Equations in the Presence of Noise

The difference equations can be modified to include the effect of noise. This is done by perturbing the period of the input data signal at each iteration cycle as shown in Fig. 8. The period of the input data at time instant k is made equal to

$$T_k = T_i + (\Delta T_{k+1} - \Delta T_k) \tag{14}$$

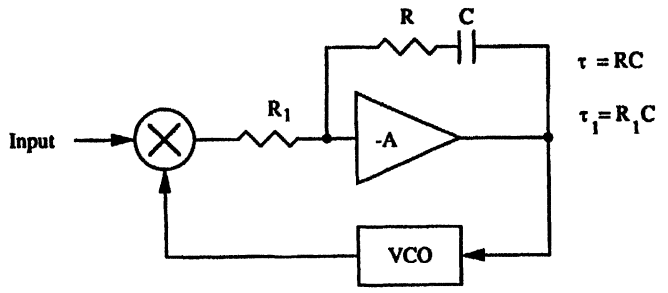
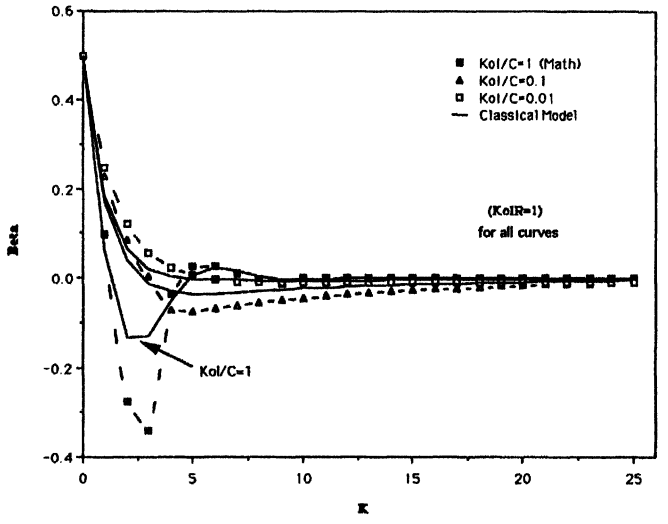
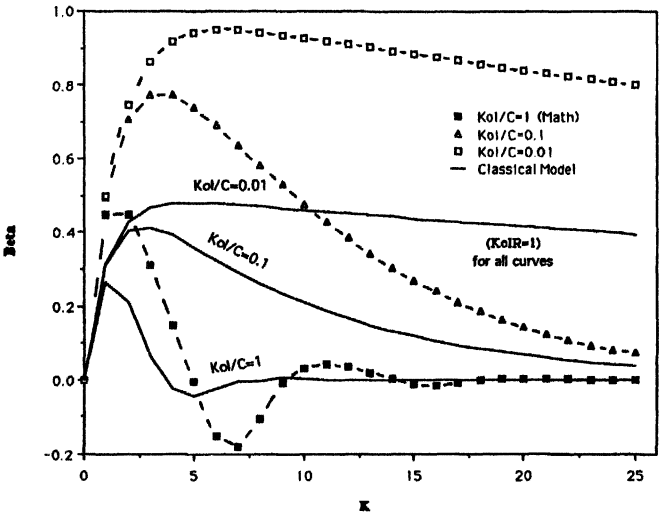


Fig. 6. Linear model of a second order loop.



(a)



(b)

Fig. 7. (a) Phase step input for math model (broken line) and classical model (solid line). (b) Frequency step input for math model (broken line) and classical model (solid line).

where T_i is the nominal period of the input data, and ΔT_k and ΔT_{k+1} are the jitter components of the input data at time instants k and $k+1$ respectively. The tracking error β_k is still the time difference between the positive transitions of the input data signal and the VCO signal, but the jitter on the VCO signal is now equal to $\beta_k + \Delta T_k$. Thus, the modification of the

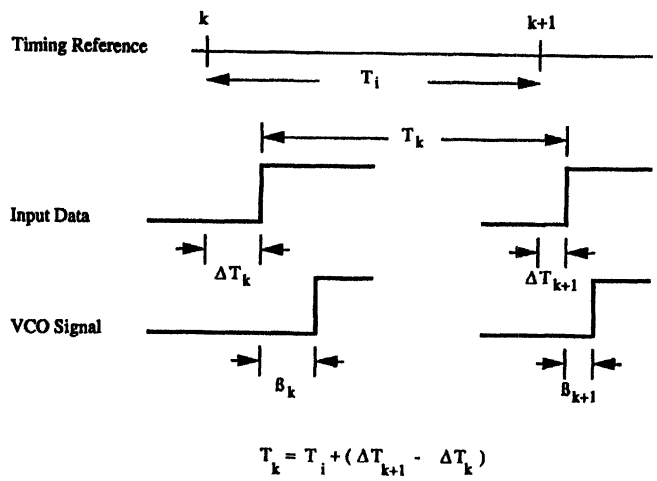


Fig. 8. Signal waveforms in the presence of noise.

difference equations to include the effect of noise is achieved by replacing T_i (a constant) in the original equations by T_k. The modified equations constitute a model for the PLL which is a principal result of this paper.

The statistics of the input data signal jitter ΔT_k are a function of the signal to noise ratio of the input data signal, the characteristics of the transmission medium, the particular data pattern (intersymbol interference), and the specific line code. A detailed investigation of the statistics of ΔT_k is beyond the scope of this paper. It can be shown, however, that if the data pattern is sufficiently random and the transmission channel is sufficiently narrowband, ΔT_k can be adequately modelled as a white noise process with uniform distribution.

Direct computer simulation can be performed on the difference equations if one is interested in finding an optimal design. For the purposes of the simulations, a random number generator (1000 samples) which is uniformly distributed with a peak value of 0.2 (of the normalized bit period) is used to represent the jitter component ΔT_k of the input data signal. The simulations obtained values of the rms VCO jitter (σ_n) and the rms tracking error (σ_t).

Fig. 9 shows the variation of σ_n, the rms VCO jitter (expressed as a fraction of the normalized bit period), as a function of the normalized loop parameters $K_0 I/C$ and K_0IR. It is observed that the VCO jitter is a minimum at $K_0 I/C = 0.001$. No appreciable improvement can be obtained by further decreasing $K_0 I/C$. For each $K_0 I/C$, there is a particular value of K_0IR that corresponds to a local minimum. Fig. 10 shows the variation of σ_t, the rms tracking error (expressed as a fraction of the normalized bit period), as a function of the normalized loop parameters. The tracking error is a minimum at $K_0 I/C = 0.001$. Similarly, no appreciable improvement can be obtained by further decreasing $K_0 I/C$. Again, for each $K_0 I/C$ there is a particular value of K_0IR that corresponds to a local minimum. Further simulations were conducted for different noise levels, i.e., 0.1 peak and 0.3 peak; similar behavior and the same optimum conditions were obtained. Thus, the parameter $K_0 I/C = 0.001$ yields a global optimum because it minimizes both the VCO jitter and the tracking error. There is a particular value of K_0IR that corresponds

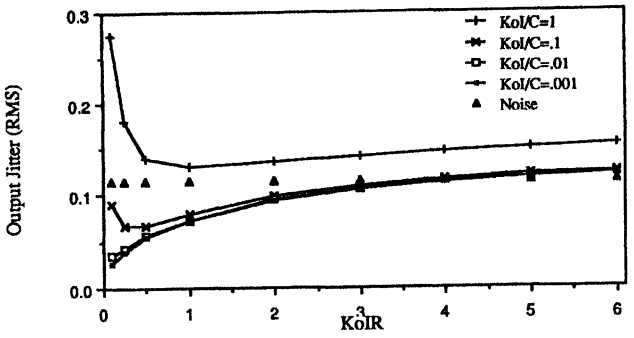


Fig. 9. Plot of output (clock) jitter versus loop parameters (noise = 0.2 peak). (Expressed as fraction of normalized bit period).

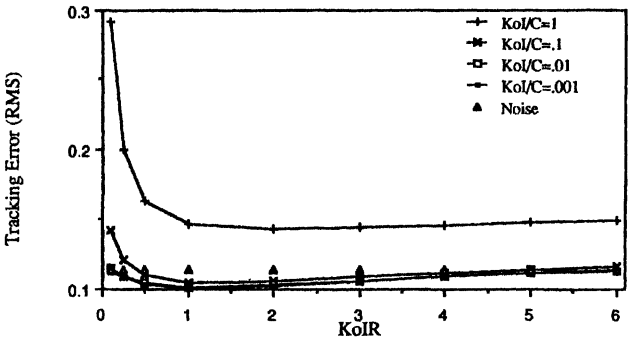


Fig. 10. Plot of tracking error versus loop parameters (noise = 0.2 peak). (Expressed as fraction of normalized bit period).

to a global minimum for VCO jitter and another value that corresponds to a global minimum for tracking error.

D. Dynamical Equations of Different Data Patterns

So far the difference equations that are derived are valid only for a data one followed by a data one pattern, i.e., a data pulse followed by a data pulse. There are three other possible patterns. The difference equations that describe these patterns are derived as follows.

Data Zero Followed by Data Zero Pattern: Under this condition, the charge pump is inhibited and the VCO is free-running at an oscillation frequency determined by the voltage across the capacitor in the loop filter.

a) VCO lags Input at time k.

$$\begin{aligned} V_{k+1} &= V_k \\ \beta_{k+1} &= (\beta_k + 2t_A) - T_i \\ &= \beta_k + T_0/(1 + K_0V_k) - T_i. \end{aligned} \tag{15}$$

β_{k+1} is positive if the VCO lags the input signal at time $k+1$, and β_{k+1} is negative if the VCO leads the input signal at time $k+1$.

b) VCO leads Input at time k.

$$\begin{aligned} V_{k+1} &= V_k \\ \beta_{k+1} &= (\beta_k + T_i) - 2t_A \\ &= \beta_k - T_0/(1 + K_0V_k) - T_i. \end{aligned} \tag{16}$$

Data Zero Followed by Data One Pattern:

a) VCO lags Input at time k.

$$V_{k+1} = V_k. \tag{17}$$

i) For $T_i < (\beta_k + 2t_A)$, VCO lags input at time $k+1$.

$$T_i = \beta_k + t_A + t_B \qquad \text{(see Fig. 2).}$$

And β_{k+1} is the positive real solution of

$$\begin{aligned} &(K_0/2)(I/C)[\beta_{k+1}^2] \\ &\quad + (1 + K_0V_k + K_0\mathrm{IR})[\beta_{k+1}] \\ &\quad + (T_i - \beta_k)(1 + K_0V_k) - T_0 = 0. \end{aligned}$$

ii) For $T_i > (\beta_k + 2t_A)$, VCO leads input at time $k+1$.

$$\begin{aligned} \beta_{k+1} &= T_i - (\beta_k + 2t_A) \\ &= T_i - \beta_k + T_0/(1 + K_0V_k). \end{aligned}$$

b) VCO leads Input at time k.

$$V_{k+1} = V_k. \tag{18}$$

i) For $(\beta_k + T_i) < 2t_A$, VCO lags input at time $k+1$.

$$\begin{aligned} (\beta_k + T_i) &= t_A + t_D^* \qquad \text{(see Fig. 3)} \\ &= [T_0 - \beta_{k+1} - K_0V_k\beta_{k+1} \\ &\quad - K_0\mathrm{IR}\beta_{k+1} - (K_0/2) \\ &\quad \times (I/C)\beta_{k+1}^2]/(1 + K_0V_k). \end{aligned}$$

β_{k+1} is the positive real solution of:

$$\begin{aligned} &(K_0/2)(I/C)[\beta_{k+1}^2] \\ &\quad + (1 + K_0V_k + K_0\mathrm{IR})[\beta_{k+1}] \\ &\quad + (1 + K_0V_k)(\beta_k + T_i) - T_0 = 0. \end{aligned}$$

ii) For $(\beta_k + T_i) > 2t_A$, VCO leads input at time $k+1$.

$$\begin{aligned} \beta_{k+1} &= (\beta_k + T_i) - 2t_A \\ &= \beta_k + T_i - T_0/(1 + K_0V_k). \end{aligned}$$

Data One Followed by Data Zero Pattern:

a) VCO lags Input at time k.
Referring to Fig. 2,

$$\begin{aligned} V_{k+1} &= V_k + \frac{I}{C}\beta_k \\ \beta_{k+1} &= (\beta_k + 2t_A) - T_i \\ &= \beta_k + T_0/(1 + K_0V_k) - T_i. \end{aligned} \tag{19}$$

β_{k+1} is positive if the VCO lags the input signal at time $k+1$, and β_{k+1} is negative if the VCO leads the input signal at time $k+1$.

b) VCO leads Input at time k.
Referring to Fig. 3,

$$\begin{aligned} V_{k+1} &= V_k - \frac{I}{C}\beta_k \\ \beta_{k+1} &= T_i - (t_C + t_D). \end{aligned} \tag{20}$$

β_{k+1} is positive if the VCO leads the input signal at time $k+1$, and β_{k+1} is negative if the VCO lags the input signal at time $k+1$.

IV. System Design Concepts

There are two stages in the data recovery process. The first is the acquisition stage, and the second is the tracking and data detection stage. Acquisition is considered to be complete when the error signal has reached the steady-state (value of zero) and the VCO frequency is equal to the incoming data signal frequency. This is, however, a very stringent definition. Noise is invariably present in the system, and therefore, the error signal is never reducible to zero. An operational definition of acquisition is to consider it to have been completed when the error signal and the VCO control voltage have become equal to or less than a prescribed percentage (usually taken as 2–5%) of the bit period and the steady-state VCO control voltage, respectively. Under this condition, the PLL can start detecting (by sampling) the received signal.

A. Acquisition Performance

To assist in the acquisition of the incoming data signal, the VCO is often pretuned to an accurate local frequency reference, such as provided by a crystal oscillator. This prevents the PLL from false locking to the harmonics or the subharmonics of the incoming data signal. When the data transitions start to occur, the PLL input is switched from the local frequency reference to the incoming data signal. The voltage which is necessary to pull the VCO frequency into the incoming data signal frequency is held temporarily in the storage elements of the loop filter. With a clever design of the VCO, it can be started (so called zero phase start) so that its transitions would immediately coincide with the transitions of the incoming data signal, thereby totally eliminating the acquisition problem.

In some communication systems, information is transmitted as a group of data bits known as packets [3]. The data packet is usually preceded by a preamble. The preamble is a periodic data pattern with distinct harmonic content so that the phase-locked loop can pull into the incoming data signal even without the benefit of pretuning to a local frequency reference. For these particular systems, the acquisition time is limited to the length of the preamble.

The PLL acquisition performance is determined by examining the transient response of the loop subjected to phase step and frequency step inputs. It can be improved by having large values of K_0I/C and $K_0\mathrm{IR}$. A large value of K_0I/C, however, tends to produce an undesirable damped oscillatory transient response, whereas, a large value of $K_0\mathrm{IR}$ tends to reduce the oscillatory tendency and indeed can yield a monotonic transient response.

B. Noise and Tracking Performance

When acquisition is complete, the PLL goes into the second stage of data tracking (and detection). The noise performance is measured by the jitter on the recovered clock, whereas the tracking performance is measured by the tracking error. The latter determines the bit error rate performance of the system.

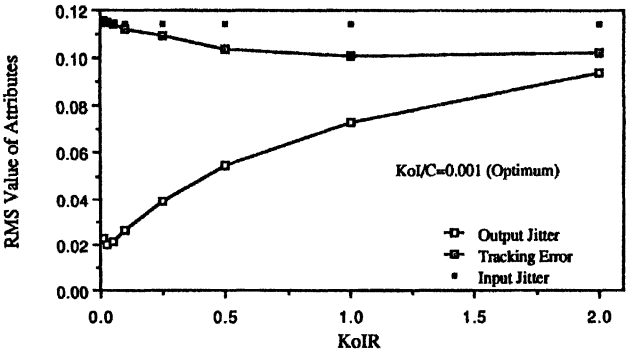


Fig. 11. Dependence of loop performance on $K_0\mathrm{IR}$ at optimum K_0I/C (expressed as fraction of normalized bit period).

It is highly desirable to have minimum values for these two attributes.

It has been found from the extensive simulations that the normalized loop parameter $K_0I/C = 0.001$ yields a global optimum, since it minimizes both the clock (VCO) jitter and the tracking error. Given the optimum K_0I/C, variation of the clock jitter and the tracking error versus the loop parameter $K_0\mathrm{IR}$ is shown in Fig. 11. The graph indicates that the clock jitter is a minimum at $K_0\mathrm{IR} = 0.025$, and the tracking error is a minimum at $K_0\mathrm{IR} = 1$. Thus, a tradeoff would have to be made in the selection of $K_0\mathrm{IR}$. A smaller value of $K_0\mathrm{IR}$ improves the clock jitter but degrades the tracking performance; whereas, a larger value of $K_0\mathrm{IR}$ improves the tracking performance but degrades the clock jitter. However, the loop bandwidth cannot be made arbitrarily small ($K_0\mathrm{IR} \rightarrow 0$) to reduce the clock jitter to zero. With an infinitely small loop bandwidth, the loop has difficulty maintaining lock to an input signal with jitter, and the clock would just drift and wander. The loop bandwidth cannot be made arbitrarily large ($K_0\mathrm{IR} \rightarrow \infty$) either to reduce the tracking error to zero. Aliasing effects become dominant as soon as the loop bandwidth exceeds half the clock frequency.

As a further note, the loop performance is not sensitive to variations in K_0I/C. Actually, there is a wide range of K_0I/C(0.01 to 0.0001) that can be considered to yield optimum performance. The choice of $K_0\mathrm{IR}$ however, is very critical.

V. Performance Optimization

A. Definition of Optimization Problem

An objective function which can be used in the optimization of the PLL performance is

$$\mathcal{F} = \alpha_a t_a + \alpha_t \sigma_t + \alpha_n \sigma_n \tag{21}$$

where α_a, α_t, and α_n are weighting factors, σ_t and σ_n are the tracking and noise performance measures, and t_a is the acquisition performance measure. The optimization problem is the minimization of the objective function $\mathcal{F}$ subject to the conditions that the requirement for each of the performance attributes is met.

Two alternative definitions of the acquisition performance measure t_a have been considered. The *first method* consists of counting the number of bits (clock cycles) until the error

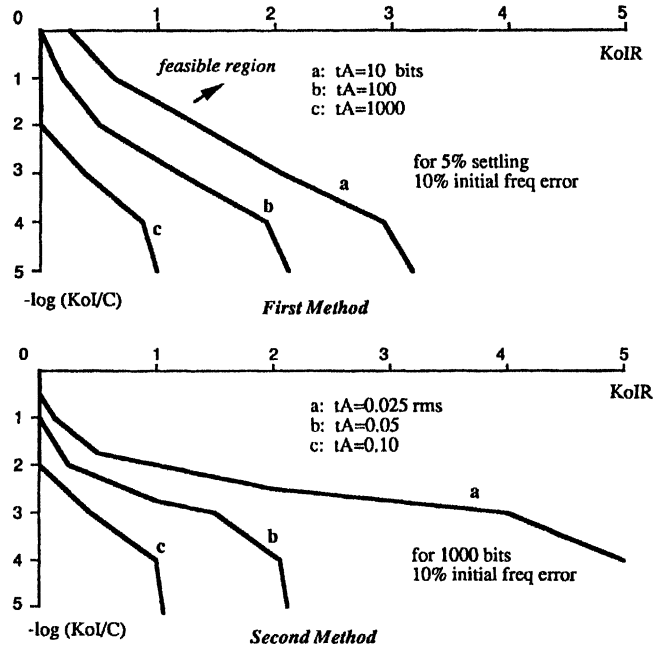


Fig. 12. Acquisition performance objective function contours.

signal and the VCO control voltage have settled to a prescribed amount. The performance measure assigned for t_a using this method has the disadvantage, however, that it has a different dimension compared to the other two attributes (σ_t and σ_n). With this approach, the acquisition time would typically be expressed as a number of bits (clock cycles), whereas the tracking error and clock jitter are expressed as rms values of fractional change of the clock period.

The *second method* consists of minimizing the rms value of the error signal samples up to the desired acquisition time and normalizing to the bit period. This measure has the advantage that the result has the same dimension as the measure of tracking and noise performance. In addition, optimization of the loop parameters with this method results in the optimization of the acquisition performance in the mean square sense. Examples of the plots of the objective function contours based on these two methods are shown in Fig. 12.

B. Existence of a Solution

For most applications, the requirements set on the performance attributes have the form:

$$\begin{aligned} \text{acquisition time} &< A \text{ bits} \\ \text{tracking error} &< T \text{ ns} \\ \text{clock jitter} &< N \text{ ns}. \end{aligned} \tag{22}$$

These restrictions impose a bound on the optimization space, i.e., the loop parameters K_0IR and $K_0 I/C$, for each of the attributes. Each attribute, therefore, has its own *feasible region* [16]. A solution exists if an intersection of the feasible regions can be found.

C. The Design Process

Classes of Design Problems: Several classes of design problems are encountered in practice. These can be summarized as the following distinct possibilities:

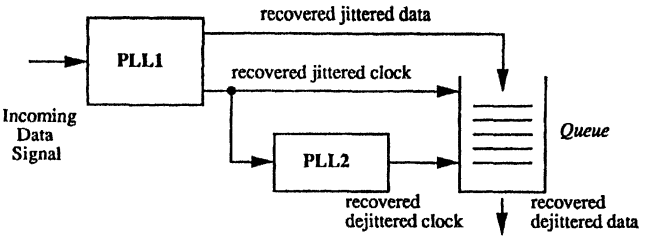


Fig. 13. A queue structure for optimum tracking and jitter.

a) The same set of loop parameters are used for acquisition, tracking, and noise.
b) A set of loop parameters is used for acquisition, and another set is used for tracking and noise. This approach permits independent optimization of the acquisition process using one set of loop parameters, and the optimization of noise and tracking (suitably weighted) with the other set.
c) The VCO is pretuned to a local frequency reference. Only optimization of the loop parameters for data tracking and clock jitter is necessary.
d) A queue is available as shown in Fig. 13 for temporary storage of jittered data. This approach permits optimization of data tracking and clock jitter independently of one another. PLL_1 is optimized for tracking performance for successful recovery of data, and PLL_2 is optimized for jitter performance to provide a jitter-free clock. PLL_1 is used to clock the data into the queue, and PLL_2 is used to clock the data out of the queue.

The first item in the list represents a general problem, and it is the most difficult one to solve. Although the design procedure to be described pertains specifically to this problem, it is also applicable to the rest since they are special cases of the first.

Design Procedure: The design problem is to choose circuit configurations and values of K_0, I, R, and C so that system specifications on acquisition time, tracking error, and clock jitter are satisfied given quantities such as the data rate and the jitter on the incoming data. The procedure includes provision for a trade-off among acquisition, tracking, and jitter performance using preassigned weighting factors (α_a, α_t, and α_n).

For convenience, the procedure is presented in several specific steps:

Step 1: Convert given requirements/constraints into the design variables t_a, σ_t, σ_n, and T_i.

Step 2: Assign values for K_0 and $I \cdot K_0$ is the VCO gain, defined in Section III as the fractional change in VCO center frequency ($1/T$ Hz) per volt of input voltage. The value is based on the designer's estimate of the change that will occur in frequency once a practical oscillator configuration has been selected. I, the current in the charge-pump, is a free variable. Its maximum value is determined by considerations of noise and other fluctuations which limit the accuracy of its specification and realization.

Step 3: Examine performance optimization in terms of the objective function $\mathcal{F}$ and the weights α_a, α_t, and α_n. The feasible region is determined for each of the performance

attributes from the design requirements. The intersection of the feasible regions is the set of possible solutions. A solution is not possible if an intersection cannot be found. If a solution is possible, the intersection of the feasible regions is applied as a set of boundary constraints to the system objective function $\mathcal{F}$. The preassigned weighting factors for the performance attributes are used to find the optimum normalized loop parameters ($K_0IT_i/C = p$ and $K_0\text{IR} = q$).

An alternate to this step is to create more contours for each of the attributes that have smaller values than the desired constraints. Creating a contour with smaller value for a particular attribute narrows the optimization space, and in effect, imposes a more stringent requirement for that particular attribute. Subsequently, an optimal solution will be found that favors the attribute whose objective function contour is reduced.

Step 4: Determine values of R and C using the equations

$$C = (K_0IT_i)/p \qquad \text{and} \qquad R = q/(K_0I).$$

The former equation results from the fact discussed in Section III upon the introduction of the normalized parameter K_0IT_i/C that use of the normalizing condition $T_i = 1$ yields K_0I/C which has been utilized throughout the ensuing development.

Step 5: Compute the state trajectories (see Fig. 4 and 5) over the expected range of initial conditions to ascertain that the loop design is stable and that the transient response is satisfactory.

Illustrative Example: Consider a data recovery system design with the following specifications:

Data Rate:	10 Mbps ($T_i = 100$ ns)
Data Jitter:	±20 ns peak (uniform distribution)
Acquisition Time:	1000 bits
Tracking Error:	11.5 ns rms
Clock Jitter:	8.1 ns rms
Acquisition Performance Measure: (using second method)	5.0 ns rms
Performance Weightings:	$\alpha_a = \alpha_t = \alpha_n = 1$.

The tracking error is chosen such that it is equal to the rms value of the data jitter, i.e., 20 ns peak / $\sqrt{3}$ = 11.5 ns rms (for a uniform distribution, rms value = peak value/$\sqrt{3}$).

The first parameter to be chosen in the design is the VCO gain K_0. It should be large enough to pull the VCO frequency into the incoming data signal frequency. A K_0 of 50% per volt is chosen as the design value based on the choice of a ring oscillator for the VCO. It is assumed that the VCO has a center frequency with 10% accuracy. This implies that the PLL can have an initial frequency step (error) of 10%. Objective function contours for acquisition performance based on the two alternative measures discussed above are shown in Fig. 12. They were computed from the PLL model for a 10% initial error in frequency based on this assumption. The second method will be used in this example. An acquisition performance measure of 5 ns rms implies a normalized rms value of 5/100 = 0.05. Therefore, *contour b* of Fig. 12 is selected to determine the relation between K_0I/C and $K_0\text{IR}$ to achieve the desired acquisition performance.

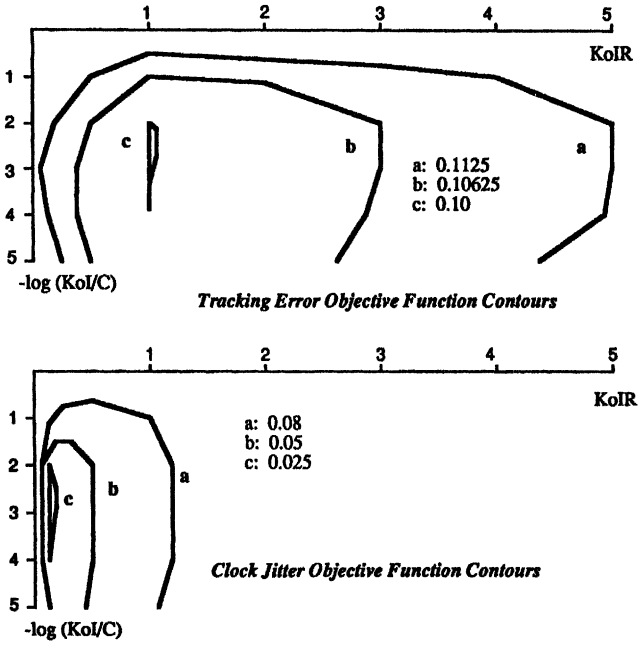


Fig. 14. Tracking and noise performance objective function contours.

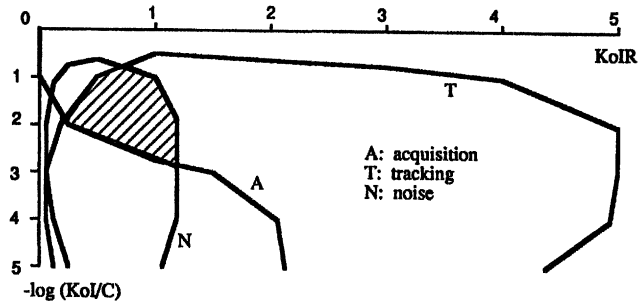


Fig. 15. Intersection of feasible regions for the illustrative example.

Plots of the objective function contours for tracking and noise performance are shown in Fig. 14. *Contour a* for tracking error provides a basis for estimating the corresponding relation between K_0I/C and $K_0\text{IR}$ to meet the 11.5 ns requirement. *Contour a* for clock jitter also provides a comparable estimate of the relation needed to meet the 8.1 ns requirement.

The three selected contours are superimposed and are shown in Fig. 15. The shaded region which is the intersection of the three contours is the set of feasible solutions. This set of feasible solutions is applied as boundary constraints in the minimization of the system objective function $\mathcal{F}$.

The system objective function $\mathcal{F}$ for $\alpha_a = \alpha_t = \alpha_n = 1$ has a minimum at $K_0I/C = 0.01$ and $K_0\text{IR} = 0.5$. These values fall within the boundary constraints depicted in Fig. 15. Hence, $K_0IT_i/C = p = 0.01$ and $K_0\text{IR} = q = 0.5$.

With a tentative circuit implementation of the charge pump in mind, a current of 1.0 mA is selected as the value of I. The component values can now be calculated using

$$C = (0.5)(0.001)(100 \times 10^{-9})/0.01 = 0.005\ \mu\text{F}$$

and

$$R = 0.5/(0.5)(0.001) = 1\ \text{K}\Omega.$$

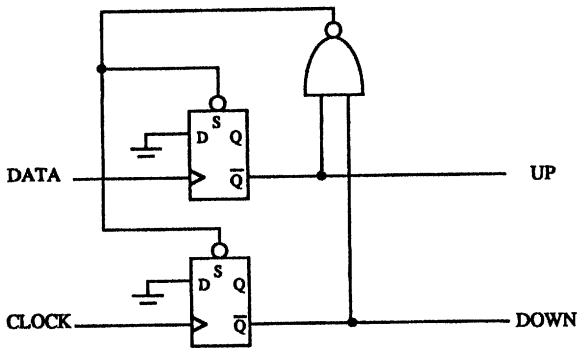


Fig. 16. Phase/frequency detector logic diagram.

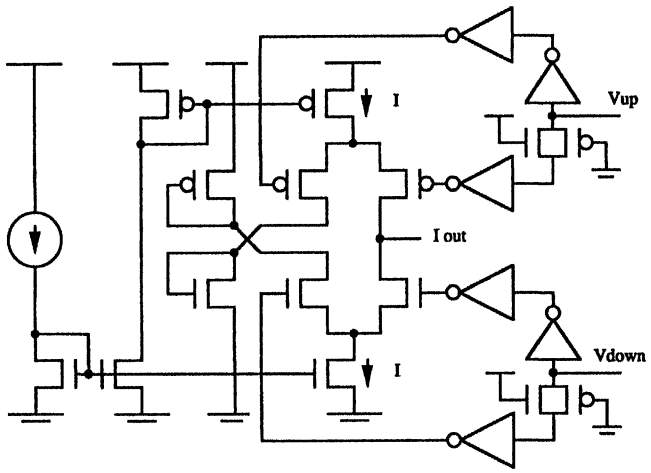


Fig. 17. Charge pump circuit schematic.

The remaining step in the design process is to compute the state trajectories to determine that the loop design is stable and that the transient response is satisfactory. The initial conditions to which the loop will be subjected to are an initial frequency error of 10% and an initial phase error which can be as much as one bit interval.

VI. Validation of Results

The validity of the applicability of the mathematical model for practical system implementation was investigated using a CMOS electronic circuit having the form of Fig. 1(a). The level 7 Lattin-Jenkin-Grove MOSFET model [17] for a 1.25 μm feature size CMOS was used as a transistor model in SPICE to simulate the operation of the circuit. The logic diagram of the phase/frequency detector PFD is shown in Fig. 16, the charge-pump circuit (generator I of Fig. 1(a) schematic is shown in Fig. 17, and the VCO of Fig. 1(a) (implemented as ring oscillator circuit) is shown in Fig. 18. Note that the particular circuit configuration shown in Fig. 16 is a conventional sequential phase detector which is not suited for random data. The circuit is used in the simulation for illustrative purposes only assuming that the input is a periodic pattern.

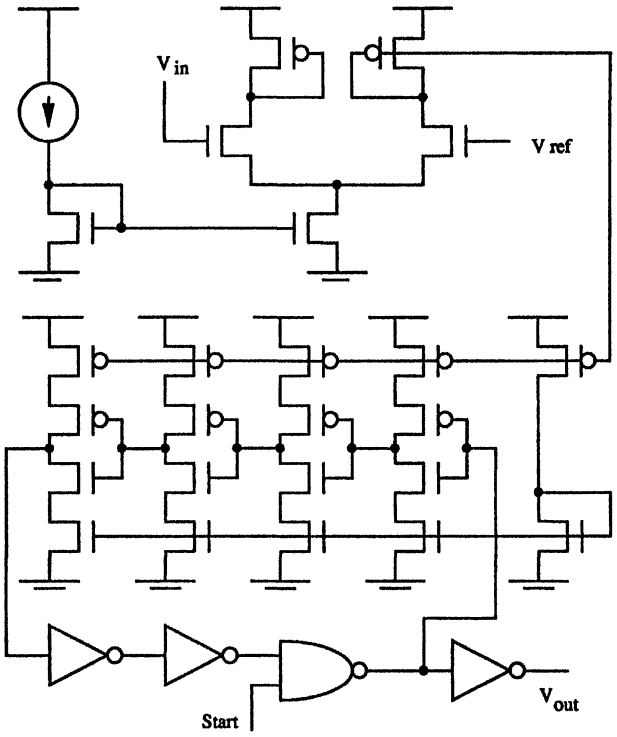


Fig. 18. Ring oscillator circuit.

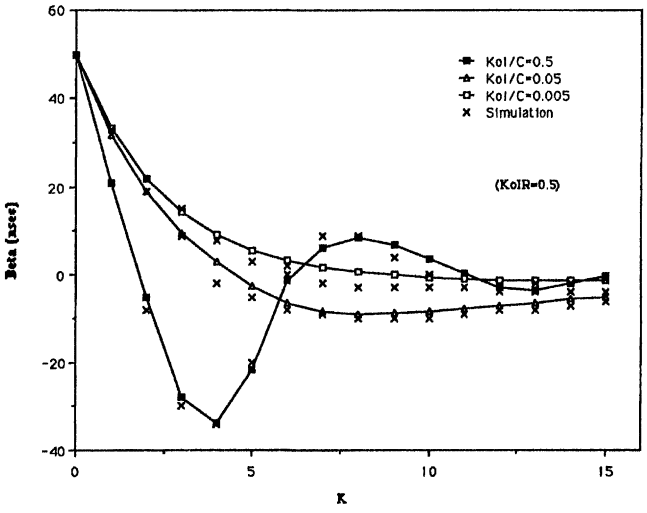


Fig. 19. Phase step input for math model and circuit simulation.

The transient response of the PLL using the CMOS circuitry was simulated for a phase step input and a frequency step input. The results are shown in Fig. 19 and 20. In the figures, the VCO frequency is 10 MHz which corresponds to a period of 100 ns. Excellent agreement between the simulated circuit and the mathematical model was obtained in all the cases.

Jitter was added into the circuit simulations by perturbing the transitions of the data signal, and the tracking error and the clock jitter were noted. These operations were performed for different noise levels and different loop parameters, including the illustrative design example presented above. Excellent agreement between the simulated circuit and the mathematical model was also obtained.

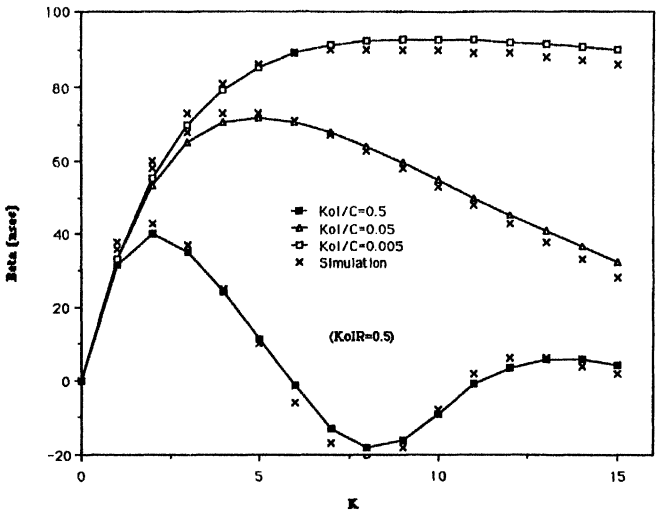


Fig. 20. Frequency step input for math model and circuit simulation.

VII. Conclusion

A mathematical model has been derived for the phase-locked loop which is commonly used in data recovery systems. The formulation of the model is exact and takes into account that a square wave voltage output of the VCO is required. The mathematical model is a difference equation which relates the tracking (zero crossing) error and the voltage across the capacitor of the loop filter at the present bit timing k to the next bit timing $k+1$. The model includes the effect of noise (or jitter) on the input signal.

The three performance attributes of a data recovery system, namely, acquisition, tracking, and noise, are discussed in detail. The optimum loop parameters for each of these attributes are determined from the numerical solutions to the mathematical model. The solutions indicate that the normalized loop parameter $K_0I/C = 0.001$ is optimum for both the tracking performance and the noise performance over a wide practical range of input noise levels. Wide bandwidth (K_0IR ≈ 1) is required for optimum tracking, and narrow bandwidth (K_0IR ≈ 0.025) is required for optimum jitter rejection.

Finally, a systematic design procedure for a PLL for use in data recovery systems is presented. An objective function is given which can be utilized for trade-offs among acquisition, tracking, and jitter performance. A method is also shown for determining the existence of a solution given the design conditions. The mathematical model as well as the design procedure were validated through simulations on a practical CMOS circuit implementation using the SPICE circuit simulation program.

References

[1] F. F. E. Owen, *PCM and Digital Transmission Systems*. New York: McGraw-Hill, 1982.

[2] Pub 62411, "High capacity digital service channel interface specification," *Bell System Techn. Ref.*, Sept. 1983.

[3] A. S. Tanenbaum, *Computer Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1981.

[4] ANSI/IEEE Std 802.3-1985, *Carrier Sense Multiple Access with Collision Detection (CSMA/CD)*. New York: Wiley, 1985.

[5] ANSI/IEEE Std 802.5-1985, *Token Ring Access Method*. New York: Wiley, 1985.

[6] W. D. Llewellyn, M. H. Wong, G. W. Tietz, and P. A. Tucci, "A 33 Mb/s data synchronizing phase-locked loop circuit," *Int. Solid-State Circuits Conf. Tech. Dig.*, San Francisco, CA., Feb. 1988, pp. 12–13.

[7] R. H. Leonowich and J. M. Steininger, "A 45-MHz CMOS phase/frequency-locked loop timing recovery circuit," *Int. Solid-State Circuits Conf. Tech. Dig.*, San Francisco, CA, Feb. 1988, pp. 14–15.

[8] B. Saltzberg, "Timing recovery for synchronous binary data transmission," *Bell Syst. Tech. J.*, pp. 593–622, March 1967.

[9] E. Roza, "Analysis of phase-locked timing extraction circuits for pulse code transmission," *IEEE Trans. Commun.*, vol. 22, pp. 1236–1249, Sept. 1974.

[10] D. L. Duttweiler, "The jitter performance of phase-locked loops extracting timing from baseband data waveforms," *Bell Syst. Tech. J.*, pp. 37–58, Jan. 1976.

[11] J. P. Hein and J. W. Scott, "z-domain model for discrete-time PLL's," *IEEE Trans. Circuits Syst.*, vol. 35, Nov. 1988.

[12] F. M. Gardner, "Charge-pump phase-lock loops," *IEEE Trans. Commun.*, vol. 28, pp. 1849–1858, Nov. 1980.

[13] F. M. Gardner, *Phaselock Techniques*, 2nd Ed. New York: Wiley, 1979.

[14] A. A. Abidi and R. G. Meyer, "Noise in relaxation oscillators," *IEEE J. Solid-State Circuits*, vol. 18, pp. 794–802, Dec. 1983.

[15] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 2nd Ed. New York: Wiley, 1984.

[16] R. L. Fox, *Optimization Methods for Engineering Design*. Reading, MA: Addison-Wesley, 1971.

[17] *HSPICE Users' Manual*, Meta-Software, Inc., Campbell, CA, June 1987.

Noise Properties of PLL Systems

VENCESLAV F. KROUPA

Abstract—**This is a survey paper which begins by the derivation of the general PLL noise equation and by dividing the additive noises into the passband group and the stopband group.**

In the following paragraphs the behavior of all the major sources of additive noises is investigated and the practical numerical values of the respective power spectral densities are given.

In the terminating sections, guidelines for minimizing the additive noises in PLL systems and PLL frequency synthesizers are emphasized, and finally, phase-noise power spectral densities of several actual PLL frequency synthesizers are plotted in the normalized form.

The paper is accompanied by a copious bibliography.

I. INTRODUCTION

THE steadily increasing congestion in communications bands of the electromagnetic spectrum results in efforts for utilization of new ranges (e.g., microwave and optical frequencies) on one hand and better exploitation of the existing frequency allocations on the other hand. The latter task is met by adopting SSB modulation, telegraph multiplexing techniques, etc., and often by sharing the same communication channel by several services. In these cases the mutual interference is the major problem and it can be reduced only by increasing both the short-term and the long-term frequency stability of respective exciters and local oscillators.

Nowadays the long term stability is effectively solved with the assistance of frequency synthesis [1]-[4]. However, the situation with the short-term frequency stability is not so simple. Even in the ideal case when the frequency synthesizer is assumed to be a noiseless frequency transformer, we face an additive frequency noise [1] and often a too high phase-noise level after multiplication. The remedy to this difficulty may be a carefully designed phase-lock loop (PLL). However, this latter technique is more often used in frequency synthesizers only since it makes possible a substantial hardware simplification. And here the troubles start as nearly all PLL building blocks may add a sometimes substantial noise power to the useful signal. The problem is not yet generally understood and, in addition, is often underestimated. We shall therefore, in the following paragraphs, discuss theoretical backgrounds and summarize all the accessible experimental results to provide the leading lines for the design of low-noise PLL systems.

Manuscript received June 5, 1981; revised April 12, 1982. This paper was presented in part at the Third Symposium on Electromagnetic Compatibility, Rotterdam, The Netherlands, May 1979, and at the Conference on Precision Electromagnetic Measurements, Braunschweig, West Germany, June 1980.

The author is with the Institute of Radio Engineering and Electronics, Czechoslovak Academy of Sciences, 182 51 Prague 8, Czechoslovakia.

II. PLL NOISE EQUATION

In the following paragraphs we shall limit ourselves only to noises generated in the PLL building blocks and in a "low-noise" reference, leaving out the large class of cases with the reference frequency embedded in atmospheric or man-made noise which have been extensively dealt with in earlier works, from which many we shall mention only [5]-[7].

In Fig. 1 we have drawn a fairly general PLL arrangement with a phase detector (PD), a low-pass filter $F_L(s)$, and a voltage-controlled oscillator (VCO) in the forward path and a mixer (—), an IF filter with the effective modulation transfer function $F_M(s)$ [1, p. 234] or [5, p. 147], and a divider ($\div N$) in the feedback path. For completeness we have placed a divider ($\div Q$) between the reference generator (RG) and the phase detector and a multiplier (XM) between RG and the second input to the mixer. However, we have to keep in mind that these two latter blocks, in actual PLL systems, are often replaced by more complicated frequency synthesis circuits.

Since all the noises generated or added in individual blocks in Fig. 1 are small compared with the useful signals, we have applied the rule of superposition and simply add them either at the respective inputs or outputs. Note that the subscript "n" indicates the noise signal throughout. Furthermore, the small signal theory makes it possible to use the Laplace transform approach to find the output noise of the considered PLL system or, more exactly, the respective power spectral densities.

By assuming a locked loop and by considering Fig. 1, we may write for the forward path of the loop

$$\phi_{o,n} = [(\phi_{i,n} - \phi_{o,n}')K_d + V_{\mathrm{PD},n} + V_{F,n}]F_L(s) \cdot \frac{K_0}{s} + \phi_{\mathrm{osc},n} \tag{1}$$

and for the feedback path

$$\phi_{o,n}' = (\phi_{o,n} - \phi_{m,n} + \phi_{MI,n})\frac{F_M(s)}{N} + \phi_{DN,n} \tag{2}$$

where

$$\phi_{i,n} = \frac{\phi_{r,n}}{Q} + \phi_{DQ,n} \tag{3}$$

and

$$\phi_{m,n} = M\phi_{r,n} + \phi_{MU,n}. \tag{4}$$

Reprinted from *IEEE Trans. Comm.*, vol. COM-30, pp. 2244-2252, October 1982.

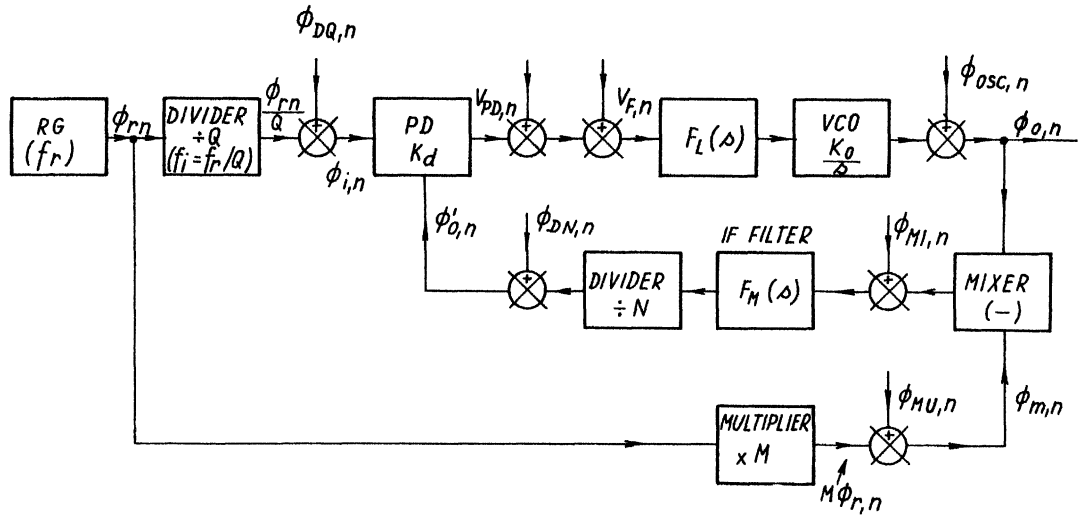


Fig. 1. Block diagram ot a general PLL system with additive noise sources.

Note that all noise components $\phi \cdots, n$ and $V \cdots, n$ are Laplace transformed quantities, i.e., $\phi \cdots, n(s)$, etc.

After introducing (2) into (1) we get for the output phase noise $\phi_{o,n}$

$$\phi_{o,n} = \left[\left(\phi_{i,n} - \phi_{DN,n} + \frac{V_{PD,n} + V_{F,n}}{K_d}\right)\frac{N}{F_M(s)} + \phi_{m,n} - \phi_{MI,n}\right]\frac{F_M(s)F_L(s)K_dK_0/Ns}{1 + F_M(s)F_L(s)K_dK_0/Ns} + \phi_{osc,n}\frac{1}{1 + \frac{F_M(s)F_L(s)K_dK_0}{Ns}} \quad (5)$$

which can be simplified with the assistance of the effective loop transfer function $H'(s)$ [1, ch. 7] and (3) and (4) into

$$\phi_{o,n} = \left[\phi_{r,n}\left(M + \frac{N}{Q}\frac{1}{F_M(s)}\right) + \left(\phi_{DQ,n} - \phi_{DN,n} + \frac{V_{PD,n} + V_{F,n}}{K_d}\right)\frac{N}{F_M(s)} + \phi_{MU,n} - \phi_{MI,n}\right] \cdot H'(s) + \phi_{osc,n}[1 - H'(s)] \quad (6)$$

where

$$H'(s) = \frac{\frac{F_L(s)F_M(s)K_dK_0}{Ns}}{1 + \frac{F_L(s)F_M(s)K_dK}{Ns}}. \quad (7)$$

We feel that it is time to discuss the results we have arrived at until now. Let us start with the last equation.

First, we see that the loop gain $K_dK_0 = K$ is reduced in proportion to the division factor N to a new value

$$K' = K/N = K_dK_0/N \quad (8)$$

and as a consequence we face a reduced effective natural frequency ω_n' and a prolongated settling time–the remedy might be a compensating dc amplifier incorporated into the $F_L(s)$ block.

The second difficulty arrises from the IF filter in the feedback path, respectively, from its effective modulation transfer function $F_M(s)$. In many instances it may be replaced by a mere time delay function which causes a reduction of the pull-in range $\Delta\omega_P$ [8]. However, when the IF filter is a more complicated circuit we face the danger of false locks [5, p. 151] and the degradation of the loop stability.

Finally, by investigating the behavior of the PLL transfer function $H'(s)$ in the frequency domain we easily arrive at the following conclusions: for $\omega < \omega_n'$

$$|H'(s)| \approx 1 \quad \text{and} \quad |1 - H'(s)| \approx 0 \quad (9a)$$

whereas for $\omega > \omega_n'$

$$|H'(s)| \approx 0 \quad \text{and} \quad |1 - H'(s)| \approx 1. \quad (9b)$$

The consequence is that in the PLL passband the output noise is given by

$$\phi_{o,n} \approx \phi_{r,n}\left(M + \frac{N}{Q}\right) + \left(\phi_{DQ,n} - \phi_{DN,n} + \frac{V_{PD,n} + V_{F,n}}{K_d}\right)N + \phi_{MU,n} - \phi_{MI,n} \quad (10)$$

(note that we have neglected $F_M(s)$, the influence of which in the passband is small) and in the stopband by

$$\phi_{o,n} \approx \phi_{osc,n} \quad (11)$$

i.e., the PLL output noise is equal to the voltage-controlled oscillator noise. However, we shall see later that this rule of thumb is often invalidated by the improper choice of ω_n' or of the loop filter $F_L(s)$.

Now reverting to the analysis of the PLL output noise in the passband we see that the first term on the right-hand side of (10) is inevitable since it is merely a multiplied reference generator noise. The situation with the second term is not so simple; here we encounter both divider noises, the phase detector noise and the "filter noise" $V_{F,n}$, all multiplied by the division factor N. Finally, with the third term we add the multiplier and the mixer noises; however, they will be generally small compared with the second term.

As all the considered noises are random by nature and uncorrelated, we may sum the respective spectral densities.

$$S_{\varphi o,n}(f) = S_{\varphi r,n}(f)\left[M + \frac{N}{Q}\right]^2 + \left[S_{\varphi DQ,n}(f) + S_{\varphi DN,n}(f) + \frac{S_{V_{PD},n}(f) + S_{V_F,n}(f)}{K_d^2}\right] \cdot N^2 + S_{\varphi MU,n}(f) + S_{\varphi MI,n}(f). \quad (12)$$

III. NOISES GENERATED IN INDIVIDUAL PLL BUILDING BLOCKS

In the following sections we shall investigate the quantitative and qualitative share of the individual terms, in the above equation, to the total effective "reference noise."

A. Loop Filter Noise

When the loop filter $F_L(s)$ is a passive one (see Fig. 1), i.e., a simple RC lag or lag-lead network, there are two major sources of noise, namely, some types of capacitors and resistors (carbon resistors) which can generate appreciable amounts of $1/f$ noise. As a consequence the low-noise design requires their individual selection (the use of metal-film type resistors is a necessity). The second source may be the decoupling resistor R_{dc}, separating the varactor circuit from the loop filter and the phase detector. The respective noise power density is

$$S_{V,F} = 4kT \cdot R_{dc} \approx R_{dc} \times 1.66 \times 10^{-20} \ [V^2/Hz] \quad (13)$$

and for the typical value of the phase detector gain, $K_d = 0.3$, we get

$$S_{\varphi,F} = R_{dc} \times 1.84 \times 10^{-19} \quad [rad^2/Hz]. \quad (14)$$

B. Phase Noise in dc Amplifiers

In many instances we need to introduce either an active lag-lead filter [1], [5] or merely a dc amplifier.

The design of a low-noise dc amplifier is not an easy task [9] even in instances where the tested circuit diagram is used [10]-[14]. Typical equivalent input noise voltage is only several $nV/\sqrt{Hz}$ with the corner-low-frequency between 10-100 Hz. Similar performance is also achieved with some modern IC operational amplifiers [9], [15].

C. Phase Noise in HF Amplifiers

Here, our investigations must start with the famous paper by Halford *et al.* [16]. They found that power spectral density of the flicker phase noise close to the carrier was approximately the same, i.e.,

$$S_\varphi(f) \approx \frac{10^{-11.2}}{f} + S_\varphi(f)_{white} \quad (15)$$

for the surveyed range from 5 to 100 MHz, quite independent of the transistor type and even of the multiplication factor. Laboratory experiments proved that the intrinsic, direct phase modulation of the RF carrier by transistors was responsible for the phenomenon. The improvement—typically more than 30 dB (and up to 40 dB in some cases)—has been achieved by applying local RF negative feedback (small unbypassed resistor in the emitter—typically from 10 to 100 Ω). These findings were later supported theoretically by Healey [17] and experimentally by other authors [12], [18], [19]. Low amplifier currents and high voltages help to keep the $1/f$ noise current low. The best white noise levels $S\varphi(f)_{white}$ reported [18] are of the order of 10^{-17}.

D. Phase Noise in Phase Detectors and Mixers

The experience with measurement of $S_{\varphi,n}(f)$ of low-noise crystal oscillators has taught that the best phase detectors are double-balanced mixers with Schottky barrier diodes in the ring configuration. A further improvement may be achieved by placing two diodes in each arm [11], [18]. Measurements performed by different authors [11], [13], [14], [18], [19] reveal

$$S_\varphi(f) \approx \frac{10^{-14\pm1}}{f} + 10^{-17}. \quad (16)$$

On the other hand, there is not yet fully proved evidence that the logic circuit or digital phase detectors are much noisier [19], [20]. Šojdr [19] has measured noise properties of popular digital phase-frequency detectors in the range from 0.1 to 1 MHz and found

$$S_\varphi(f) = \frac{10^{10.6\pm0.3}}{f}. \quad (17)$$

Furthermore, he verified the statement by Underhill *et al.* [20] that phase detectors built of ECL and CMOS logic families exhibit a better noise behavior up to about −22 dB in the flicker noise region.

E. Phase Noise in Digital Frequency Dividers

Since the frequency or phase modulation index decreases proportionally to the division factor N, the ideal noise figure is

$$F_{div} = -20 \log N. \quad (18)$$

However, there is an additional noise generated in the divider

itself; thus, the output phase noise is given by

$$S_{\varphi,D,n}(f) = \frac{S_{\varphi,\text{in}}(f)}{N^2} + S_{\varphi,\text{add}}(f). \tag{19}$$

To find out the properties of the additive term we have normalized the above equation with respect to the output frequency, i.e. (cf. Section III-G),

$$\frac{S_{\varphi,D,n}}{f_{\text{out}}^2} = \sum_{k=-2}^{2} h_{k,D,n}$$

$$= \sum_{k=-2}^{2} h_k f^{k-2} + \frac{h_{1,\text{add}}}{f} + h_{2,\text{add}} \tag{20}$$

and plotted the respective coefficients $h_{1,\text{add}}$ and $h_{2,\text{add}}$, computed from accessible experimental results [3], [19], [21]-[23] in Fig. 2. We see that the overall behavior is the same, namely, for large division factors (small output frequency) the data fit a straight line with the slope of 20 db/dec. Be referring to the lowest noise points we may write for the additive term

$$S_{\varphi,\text{add}} \approx \frac{10^{-14.7}}{f} + 10^{-16.5}. \tag{21}$$

On the other hand, for higher output frequencies (above 1 MHz) both coefficients $h_{1,\text{add}}$ and $h_{2,\text{add}}$ become constants which indicate that the input noise predominates.

F. Phase Noise in Frequency Multipliers

We have to refer first to the above mentioned paper by Halford [16] and further to that by Baugh [24] where guidelines for the design of low-noise frequency multipliers are given (RF negative feedback—see Section III-C—and steep zero crossings). To get more information we have collected published results about noise properties of transistor frequency multipliers [12], [24]-[29] and diode frequency multipliers [12], [28], [30], [31] as well and found that both noise constants $a_{-1,\text{inp}}$ and $a_{0,\text{inp}}$ are nearly the same for properly designed transistor frequency multipliers, irrespective of the frequency, i.e.,

$$S_{\varphi MU,\text{inp}}(f) \approx \frac{10^{-14}}{f} + 10^{-16.5}. \tag{22}$$

In diode frequency multipliers the flicker phase noise level is higher, typically

$$10^{-12.9}/f. \tag{23}$$

We shall see later that the flicker and white phase-noise spectral densities of the best crystal oscillators are of the same order or rather worse as the respective additive terms in frequency multipliers; thus, we can conclude that the phase noise is not appreciably deteriorated by passing the signal through a properly designed frequency multiplier. (Note that this is hardly true with the diode frequency multipliers.)

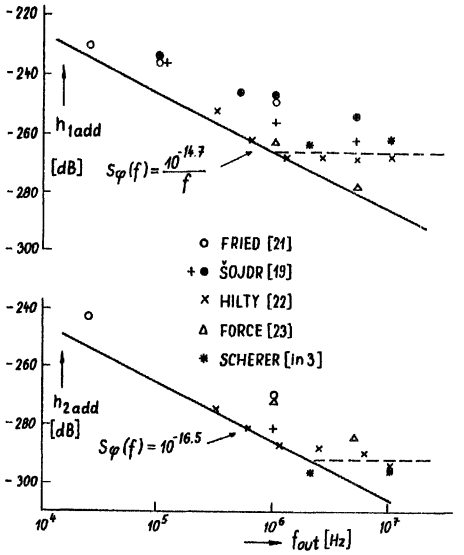


Fig. 2. Plot of the normalized additive phase-noise coefficients $h_{1,\text{add}}$ and $h_{2,\text{add}}$ in respect to the divider output frequency as measured by different authors.

G. Phase Noise in Oscillators[1]

The theory of noise in free-running oscillators has been dealt with by many authors for nearly half a century and we feel that the best we can do is to mention only a few of them [33]-[36]. However, they do not provide the information the designer of the PLL systems is looking for, at least by the first fast reading. On the other hand, when we start from the condition of the zero phase shift around the oscillating loop [37] we easily arrive at the heuristic oscillator phase-model suggested by Leeson [38], [17].

In accordance with this model any oscillator can be simplified into a loop containing a resonator and an amplifier-limiter. As a consequence its output spectral density is given by

$$S_\varphi(\omega)_{\text{osc},n} = S_\varphi(\omega)_{A,n}[1 + (\omega_0/2Q_L\omega)^2] \tag{24}$$

where the amplifier-limiter noise is

$$S_\varphi(f)_{A,n} = a_{-1}/f + a_0. \tag{25}$$

The magnitude of the flicker noise constant a_{-1} has been found experimentally [16] in the range from 5 to 100 MHz to be

$$a_{-1} = F_{-1} \cdot 10^{-11.2} \qquad [\text{rad}^2]. \tag{26}$$

The noise factor F_{-1} depends on the emitter RF feedback and may be made as small as 10^{-3}.

The white noise constant is the ratio of the noise power $\varphi_{\text{noise}}^2(f)$ to the oscillator power P_0 reduced to 1 Hz bandwidth and multiplied by a noise factor F_0:

$$a_0 = F_0 \cdot kT/P_0 \approx 4 \times 10^{-21} F_0/P_0 \quad [\text{rad}^2/\text{Hz}]. \tag{27}$$

[1] This section is based on the paper read by the author at the EMC-79 Symposium, Rotterdam, The Netherlands [32].

By introducing (25) into (24) we arrive at a power law relation

$$S_{\varphi \mathrm{osc},n}(f) = \left(\frac{f_0}{f}\right)^2 \cdot \frac{a_{-1}}{f \cdot 4Q_L{}^2} + \left(\frac{f_0}{f}\right)^2 \cdot \frac{a_0}{4Q_L{}^2} + \frac{a_{-1}}{f} + a_0. \tag{28}$$

To compare oscillators with different output frequencies we face the difficulty that the noise sidebands close to the carrier are proportional to the square of the resonant frequency f_0. This problem is often solved by referring to the fractional-frequency power spectral density

$$S_y(f) = (f/f_0)^2 S_\varphi(f), \tag{29}$$

i.e.,

$$S_{y,\mathrm{osc},n}(f) = \frac{a_{-1}}{f \cdot 4Q_L{}^2} + \frac{a_0}{4Q_L{}^2} + \frac{a_{-1}}{f_0{}^2} f + \frac{a_0}{f_0} f^2 \tag{30a}$$

$$= \frac{h_{-1}}{f} + h_0 + h_1 f + h_2 f^2 \tag{30b}$$

or to the spectral density of the phase-noise time [14]

$$S_x(f) = S_y(f)/(2\pi f)^2. \tag{31}$$

The advantage of this second normalization is the close resemblance to the actual phase noise characteristic and the ease with which $S\varphi(f)$ is calculated–the proportionality factor being $(2\pi f_0)^2$. By dropping the factor $(2\pi)^2$ we arrive at a very useful simplification:

$$S_{\varphi \mathrm{osc},n}(f)/f_0{}^2 = \frac{h_{-1}}{f^3} + \frac{h_0}{f^2} + \frac{h_1}{f} + h_2 \tag{32}$$

where

$$h_{-1} = a_{-1}/4Q_L{}^2; \qquad h_0 = a_0/4Q_L{}^2; \tag{33a}$$

$$h_1 = a_{-1}/f_0{}^2; \qquad h_2 = a_0/f_0{}^2. \tag{33b}$$

We have verified the validity of (28)-(32) by computing coefficients h_{-1} and h_0 for different types of oscillators from a wealth of published data and plotted them as functions of quoted Q_L. The results are shown in Fig. 3 and indicate a good agreement with the simplified oscillator noise theory. The mean value of the flicker noise constant a_{-1} is

$$a_{-1} \approx 10^{-11} \tag{34}$$

and is practically independent of the oscillator type in the whole frequency range from 5 MHz to 100 GHz; the same is true also for the white noise constant a_0, the mean value of which is

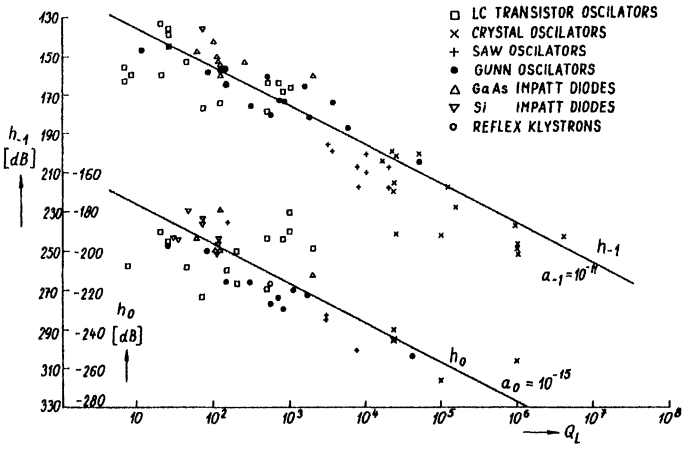


Fig. 3. Plot of the normalized phase-noise coefficients h_{-1} and h_0 in respect to the loaded Q_L of different oscillators.

$$a_0 \approx 10^{-15}. \tag{35}$$

Similar experimental verification of the dependence of h_1 and h_2 on $1/f_0{}^2$ is prevented by the lack of data for the noise power at higher Fourier frequencies. Some insight provides the investigation of noise properties of crystal oscillators discussed in the following section.

Summarizing all the above results we can write a fairly general oscillator noise equation

$$\frac{S_{\varphi \mathrm{osc},n}(f)}{f_0{}^2} = \frac{1}{f^3} \cdot \frac{10^{-11.6}}{Q_L{}^2} + \frac{1}{f^2} \frac{10^{-15.6}}{Q_L{}^2} + \frac{1}{f} \cdot \frac{10^{-11}}{f_0{}^2} + \frac{10^{-15}}{f_0{}^2} \tag{36}$$

and plot the normalized oscillator phase-noise characteristics which consist of two sets of straight lines with parameters Q_L and f_0; see Fig. 4.

H. Phase Noise in Reference Frequency Generators

The reference generator in low-noise PLL systems (frequency synthesizers) is a spectrally pure crystal oscillator. By inspecting (36) or Fig. 4 we see that a low close-to-carrier noise requires the use of resonators with the highest possible Q. It has been found earlier [39] that for the AT-cuts the intrinsic losses in the quartz crystal material are related to the resonant frequency by

$$f_0 Q \approx 1.5 \times 10^{13}. \tag{37}$$

This product is slightly lower for the advantageous SC-cuts and nearly two times larger for BT-cuts. Consequently, the low flicker and white frequency noise, i.e., small h_{-1} and h_0 in (32) and (33), requires the lowest possible frequency f_0. However, to keep dimensions of the crystal resonators in practical limits, we hardly can go below 5 MHz (cf. [39, Fig. 2]). Very often 10 MHz crystal oscillators are used as reference frequency generators.

To find practical values of h_k coefficients in (32) we have recently investigated noise characteristics of about 60 crystal oscillators in the range from 5 to 170 MHz [40] and

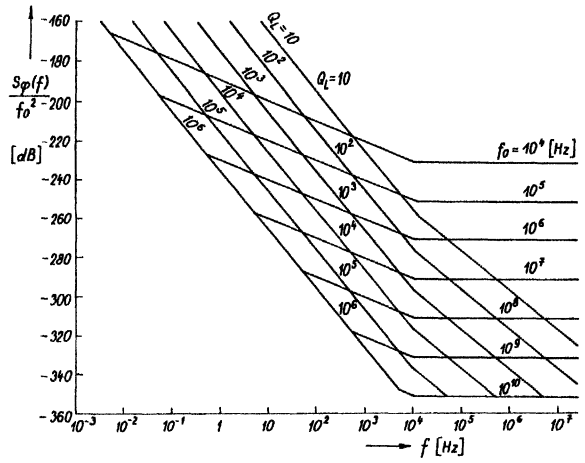


Fig. 4. Normalized phase-noise characteristics of oscillators: parameters are the loaded Q_L of the resonator and the output frequency f_0.

get for an average crystal oscillator the following noise equation.

$$\frac{S_{\varphi osc,n}(f)}{f_0^2} = \frac{1}{f^3}\,10^{-37.25} f_0^2 + \frac{1}{f^2}10^{-39.4} f_0^2 + \frac{1}{f}\,\frac{10^{-12.15}}{f_0^2} + \frac{10^{-14.9}}{f_0^2}. \tag{38}$$

By investigating the dispersion of the h_k values, particularly of the best crystal oscillators we have arrived at the conclusion that in the $1/f^3$ and $1/f^2$ regions the crystal resonator noise [41], rather than the transistor noise, is the limiting factor [42].

IV. PLL LOOP FILTERS FOR LOW NOISE BEHAVIOR

The problem has been discussed earlier by the author [43] and here only the results will be summarized.

By formulating the rule of thumb in (9) in Section II we have assumed that the transfer function $H'(s)$ has a rectangular behavior. However, this is not true in real systems and a closer investigation of the second- and third-order systems gives

$$H'(s) = \frac{s\omega_n'(2\xi - \omega_n'/AK') + \omega_n'^2}{s^2 + 2\xi\omega_n' s(1 + s^2\kappa\omega_n'^2) + \omega_n'^2} \tag{39}$$

where

$$\omega_n' = \sqrt{K'/T_1}; \quad \xi = \frac{\omega_n'}{2}T_2 + \frac{1}{AK'};$$
$$\kappa = T_3/T_2; \quad K' = K_d K_0/N \tag{40}$$

and A is the gain of the operational amplifier used.

Since the asymptotic approximation will generally supply sufficient information for the noise behavior of the studied PLL system, we shall consider the four most important configurations.

1) Simple RC filter, i.e., $A = 1$, $\xi = \omega_n'/2K'$, and $\kappa = 0$. After introducing the normalization

$$\frac{\omega}{\omega_n} = x \tag{41}$$

we shall find

$$H(jx) \approx -1/x^2 \qquad x \gg 1 \tag{42a}$$

and

$$1 - H(jx) \approx 2j\xi x \qquad x \ll 1. \tag{42b}$$

2) Passive lag-lead filter, i.e., $A = 1$, and $\kappa = 0$.

$$H(jx) \approx -j(2\xi - \omega_n'/K')/x \qquad x \gg 1 \tag{43a}$$

$$1 - H(jx) \approx jx\omega_n'/K' \qquad x \ll 1. \tag{43b}$$

3) Active lag-lead filter, i.e., $A \to \infty$, and $\kappa = 0$.

$$H(jx) \approx -2j\xi/x \qquad x \gg 1 \tag{44a}$$

$$1 - H(jx) \approx -x^2 \qquad \omega_n'/AK' \ll x \ll 1. \tag{44b}$$

4) Active lag-lead filter with an additional RC section (it has been shown [43] that for practical applications $\kappa < 0.3$).

$$H(jx) \approx -1/\kappa x^2 \qquad x \gg 1 \tag{45a}$$

$$1 - H(jx) \approx -x^2 \qquad \omega_n'/AK' \ll x \ll 1. \tag{45b}$$

To get a better insight we shall consider the problem of phase-locking a 100 MHz crystal oscillator and a low Q LC-oscillator to a 5 MHz reference signal. All normalized phase noise characteristics are plotted in Fig. 5. By considering first the 100-to-5 MHz PLL system we find the crossover point to be approximately 200 Hz. In instances where f_n is smaller than 200 Hz we face a large amount of additive noise, the origin of which is the attenuated 100 MHz oscillator noise; the dashed lines indicate the situation with passive filters and dot-and-dash lines indicate the improvement when the active lag-lead filter is used. Similarly, we encounter an unnecessary additive noise caused by the attenuated 5 MHz oscillator noise in cases where $f_n > 200$ Hz (dashed line). A remedy can be provided by additional filtering; however, the stability of the loop deteriorates.

In the second example of phase-locking a low Q oscillator to a crystal reference, the use of the active lag-lead filter is necessary since even for f_n equal to the crossover frequency we face a large additive noise with passive filters only (see again the dashed line in Fig. 5).

V. NOISE IN PLL FREQUENCY SYNTHESIZERS

By considering the basic PLL configuration, as shown in Fig. 1, we easily arrive, with the assistance of (12) and the condition of only a 3 dB noise increase, at the conclusion that

$$S_{\varphi,r,n}(f)\left[M + \frac{N}{Q}\right]^2 \leqq N^2 \cdot S_{\varphi,\mathrm{PLL}}(f) \tag{46}$$

where

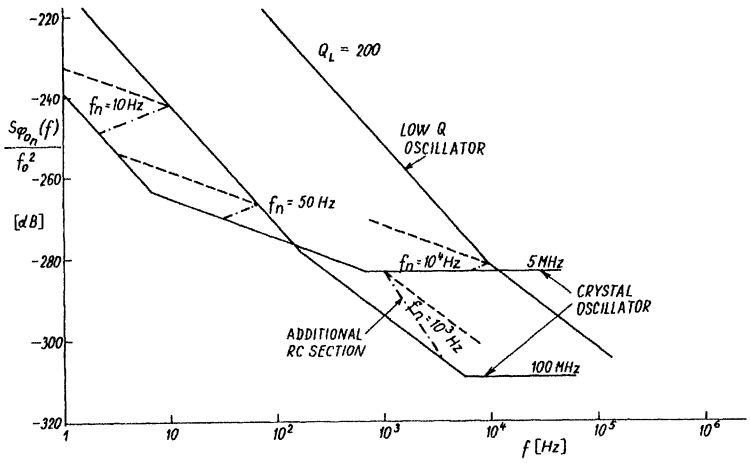


Fig. 5. Normalized phase noise of the PLL system of the 100 MHz crystal oscillator locked to a 5 MHz reference crystal oscillator and a low Q oscillator locked to the same reference oscillator.

$$S_{\varphi,\mathrm{PLL}}(f) = 2S_{\varphi,D}(f) + S_{\varphi,\mathrm{PD}}(f) + S_{\varphi,A,n}(f). \quad (47)$$

The power spectral densities of individual noise sources in the above equation have been evaluated in Section III. In a large majority of cases where active filters are used the last term on the right-hand side of (47) will dominate. With the assistance of [9, Fig. 12] and a reasonable value for K_d, $K_d = 0.3$, we find

$$S_{\varphi,\mathrm{PLL}}(f) \approx \frac{10^{-14}}{f} + 10^{-15.5}. \quad (48)$$

Since the flicker and white noise spectral densities of good reference oscillators in the 5 and 10 MHz frequency ranges are of the same order we arrive at the rule of thumb that the division factor N should not exceed the multiplication factor M, i.e.,

$$N \lessapprox M. \quad (49)$$

However, this condition can hardly be met in instances where small frequency steps are desired at the synthesizer output. One solution is provided by the application of fractional frequency dividers [2, p. 74], [44], [45]. The spurious phase modulation which is often quite large but predictable [46] is suppressed by compensation.

Another solution may be provided by a subtracting PLL-system, the principle of which will be explained with the assistance of Fig. 6. In the case where N in Fig. 1 is much larger than M, the output noise of the first digital PLL in its passband is given approximately by

$$S_{\varphi,01,n}(f) \approx N^2 \cdot S_{\varphi,\mathrm{PLL},1}(f). \quad (50)$$

With the assistance of (12) we find for the output noise (in the passband) of the subtracting loop

$$S_{\varphi,02,n}(f) \approx \frac{N^2 S_{\varphi,\mathrm{PLL},1}(f)}{P^2} + M^2 S_{\varphi r,n}(f) + S_{\varphi,\mathrm{PLL},2}(f). \quad (51)$$

Generally, the last term on the right-hand side of the above equation may be neglected and the application of the 3 dB noise-increase condition reveals the second rule of thumb, i.e.,

$$\frac{N}{P} \lessapprox M. \quad (52)$$

By a judicious combination of both these techniques one can expect a very low additive phase noise from frequency synthesizers at frequency ranges above 100 MHz [44], [45], [47]. For generation of lower frequencies the use of dividers instead of output mixers should be preferred [21], [44], [45].

To demonstrate the state of the art we have plotted in Fig. 7 the normalized power spectral densities of the output phase noise of several commercial PLL frequency synthesizers. The advantage of the normalization is the possibility to compare the noise properties of frequency synthesizers with different output frequencies and that of the reference generator together in the same figure. The progress achieved in the last ten years is impressive.

VI. CONCLUSIONS

In this survey paper we have called the readers' attention to all major sources of additive noise in the PLL system. Furthermore, we have shown that, in the first approximation, these noises add to the input or the reference noise.

In the second we have investigated noises generated in individual PLL building blocks and tried to find numerical values for the respective power spectral densities with the assistance of the experimental findings published by different authors all over the world.

The major guideline rules for minimizing the additive noises in complicated PLL systems or PLL frequency synthesizers have been discussed in the last two sections. Finally, phase-noise power spectral densities of several commercial PLL frequency synthesizers have been plotted in normalized form in Fig. 7. On one hand, this figure demonstrates the state of the art and, on the other hand, the progress achieved in the last decade.

For readers who intend to go deeper into the problem we have collected a copious bibliography.

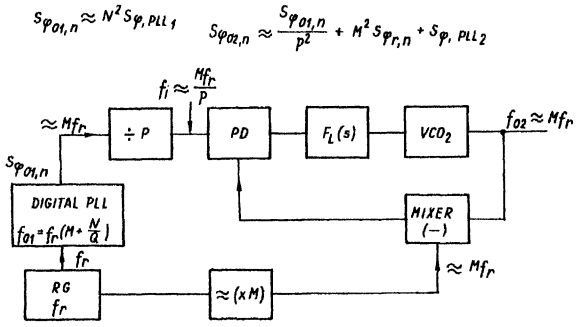


Fig. 6. A subtracting PLL system.

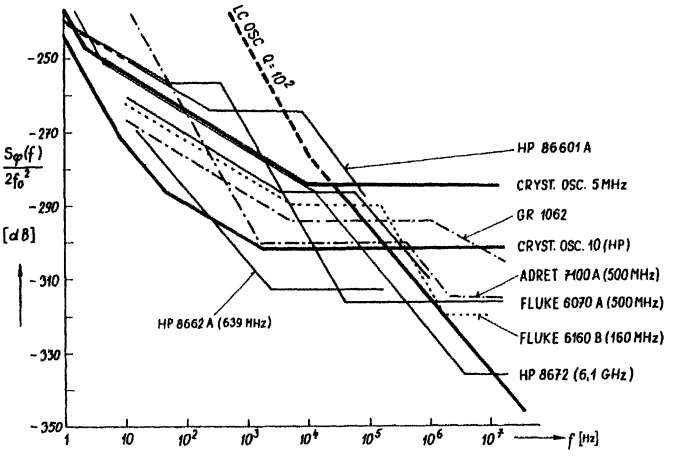


Fig. 7. Normalized phase noise of several commercial PLL frequency synthesizers, two reference crystal oscillators, and one typical LC oscillator (VCO).

REFERENCES

[1] V. F. Kroupa, *Frequency Synthesis, Theory, Design and Applications*. London, England: Griffin, 1973; New York: Wiley, 1973.

[2] J. Gorski-Popiel, *Frequency Synthesis: Techniques and Applications*. New York: IEEE Press, 1975.

[3] V. Manassewitsch, *Frequency Synthesizers, Theory and Design*. New York: Wiley, 1980.

[4] W. F. Egan, *Frequency Synthesis by Phase Lock*. New York: Wiley, 1981.

[5] F. M. Gardner, *Phaselock Techniques*, 2nd ed. New York: Wiley, 1979.

[6] W. C. Lindsey, *Synchronization Systems in Communication and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

[7] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[8] J. A. Develet, Jr., "The influence of time delay on second-order phase-lock loop acquisition range," in *Proc. Int. Telemetering Conf.*, vol. 1, Sept. 23–27, 1963, pp. 432–437; see also, W. C. Lindsey and M. K. Simon, Eds., *Phase-Locked Loops and Their Application*. New York: IEEE Press, 1978.

[9] Y. Netzer, "The design of low-noise amplifiers," *Proc. IEEE*, vol. 69, pp. 728–741, June 1981.

[10] B. E. Blair, Ed., *Time and Frequency: Theory and Fundamentals*, Nat. Bur. Stand., Boulder, CO, Monog. 140, May 1974, p. 169.

[11] D. G. Meyer, "A test set for the accurate measurement of phase noise on high-quality signal sources," *IEEE Trans. Instrum. Meas.*, vol. IM-19, pp. 215–227, Nov. 1970.

[12] S. G. Andresen and J. K. Nesheim, "Phase noise of various frequency doublers," *IEEE Trans. Instrum. Meas.*, vol. IM-22, pp. 185–188, June 1973.

[13] K. Hilty, "Messtechnik zur Bestimmung des Phasenrauschspektrums," *Technische Mitteilungen PTT*, pp. 1–15, Jan. 1980.

[14] P. Kartaschoff, *Frequency and Time*. London, England: Academic, 1978.

[15] G. Erdi *et al.*, "Op amps tackle noise—And for once, noise loses," *Electron. Design*, vol. 28, pp. 65–71, Dec. 20, 1980.

[16] D. Halford, A. E. Wainright, and J. A. Barnes, "Flicker noise of phase in RF amplifiers and frequency multipliers: Characterization, cause, and cure," in *Proc. 22nd Annu. Freq. Contr. Symp.*, Atlantic City, NJ, Apr. 1968.

[17] D. J. Healey, "Flicker of frequency and phase and white frequency and phase fluctuations in frequency sources," in *Proc. 26th Annu. Symp. Freq. Contr.*, Atlantic City, NJ, June 1972, pp. 29–49.

[18] F. L. Walls, S. R. Stein, J. E. Gray, and D. J. Glaze, "Design considerations in state-of-the-art signal processing and phase noise measurement systems," in *Proc. 30th Annu. Symp. Freq. Contr.*, Atlantic City, NJ, June 1976, pp. 269–274.

[19] L. Šojdr, "Phase noise in oscillators and synthesizers"(in Czech), Dissertation, Prague, Czechoslovakia, June 1979.

[20] M. J. Uderhill, P. A. Jordan, M. A. G. Clark, and R. I. H. Scott, "A general purpose LSI frequency synthesizer system," in *Proc. 32nd Annu. Symp. Freq. Contr.*, Atlantic City, NJ, May–June 1978, pp. 365–372.

[21] R. Fried, "A wide-band non-heterodyned frequency synthesizer," John Fluke, Inc., Seattle, WA, 1974.

[22] K. Hilty, private communication at the occasion of the CPEM-80 Conf., Braunschweig, West Germany, 1980.

[23] M. Force, "Low-noise interface circuits format crystal-oscillator signals," *EDN*, vol. 25, pp. 125–129, 131, Oct. 20, 1980.

[24] R. A. Baugh, "Low noise frequency multiplication," in *Proc. 26th Annu. Symp. Freq. Contr.*, Atlantic City, NJ, June 1972, pp. 50–54

[25] D. J. Glaze, "Improvements in atomic cesium beam frequency standards at the National Bureau of Standards," *IEEE Trans. Instrum. Meas.*, vol. IM-19, pp. 156–160, Aug. 1970.

[26] D. J. Healey, "$\mathcal{L}(f)$ measurements on UHF sources comparising VHF crystal controlled oscillator followed by a frequency multiplier," in *Proc. 28th Annu. Symp. Freq. Contr.*, Atlantic City, NJ, May 1974, pp. 190–202.

[27] F. L. Walls and A. De Marchi, "RF spectrum of a signal after frequency multiplication; Measurement and comparison with a simple calculation," *IEEE Trans. Instrum. Meas.*, vol. IM-24, pp. 210–217, Sept. 1975.

[28] E. Bava, A. De Marchi, and A. Godone, "Spectral analysis of synthesized signals in the mm wavelength region," *IEEE Trans. Instrum. Meas.*, vol. IM-26, pp. 128–132, June 1977.

[29] A. L. Lance, W. D. Seal, F. G. Mendoza, and N. W. Hudson, "Automating phase noise measurements in the frequency domain," in *Proc. 31st Annu. Symp. Freq. Contr.*, Atlantic City, NJ, June 1977, pp. 347–358.

[30] E. Bava, A. De Marchi, and A. Godone, "A narrow output linewidth multiplier chain for precision frequency measurements in the 1 THz region," in *Proc. 31st Annu. Symp. Freq. Contr.*, Atlantic City, NJ, June 1977, pp. 578–582.

[31] M. B. Bloch and A. Vulcan, "Low noise measurement techniques," in *Proc. 28th Annu. Symp. Freq. Contr.*, Atlantic City, NJ, May 1974, pp. 184–189.

[32] V. F. Kroupa, "Electromagnetic compatibility of frequency synthesizers," in *Proc. 3rd Symp.*, Rotterdam, The Netherlands, May 1979, pp. 335–340.

[33] I. Berstein, "On fluctuations in the neighbourhood of periodic motion of an auto-oscillating system," *Comptes Rendus de l'Academie des Sciences de l'URSS*, vol. XX, no. 1, 11–16, 1938.

[34] W. A. Edson, "Noise in oscillators," *Proc. IRE*, vol. 48, pp. 1454–1466, Aug. 1960.

[35] K. Kurokawa, "Noise in synchronized oscillators," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16, pp. 234–240, 1968.

[36] K. F. Schünemann and K. Behm, "Nonlinear noise theory for synchronized oscillators," *IEEE Trans Microwave Theory Tech.*, vol. MTT-27, pp. 452–458, May 1979.

[37] E. J. Baghdady, R. N. Lincoln, and B. D. Nelin, "Short-term frequency stability: Characterization, theory, and measurement," *Proc. IEEE*, vol. 53, pp. 704–722, July 1965.

[38] D. B. Leeson, "A simple model of feedback oscillator noise spectrum," *Proc. IEEE*, vol. 54, pp. 329–330, Feb. 1966.

[39] A. W. Warner, "Design and performance of ultraprecise 2.5-mc quartz crystal units," *Bell Syst. Tech. J.*, vol. 39, pp. 1193–1217, Sept. 1960.

[40] V. F. Kroupa, "Noise of crystal oscillators" (in Czech), Rep. Z-1042/I, Prague, Czechoslovakia, Nov. 1979.
[41] F. L. Walls and A. E. Wainwright, "Measurement of the short-term stability of quartz crystal resonators and the implications for crystal oscillator design and applications," *IEEE Trans. Instrum. Meas.*, vol. IM-24, pp. 15–17, Mar. 1975.
[42] V. F. Kroupa, J. Pavlovec, and L. Sojdr, "Noise in standard frequency sources," to be published.
[43] V. F. Kroupa, "Spectral properties of third order phase-locked loops," in *Proc. 5th Summer Symp. Circuit Theory*, Kladno, Sept. 1977, pp. 201–215.
[44] K. L. Astrof, "Frequency synthesis in a microwave signal generator," *Hewlett-Packard J.*, vol. 29, pp. 8–15, Nov. 1977.
[45] Special Issue on PLL Frequency Synthesizer Type 8662A, *Hewlett-Packard J.*, vol. 32, Feb. 1981.
[46] V. F. Kroupa, "Spectra of pulse rate frequency synthesizers," *Proc. IEEE*, vol. 67, pp. 1680–1682, Dec. 1979.
[47] G. Mackiw and G. W. Wild, "Microwave frequency synthesis for satellite communications ground terminals," in *Proc. 30th Annu. Symp. Freq. Contr.*, Atlantic City, NJ, June 1976, pp. 420–437.

PLL/DLL System Noise Analysis for Low Jitter Clock Synthesizer Design

Beomsup Kim*, Todd C. Weigandt, Paul R. Gray

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

**Department of Electrical Engineering*
Korea Advanced Institute of Science and Technology

Abstract

This paper presents an analytical model for timing jitter accumulation in ring-oscillator based phase-locked-loops (PLL). The timing jitter of the system is shown to depend on the jitter in the ring-oscillator and an accumulation factor which is inversely proportional to the bandwidth of the phase-locked-loop. Further analysis shows that for delay-locked-loops (DLL), which use an inverter delay chain that is not configured as a ring-oscillator, there is no noise enhancement since noise jitter events do not contribute to the starting point of the next clock cycle. Finally, theoretical predictions for overall jitter are compared to behavioral simulations with good agreement.

I. Introduction

Higher clock rates in many applications such as video, audio, and data processors, requires increasingly higher performance from the clock synthesizers used to drive them. In clock recovery applications, such as data communications and disk drive read channels, as well, higher speeds require better performance from the VCOs and the overall timing recovery phase-locked-loop itself. In both types of applications clocks are generated to drive mixers or sampling circuits in which the random variation of the sampling instant, or jitter, is a critical performance parameter. The goal of this paper is to predict the timing jitter of phase-locked-loop (PLL), and delay-locked-loop (DLL) systems from the parameters of the loop and the jitter in the VCO itself. Of particular interest are ring-oscillator VCOs which are attractive from an integration and cost point of view, but suffer from larger timing jitter than traditional tuned LC-tank oscillators.

In most clock synthesis applications a VCO is locked to a low-jitter reference, often in the form of a crystal, using a phase-locked-loop (figure 1). Most of the output jitter results from noise sources in the phase detector, loop filter, and VCO. With careful PLL design, however, the jitter in the VCO is usually the dominant contributor. In clock and data recovery applications there is often a significant amount of jitter from the input source as well as the VCO. In this case, it will be shown that there is a trade-off involved in selecting the bandwidth of the PLL. A narrow bandwidth PLL rejects input jitter but does not correct VCO timing errors as quickly, leaving the total output jitter, VCO noise limited. A wide bandwidth PLL can correct VCO errors more quickly but if made too wide, leaves the system input jitter limited.

The ring-oscillator VCO in figure 1 is popular choice in many applications. The jitter per cycle of oscillation is determined by the sum of the timing error contributions of each inverter stage in the ring [1][1]. With each cycle of oscillation the jitter variance, relative to a reference transition in the past continues to grow, unless the oscillator is configured in a PLL. In a ring-oscillator PLL, however, the total timing error is the sum of all past errors weighted by the corrective action of the loop. The total jitter is made up of the errors in the most recent cycles of oscillation, yet to be corrected by the PLL, and therefore improves for higher loop bandwidths.

Another structure popular in many applications is a DLL in which a voltage controlled delay line, (VCD) is used in place of a VCO. In this case the jitter accumulated by the end of the delay chain does not contribute to the starting point of the next cycle since the delay chain is not configured as an oscillator. The reference determines the next transition point instead. This type of system has superior jitter performance, but is only usable in some applications.

In this paper, the total output jitter for PLL and DLL systems will be determined and compared to the results of behavioral simulations. The analysis is the scope of section II and some design examples and simulations will be shown in section III.

II. PLL/DLL Jitter Analysis

Timing jitter in a ring-oscillator PLL depends on the interaction of noise in the oscillator with the dynamics of the phase-locked loop. It has been shown in [1] that the timing jitter variance at the end of a chain of inverters is given by the sum of the contributions of each stage. If each stage contributes a timing error with variance $\overline{\Delta t_n^2}$, then the total jitter at the end of N stages is $N \times \overline{\Delta t_n^2}$. In a ring-oscillator this timing error determines the starting point of the next cycle and therefore creates a permanent phase shift in the output signal. If the ring-oscillator is configured in a phase-locked-loop, however, the phase difference between the reference

1. The paper analyzes the output jitter of the delay cells. Each jitter source is identified and its contribution to the output jitter is calculated. It also gives guidelines for design of low jitter delay lines.

Program Supported by NSF, ARPA, and California MICRO Program

Reprinted from *Proc. of ISCAS*, June 1994.

clock and the oscillator output is detected and compensated for by the dynamics of the loop. The phase detector will sense the shift and create an error signal to change the frequency of the ring-oscillator VCO in a way which moves the phase of the output in the right direction.

Since the amount of phase adjustment is usually small, the phase error is not corrected in one clock cycle, but it is reduced gradually over the course of several cycles. The phase error may remain for up to several hundreds of cycles, depending on the bandwidth of the loop filter in the PLL.

Analysis of the accumulated phase jitter and its relation to the loop bandwidth is important for both clock synthesis and clock recovery applications. In most PLL clock synthesizer designs, the reference clock comes from a very low jitter source such as crystal oscillator. Therefore, the jitter in the ring-oscillator is the main source of the phase error in the synthesized clock. In this case the bandwidth of the loop filter determines how large the accumulated timing jitter gets. For clock recovery applications there is a trade-off involved in the choice of the loop bandwidth since the input signal that is being locked to is not ideal, but has timing jitter associated with it as well. A narrow loop-bandwidth will reduce the impact of jitter in the input signal since the loop will not try to track input fluctuations as strongly. On the other hand, this means that it will take more time to compensate for jitter events in the ring-oscillator. Previously, more attention has been paid to the first effect than the second, but both are important for high performance clock recovery applications. So for both clock synthesis and clock recovery applications, a thorough analysis of the output jitter due to the internal jitter sources is important [2].

To find the accumulated rms jitter, a PLL which uses a sequential phase detector and a charge-pumping circuit is represented by a simple discrete-time model as shown in Figure 2 [3]. The transfer function for jitter in the PLL due to the internal jitter sources is represented by (EQ 1) in z-transform domain.

$$\Theta_{on}(z) = \frac{\Theta_n(z)}{1 + K_d K_w Z_F(z) z^{-1}} \qquad \text{(EQ 1)}$$

Here the phase detector gain, $K_d = \frac{I_S}{2\pi}$ and VCO gain, $K_w = \frac{dw}{dv}$ respectively, and I_S indicates the charge pumping current. $Z_F(z)$ is the z-transform $H(s)/s$, where $H(s)$ is the transfer function of the PLL loop filter in s-domain. In most PLL design, the second order loop filter is used and the transfer function is given by (EQ 2).

$$H(s) = \frac{a(s+n_1)}{s(s+p_1)} \qquad \text{(EQ 2)}$$

where the DC filter gain, $a = \frac{1}{C_P}$, a zero, $n_1 = \frac{1}{C_I R}$, and a pole, $p_1 = \frac{C_I + C_P}{C_I C_P R}$. In most cases, the capacitor C_P does not affect the bandwidth of a PLL and can be ignored for simplicity. Then, the loop filter is configured as a lead-lag filter composed of a resistor R, in series with a capacitor C_I. In this case, $H(s) = \frac{a(s+n_1)}{s}$, $a = R$, and $n_1 = \frac{1}{RC_I}$. Since $n_1 T \ll 1$ in most PLL designs, (EQ 1) can be re-written as (EQ 3),

$$\Theta_{on}(z) = \frac{(1-z^{-1})}{1-(1-\varepsilon)z^{-1}}\Theta_n(z) \qquad \text{(EQ 3)}$$

where $K = K_d K_w a T$ and is actually replaced with the term ε since $K \ll 1$.

The phase jitter from the ring oscillator can be modeled as a sequence of unit step phase jumps with random magnitude. A single phase jump at time nT can be represented by (EQ 4) in the z-domain.

$$\Theta_n(z) = \frac{2\pi\Delta t_n}{T(1-z^{-1})}, \qquad \text{(EQ 4)}$$

Here the magnitude of the error step is Δt_n. The variance of this error is shown in [1] to be proportional to the number of stages in the ring-oscillator, and the timing jitter variance contributed by each stage. Hence the output jitter in z-domain is,

$$\Theta_{on}(z) = \frac{2\pi\Delta t_n}{T(1-(1-\varepsilon)z^{-1})} \qquad \text{(EQ 5)}$$

For all events up to time nT, the sum of output phase shifts is represented by (EQ 6).

$$\Theta_{tot}(nT) = \sum_{k=-\infty}^{n} \frac{2\pi\Delta t_k}{T}(1-\varepsilon)^{n-k} \qquad \text{(EQ 6)}$$

To find the rms output jitter, the expectation of the square of the sum is calculated and given by (EQ 7). Since Δt_k and Δt_l are not correlated, the $E[\Delta t_k \Delta t_l] = 0$ when $k \neq l$. When $k = l$, $E[\Delta t_k \Delta t_l]$ can be replaced by $\Delta\tau_N^2$.

$$E\left[\Theta_{tot}^2(nT)\right] = \left(\frac{2\pi}{T}\right)^2 \frac{\Delta\tau_N^2}{\varepsilon(2-\varepsilon)} \approx \left(\frac{2\pi}{T}\right)^2 \left(\frac{\Delta\tau_N^2}{2\varepsilon}\right) \qquad \text{(EQ 7)}$$

Note that the expectation of the phase jitter is independent of nT, the time instant. Hence the r.m.s. phase jitter is,

$$\sqrt{E\left[\Theta_{tot}^2(nT)\right]} \approx \sqrt{\frac{1}{2\varepsilon}}\frac{2\pi\Delta\tau_{rms}}{T} = \alpha\frac{2\pi\Delta\tau_{rms}}{T} \qquad \text{(EQ 8)}$$

where $\Delta\tau_{rms}$ is $\sqrt{\Delta\tau_N^2}$, and $\alpha = \sqrt{\frac{1}{2K_d K_w a T}}$ is defined as the accumulation factor. The result in (EQ 8) is the r.m.s. phase jitter for a ring-oscillator PLL. From the result, the rms timing jitter in a phase-locked-loop is seen to be α times larger than the intrinsic jitter in the delay chain. The accumulation factor α is inversely proportional to the square-root of $K_d K_w a T$ and in this case shows little dependency on C_I and C_P. Therefore, as long as $n_1 T \ll 1$, $C_P \ll C_I$ and stability requirements are met [3], the jitter accumulation factor can be lowered by increasing the bandwidth of the loop filter, $w_L \approx K_d K_w a$.

An alternative scheme for clock synthesis is to use a delay-locked-loop [4]. In this case, the reference clock is fed to the input of the delay line, and the rising edge of the output of the delay line is compared to that of the reference clock. Since the rising edge of the reference clock reaches the output of the delay line after passing through all delay cells, the total delay is driven to be the same as one period of the reference clock. Also, since the output of the loop filter just changes the phase of the output of the delay line, the loop does not have any extra poles as a PLL does.

Therefore, the stability problem is relaxed and a simple capacitor loop filter can be used without any stability consideration.

In a DLL, phase jitter is not passed on from one period of the clock to the next since the output of the delay-line is not fed back to the input. Therefore we expect the jitter in a DLL to be much smaller than in a ring-oscillator based PLL. To show this quantitatively we proceed with an analysis similar to that in the previous section but with the simplified discrete time DLL model. In this case, the transfer function for output phase noise in terms of the internal jitter from the delay line is represented by (EQ 9).

$$\Theta_{on}(z) = \frac{\Theta_n(z)}{1 + K_d K_p T Z_F(z) z^{-1}} \qquad \text{(EQ 9)}$$

where K_d is phase gain and given by $\frac{I_S}{2\pi}$, and K_p is phase gain and given by $\frac{d\theta}{dv}$ when voltage controlled delay line is assumed. If the loop filter in the DLL is a single capacitor and given by $\frac{1}{sC} = \frac{a}{s}$, the transfer function (EQ 9) becomes (EQ 10).

$$\Theta_{on}(z) = \frac{(1 - z^{-1})\Theta_n(z)}{1 + (\varepsilon - 1) z^{-1}} \qquad \text{(EQ 10)}$$

where $K_d K_p a T \ll 1$, and is replaced by the constant ε, The jitter introduced by the delay line is represented by (EQ 11) in the z-domain since in the time domain the effect of one pass down the chain is just an error impulse.

$$\Theta_n(z) = \frac{2\pi \Delta t_n}{T} \qquad \text{(EQ 11)}$$

Therefore, the variance of the total output jitter can be shown to be

$$E\left[\Theta_{tot}^2(nT)\right] = \left(\frac{2\pi}{T}\right)^2 \overline{\Delta\tau_N^2}\left(1 + \frac{\varepsilon}{2 - \varepsilon}\right) \approx \left(\frac{2\pi}{T}\right)^2 \overline{\Delta\tau_N^2} \qquad \text{(EQ 12)}$$

and the rms output jitter is therefore given by (EQ 13).

$$\sqrt{E\left[\Theta_{tot}^2(nT)\right]} \approx \frac{2\pi \Delta\tau_{rms}}{T} \qquad \text{(EQ 13)}$$

This expression is very similar to the result for the PLL, given in (EQ 8), except now there is no noise enhancement factor α. Therefore a DLL provides superior timing jitter performance. How much better depends on the size of α which will be discussed in the next section.

III. Design Examples and Simulation

To verify the theoretical predictions for PLL/DLL jitter performance given in the previous sections, monte-carlo simulations were performed using a behavioral model built around the basic functional blocks pictured in Figure 3. The timing jitter generated by the noisy inverter cells in the ring-oscillator is modeled by a phase jitter noise source which adds an error phase to the ideal phase coming out of the VCO. This phase jitter is assumed white and its variance is proportional to the value of $2N(\Delta\tau_1)^2$ determined for a given ring-oscillator design using the results of [1]. The jitter is normalized to the period of the delay in order to determine the phase noise variance.

If a phase noise source is applied to a PLL with design parameters $K_d = 8.4/2\pi$ μA, $K_w = 2\pi$ x 20 MHz/volt, $R = 90\Omega$ and $T = 50$ nsec (same parameters as in [6]), then the total PLL jitter is shown in Figure 4. The total phase error wanders over a wider range than the unit variance input source. In this example the predicted accumulation factor is 25.7 and for a unit variance input noise, should yield a total PLL jitter of around 25.7. This is close to the result which was extracted from the simulation to be about 26. The simulation also indicates that the PLL shapes the free running ring oscillator phase fluctuation to a finite values with a variance (σ_P^2=676). Here, the input jitter variance is normalized to 1.

To reduce the jitter accumulation effect, a new design is simulated. In this design, the bandwidth of the PLL is increased by using a larger value for R (R=900). In this case, the calculated jitter accumulation factor α becomes 8.1. Figure 5 shows the simulation results. Now the PLL phase error wanders over a smaller range, and changes more rapidly since the loop bandwidth is higher. The noise enhancement factor for this data is 8.3 which is very close to the predicted result of 8.1.

Figure 6 shows the jitter accumulation factor in respect to the loop bandwidth, $w_L \approx K_d K_w a$ for two input jitter values. This figure shows that the PLL output jitter decreases as the loop bandwidth increases until the external jitter becomes dominant. Therefore, the optimum bandwidth is given by the minimum point of the curve. Smaller bandwidth is preferred when the input jitter is dominant.

Using the same input noise from the delay cells, a DLL output jitter is simulated. Figure 7 shows the output jitter for the same time period. This simulation shows that DLL does not accumulate jitter and performs better for the internal jitter sources such as delay cell jitter. In this case, parameter values are taken from [2] and given as C = 0.039 μF, K_d = 8.4/2p μA, K_P = 2π rads/volts and T = 50 nsec.

For a rough experimental verification, this model for timing jitter was compared to the jitter observed in the ring-oscillator PLL described in [5]. This PLL was fabricated in a 2 μm CMOS technology and the ring-oscillator was comprised of 16 inverter delay stages running at a frequency of 30-MHz. The jitter for this case can be calculated using (EQ 14) and is a function of the jitter contribution per stage ($\Delta\tau_1$), the number of taps in the oscillator (N), and the PLL accumulation factor (α). In [1], it is shown that for the circuit parameters used in the delay cells in [5], the jitter contribution per cell was $\Delta\tau_1$=2.09ps. The parameters for the PLL in this design were $K_d = 20/2\pi$ μA, $K_w = 2\pi$ x 5 MHz/volt, R = 200 Ω, and T = 33 nsec, giving a jitter accumulation factor of 38.9. Therefore, the r.m.s. timing jitter for this PLL, is predicted by (EQ 14) to be 81.30 ps. This agrees well with the experimental result in [5], which was not measured exactly, but determined to be somewhere in the range of 50-100 ps.

IV. Conclusion

This analysis has shown that, including the results of [1], the jitter in a ring-oscillator is proportional to three factors; the number of stages, the jitter contribution per stage, and a PLL accumulation factor α, which is inversely proportional to the square-root of the bandwidth of the PLL. For a DLL the result is the same, except the noise enhancement factor is 1. Therefore in applications such as clock synthesis, where a DLL can be used, it is the better choice for jitter performance. To reduce the jitter enhancement in a PLL a larger loop bandwidth should be used. For applications such as clock-recovery, however, this bandwidth cannot be increased too much or it will enhance the jitter seen in the input signal.

References

[1] Todd C. Weigandt, Beomsup Kim, Paul R. Gray, "Timing Jitter Analysis for High-Frequency, Low-Power CMOS Ring-Oscillator Design", *ISCAS*, June 1994.

[2] Beomsup Kim, "High Speed Clock Recovery in VLSI Using Hybrid Analog/Digital Techniques", *UCB/ERL Memorandum*, June 1990.

[3] Floyd M. Gardner, "Charge-Pump Phase-Locked Loops", *IEEE Trans. on Communications*, vol. COM-28, no. 11, Nov. 1980.

[4] J. Sonntag, R. Leonowich, "A Monolithic CMOS 10 MHz DPLL for Burst-Mode Data Retiming", *ISSCC*, vol. 33, pp. 104-105, Feb. 1990.

[5] Beomsup Kim, David H. Helman, Paul R. Gray, "A 30 MHz High Speed Analog/Digital PLL in 2um CMOS", *ISSCC*, vol. 33, pp.104-105, Feb. 1990.

[6] National Semiconductor, *Mass Storage Handbook*, pp 2.49-2.51, 1989.

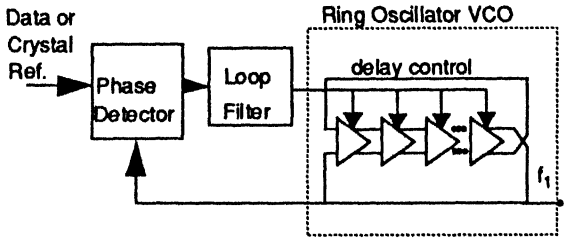


Figure 1 Ring-oscillator phase-locked-loop

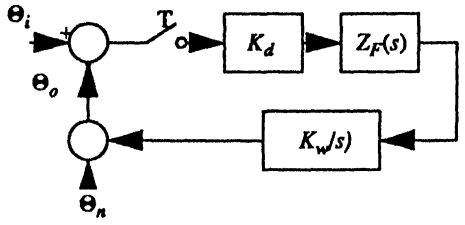


Figure 2 Simplified PLL Discrete Time Model

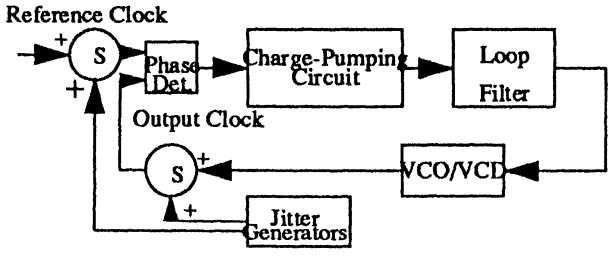


Figure 3 Jitter Simulation Setup

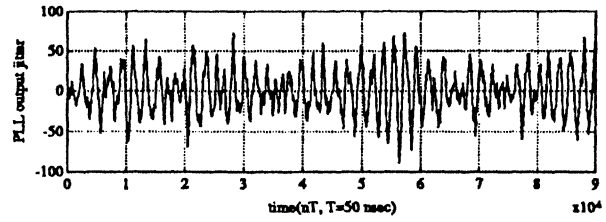


Figure 4 PLL Jitter Accumulation Effect Simulation (w_L = 15 KHz)

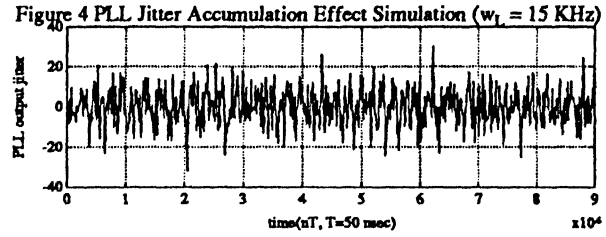


Figure 5 PLL Jitter Accumulation Effect Simulation (w_L = 150 KHz)

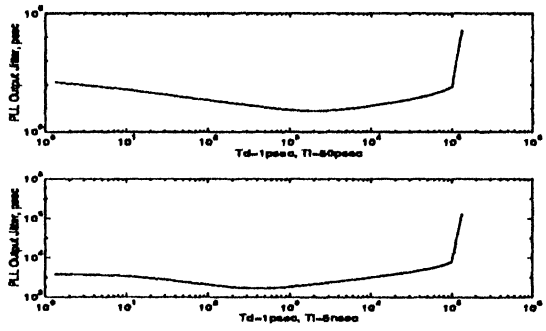


Figure 6 Jitter vs. Loop Bandwidth, VCO Noise Dominant Case (Upper Figure) and Input Noise Dominant Case (Lower Figure)

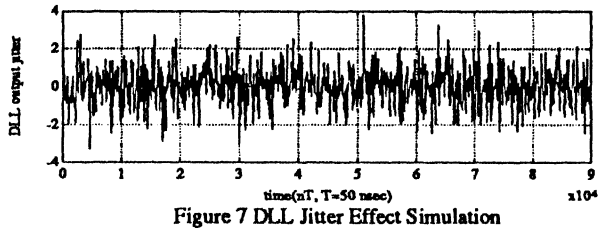


Figure 7 DLL Jitter Effect Simulation

PRACTICAL APPROACH AUGURS PLL NOISE IN RF SYNTHESIZERS

By following a graphical analysis routine, phase noise can be predicted accurately.

NOISE and spurious signals can be analyzed for even the most complex of phase-locked loop (PLL) synthesizer architectures. Even synthesizers utilizing dual-modulus prescalers, fractional-N dividers, and translation loops can be understood.[1,2]

For example, the presence of the feedback divider makes the PLL a sampled data system.[3] However, the simpler continuous approximation is accurate if the loop bandwidth is less than 20 times the sampling rate (the comparison frequency). With this approximation, a Laplace transform analysis of the basic synthesizer yields the following results for the open-loop gain and the transfer function:

$$G(s) = \frac{K_\alpha K_0 F(s)}{s N} \qquad (1)$$

where:

G(s) = the open-loop gain, and

$$H(s) = \frac{\phi_0(s)}{\phi_i(s)} = \frac{G(s)}{1 + G(s)} =$$

$$\frac{K_\alpha K_0 F(s)}{s N + K_\alpha K_0 F(s)} \qquad (2)$$

where:

H(s) = the transfer function.

For a second-order, type-II loop:

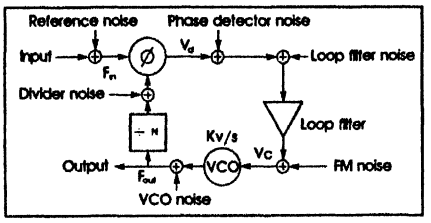
$$F(s) = \frac{s \tau_2 + 1}{s \tau_1} \Rightarrow H(s)$$

$$= \frac{2 s \xi w_n + w_n^2}{s^2 + 2 s \xi w_n + w_n^2} \qquad (3)$$

where:

$$w_n = \sqrt{\frac{K_\alpha K_0}{\tau_1 N}} \qquad (4a)$$

and,



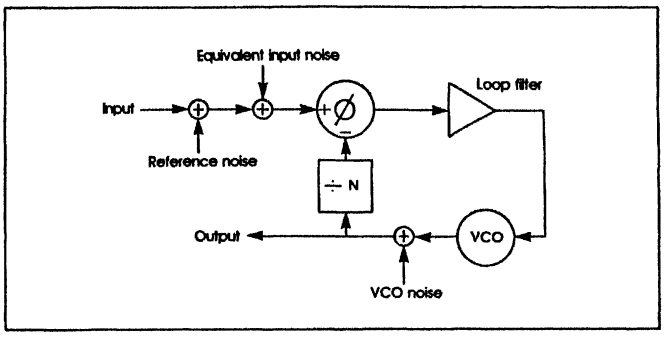
1. This PLL noise model shows all of the relevent individual noise sources.

$$\xi = \frac{\tau_2}{2} \sqrt{\frac{K_\alpha K_0}{\tau_1 N}} \qquad (4b)$$

A well-known PLL noise model (Fig. 1) includes the individual sources of noise within the synthesizer.[4] Each of these sources arises from a different mechanism. Loop-filter noise arises from the equivalent input noise sources of the DC amplifier, if one is used, and from logic circuit and current source noise, if a pure switching charge pump is utilized. Phase-detector noise results whether a digital or analog PLL is used. In either case, the phase detector will degrade the phase difference signal by adding white and flicker noise.

VCO noise, another oscillator noise source, is measured while the oscillator is free-running under laboratory conditions.[5,6] FM noise occurs when the VCO is operating in a synthesizer and it is subjected to several noise sources that modulate its frequency. These sources include

(continued on next page)



2. A consolidated noise model shows both the high-pass and low-pass terms in Eq. 6.

MARK O'LEARY, Consultant, Commdesign, 21213B Hawthorne Blvd. Ste. 5576, Torrance, CA 90509; (213) 370-3298

Reprinted with permission from *Microwaves & RF*, M. O'Leary, "Practical Approach Augurs PLL Noise in RF Synthesizers," pp. 185-194, September 1987.

control-voltage pickup (for example, capacitive coupling from nearby digital circuits and audio oscillators), noise on the VCO supply voltage (all VCO's have some frequency sensitivity to supply voltage), and vibration (through frequency sensitivity to mechanical stress). These effects are modeled by a single noise source summed with the VCO control voltage.

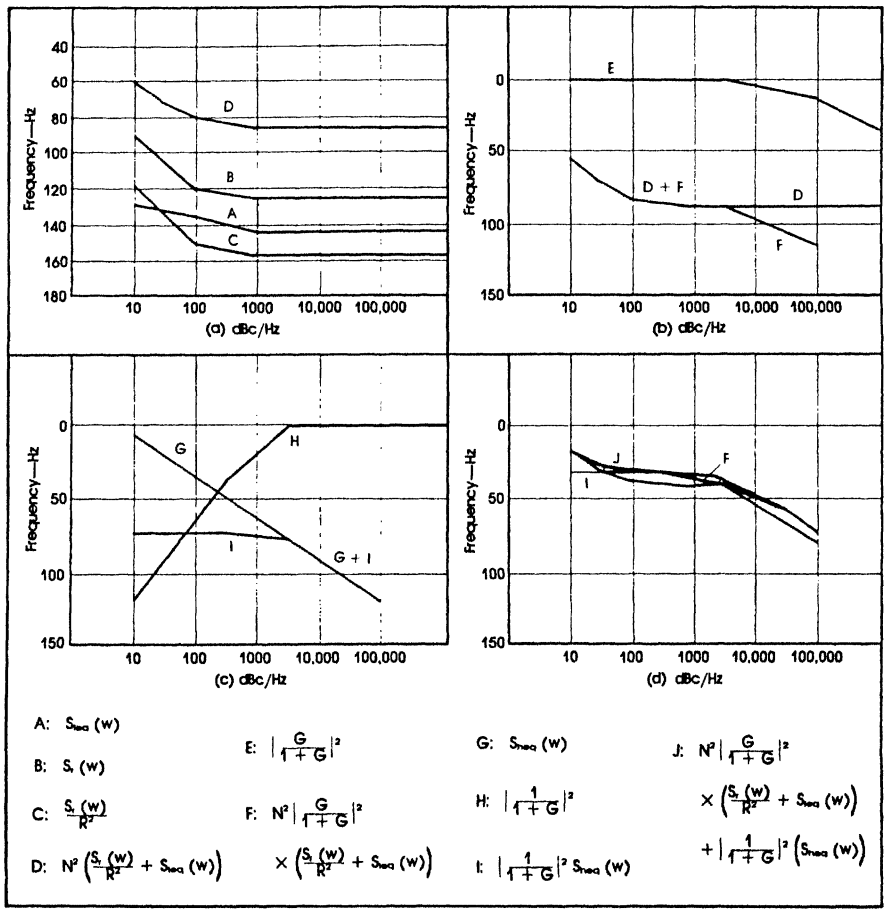
Divider noise occurs when a divider's output contains phase information in the position of its rising edges, which are influenced by electrical fluctuations within the divider. The divider noise is modeled as a source summed at the divider output, since contributions from higher stages are reduced by the factor, 20 log N, where N is the divide ratio. By modeling divider noise at the output, this source becomes nearly independent of N and constant for any given logic family.

All of the above noise sources can be measured and evaluated. For example, op-amp input noise data can be gleaned from data sheets; such publications contain much information on divider and phase detector noise.[7,8] However, this approach is tedious and time-consuming. The only information necessary is the synthesizer's output phase noise. Once this information is known, the PLL can be made to meet the system specification. A consolidated noise model is used to analyze the phase noise (Fig. 2).

Applying a Laplace transform analysis to this model results in the expression for output phase noise:

$$S_0(w) = N^2 \left| \frac{G(s)}{1 + G(s)} \right|^2 \left(\frac{S_r(w)}{R^2} + S_N(w) + \frac{S_{VLF}(w)}{K_D^2} + \frac{S_{Vpd}(w)}{K_D^2} \right) + \left| \frac{1}{1 + G(s)} \right|^2 \left(S_{vco}(w) + \frac{S_{Vfm}(w)\, K_0^2}{S^2} \right) \quad (5)$$

Eq. 5 can be rewritten as,



3. A graphical solution is available, which gives phase-noise output for a PLL synthesizer.

■

$$S_0(w) = N^2 \left| \frac{G(s)}{1 + G(s)} \right|^2 \times \left(\frac{S_r(w)}{R^2} + S_{leq}(w) \right) + \left| \frac{1}{1 + G(s)} \right|^2 \left(S_{heq}(w) \right) \quad (6)$$

In Eq. 6, two kinds of noise sources have been combined. The first type of noise source, to which the loop response is low pass, is called the equivalent input noise, $S_{leq}(w)$. The second source, to which the loop response is high pass, is denoted in-circuit VCO noise, as $S_{heq}(w)$.

Note that S_{leq} is independent of the feedback divider ratio and is characteristic for a given synthesizer technology—the type of phase detector or loop filter, or the logic family of divider output stages.

The consolidated noise model is useful in that the equivalent input is easily measured, as is the VCO noise (Fig. 3). To measure equivalent input noise for a given technology, the user simply builds a synthesizer with a high division ratio and wide-loop bandwidth, and locks it to a very clean reference frequency. Equivalent input noise is then the output phase noise reduced by 20 log N.

With this model, the importance of N in determining the output phase noise of a given synthesizer is immediately apparent. The output phase noise increases in dB as the value of 20 log N increases for offset frequencies within the loop bandwidth. Therefore, a constraint on any synthesizer design is to keep N low.

The equivalent input noise also constitutes a noise floor for the synthesizer. Regardless of how quiet the input signal is, the output phase noise will be at least 20 log N × $S_{leq}(w)$ for frequency offsets within the loop bandwidth. One effect is that output noise of PLL synthesizers is not necessarily a multiplied version of the reference, as often believed.

Another implication of Eq. 2 is the strong dependence of phase-noise performance on loop response. Both the loop bandwidth and the sharpness of loop roll-off are very important. To illustrate these concepts, Fig. 3 demonstrates a graphical so-

lution to Eq. 2 for a hypothetical synthesizer.

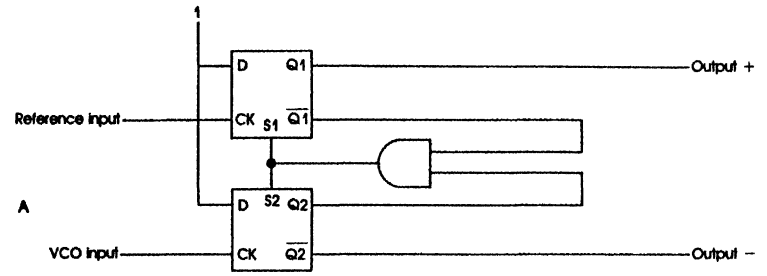
COMPARISON FREQUENCY

Generally, the synthesizer output will have sidebands offset from the carrier frequency by the comparison frequency and its harmonics. For the simplest LO synthesizers, the comparison frequency corresponds to the step size. Therefore, comparison-frequency sidebands will limit the receiver signal-to-noise (S/N) ratio by mixing adjacent channel signals into the fixed intermediate frequency (IF) along with the desired signal.[9] If the total power in each channel slot is the same, the receiver S/N ratio is limited to B – 3 dB, where B (in dBc units) is the level of the comparison-frequency sidebands. Of course, comparison frequency does not correspond to step size for fractional-N synthesizers. Nevertheless, the sidebands must still meet certain spurious-signal specifications.

Sideband levels depend on the details of phase-detector and loop-filter implementation. For example, consider a synthesizer that is a mixed-sampled and continuous-time system. Almost always, the transition from sampled to continuous systems is done at the phase-detector output. This is because the loop filter is most easily implemented as an analog integrator, while sampled phase detectors are preferred to continuous ones. Continuous phase detectors, such as multipliers and switched mixers, are inherently low-gain, high-offset, high-drift devices. Also, their good threshold performance is not advantageous in the high S/N environment of a synthesizer. On the other hand, digital phase detectors easily interface to dividers and are less expensive.

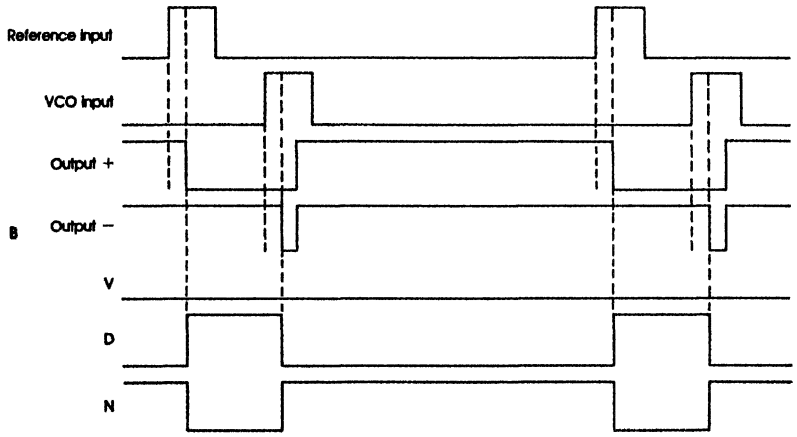
The class of sequential digital phase detectors includes very simple circuits such as gates and flip-flops, as well as the phase/frequency detector (Fig. 4).

The phase/frequency detector (PFD) has a wide linear range (that



PFD locked-state table

Q1	Q2	V	D	N
0	0	0	0	1
0	1	0	1	0
1	0	1	0	0
1	1	0	0	1



4. A common version of the phase/frequency detector (a) operates as illustrated by the characteristic waveforms (b).

■

is, $\pm 2\pi$) and possesses the remarkable feature of functioning as a frequency comparator when it is out of phase lock.[10] Fig. 4 illustrates the locked-condition operation of a common version of the phase frequency detector. The leading edge of the signal causes its corresponding output to be set; the other input edge, after setting its output, allows both outputs to be reset after two gate delays. This creates a linear phase detector with a range of $\pm$ 1 cycle, the phase being encoded into the pulse width of the difference of the two outputs.

A device called a charge pump takes the digital output of the phase/frequency detector and produces an analog signal suitable to drive the loop filter.[11] The charge pump recognizes three independent states: pump up (U), pump down (D), and neutral (N). It sources current to the loop filter when U is true, sinks current when D is true, and is isolated when N is true. If U, D, and N are derived from the PFD output states, as shown in Fig. 4, the charge pump will provide the following average current to the loop filter each cycle:

$$\overline{i_d(t)} = \frac{I_p \theta_e}{2\pi} \qquad (7)$$

Assuming the loop bandwidth is much less than the comparison frequency, the sampling period average describes the system:

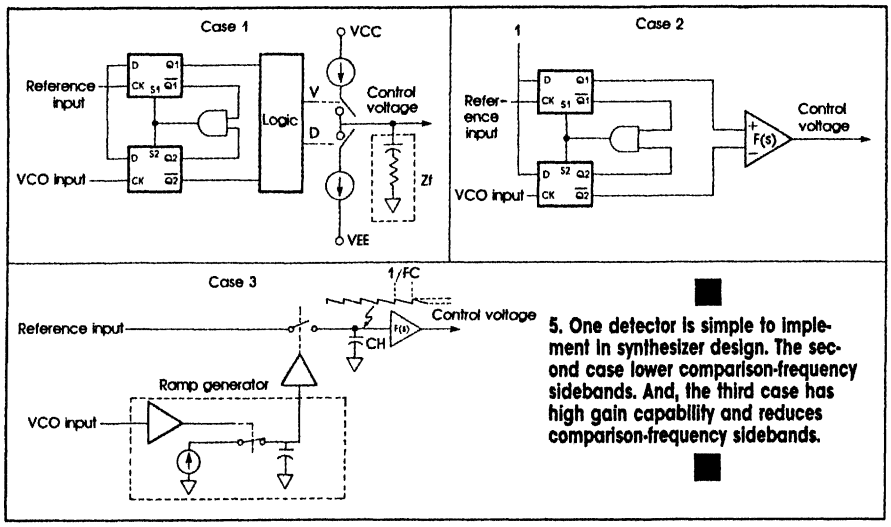
$$I_d(s) = \frac{I_p \theta_e(s)}{2\pi} \qquad (8)$$

The loop filter consists of an impedance, $Z_F(s)$, so that:

$$V_c(s) = I_d(s) Z_F(s)$$

$$= \frac{I_p}{2\pi} \theta_e(s) Z_F(s) \qquad (9)$$

A charge pump may produce a pump voltage instead. When this voltage is averaged over one cycle and an active loop filter with a transfer function, F(s), is used, this results in:



5. One detector is simple to implement in synthesizer design. The second case lower comparison-frequency sidebands. And, the third case has high gain capability and reduces comparison-frequency sidebands.

$$V_c(s) = K_d\,\theta_e(s)\,F(s) \qquad (10)$$

To summarize, the PFD accepts digital inputs, produces an analog output in conjunction with a charge pump, and causes comparison-frequency sidebands because of the pulsed nature of its output.

The other type of phase detector of practical importance is the sample-and-hold phase detector. One input clocks a ramp (if a linear characteristic is desired) or another waveform. The other input samples the waveform. The sampled voltage is held until the next cycle. This process implements a sampled phase detector and zero-order hold with a linear range of $\pm\pi$. The sample-and-hold phase detector has several advantages and disadvantages relative to phase/frequency detectors.

Advantages include reduced comparison-frequency sidebands; high gain capability resulting in noise advantages; and accurate Z-transform analysis and design. Disadvantages include the lack of built-in frequency-acquisition capability and the need for analog circuitry.

The comparison-frequency sideband levels for the three cases can now be computed (Fig. 5). The first case is the phase/frequency detector with a switching charge pump. Leakage current out of the loop-filter impedance node results in the creation of comparison-frequency sidebands. In the phase-locked steady-state case, this current, which may include charge-pump switch varactor bias and filter-capacitor leakage currents, must be cancelled by the average charge-pump current each cycle to maintain the filter capacitor at the correct voltage. The charge-pump supplies current in pulses of peak value, I. These pulses modulate the VCO to make the sidebands. (Assume that the loop filter contains a capacitor, as is the case with the usual type-II loop.)

The level of the sidebands can be calculated as follows. In the steady state, the charge pump supplies current in pulses of peak value, I_p, and width:

$$D = \frac{I_l}{I_p} \qquad (11)$$

From Fourier analysis, the RMS level of the pulse train at the nth harmonic of the comparison frequency is as follows:

$$I_{RMS} = \frac{I_p D}{\sqrt{2}} \quad \text{for } n \ll \frac{1}{D} \qquad (12)$$

or,

$$I_{RMS} = \frac{I_l}{\sqrt{2}} \quad \text{for } n \ll \frac{1}{D} \qquad (13)$$

To compute the output phase modulation, it is easiest to let the phase-detector output current harmonics be the input of the loop and use the ordinary closed-loop transfer function. The following equation relates PFD output, I_d, to input Θ_e:

$$I_d = \frac{\theta_e\, I_p}{2\,\pi} \qquad (14)$$

so,

$$\theta_e = I_l \frac{\sqrt{2}\,\pi}{I_p} \qquad (15)$$

The output response to this equivalent input (which arose due to the pulsed nature of the charge pump) is:

$$\theta_0^n = \frac{N\,H\,(nf_c)\,I_l\,\sqrt{2}\,\pi}{I_p} \qquad (16a)$$

and,

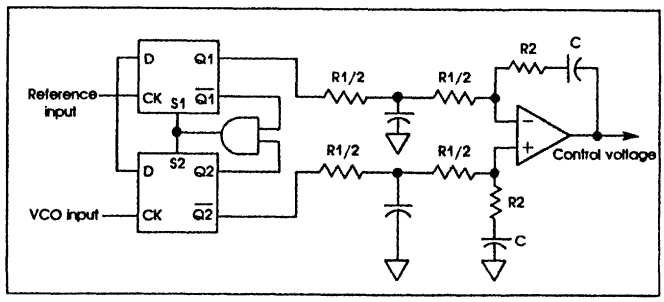
$$B = 20 \log\left[\frac{N\,H\,(f_c)\,I_l\,\sqrt{2}\,\pi}{I_p}\right] \text{dBc},$$

$$\text{for } B < -20 \text{ dBc} \qquad (16b)$$

Assuming that N is constrained by the architecture, the sidebands can still be reduced. $H(f_c)$—where f_c equals the carrier frequency—can be decreased either by lowering the loop bandwidth or by adding a pole to the loop filter. The extra pole must be no lower than about 10 times the loop bandwidth to maintain adequate stability margin. Next, the charge pump can be improved. As evident in the above equation, I_p/I_l is a figure-of-merit for the comparison-frequency sideband performance of a charge pump. The value of f_c should be maximized in order to minimize the sidebands.

The second case, for which com-

6. A differential amplifier charge pump eliminates finite-switching-time distortion.



parison-frequency sidebands can be computed, is a phase/frequency detector with an amplifier charge pump. The previous scheme—a PFD with a switching charge pump—has the advantages of simple implementation, rapid acquisition, and no need for a DC amplifier. However, the charge-pump switching time imposes a limitation on input zero phase-error resolution. The phase detector cannot distinguish input edges whose arrival time differences are less than the switching time of the charge pump.

Referring to Fig. 4, the transformation between the PFD'S 2-bit output and the charge pump's three-state input can be accomplished using subtraction with a linear circuit instead of a digital one. The circuit for which a differential amplifier implements the charge pump and loop filter is shown in Fig. 6.

There is no limitation on the zero-phase resolution with the circuit in Fig. 6. Comparison-frequency sideband levels are calculated from the formula derived above, using the op-amp input offset current in place of I_I. The op-amp input voltage does not affect the sidebands unless a resistive buffer stage is used in the circuit before a single-ended loop filter. This is because the capacitors will hold the correct voltage to cancel the input offset voltage and obtain the necessary VCO input. If a resistive buffer stage is used, the sideband levels are:

$$B = 20 \log \left[\frac{N \mid H(f_c) \mid V_{os}}{2\sqrt{K_D}} \right] \text{dBc}, \quad \text{for } B < -20 \text{ dBc} \qquad (17)$$

where:

$K_D = V_{cc}/4\pi$.

However, this is unnecessary since the charge pump, besides solving the phase-resolution problem mentioned, is also capable of a high figure-of-merit:

$$\frac{I_p}{I_I} = \frac{V_{cc}}{R\, I_{os}} \qquad (18)$$

This value can be greatly increased by using a FET input amplifier. Disadvantages to the PFD/difference-amplifier approach are that it requires an op amp, op-amp noise is not negligible in some cases, and the loop bandwidth is limited by amplifier slew rate. However, these aren't normally important factors in commercial areas of the LO synthesizer performance envelope.

The third case is the sample-and-hold phase detector. In a type-II loop with a sample-and-hold phase detector, ripple due to hold-capacitor droop results in the comparison-frequency sidebands. Assuming the phase detector is well designed so that switching transients are negligible the sidebands can be calculated as follows. The phase detector output voltage, V_d, will be a sawtooth waveform (Fig. 5). From Fourier analysis, the level of V_d at the nth harmonic is:

$$V_d = \frac{A}{N\pi} \qquad (19)$$

And, since:

$$A = \frac{1}{F_c} \frac{d V_d}{d t} = \frac{I_I}{C_H f_c} \qquad (20)$$

Then:

$$V_d = \frac{I_I}{C_H f_c\, n} \qquad (21)$$

where:

I_I = the leakage current out of the hold-capacitor node, between sampling instants.

Using $V_d = \Theta_i K$, the ripple can be sent to the input for ease of computation. This gives a fictitious input signal at the nth harmonic of:

$$\theta_I^n = \frac{I_I}{C_H f_c\, n\, \pi\, K_D} \qquad (22)$$

Then, using Eq. 21, the output phase is:

$$\theta_0^n = \frac{N\, H(nf_c)\, I_I}{C_H f_c\, n\, \pi\, K_D} \qquad (23a)$$

And, the comparison-frequency sidebands are "D" dB down from the carrier, where:

$$B = 20 \log \left[\frac{N\, H(f_c)\, I_I}{C_H f_c\, \pi\, K_D} \right] \qquad (23b)$$

The figure-of-merit for sideband performance of the sample-and-hold phase detector is slow and unreliable. A convenient way to obtain frequency-aided acquisition is to use a PFD and switching charge pump as a parallel phase detector. Outputs are then summed for both the PFD and the switching charge-pump circuits, with the PFD gain much lower so that performance is not harmed while both circuits are in phase lock. Some CMOS synthesizer parts provide both types of phase detectors for this reason. ••

References

1. Floyd M. Gardner,"Phase-lock Techniques," *Wiley Interscience*, 2nd Ed., 1979, pp. 8-16.
2. D. Brewerton and N. Urbaneta, "Defining the elements of good design," *Microwaves & RF*, Vol. 23, Jun. 1984, pp. 79-125.
3. James A. Crawford, "Understanding the specifics of sampling in synthesis," *Microwaves & RF*, Vol. 23, Aug. 1984, pp. 120-144.
4. L.S. Cutler and C.L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," *Proc. IEEE*, Vol. 54, Feb. 1966, pp. 136-154.
5. Gerald Sauvage, "Phase noise in oscillators: a mathematical analysis of Leeson's model," *IEEE Transactions on Instruments and Measurements*, Vol. IM-26, Dec. 1977, pp. 408-410.
6. K. Kurokawa, "Noise in synchronized oscillators," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-16, Apr. 1968, pp. 234-240.
7. V.F. Kroupa, "Noise properties of PLL systems," *IEEE Transactions on Communications*, Vol. COM-30, Oct. 1982, pp. 2244-2252.
8. Dieter Scherer, "Design principles and test methods for low-phase-noise RF and microwave sources," Hewlett-Packard.
9. C. John Grebenkemper, "Local oscillator phase noise and its effect on receiver performance," Watkins-Johnson Co. Technical Notes, Vol. 8, Nov./Dec. 1981, pp. 1-13.
10. J.I. Brown, "A digital phase and frequency-sensitive detector," *Proc. IEEE*, Vol. 56, Apr. 1971, pp. 717-718.
11. Floyd M Gardner, "Charge-pump phase-lock loops," *IEEE Transactions on Communications*, Vol. COM-28, Nov. 1980, pp. 1849-1858.

The Effects of Noise in Oscillators

ERICH HAFNER, MEMBER, IEEE

Abstract—An explicit expression for the output signal from an oscillator with several noise sources in the circuit is derived. This formula describes qualitatively and quantitatively the manner in which thermal and shot noise act to corrupt the performance of an ideal oscillator. The statistical properties of the signal are then evaluated, as it emerges from the oscillator stage, after passage through an output filter and after being operated on by an ideal n-times multiplier. Expressions are derived for the short term frequency stability, the power spectral density, and the power spectrum of the signal, as well as for the spectral density of the signal phase.

The key to the results reported is an apparently novel perturbation technique which does not require smoothing of the instantaneous nonlinearity in the basic differential equation. Discussion of the solutions shows that the instantaneous nonlinearities cause the device to act simultaneously like a linear AGC oscillator and like a high Q passive tuned circuit, with each aspect accorded one half the total noise excitation. Possible implications of this effect for other types of transient conditions in oscillators are indicated briefly.

Introduction

THE THEORY of noise in nearly harmonic oscillators has received considerable attention in the past, and a number of properties of the noise-perturbed signal are firmly established. However, it is very well recognized that the existing theories provide only partial descriptions of a many-sided phenomenon, and that several important questions have remained unanswered.

A common feature of the literature on the subject is the derivation of the power spectral density of the signal voltage in a noise-perturbed oscillator. This provides a convenient reference for discussing the most essential aspects of the earlier work. Some investigators [1]–[4] chose to consider the oscillator simply as a linear noise filter of very narrow bandwidth; and they arrived at results which generally agreed (a factor of two which often occurred is now known to be extraneous [5], [6]) with those derived by far more sophisticated techniques [8]–[13]. Since these latter techniques are based on the nonlinear differential equations for the noise, perturbed oscillator, a satisfactory physical explanation for this agreement could not readily be offered, primarily because the significance of the approximations involved was rather difficult to assess.

It does not appear to have been fully realized that, as is shown in Section III-C of this paper, the linear noise-filter approach contains implicit assumptions which can be met only when the device considered is indeed a linear oscillator, equipped with an external mechanism (AGC) which automatically regulates the gain of the active device, or the circuit losses. On the other hand, the most important step, consistently taken in all nonlinear analyses reported so far, is the application of the averaging principle [14] which involves smoothing of the instantaneous nonlinearity in the circuit [12]. It is shown in the Appendix that this approximation is equivalent to replacing the actual nonlinear oscillator again by a linear oscillator with AGC. The agreement just mentioned is thus to be expected, even though explicit equations describing the behavior of a linear AGC oscillator have only recently become available [15].

Whereas the power spectral density of the oscillator signal is, of course, a very useful piece of information, the derivations of the expressions for it do not contain enough parameters to provide a clear picture of what the oscillator signal is really like, and just why real oscillators do not behave as they are supposed to according to these theories. When investigating the performance of systems fed by a signal from a noise-perturbed oscillator, instead of by a pure sinusoid, it has been necessary [16] to invent working models for this signal, making numerous and varied ad hoc assumptions in the process. Also, only a limited amount of useful guidance for the development of improved devices could be extracted from the analytic investigations of the oscillator itself.

To extend the results of the earlier work, an oscillator model was chosen for the present analysis which closely resembles an actual quartz crystal oscillator. The several noise sources in the circuit are assumed at first to generate impulses of random strength at randomly spaced intervals, and the effects of white noise on the oscillator signal are obtained by appropriate summation over the disturbances caused by the individual impulses. The result is an explicit expression for the output signal from the noise perturbed oscillator. It reveals that, regardless of their location in the circuit, the effects of all noise sources are essentially equivalent. However, the source which appears directly across the output of the oscillator also contributes an additive white noise component to the output signal, which plays a very significance role in high Q oscillators.

Although the statistical properties of this signal are computed after the series of impulses are replaced in Section III-A by continuous random functions, the response of the oscillator to a single noise impulse is found to be a very powerful tool, peculiarly well suited for the investigation of heretofore unexplored areas of oscillator behavior. Only those aspects which pertain to noise effects are discussed in detail.

The key to the advances reported here is an appar-

Manuscript received October 29, 1965; revised November 29, 1965.

The author is with the U. S. Army Electronics Command, Ft. Monmouth, N. J.

Reprinted from *Proc. of IEEE*, vol. 54, pp. 179-198, February 1966.

ently novel technique for the solution of the perturbation equation. It is developed in the Appendix. Unlike the earlier techniques, it fully recognizes the instantaneous character of the oscillator nonlinearity, and does not employ the smoothing concept. It also provides a clear appreciation of the significance of the approximations which are made to arrive at reasonably compact expressions.

The major consequence of the presence of instantaneous nonlinearities in the circuit is that they cause the oscillator to act simultaneously like a linear AGC oscillator and like a passive tuned circuit whose effective quality factor is inversely proportional to the nonlinearity parameter, with each aspect accorded one half the total perturbing excitation. While, with white noise excitation, the former aspect is responsible for the familiar random walk phenomenon in the oscillator phase, the latter aspect contributes with each impulse a phase disturbance that decays slowly to zero, usually with a very long time constant.

In regards to white noise excitation of oscillators with time invariant circuit parameters, the existence of the tuned circuit aspect is perhaps of limited significance for practical applications. This is so primarily because the noise effects from within the oscillator loop are in many cases less important than the additive white noise from the source across the oscillator output. However, the virtual tuned circuit is also excited by sudden changes in the circuit parameters of the oscillator; and it is quite likely that, so far, the behavior of the signal phase in oscillators with time variable circuit elements has largely defied theoretical description, just because the slow decay of this excitation has not been reckoned with. It is strongly suspected that future work in this area will be considerably more successful if the presence of the virtual tuned circuit is admitted.

The impetus to the work reported here was provided by the need to determine in detail the effects of noise in lumped parameter oscillators, particularly crystal oscillators, without resort to the many intuitive concepts which are often employed. The results, however, are applicable to all major classes of oscillators, and this includes masers and lasers.

While oscillators of this latter type are governed by the laws of quantum mechanics and electrodynamics, those laws which are of consequence here can quite generally be recast in classical form [17], and the description of the device in terms of a van der Pol oscillator becomes possible. In fact, Lamb's equations [18] that pertain to lasers in single mode operation are essentially the equations for a linear AGC oscillator. The analysis there is concerned primarily with the steady state and the question of whether or not the nonlinearity is instantaneous—that is, whether or not the tuned circuit aspect exists in lasers, too—does not arise. This question could conceivably be quite important in explaining the role of frequency pulling and entrainment phenomena in multimode oscillators during build-up and transient phases of operation.[1] Evidently, it cannot yet be answered with certainty.

I. The Basic Differential Equations

An oscillator model which closely resembles a quartz crystal oscillator and yet is still manageable analytically is shown in Fig. 1. All elements of the feedback network are considered constant unless specifically stated otherwise. The active device in the oscillator is assumed to have infinitely high input and output impedances, and to generate a current i_1, which is a nonlinear function of the output voltage of the feedback network:

$$i_1 = f(e_0). \tag{1}$$

The current generator i_s and the voltage generators v_1, v_2, and v_3 inject extraneous signals into the oscillator whose effects are to be evaluated.

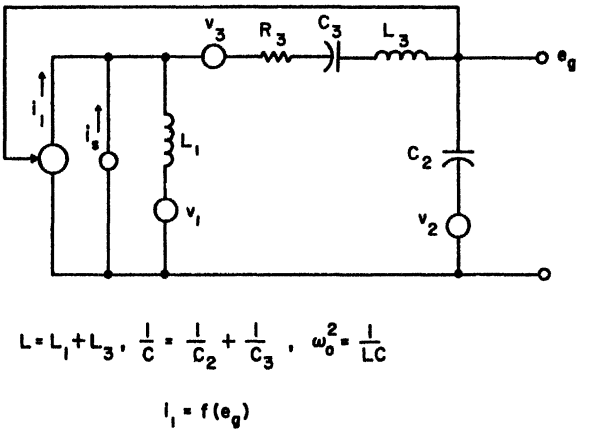


Fig. 1. Oscillator Model. The feedback network is fed from an ideal current generator whose strength i_1 is a nonlinear function of the output voltage e_0. The generators i_s, v_1, v_2, and v_3 are assumed to be white noise sources. The R_3, L_3, C_3 branch approximates, by proper choice of the parameter values, the action of a quartz crystal unit.

The differential equation describing the form and behavior of the output voltage e_0 is best derived by starting with the equation

$$e_0 = \frac{Z_1 Z_2}{Z_1 + Z_2 + Z_3}(i_1 + i_s) + \frac{Z_2}{Z_1 + Z_2 + Z_3}(v_1 + v_3) + \frac{Z_1 + Z_3}{Z_1 + Z_2 + Z_3} v_2, \tag{2}$$

which follows readily from Fig. 1. Considering the Z_j as operational impedances [19]

$$Z_1 = pL_1, \quad Z_2 = 1/pC_2, \quad Z_3 = R_3 + pL_3 + 1/pC_3$$

and replacing, after some rearrangements the differentiation operator p by d/dt, one finds with (1)

$$\left[\frac{d^2}{dt^2} + \frac{L_1}{LC_2}\left(\frac{R_3 C_2}{L_1} - \frac{df(e_0)}{de_0}\right)\frac{d}{dt} + \omega_0^2\right] e_0 = F(t) \tag{3}$$

[1] Multimode laser oscillators can be represented, equivalently as shown by Lamb, by an equal number of van der Pol oscillators, weakly coupled to one another.

whereby

$$L = L_1 + L_3, \quad \frac{1}{C} = \frac{1}{C_2} + \frac{1}{C_3}, \quad \omega_0^2 = \frac{1}{LC} = \frac{1}{L_1 C_2} \tag{4}$$

and

$$F(t) = \frac{L_1}{LC_2}\frac{di_s}{dt} + \omega_0^2 \frac{C}{C_2}(v_1 + v_3) + \left(\frac{d^2}{dt^2} + \frac{R_3}{L}\frac{d}{dt} + \frac{C}{C_3}\omega_0^2\right) v_2. \tag{5}$$

Whereas the circuit in Fig. 1 contains four reactive components compared to the two customarily assumed in noise analyses of oscillators, it is important to realize that the two additional components do not cause a qualitatively more complicated behavior of the oscillator. The basic differential equation (3) is only of second order, and its essential features are not at all altered when these components are eliminated from the circuit. The ratios L_1/L and C/C_2 revert to unity for a simple LC oscillator in which $L_3 = 0$ and $C_3 \to \infty$. They carry the information required to describe the effects of the simultaneous presence of a high Q and a low Q element in the oscillator, without adding to the labor in the analysis. This information is useful when dealing with quartz crystal oscillators or with masers and lasers.

For simplicity, and to make the general applicability of the results more apparent, the following symbols will be used:

$$\frac{\omega_1 L}{R_3} \equiv Q_T \equiv \frac{1}{\gamma_T}; \quad \frac{\omega_1 L_1}{R_3} \equiv Q_N \equiv \frac{1}{\gamma_N}; \quad \frac{L}{L_1} \equiv \bar{Q}\left(= \frac{Q_T}{Q_N}\right). \tag{4a}$$

Q_T is the effective quality factor of the entire passive feedback network; Q_N is its quality factor when the high Q element is replaced by its series resistance at ω_1. $\bar{Q}$, the ratio of these two, shall be called the reduced quality factor of the oscillator. Later on, an additional symbol Q_C, will be introduced for the quality factor of the filter following the oscillator. No Q symbol is used for the virtual tuned circuits representing certain performance aspects of the disturbed oscillator; their effective relative bandwidths will be characterized by γ, to be defined in (22).

Of major concern in this paper are the properties of the solutions of (3) when the extraneous signals are random noise. Stable, nearly harmonic solutions are assumed to exist; and we restrict ourselves to the stationary-state properties of such solutions.

If the noise sources in Fig. 1 are quiescent, (3) becomes completely deterministic, and it is well known that, with $F(t) = 0$, there are steady-state solutions which are oscillatory in nature, provided certain conditions are met [20], [21]. The presence of the noise sources is not essential for these solutions to exist; rather, with increasing intensity, the noise causes progressively more severe random disturbances of the deterministic solutions.

The solutions of (3) with $F(t) = 0$ are perfectly periodic in the steady state and can be represented in the form

$$e_{g0} = \sum_i A_{i0} \cos(i\omega_1 t + \varphi i). \tag{6}$$

For a harmonic oscillator capable of steady-state oscillations, the A_{i0} are not all zero and ω_1 is in the neighborhood of ω_0 defined in (4). With all transients in the infinite past, the A_{i0} and φ_i are constants whose values, as well as the value of ω_1, can be determined at least in principle to any desired accuracy.[2]

The noise sources in Fig. 1 introduce transients into the system and the steady state cannot be maintained exactly; the various harmonics of the solution of (3) never have sharply defined amplitudes and phases.

Consequently we write e_g in the form

$$e_g = \sum_i (A_{i0} + a_i(t)) \cos(i\omega_1 t + \varphi_i + \phi_i(t)), \tag{7}$$

where $a_i(t)$ and $\phi_i(t)$ are stochastic variables representing the amplitude and phase disturbances caused by the noise. Equation (7), in turn, can always be written as

$$e_g = e_{g0} + u(t) \tag{8}$$

where e_{g0} is the solution (6) of the unperturbed oscillator and $u(t)$ represents all disturbances and only the disturbances.

When (8) is substituted into (3) and the fact that e_{g0} satisfies the unperturbed equation identically is considered, one finds an equation for $u(t)$ alone. If the disturbances are small, so that terms of higher order in $u(t)$ are negligible, the following perturbation equation, linear in $u(t)$ is obtained:

$$\left[\frac{d^2}{dt^2} + \frac{L_1}{LC_2}\left(\frac{R_3 C_2}{L_1} - \frac{df(e_{g0})}{de_{g0}}\right)\frac{d}{dt} + \left(\omega_0^2 - \frac{L_1}{LC_2}\frac{df(e_{g0})}{de_{g0}}\right)\right] u(t) = F(t). \tag{9}$$

The time variable coefficients of (9) depend upon e_{g0}, the steady-state solution of the undisturbed oscillator, and hence are known.

Assuming higher order terms in $u(t)$ to be negligible already implies that $x_i(t)$ and $\phi_i(t)$ in (7) are small; and $u(t)$ can be written to within the same degree of approximation required to obtain (9) as

$$u(t) = \sum_i [a_i(t)\cos(i\omega_1 t + \varphi_i) - A_{i0}\phi_i(t)\sin(i\omega_1 t + \varphi_i)]. \tag{10}$$

[2] When (6) is substituted into (3) with $F(t) = 0$ and the principle of the harmonic balance [22] is applied, an infinite set of nonlinear algebraic equations results which can be solved by an iteration procedure. The effects of harmonic content on oscillator frequency have been studied extensively by Groszkowski [23].

The fact that this approximation is justified, in particular that

$$|\phi_i(t)| \ll 1, \tag{11}$$

remains to be verified once $u(t)$ is computed for any given situation.

Evidently, a basic problem which must be dealt with herein is to determine the solution $u(t)$ of (9) when $F(t)$ depends upon the white noise sources in the oscillator according to (5).

The most realistic representations of the noise current i_s and the noise voltages v_1, v_2, and v_3 in $F(t)$ are series of delta functions of variable strength and occurrence. These are also the representations for which the solution of (9) is readily found and conveniently interpreted.

We therefore assume

$$i_s = \sum_k a_{sk}\delta(t - t_k)$$

$$v_r = \sum_k a_{rk}\delta(t - t_k) \qquad (r = 1, 2, 3). \tag{12}$$

Once e_{g0} is determined and (9) has been solved for a single impulse from each one of the noise sources in the circuit, the stationary-state solution of (3) is obtained with (8) according to the linear superposition principle. We confine ourselves to an approximate solution.

II. The Disturbances of the Fundamental Component

A. The Approximations

When the nonlinear terms in the current voltage characteristic (1) of the active device are small and/or the feedback network in Fig. 1 is highly selective, the second and higher harmonics in (6) are much smaller than the fundamental and can be considered negligible to a first approximation. Accordingly, (6) becomes

$$e_{g0} = A_1 \cos(\omega_1 t + \varphi). \tag{13}$$

A first approximation to the disturbances of the fundamental frequency component of e_g can now be evaluated from (9) by letting $u(t)$ in (8) become

$$u(t) = a_1(t) \cos(\omega_1 t + \varphi) + y_1(t) \sin(\omega_1 t + \varphi) \tag{14}$$

where

$$y_1(t) = -A_1\phi_1(t). \tag{15}$$

Better approximations to x_1 and y_1 can still be determined from the linear perturbation equation (9) if the noise sources are weak. Successively higher harmonics must be included in e_{g0} and $u(t)$, whereby the first significant improvement should not be expected until at least the third harmonic is considered.[3] When any one or all of the extraneous sources i_s, v_1, v_2, and v_3 are strong, higher order terms in $u(t)$ are no longer negligible, even if $u(t)$ is approximated by (14) and e_{g0} by (13); (9) is then no longer adequate, and a nonlinear perturbation analysis becomes necessary. However, this latter case can be of significance in practical oscillators only when the perturbing forces are signals other than thermal or shot noise.

The two approximations of major significance for the following developments are, therefore: First, the noise sources in the oscillator are assumed weak so that the terms of higher order in the disturbances $u(t)$ are negligible. This assumption led in Section I to the linear differential equation (9) for $u(t)$. Second, the harmonic content of the output signal from the undisturbed oscillator is assumed to be very low. This assumption permits approximating e_{g0} by (13) and $u(t)$ by (14). Together the two assumptions imply that all amplitude disturbances of the fundamental component of the signal are represented by $x_1(t)$ in (14), all phase disturbances by $-y_1/A_1$.

To assure that the second approximation is reasonable, it will be assumed later that the active device in the oscillator is only weakly nonlinear and/or that the oscillator is operated at low signal levels. This assumption in turn entails that the quantity γ, to be defined later, is always very small. As shown in the Appendix, other approximations, which go beyond those stated here, are not required.

B. The Unperturbed Signal

Without imposing undue further restrictions, the current voltage characteristic (1) of the active device is assumed to be

$$i_1 = f(e_g) = g_{m0}e_g - \beta e_g^3. \tag{16}$$

For (13) to be a reasonable approximation to (6), $\beta e_g^2 \ll g_{m0}$ is desirable, especially when the transfer impedance of the feedback network at the harmonic frequencies is not extremely low.

When (16) and (13) are inserted into (3) with $F(t) = 0$, the values of A_1 and ω_1 can be determined to

$$A_1^2 = \frac{4}{3\beta}\left(g_{m0} - \frac{C_2R_3}{L_1}\right) \tag{17}$$

$$\omega_1 = \omega_0. \tag{18}$$

The quantity

$$g_m = g_{m0} - \tfrac{3}{4}\beta A_1^2 \tag{19}$$

will be recognized as the effective transconductance [24] of the active device for signals of amplitude A_1. With g_{m0} and β both positive, (19) indicates that g_m decreases monotonically with increasing A_1, as shown in Fig. 2. Because (17) requires [25]

$$\frac{C_2R_3}{L_1} = g_m \left(= \frac{\omega_1 C_2}{Q_N}\right), \tag{20}$$

[3] When dealing with crystal oscillators we also note that crystal units generally have an overtone response close to the third electrical harmonic of the oscillator frequency. Under high-drive conditions, the two frequencies can coincide. The presence of this crystal response in the feedback network then becomes very important and cannot be disregarded.

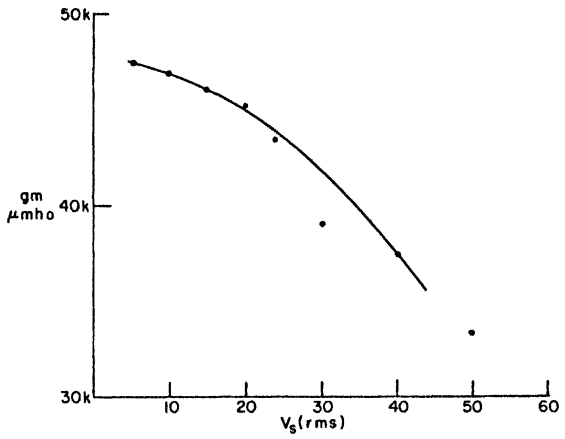


Fig. 2. Typical behavior of the effective transconductance g_m as a function of signal level. Dots are experimental points obtained on a 2N2808 transistor, rated at $i_e = 2$ mA.

the amplitude A_1 of the steady-state oscillations can be adjusted by proper choice of the circuit parameters.

In many cases g_{m0} and β can be determined from data such as plotted in Fig. 2, and the values obtained are the proper ones for use in (20). A more detailed examination of the contributing factors shows, however, that a $\beta^* \leq \beta$ should be used in the following Sections pertaining to $u(t)$, whereby β^* is the coefficient of the third-order term when the instantaneous function $f(e_g)$ in (1) is developed into a Taylor series. The coefficients of the Taylor series depend upon the bias conditions of the active device and the bias conditions change, almost invariably, with amplitude. Therefore, the g_m vs. amplitude curves, when measured under static conditions, include implicitly the effects of the even-order terms in $f(e_g)$, particularly the second-order term. The difference between β and β^* is not very significant except in oscillators with artificial level control such as AGC or lamp bridge oscillators. In AGC oscillators the change in the bias conditions is artificially magnified and utilized [25] to adjust g_{m0} and β such that (20) is satisfied for a very small value of A_1. In a lamp bridge oscillator [3] the value of R_3 or its equivalent, depending upon the actual circuit, varies with signal amplitude, while g_{m0} and β remain nominally unchanged. All amplitude dependent changes in bias conditions, or in R_3, require a finite time to become effective, governed by an RC time constant, while the first effects of a noise impulse occur instantaneously.

Therefore, for the purposes of the present analysis, the function (16) is always understood to represent the instantaneous relationship between i_1 and e_g. Any delayed action will primarily affect the envelope function of e_g and can be accounted for approximately by imposing suitable constraints on the behavior of this function after the instantaneous behavior has been established.

C. The Impulse Response

The general equation (9) for the perturbation $u(t)$ assumes, with (13), (16), and (17), the form

$$\ddot{u} + \omega_1\gamma(1 + 2\cos 2(\omega_1 t + \varphi))\,\dot{u} + \omega_1^2(1 - 4\gamma \sin 2(\omega_1 t + \varphi))u = F(t) \tag{21}$$

whereby the parameter γ is conveniently defined by either of the two equivalent expressions

$$\gamma = \frac{1}{\omega_1 C_2}\,\frac{1}{Q_N}\,\frac{3}{4}\beta A_1^2 \tag{22a}$$

$$\gamma = \frac{1}{Q}\,\frac{g_{m0} - g_m}{g_m Q_N}\,. \tag{22b}$$

It will become apparent later on that $2/\omega_1\gamma$ is the time constant which controls the decay of disturbances in the oscillator. This time constant is appreciably larger than $2/\omega_1\gamma_T$ the time constant of the passive feedback network, when the assumptions made to arrive at (21) are met: since βe_{g0}^2 in (16) should be much smaller than g_{m0}, it follows approximately from (22a) with (19) and (20) that

$$\gamma = \frac{1}{Q}\,\frac{1}{\omega_1 C_2}\,\frac{3}{4}\beta A_1^2 \ll \frac{1}{Q}\,\frac{1}{\omega_1 C_2}\,g_{m0} \approx \frac{1}{Q}\,\frac{g_m}{\omega_1 C_2} = \frac{R_3}{\omega_1 L} \quad \text{(i.e., } \gamma \ll \gamma_T\text{)}. \tag{23}$$

Equation (21) is a Mathieu equation [26]. Because it is linear, the response of the system to white noise from all sources is found by superposition of the individual impulse responses. The method used here to find these impulse responses is a cornerstone of this paper and is detailed in the Appendix.

Following the procedure used there it can be shown that the response of the system to an impulse from any one of the noise generators in Fig. 1 is, aside from differences in magnitude and phase, the same for each generator, with only one exception: the impulse from v_2 (i.e., from the source across the output) generates a system response like that caused by the other sources, and it also appears directly in $u(t)$ as an additive term. The existence and form of this additive term follows from the equations and requires no further assumptions.

For a single impulse at time t_k from each generator, i.e., when $v_1 = a_{1k}\delta(t - t_k)$, $v_2 = a_{2k}\delta(t - t_k)$, $v_3 = a_{3k}\delta(t - t_k)$ and $i_s = a_{sk}\delta(t - t_k)$, one finds as the solution of (21), except for terms with γ or γ^2 as a factor,

$$u_k(t) = \alpha_k e^{-\omega_1\gamma(t - t_k)} \cos(\omega_1 t + \varphi) - A_1\eta_k \sin(\omega_1 t + \varphi) + n_k(t) + v_2 \tag{24}$$

whereby

$$n_k(t) = [M_k \sin\omega_1(t - t_k) + N_k \cos\omega_1(t - t_k)]e^{-(\omega_1\gamma/2)(t - t_k)} \tag{24a}$$

and the constants α_k and η_k are

$$\alpha_k = -M_k \sin(\omega_1 t_k + \varphi) + N_k \cos(\omega_1 t_k + \varphi) \tag{24b}$$

$$\eta_k = -\frac{M_k}{A_1}\cos(\omega_1 t_k + \varphi) - \frac{N_k}{A_1}\sin(\omega_1 t_k + \varphi).$$

The parameters M_k and N_k define the effective strength

of the impulses as weighted by the feedback network, and, for the circuit in Fig. 1, they are

$$M_k = \frac{\omega_1}{2\bar{Q}}\,(a_{1k} + a_{3k} - a_{2k})$$

$$N_k = \frac{\omega_1}{2\bar{Q}Q_N}\left(\frac{a_{sk}}{g_m} + a_{2k}\right). \tag{24c}$$

The significance of the solution (24) of the Mathieu equation (21) can perhaps best be appreciated when it is compared to the corresponding solution of the equation

$$\ddot{u}_p + \omega_1\gamma_T\dot{u}_p + \omega_1^2 u_p = F(t), \tag{25}$$

which describes the output of the passive feedback network in Fig. 1. The solution of (25)

$$u_{p_k} = 2[M_k \sin \omega_1(t - t_k) + N_k \cos \omega_1(t - t_k)]$$
$$e^{-\omega_1\gamma_T(t-t_k)} + v_2$$

which is easily found by standard techniques, can be written as

$$u_{p_k}(t) = \alpha_k e^{-(\omega_1\gamma_T/2)(t-t_k)} \cos(\omega_1 t + \varphi)$$
$$- A_1\eta_k e^{-(\omega_1\gamma_T/2)(t-t_k)} \sin(\omega_1 t + \varphi) + n_{pk} + v_2, \tag{26a}$$

with

$$n_{p_k} = \tfrac{1}{2}(u_{p_k} - v_2). \tag{26b}$$

α_k η_k, M_k and N_k are again defined by (24b) and (24c).

Evidently, the method developed in the Appendix separates (21) into two parts, with each accorded one half the strength of the original excitation. One describes the behavior due to the low-pass equivalent of the system (around zero frequency), the other due to the band-pass equivalent (around 2ω). The equations are solved in the transformed form and the results projected back again into the range around ω_1. In (24) $n_k(t)$ is the response of the band-pass part. It differs from $n_{p_k}(t)$ in (26a) only by the value of the time constant. Hence, with regard to this part of the solution, the disturbed oscillator behaves like a passive tuned circuit with quality factor $1/\gamma$.

The first two terms in (24) and (26a) represent the responses of the respective low-pass parts; here, essential differences are noted, which are brought about by the parametric pumping action apparent in (21). The low-pass part of the solution of (21) is that of a lossless tuned circuit as far as the component in-phase[4] with the pump is concerned; but the out-of-phase component is the damped response of a tuned circuit with quality factor $1/2\gamma$. In the solution of (25) the low-pass part is, of course, identical with the band-pass part.

The practice of evaluating only the low-pass response and doubling the result, correct for the passive circuit, is seen to be inapplicable in general for the system described by (21). Its applicability to other problems—even linear problems, since (21) is linear—must evidently be carefully evaluated in every case.

The representation of the impulse response of (21) thus requires two virtual tuned circuits. To within the approximations used here, both are resonant at ω_1. Each is excited by one half the strength of the original impulse and their outputs are added linearly to the almost full strength impulse from the source across the output terminals of the network.[5] The one resulting from the low-pass equivalent shall be called the "L circuit," that from the band-pass equivalent the "B circuit." The "L circuit" performs independent operations on the two orthogonal components of its response, with the pump signal providing the reference. These can be represented by different effective quality factors and the composite shown in Fig. 3 results. The applicable Q values are indicated in the individual circuits. The circles denote the effective noise voltage generators. The distribution of the half-strength impulse to the generators in the $\|$ and $\perp$ circuits depends on the time t_k at which the impulse occurs in relation to the phase of the pump at $t = t_k$. This follows from (24). Because the pump is out of phase with the undisturbed oscillator signal $A_1 \cos(\omega_1 t + \varphi)$, the outputs of the $\|$ circuits are the phase disturbances (ϕ), those of the $\perp$ circuits the amplitude disturbances (a).

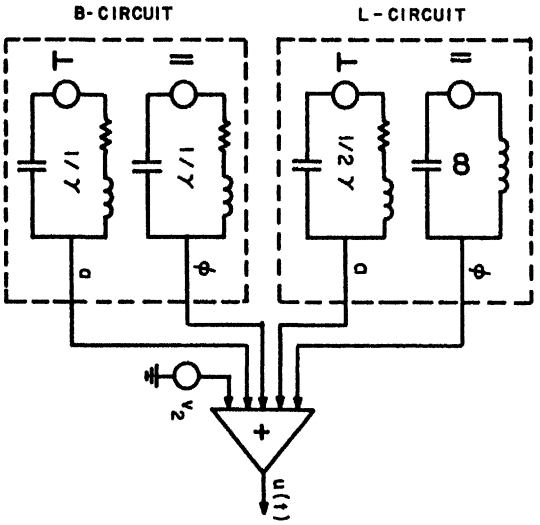


Fig. 3. Virtual circuits used to represent the components of the impulse response of the perturbed oscillator.

The $L\,\|$ circuit is lossless. Energy supplied to it by its share of the impulse strength is stored indefinitely; its output is the undamped sinusoid $-A_1\eta_k \sin(\omega_1 t + \varphi)$ in (24). The output of the $LV\,\perp$ circuit is the first term in (24), and the $B\|$ and $B\perp$ circuits deliver, respectively, the two orthogonal components

$$-A_1\eta_k \exp[-(\omega_1\gamma/2)(t - t_k)] \sin(\omega_1 t + \varphi)$$

[4] The pump phase is taken here to be that of $\sin(\omega_1 t + \varphi)$.

[5] Because γ is assumed very small, only a small portion of the energy in the impulse is accepted by the virtual circuits. Terms with γ or γ^2 as factor were neglected in (24)

and

$$\alpha_k \exp\left[-(\omega_1\gamma/2)(t - t_k)\right] \sin(\omega_1 t + \varphi)$$

of $n_k(t)$.

The structure of the disturbances $u(t)$ of the oscillator signal, due to a single impulse from each one of the noise sources, is thus established. It is most interesting that the properties of the particular oscillator model chosen in Fig. 1 enter (24) only by way of the parameters M_k and N_k. If the appropriate values are used for these parameters, this formula and, hence, the scheme in Fig. 3 evidently applies to any arbitrary oscillator. Those whose frequency depends strongly on fundamental amplitude already in the undisturbed case [14] will require additional terms.

D. The Perturbed Oscillator Signal

The approximation (13) requires the undisturbed oscillators to be represented as a single lossless tuned circuit with energy stored in it. According to (8) and the developments of the preceding Section, the perturbed oscillator is described by the virtual circuits in Fig. 3 if the $L\|$ circuit has been imparted a finite amount of energy initially and if the pump signal is orthogonal to the oscillations represented by this energy (i.e., e_{g0}) at all times.

The undamped oscillations in the $L\|$ circuit caused by the noise impulses at $t=t_k$ are out of phase with e_{g0}; added to it, they change its phase permanently. Since (8) with (10) is an approximation of (7), the response functions of the $L\|$ circuit represent phase disturbances only, and amplitude effects of any order do not occur in this circuit. All amplitude disturbances are represented by the outputs of the $L\perp$ and the $B\perp$ circuit, with the $B\|$ circuit contributing another phase disturbance which, significantly, decays with time.

It is important to realize that the energy supplied to the $L\|$ circuit by the impulses at t_k is necessary to effect the permanent phase changes in e_{g0}; although it is not "used up" thereby and remains stored in the circuit, it cannot be regained if no phase reference is available, and is nevertheless lost irreversibly.

Analytically, the oscillator output follows from (8), (13), and (24) as

$$e_g = \left[A_1 + \alpha_k e^{-\omega_1\gamma(t-t_k)}\right] \cos(\omega_1 t + \eta_k + \varphi) + n_k(t) + v_2 \tag{27}$$

or alternatively, when $n_k(t)$ is decomposed into components parallel and normal to e_g, as

$$e_g = \left[A_1 + x_k(t)\right] \cos(\omega_1 t + \eta_k + \vartheta_k(t) + \varphi) + v_2 \tag{28}$$

whereby

$$x_k(t) = \alpha_k\left(e^{-\omega_1\gamma(t-t_k)} + e^{-(\omega_1\gamma/2)(t-t_k)}\right) \tag{28a}$$

$$\eta_k + \vartheta_k(t) = \eta_k\left(1 + e^{-(\omega_1\gamma/2)(t-t_k)}\right). \tag{28b}$$

These two forms for e_g illustrate different performance aspects of the disturbed oscillator. Equation (28) conforms closely to (7) and will generally be found more convenient for further work. Equation (27) emphasizes the existence of the "B circuit." It shows that the oscillator, in addition to generating e_{g0}, disturbed by the output components of the "L circuit" [i.e., by the first two terms in (24)] also acts simultaneously like a passive tuned circuit of relative bandwidth $\Delta\omega/\omega=\gamma$.

The last term in (27), $v_2=a_{2k}\delta(t-t_k)$ is contributed to the signal by the source across the output. This term is present also as seen from (26), in the output from the passive network alone and is, thus, not peculiar to the oscillator. It still appears as an additive term in (28). Since a δ-function is infinitely large by definition, what might be its orthogonal component in reference to e_{g0} can of course not simply be taken into the argument of the cosine function. Until the oscillator signal has been acted upon by a filter in the output amplifier, v_2 must be carried as an additive term.

The solution (27) or (28) applies to oscillators whose amplitude is limited by the instantaneous nonlinearities in the circuit and whose circuit parameters are independent of time. The effects of a delayed action mechanism for amplitude control (such as changes in g_{m0} and/or the circuit losses) on this solution are complicated in the general case. An approximation is obtained rather simply, however, when the time constant T of the AGC loop and the time τ_a required for the loop to correct an amplitude disturbance obey the conditions:

$$2\pi/\omega_1 \ll T \ll \tau_a \ll 1/\omega_1\gamma.$$

Whereas the effects of the impulses at t_k on e_g are instantaneous, the AGC mechanism does not affect it immediately. The form of e_g is thus established first and can then be assumed modified by operations on the amplitude disturbances only (i.e., the concomitant effects on the signal phase are neglected). Hence, an approximation for the output of a nonlinear AGC oscillator is given by (28) if (28a) is replaced by

$$\alpha_k(t) = 2\alpha_k e^{-(t-t_k)/\tau_a}. \tag{28c}$$

The representation by virtual circuits results from Fig. 3 if the indicated Q values in both the $L\perp$ and the $B\perp$ circuits are replaced by $\omega_1\tau_a$. When τ_a goes beyond the upper limit stated above, the AGC mechanism becomes ineffectual: when T is comparable to or larger than τ_a, it becomes a disturbance. For $T\to 0$ the action becomes instantaneous. In the latter case the output signal is again given by (28) with (28a) and (28b), if the proper value for the nonlinearity is used in γ.[6]

E. The Response to White Noise

The response of the oscillator to white noise is obtained by linear superposition of the effects of all the individual impulses in (12). The undisturbed signal e_{g0}, used in deriving the response to the impulse at time t_k,

[6] Further details of AGC action in linear oscillators will be discussed in Section III-C.

was assumed in the Appendix to include all permanent effects of the impulses prior to t_k. Hence, the phase angle φ in (13) and, consequently, in (27) and (28) has the form

$$\varphi = \sum_{l}^{k-1} \eta_l \tag{29}$$

and need not be carried further. The oscillator signal as disturbed by the action of the white noise source in the circuit becomes

$$e_g = (A_1 + x(t)) \cos(\omega_1 t + \bar{\eta}(t) + \vartheta(t)) + v_2 \tag{30}$$

or alternatively, for reasonable values of γ

$$\begin{aligned} e_g &= A_1 \cos(\omega_1 t + \bar{\eta}(t)) + x(t) \cos(\omega_1 t + \bar{\eta}(t)) \\ &\quad - A_1 \vartheta(t) \sin(\omega_1 t + \bar{\eta}(t)) + v_2 \\ &= E_C + E_A + E_\vartheta + E_N \end{aligned} \tag{31}$$

with

$$x(t) = \sum_k x_k(t), \quad \bar{\eta}(t) = \sum_k \eta_k, \quad \vartheta(t) = \sum_k \vartheta_k(t) \tag{32}$$

whereby $x_k(t)$, η_k, and $\vartheta_k(t)$ are given by (28a) and (28b) together with (24b) and (24c). The summations in (32) extend over all values of k for which $t_k \leq t$. According to the developments of the preceding Section, $E_C = A_1 \cos(\omega_1 t + \bar{\eta}(t))$ in (31) is the output of the $L\|$ circuit in Fig. 3, if the latter has initially energy stored in it. E_C is thus to be considered the carrier signal. Its phase, and hence the phase of the pump signal in (21) execute a random walk due to the cumulative effect of the permanent shifts caused by the individual impulses. $E_A = x(t) \cos(\omega_1 + \bar{\eta}(t))$ is the combined output of the $L\perp$ and the $B\perp$ circuit and represents the amplitude disturbances of the carrier, whereas

$$E_\vartheta = - A_1 \vartheta(t) \sin(\omega_1 t + \bar{\eta}(t))$$

are the phase disturbances due to the $B\|$ circuit. Both E_A and E_ϑ consist of the superposition of a large number of exponentially decaying components. $E_N = v_2$ is the white noise from the source across the oscillator output.

The process indicated by (32) has considerable merit conceptually. The individual impulses in (12) are caused by the elemental phenomena involved in the transport of electrical charges and, though exceedingly numerous, are extremely weak. The addition of the effects of any one of these impulses to $x(t)$, $\bar{\eta}(t)$, and $\vartheta(t)$ in (30) or (31) (and they are to be added one by one) changes these quantities by only infinitesimal amounts. There can never be any reasonable doubt that the linear perturbation equation (21) applies and that the out-of-phase components of $u_k(t)$ can be taken into the argument of the sinusoid. It will frequently be found useful to retrace, at least mentally, the steps going from single impulses in the oscillator to the summation of their effects.

Of the various assumptions which had to be made in deriving the expressions (30) or (31) as an approximation to the solution of (3), the one regarding the absence of harmonic components, particularly of the third harmonic, in the undisturbed signal is considered to be the most serious. In general it must be expected to limit the validity of the results to oscillators operating at low signal levels. Nevertheless, the expressions do give a detailed description of the effects of noise on the oscillator signal, which becomes increasingly more accurate as the harmonic content is reduced.

III. The Statistical Properties of the Signal

A. The Continuous Noise Record

The statistical properties of e_g are more readily evaluated when the sums in (12) are replaced by continuous functions. Without discussion of the essentially philosophical questions involved thereby [27], it will be assumed from here on that the output of the noise sources in Fig. 1 is equally well described by

$$\begin{aligned} i_s &= B_s(t) \\ v_r &= B_r(t) \qquad (r = 1, 2, 3). \end{aligned} \tag{33}$$

The $B_j(t)$ can be visualized as the curves that result when a series of points Δt apart are connected by a continuous line, with each point representing the integral over all noise impulses occurring during the respective time interval Δt.

While the detailed course of any $B_j(t)$ is basically unpredictable, these functions are uniquely defined if it is specified that they are independent stationary Gaussian random variables, and if their mean and variance is given. Forcing functions of the type described have been dealt with extensively in the theory of Brownian Motion and their Gaussian property is well established [28]. It is also known that they are ergodic, have zero mean, and are delta correlated; hence,

$$\begin{aligned} \langle B_j(t) \rangle &= 0 \\ \langle B_j(t_1) B_j(t_2) \rangle &= B_j^2 \delta(t_1 - t_2) \\ \langle B_i(t_1) B_j(t_2) \rangle &= 0 \qquad i \neq j. \end{aligned} \tag{34}$$

The brackets here and in the following denote ensemble averages which, as generally understood, are the arithmetic mean of the quantities within the brackets, each formed with values from a particular noise record, averaged over a large number N of like records, with $N \to \infty$. The fact that the $B_j(t)$ are delta correlated [expressed by the second relation in (34)] means that the value of $B_1(t_1)$, for example, is completely independent of the value of $B_1(t)$ a moment before (i.e., at $t = t_1 - \Delta t$), and it is independent of the values of B_2, B_3, and B_s at the same or any other moment, as expressed by the third relation in (34).

With v_j now given by (33) instead of by (12), the a_{jk} in (24c) are to be replaced by

$$a_{jk} = B_j(t)\, dt \tag{35}$$

and the sums in (32) become integrals:

$$x(t) = \int_0^t \alpha(\xi)\left[e^{-\omega_1\gamma(t-\xi)} + e^{-(\omega_1\gamma/2)(t-\xi)}\right]d\xi$$

$$\bar{\eta}(t) = \int_0^t \eta(\xi)d\xi$$

$$\vartheta(t) = \int_0^t \eta(\xi)e^{-(\omega_1\gamma/2)(t-\xi)}d\xi. \tag{36}$$

$\alpha(\xi)$ and $\eta(\xi)$ follow from (24b) with (24c), (29) and (35) to

$$\begin{aligned}\alpha(\xi) &= -M(\xi)\sin(\omega_1\xi + \bar{\eta}(\xi)) \\ &\quad + N(\xi)\cos(\omega_1\xi + \bar{\eta}(\xi)) \\ \eta(\xi) &= -\frac{M(\xi)}{A_1}\cos(\omega_1\xi + \bar{\eta}(\xi)) \\ &\quad - \frac{N(\xi)}{A_1}\sin(\omega_1\xi + \bar{\eta}(\xi))\end{aligned} \tag{36a}$$

with

$$M(\xi) = \frac{\omega_1}{2\bar{Q}}\left[B_1(\xi) + B_3(\xi) - B_2(\xi)\right]$$

$$N(\xi) = \frac{\omega_1}{2\bar{Q}Q_N}\left[\frac{B_3(\xi)}{g_m} + B_2(\xi)\right]. \tag{36b}$$

The difference in the definitions of $M(\xi)$, $N(\xi)$, $\alpha(\xi)$, and $\eta(\xi)$ when compared to that of M_k, N_k, $\cdots$ in (24) should cause no confusion: $M(\xi)d\xi \leftrightarrow M_k$. When $B_j(\xi) = a_{jk}\delta(\xi - t_k)$ is substituted into the above expressions, they revert to (28a) and (28b), respectively.

The properties of $M(\xi)$ and $N(\xi)$ follow directly from those of the $B_j(\xi)$:

$$\begin{aligned}\langle M(\xi)\rangle &= \langle N(\xi)\rangle = 0 \\ \langle M(\xi_1)M(\xi_2)\rangle &= M^2\delta(\xi_1 - \xi_2) \\ \langle N(\xi_1)N(\xi_2)\rangle &= N^2\delta(\xi_1 - \xi_2)\end{aligned}$$

with

$$M^2 = \left(\frac{\omega_1}{2\bar{Q}}\right)^2(B_1^2 + B_2^2 + B_3^2)$$

$$N^2 = \left(\frac{\omega_1}{2\bar{Q}Q_N}\right)^2\left(\frac{B_3^2}{g_m^2} + B_2^2\right).$$

Of special importance in the later developments is the quantity $K^2 = (M^2 + N^2)/\omega_1^2$

$$K^2 = \frac{1}{4}\frac{1}{\bar{Q}^2}\left[B_1^2 + B_2^2 + B_3^2 + \frac{1}{Q_N^2}\left(B_2^2 + \frac{B^{s2}}{g_m^2}\right)\right] \tag{37}$$

which can be regarded as the weighted noise intensity of all sources in the oscillator combined.

The means of $\alpha(\xi)$ and $\eta(\xi)$ are obviously zero:

$$\langle\alpha(\xi)\rangle = \langle\eta(\xi)\rangle = 0. \tag{38a}$$

If all oscillators in the ensemble are assumed to have already been operating sufficiently long for their relative phase angles to be randomly distributed at time $t = 0$, $\bar{\eta}(\xi)$ with respect to the ensemble is a random number between zero and 2π at any value of ξ in $0 \leq \xi \leq t$. Hence, ensemble averages can be employed to compute the correlation functions of $\alpha(\xi)$ and $\eta(\xi)$. Because

$$\langle\cos(\bar{\eta}(u) + \bar{\eta}(v))\rangle = \langle\sin(\bar{\eta}(u) + \bar{\eta}(v))\rangle = 0,$$

we find

$$\begin{aligned}\langle\alpha(u)\alpha(v)\rangle &= (\omega_1^2K^2/2)\delta(u - v) \\ \langle\eta(u)\eta(v)\rangle &= (\omega_1^2K^2/2)\delta(u - v) \\ \langle\alpha(u)\eta(v)\rangle &= 0.\end{aligned} \tag{38b}$$

The relations (38a) and (38b), together with the knowledge that $\alpha(\xi)$ and $\eta(\xi)$ are Gaussian random processes, completely characterize these functions.

The explicit expressions for the fundamental frequency component of the output from a noise perturbed oscillator have thus been derived. The more general form is given by (30), with (32), when the noise is thought of as a series of random impulses, or with (36), when the noise is represented by continuous random functions. Whereas the two descriptions of white noise (12) and (33) are equivalent, the expressions resulting from the latter are obviously more abstract and do not provide the same conceptual insight into oscillator operation afforded by the impulse response functions. Nevertheless, the expression (30) with (36), (37), (38a), and (38b) is more compact, and is more easily handled when computing the statistical properties of the oscillator signal. Because the decomposition of (30) into (31) is valid in nearly all cases of practical significance, the latter can be used whenever this is desirable.

B. The Autocorrelation Function of the Signal

The power spectral density $G_{ee}(f)$ will be computed from the autocorrelation function $\Gamma_{ee}(\tau)$ of e_g according to the well-known relations [29]

$$G_{ee}(f) = 4\int_0^\infty \Gamma_{ee}(\tau)\cos\omega\tau\, d\tau \tag{39}$$

$$\Gamma_{ee}(\tau) = \langle e_g(t)e_g(t - \tau)\rangle. \tag{40}$$

The most convenient representation of the oscillator signal for the present purpose is given by (31), together with (36). Since all cross correlations between the terms there are zero, the autocorrelation function simply is

$$\Gamma_{ee}(\tau) = \Gamma_{CC}(\tau) + \Gamma_{AA}(\tau) + \Gamma_{\vartheta\vartheta}(\tau) + \Gamma_{NN}(\tau). \tag{41}$$

The individual terms in (41) are found with (38b) by established techniques [30], [5].

$$\Gamma_{CC}(\tau) = \frac{A_1^2}{2}e^{-(\omega_1^2K^2/4A_1^2)\tau}\cos\omega_1\tau$$

$$\Gamma_{AA}(\tau) = \frac{\omega_1K^2}{\gamma}\left(\frac{7}{24}e^{-\omega_1\gamma\tau} + \frac{5}{12}e^{-(\omega_1\gamma/2)\tau}\right)\cos\omega_1\tau$$

$$\Gamma_{\vartheta\vartheta}(\tau) = \frac{\omega_1K^2}{4\gamma}e^{-(\omega_1\gamma/2)\tau}\cos\omega_1\tau$$

$$\Gamma_{NN}(\tau) = B_2^2\delta(\tau). \tag{42}$$

The relations for Γ_{AA} and $\Gamma_{\vartheta\vartheta}$ are approximations which hold whenever

$$\gamma \gg 2\omega_1 K^2/A_1^2. \tag{43}$$

When AGC is used, $\Gamma_{AA}(\tau)$ in (42) is to be replaced by

$$\bar{\Gamma}_{AA}(\tau) = \frac{\omega_1^2 K^2}{2} \tau_a e^{-\tau/\tau_a} \tag{44}$$

where τ_a is, according to (28c), the time required by the AGC mechanism to correct an amplitude disturbance. All other terms in (42) remain the same.

The magnitude of $2\omega_1 K^2/A_1^2$ in (43) is in the order of 10^{-21}, with values of ω_1, K^2, and A_1 representative of a typical precision crystal oscillator, while the actual values of γ encountered in such devices are in the order of 10^{-8}. [See (47) in Section III-D.] The discrepancy is not always this large, but it is doubtful that the instantaneous nonlinearities are ever small enough to violate (43), even in laser oscillators [31]. Hence, in nearly all cases of practical significance, the autocorrelation function of the output signal is properly given by (41) with (42) and, where applicable, (44). It is this case that will be considered when the power spectral density function of the signal is computed in Section III-D.

The average power in the individual output components is obtained from the relations (42) with $\tau = 0$ [32]. That due to the B circuit in Fig. 3 [$\Gamma_{\vartheta\vartheta}(\tau) + \frac{1}{2}\bar{\Gamma}_{AA}(\tau)$ for AGC oscillators] is seen to be substantially smaller than that in the carrier for most practical oscillators. That is, when (43) holds, the tuned circuit aspect contributes only a very small fraction of the total output power of the oscillator. The practical significance of the B circuit is rather limited, therefore, when the disturbances due to thermal and shot noise in an oscillator with time invariant circuit parameters are considered. It is held, however, that its existence is vitally important for a proper explanation of the behavior of the signal phase in oscillators with time variable circuit elements, an area which is left for further research.

C. The Linear Oscillator

To aid the reader in relating earlier concepts to those developed here, the properties of signals from oscillators with linear active elements will now be discussed. Since the virtual circuits in Fig. 3 describe the oscillator output, these will be used thereby.

As γ decreases, the quality factors of the B circuit and of the $L \perp$ circuit increase. The responses to the individual impulses require a longer time to decay, and their cumulative effects become larger. Eventually the phase disturbances due to the $B\|$ circuit become too large to be considered as additive terms in the oscillator output (which in effect means to neglect their influence on the phase of the pump signal), and it becomes necessary to include them immediately into the phase of e_0; that is, the transition from (30) to (31) is no longer applicable. With $\gamma \to 0$, all circuits in Fig. 3 become lossless and the oscillator output no longer shows a performance aspect identifiable with a tuned circuit of finite Q. In this transition the B circuit becomes identical with the L circuit and the two can be fused into one, which shall be called the LB circuit. Now the amplitude disturbances, as well as the phase disturbances, are undamped sinusoids; and the amplitude of the output signal also carries out a random walk, with A_1 as its mean value.

Since E_A in (30) is, with $\gamma = 0$ in (36), $E_A = 2\int_0^t \alpha(\xi)d\xi$ and $\vartheta(t) = \bar{\eta}(t)$, the autocorrelation function of the output signal is now

$$\bar{\Gamma}_{ee} = \bar{\Gamma}_{CC} + \bar{\bar{\Gamma}}_{AA} + \Gamma_{NN}$$

whereby

$$\bar{\Gamma}_{CC} = \frac{A_1^2}{2} e^{-(\omega_1^2 K^2/A_1^2)\tau} \cos \omega_1 \tau$$

$$\bar{\bar{\Gamma}}_{AA} = \omega_1^2 K^2 (t - \tau) e^{-(\omega_1^2 K^2/A_1^2)\tau} \cos \omega_1 \tau. \tag{45}$$

This is the situation obtained in a linear oscillator without an AGC mechanism (i.e., with $T \to \infty$, where T is the time constant of the AGC loop), provided, of course $g_m = g_{m0}$ satisfies the condition (20) for steady-state oscillations precisely.

When g_{m0} is larger than is required by this condition, the oscillations of the undisturbed oscillator build up; when it is smaller they decay, as is well known. The undisturbed oscillator can thus be represented as a tuned circuit, to be called the $\overline{LB}$ circuit, whose quality factor is $1/\gamma_e$, whereby γ_e is defined by (22b) if the difference between the quantities on the rhs and lhs of (20) is used in place of $g_{m0} - g_m$.

Amplitude noise interferes with the orderly build-up or decay of the oscillations, originally set up in the $\overline{LB}$ circuit at $t = 0$, and if its average, obtained by integrating over a time T ($\ll 2/\omega_1\gamma_e$), has the proper sign and magnitude, it can, in fact, hold the amplitude at a constant level. It is the function of a properly designed AGC mechanism to steer the value of γ_e so that this condition does occur continually. In an AGC oscillator, therefore, γ_e is a random variable of zero mean $\langle\gamma_e\rangle = 0$. The average amplitude noise energy delivered by the source during T is $\omega_1^2 K 2T$; the initial gain or loss in oscillation energy during this time is $(A_1^2/2)\omega_1\gamma_e T$; since the two should balance, the root-mean-square value of γ_e is

$$\sqrt{\langle\gamma_e^2\rangle} = 2\omega_1 K^2/A_1^2.$$

It is noted that most of the oscillation energy supplied at $t = 0$ has been replaced in the process by noise energy of the proper phase after $2/\omega_1\gamma_e$ seconds, with the AGC mechanism lending or borrowing energy temporarily to smooth the random variations in the noise energy supply. The remaining amplitude variations are those occurring during the time interval $[t - \tau_a, t]$, whereby τ_a

is the time required by the AGC mechanism to remove a given amplitude disturbance. Their autocorrelation function is given by (44).

The time period τ_a depends inversely on the AGC loop gain (and directly on the reduced quality factor Q of the feedback network). When the time constant T of the loop increases beyond τ_a, the AGC action gets to be out of phase with the amplitude variations, and the slow oscillations in amplitude discussed by Golay [15] and earlier by Edson [33] will occur. To maintain minimum total disturbances requires, therefore, that the AGC loop gain be decreased as T increases. This eventually will act to limit the range of γ_e values controlled by the AGC mechanism and $\langle\gamma_e^2\rangle \to 0$ with $T \to \infty$ results. Hence, with $T \to \infty$ the $\overline{LB}$ circuit becomes lossless and identical to the LB circuit described above. Oscillations of stable amplitude, however, are obtained only for $T \ll \tau_a \ll A_1^2/\omega_1^2 K^2$. The following discussion of the signal phase is restricted to these cases.

Assume again the $\overline{LB}$ circuit supplied at $t=0$ with an amount of energy large when compared to the noise energy. The oscillations represented by this energy provide the reference for decomposing the noise impulse responses of the circuit into amplitude and phase disturbances and, in particular, for the AGC mechanism while removing the amplitude noise. Although the phase disturbances are time dependent with $\exp\left[-(\omega_1\gamma_e/(2)(t-t_k)\right]$, it is the shift in the reference phase accumulating during τ_a seconds that determines their effect on the oscillations. Since the concurrent amplitude variations are eliminated after τ_a seconds, time can start anew, but with the basic oscillations at a different phase. In spite of their time dependence, the phase disturbances in a linear AGC oscillator act as though they were undamped sinusoids, which, added out of phase to a constant amplitude signal, cause its phase to execute a random walk.[7] The output of an AGC oscillator with a linear active element is thus to be represented by the output from a lossless tuned circuit, which has stored in it the energy of the basic oscillations and is excited by a noise source delivering that half of the total noise energy which produces the out-of-phase response components, and added to it the output from another tuned circuit of effective quality factor $\omega_1\tau_a/2$ to which the second half of the noise energy is fed, producing the amplitude disturbances. That is to say, the $\overline{LB}$ circuit is equivalently represented by the LB circuit discussed above when the latter is modified to account for the AGC according to (44). The link from the scheme in Fig. 3 is thus established, even if the reverse path is not obvious. The linear model seems to give no indication that the introduction of instantaneous nonlinearities requires the LB circuit to be split into the L and B circuits shown there.

Approaching the linear oscillator from the standpoint of a passive noise filter, as was variously done in the past, leads in a most direct manner to the autocorrelation function of the oscillator output. Consider the passive feedback circuit in Fig. 1, replace R_3 by $R_3' = (\gamma_f/\gamma_T)R_3$ without affecting the strength of the noise sources, assume half the average output power removed by some appropriate mechanism, and demand that the remainder of the average output power equal $A_1^2/2$. One finds for the only adjustable parameter γ_f the value $\gamma_f = 2\omega_1 K^2/A_1^2$; that is, $\gamma_f = \sqrt{\langle\gamma_e^2\rangle}$. With it the autocorrelation function of the output becomes

$$\bar{\Gamma}_{ee} = \bar{\Gamma}_{CC} + \bar{\Gamma}_{AA} + \Gamma_{NN}$$

with $\bar{\Gamma}_{CC}$ given in (45) and $\bar{\Gamma}_{AA}$ by (44) exactly.

Hidden in these assumptions, however, is the full description of the linear oscillator with AGC as given above. If the AGC mechanism is visualized as an amplifier following the filter, the amplifier provides the reference for decomposing the noise responses into orthogonal components. As this reference is fixed, no random walk of the output phase will occur. However, the amplitude of a noise filter output can go to zero temporarily. To maintain constant amplitude at all times, therefore, requires the AGC mechanism to be part of the feedback loop and the oscillator discussed before results. Also, it is noted that the amplitude noise is not removed by the AGC mechanism as one might be led to believe by the above assumptions. It is converted at the proper rate into the basic oscillation as mentioned before. The phase noise energy remains stored in the random phase walk and is irreversibly lost, since no absolute phase reference is available.

Whereas the passive noise filter is not the proper physical model for the linear AGC oscillator, it does lead in the most simple and direct manner to the correct autocorrelation function and, hence, power spectral density of the output from these devices. However, it is of no apparent assistance in dealing with instantaneous nonlinearities in the oscillator.

D. The Power Spectral Density of the Signal

Proceeding now to the power spectral density (PSD) of the signal, one finds $G_{ee}(f)$ from (39) with (41) and (42) as the sum of the following components:

$$G_{CC}(f) = \frac{4A_1^4/\omega_1^2K^2}{1 + (4A_1^2/\omega_1K^2)^2(1 - \omega/\omega_1)^2}$$

$$G_{AA}(f) = \frac{(7/12)(K^2/\gamma^2)}{1 + (1/\gamma)^2(1 - \omega/\omega_1)^2} + \frac{(5/3)(K^2/\gamma^2)}{1 + (2/\gamma)^2(1 - \omega/\omega_1)^2}$$

$$G_{\vartheta\vartheta}(f) = \frac{K^2/\gamma^2}{1 + (2/\gamma)^2(1 - \omega/\omega_1)^2}$$

$$G_{NN}(f) = 4B_2^2. \qquad (46)$$

[7] This of course does not apply to the phase disturbances due to the $B\|$ circuit when γ is large enough to satisfy (43). The carrier provides the reference phase in that case, and the phase disturbances in response to an impulse excitation relax towards it exponentially.

As a numerical example let

$$\omega_1 L_1 = 100\ \Omega \qquad \omega_1 = 2\pi 5 \times 10^6 \text{ rad/sec} \qquad g_{m0} = 1/20 \text{ mho}$$
$$R_3 = 100\ \Omega \qquad \overline{Q} = 10^6 \qquad \beta = 10 \text{ mho/volt}^2$$
$$\frac{1}{\omega_1 C_2} = 20\ \Omega \qquad A_1^2/2 = 4 \times 10^{-5} \text{ volt}^2 \qquad \gamma = 1.2 \times 10^{-8}$$
$$P_3 = (A_1^2/2)R_3(\omega_1 C_2)^2 = 10^{-5} \text{ watts}, \qquad K^2 = 1.75 \times 10^{-31} \text{ watts/cycle}. \tag{47}$$

The parameters for the active device are typical for a 2N2808 transistor; the values for R_3 and $\overline{Q}$ could apply to a precision crystal unit. The power spectral density of Johnson noise is $4kTR$, that of shot noise in a transistor $2kTg_m$; hence, because of (42) and (39),

$$B_j^2 = kTR_j$$
$$B_s^2 = \tfrac{1}{2}kTg_m. \tag{48}$$

The resistive components R_1 and R_2 have not been considered so far because, other than affecting the values of Q_T and Q_N, their effect upon the results does not extend significantly beyond their action as noise generators. Rather pessimistic assumptions [34] about the effective values of R_1 and R_2 in a transistor oscillator lead to 10 ohms each. With the above values and $kT = 5\times10^{-21}$ W/s, one finds, for the peak densities (in watts per cycle) and normalized half widths at midheight (in cycles/cycle) of the individual terms in (46), respectively

$G_{CC}[1.4\times10^{8}, 3.5\times10^{-20}]$, $G_{AA1}[7.1\times10^{-16}, 3.5\times10^{-8}]$,

$G_{AA2}[2\times10^{-15}, 1.75\times10^{-8}]$, $G_{\vartheta\vartheta}[1.2\times10^{-15}, 1.75\times10^{-8}]$;

and $G_{NN} = 2\times10^{-20}$ watts/cycle.

The power spectral density function cannot be observed directly. The power spectrum $P(\bar{f})$ which is observable can be defined as

$$P(\bar{f}) = \int_0^\infty |H(\bar{f}, f)|^2 G_{ee}(f)\, df \tag{49}$$

whereby $H(\bar{f}, f)$ is the transfer function of the filter being used in the observation of the spectrum and $\bar{f}$ its mean frequency. The output power from the filter, when it is tuned to the carrier frequency, i.e., when $\bar{f} = f_1$, must then be considered the signal power and the definition of the noise-to-signal power ratio at the frequency $\bar{f}$ becomes

$$N/S = P(\bar{f})/P(f_1). \tag{50}$$

It is obvious that the noise-to-signal power ratio depends upon the bandwidth of the filter. This parameter is implicit in $H(\bar{f}, f)$. The form factor of the filter determines whether $P(\bar{f})$ (i.e., the output power from the filter when it is tuned to $\bar{f}$), is indeed due to $G_{ee}(f)$ in the neighborhood of $f = \bar{f}$, or instead is due to the carrier in the tail of the filter characteristic.

When evaluating the ratio (50) analytically, it is frequently possible to approximate the PSD curves given in (46) by their limiting curves, truncated at their peak height, as indicated in Fig. 4. The limiting curves are

$$G_{CC}(f) \to (K^2/4)(1 - \omega/\omega_1)^{-2}$$
$$G_{AA}(f) \to K^2(1 - \omega/\omega_1)^{-2}$$
$$G_{\vartheta\vartheta}(f) \to (K^2/4)(1 - \omega/\omega_1)^{-2}$$
$$G_{NN}(f) \to 4B_2^2. \tag{51}$$

They are, in conformance with a general property of Lorentzian curves, independent of A_1^2 and γ^2, respectively.

Because G_{NN} does not decrease with increasing frequency separation from the carrier, it crosses the other curves at a certain value of $F = f - f_1$, and from there on dominates the PSD of the oscillator output. The crossover point depends, according to (37), on the strength of the noise sources and on the reduced quality factor $\overline{Q}$ of the oscillator. In an LC oscillator $\overline{Q} = 1$, and the crossover occurs too far out from the carrier frequency to be of practical significance. The white noise from the source across the output has no noticeable effect on the PSD of the LC oscillators. When $\overline{Q}$ is large, however, such as in quartz crystal oscillators, the crossover is very close to the carrier (about 3 parts in 10^6 in the above numerical example), and it is the white noise component that determines the N/S ratio of the oscillator in many applications. A very significant improvement in this ratio can often be achieved in this case by using an output filter of sufficiently narrow bandwidth [35]. The relations pertaining to this case are given in the following Section. In all cases, it is obvious that the most effective way to improve the N/S ratio is to operate the oscillator at as high a power level as is possible, subject to limitations by higher order effects.

E. The Effects of the Output Filter

Up to now it was assumed implicitly that the signal at the output of the oscillator stage is available directly for observation or actual use. This however, is rarely the case. Normally the oscillator signal is fed to an isolation amplifier or to some other devices which contain tuned elements. It is the output of these devices that is actually observed, and the modifications of the signal in passage through them must be considered. Of primary interest is the effect of the tuned elements.

Although it is obvious that more effective filters can be used, consider as a simple example the circuit in Fig. 5. When the filter is tuned to ω_1 (i.e., $L_c C_c = 1/\omega_1^2$), and it is assumed that, because of (23),

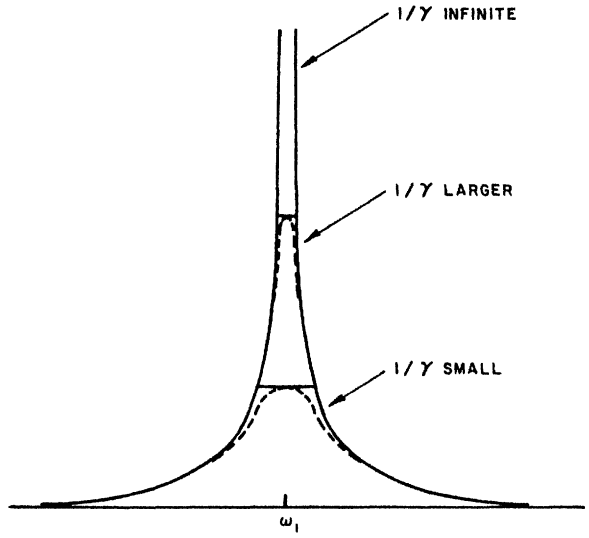


Fig. 4. Illustration of general property of PSD curves. Truncated limiting curves are useful in approximate calculations involving PSD functions.

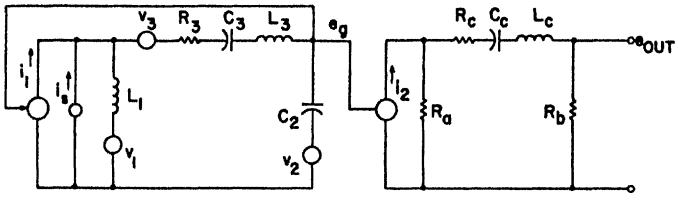


Fig. 5. Model of oscillator with output amplifier. The filter in the output amplifier serves primarily to shape the white noise component in e_g, which stems from v_2. The effect of the filter on the carrier and on the noise response of the oscillator loop is negligible in a high Q oscillator.

$$Q_c \equiv 1/\gamma_c \equiv \frac{\omega_1 L_c}{R_a + R_b + R_c} \ll 1/\gamma, \tag{52}$$

the equivalent to (30) becomes to a very good approximation

$$e_{out} = (A_1 + \bar{x}(t)) \cos(\omega_1 t + \phi(t)) \tag{53}$$

with

$$\bar{x}(t) = x(t) + x_n(t)$$
$$\phi(t) = \bar{\eta}(t) + \vartheta(t) + \vartheta_n(t), \tag{53a}$$

and

$$x_n(t) = \frac{\omega_1}{Q_c}\int_0^t B_2(\xi) e^{-(\omega_1\gamma_c/2)(t-\xi)} \cos[\omega_1\xi + \bar{\eta}(\xi)]\, d\xi$$
$$\vartheta_n(t) = \frac{\omega_1}{Q_c}\int_0^t B_2(\xi) e^{-(\omega_1\gamma_c/2)(t-\xi)} \sin[\omega_1\xi + \bar{\eta}(\xi)]\, d\xi \tag{53b}$$

where $x(t)$, $\bar{\eta}(t)$ and $\vartheta(t)$ are still defined by (36). The equivalent to (21) becomes

$$e_{out} = E_C + E_A + E_\theta + E_N \tag{54}$$

whereby only the definition of E_N is changed from $E_N = B_2(t)$ to

$$E_N = \frac{\omega_1}{Q_c}\int_0^t B_2(\xi) e^{-(\omega_1\gamma_c/2)(t-\xi)} \cos\omega_1(t-\xi)\, d\xi. \tag{54a}$$

The expressions (53) and (54) then are explicit forms of the oscillator signal after it has passed through the tuned output stage.

All relations in Section III-B for e_g apply equally for e_{out}, if only $\Gamma_{NN}(\tau)$ and $G_{NN}(f)$ in (42) and (46), respectively, are replaced by

$$\Gamma_{NN}(\tau) = \frac{\omega_1}{Q_c}\frac{B_2^2}{2}(1 + Q_c/Q_T) e^{-(\omega_1\gamma_c/2)\tau} \cos\omega_1\tau \tag{55a}$$

$$G_{NN}(f) = \frac{2B_2^2(1 + Q_c/Q_T)}{1 + (2Q_c)^2(1 - \omega/\omega_1)^2}. \tag{55b}$$

The limiting curves for $G_{NN}(f)$ are given by

$$G_{NN}(f) \to (B_2^2/2Q_c^2)(1 + Q_c/Q_T)(1 - \omega/\omega_1)^{-2}. \tag{56}$$

F. The Properties of the Signal Phase

When evaluating the properties of the signal phase it is important that the effects of the output filter be included in the considerations. This eliminates the difficulties with the additive white noise component in (30) which, according to the comments in Section II-D, cannot be decomposed into orthogonal components until acted upon by the filter. Hence, the signal representation (53) must be used. Because of $\bar{\eta}(t)$ in (53a), the phase of the signal is only a weakly stationary random process; however, it does have stationary independent increments, and the autocorrelation function of $\phi(t)$ can be readily computed to

$$\langle\phi(t)\phi(t-\tau)\rangle$$
$$= (\omega_1^2K^2/2A_1^2)\left(t - \tau + \frac{2}{\omega_1\gamma} + \frac{3}{\omega_1\gamma} e^{-(\omega_1\gamma/2)\tau}\right)$$
$$+ (\omega_1B_2^2/2A_1^2Q_c)[Q_c/2Q_T + (1 + Q_c/Q_T)e^{-(\omega_1\gamma_c/2)\tau}]. \tag{57}$$

With some care in dealing with the first term, (57) can be used in (39) to obtain the spectral density of the signal phase. Except for some δ functions at the origin, it is

$$G_{\phi\phi}(f) = (2K^2/A_1^2)(\omega_1/\omega)^2$$
$$+ (12K^2/A_1^2\gamma^2)[1 + (2/\gamma)^2(\omega/\omega_1)^2]^{-1}$$
$$+ (4B_2^2/A_1^2)(1 + Q_c/Q_T)[1 + (2/\gamma_c)^2(\omega/\omega_1)^2]^{-1}. \tag{58}$$

A sketch of the three terms in (58) is shown in Fig. 6. The respective limiting curves for the second and third term are

$$(3K^2/A_1^2)(\omega_1/\omega)^2;\quad (B_2^2/A_1^2Q_c^2)(1 + Q_c/Q_T)(\omega_1/\omega)^2 \tag{58a}$$

and, as before, it may be adequate for practical purposes to approximate these terms by their truncated limiting curves as indicated by the dotted lines in Fig. 6.

The expression (58) for the spectral density of the signal phase can be compared to the expressions for G_{ee} or for $G_{CC} + G_{\vartheta\vartheta} + \frac{1}{2}G_{NN}$ [using the relations (46) with (55b)]. The latter is the PSD of the output signal with all amplitude disturbances removed. It will be noted that, in spite of the similarities of the corresponding terms, there is no simple relationship between these expressions. In the most general case the seven parameters entering them must be evaluated from independent

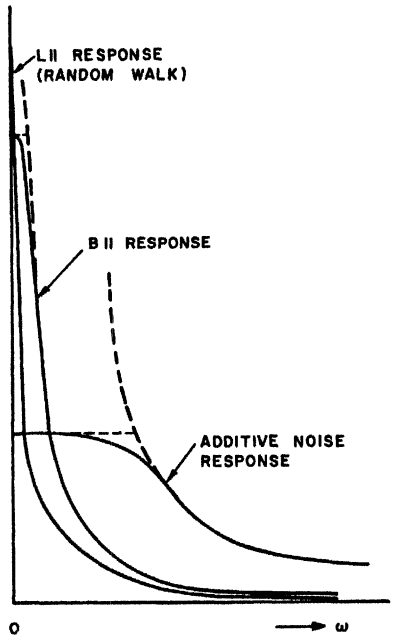


Fig. 6. Sketch of spectral density of oscillator phase disturbance components.

measurements to specify the different aspects of the signal. These parameters are A_1, ω_1, Q_c, Q_T, γ, K^2, and B_2^2. Only when the additive noise component due to the source across the output of the oscillator stage predominates over all other terms, can the properties of the power spectrum of the output signal be related directly to those of the spectrum of the phase and vice versa. In all cases it is important to recognize the distinction between these two spectra.

IV. The Short Term Frequency Stability

The frequency of a periodic or quasi-periodic signal can be obtained by integrating the phase θ over a time τ and dividing the result by τ:

$$\omega_\tau = \frac{1}{\tau}\int_{t-\tau}^{t} d\theta = \frac{\theta(t) - \theta(t-\tau)}{\tau}. \tag{59}$$

In general, ω_τ is a function of the time t at which the integration is carried out. When $\theta(t)$ can be assumed to have the form

$$\theta(t) = \omega_1 t + \phi(t), \tag{60}$$

whereby ω_1 is a constant and $\phi(t)$ a random variable of zero mean, the short term frequency stability of the signal for integration times of τ seconds can be defined as

$$S(\tau) = \sqrt{\left\langle 4\,\frac{(\omega_\tau - \omega_1)^2}{\omega_1^2}\right\rangle}. \tag{61a}$$

With (59) and (60) this becomes

$$S(\tau) = \frac{2}{\omega_1\tau}\left\{\langle(\phi(t))^2\rangle + \langle(\phi(t-\tau))^2\rangle - 2\langle\phi(t)\phi(t-\tau)\rangle\right\}^{1/2}. \tag{61b}$$

When $\phi(t)$ is known, ω_τ and $S(\tau)$ are readily evaluated; however, the reverse process, that is the determination of the characteristics of the signal phase from the properties of ω_τ, can be carried out only in a very restricted sense.

The form of the output signal of the oscillator appropriate for use in this Section is given by (53), since the effects of the output filter must be included in the analysis. The autocorrelation function of $\phi(t)$ required in (61b) was derived in Section III-F. With (57) one finds the short term frequency stability of the signal as

$$S(\tau) = \left\{(2K^2/A_1^2\tau^2)[\tau + (6/\omega_1\gamma)(1 - e^{-(\omega_1\gamma/2)\tau})] + (4B_2^2/\omega_1 Q_c A_1^2\tau^2)(1 + Q_c/Q_T)(1 - e^{-(\omega_1\gamma c/2)\tau})\right\}^{1/2}. \tag{62}$$

K^2 in (62) is given by (37), with (4a) and (48), γ by (22); and $Q_c = 1/\gamma_c$ is the quality factor of the output filter in in Fig. 5. The conditions (23) and (52) are assumed satisfied.

The first term in (62), $(2K^2/A_1^2\tau)^{1/2}$, corresponds qualitatively to the familiar result obtained when only the random walk in the signal phase is considered, while the second term in the first bracket, with

$$(1 - \exp[-(\omega_1\gamma/2)\tau])$$

as factor, is due to the output from the $B\|$ circuit in Fig. 3. With $\gamma \to 0$ the first component of (62) becomes $(8K^2/A_1^2\tau)^{1/2}$, in full agreement with the corresponding expressions found in the literature.

The second component of (62) is due to the additive white noise component in e_g and depends strongly on the properties of the output filter.

With the numerical values for the various parameters as chosen in (47) and a $Q_c = 10$, (62) becomes

$$S^2(\tau) = \frac{44 \times 10^{-28}}{\tau^2}\left[\tau + \frac{3}{0.2}(1 - e^{-0.2\tau})\right] + \frac{8 \times 10^{-24}}{\tau^2}(1 - e^{-1.6x10^6\tau}). \tag{63}$$

The two components of $S(\tau)$ as given by (63) are drawn in Fig. 7. The overall character of the first component shows a $\tau^{-1/2}$ dependence on integration time. The transition of the curve from a higher to a lower level occurs at a value of τ, which depends upon γ and, hence, upon the nonlinearity of the active device. A large value of γ is desirable to push this transition to shorter integration times. The overall level of this component of $S(\tau)$, in an oscillator whose noise generators have a given strength, depends primarily upon the reduced quality factor $\bar{Q}$ and upon the signal amplitude.

The contribution to $S(\tau)$ of the second component in (63) varies with τ^{-1} for integration times larger than the time constant of the output filter. In the above example the latter is less than one microsecond; and at $\tau = 1$ sec this contribution is still greater than that of the first component by more than one order of magnitude. It can be reduced, as apparent from (62) and as indicated

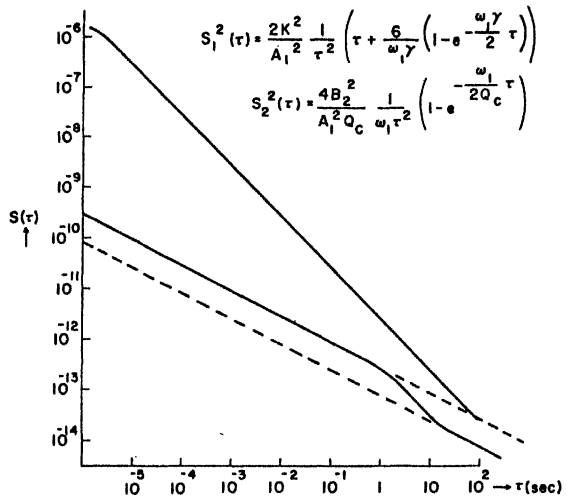


Fig. 7. The short term frequency stability of a crystal oscillator as a function of integration time is the sum of the two solid curves. The lower curve, given by $(S_1^2)^{1/2}$, is due to effects inside the oscillator loop. It includes the effects of the random walk phenomenon in the carrier phase and of the noise response of the oscillator. The upper curve, given by $(S_2^2)^{1/2}$, is due to the additive white noise from the source in the oscillator's output (v_2 in Fig. 5). Its significance can be reduced by the use of high Q filters in the output amplifier. The numerical values of (47) apply.

previously, by increasing the quality factor of the output filter. When Q_c is larger than 10^4, the short term stability of the signal frequency at $\tau = 1$ sec is in this example essentially given by the first term in (63).

An important conclusion can be drawn from the fact that the last term in (62) is for $Q_c \ll Q_T$ independent from the properties of the oscillator feedback network and the active device characteristics. If this term is much larger than the first, the short term frequency stability of the oscillator can be expressed approximately by

$$S(\tau) \doteq \frac{1}{\tau}\sqrt{4kTR_3/\omega_1 Q_c A_1^2} = (2/\omega_1\tau)\sqrt{V_N^2/V_S^2} \tag{64a}$$

whereby

$$V_N^2 = \omega_1 B_2^2/2Q_c \tag{64b}$$

is the mean square noise voltage measured at the amplifier output when the crystal unit is disconnected, and V_S^2 the mean square output voltage when the oscillator is operating normally. The advantage of (64) for practical design work is obvious.

The relationship between the power spectrum of the signal or the spectrum of the phase and the short term frequency stability is rather simple only when $S(\tau)$ in (62) is dominated by the last component, i.e., when the signal disturbances of interest are due to the additive white noise from the oscillator output. In the general case such a relationship is again best established, as in Section III-F, via the basic parameter A_1, ω_1, Q_c, Q_T, γ, K^2, and B_2^2.

V. The Signal After Frequency Multiplication

For many applications it is necessary to multiply the frequency of the oscillator signal, and the properties of the signal at the output of the multiplier have to be known. The short term frequency stability of the multiplied signal is still properly represented by the expressions derived in Section IV, provided the multiplier itself does not contribute significant noise components. The spectral characteristics of the signal, however, are modified by the multiplication process, and the nature of these modifications under idealized conditions will now be determined.

An ideal n times multiplier removes all amplitude disturbances and multiplies the phase of the signal by n, hence, with the signal as given by (53) as an input, the multiplier output is

$$e_m = A_1 \cos n[\omega_1 t + \phi(t)]. \tag{65}$$

It can be approximated for small $\vartheta(t)$ by

$$\begin{aligned} e_m &= A_1 \cos n[\omega_1 t + \tilde{\eta}(t)] \\ &\quad - A_1[\vartheta(t) + \vartheta_n(t)] \sin n(\omega_1 t + \tilde{\eta}(t)) \\ &= E_{mC} + E_{m\theta}, \end{aligned} \tag{66}$$

and the power spectral density of e_m can be computed following the procedure used in Section III-D. One finds

$$\begin{aligned} G_{e_m e_m} = {} & \frac{4A_1^4/n^2\omega_1^2K^2}{1 + (4A_1^2/n\omega_1K^2)^2(1 - \omega/n\omega_1)^2} \\ & + \frac{n^2K^2/\gamma^2}{1 + (2n/\gamma)^2(1 - \omega/n\omega_1)^2} \\ & + \frac{n^2B_2^2(1 + Q_c/Q_T)}{1 + (2n/\gamma_c)^2(1 - \omega/n\omega_1)^2}. \end{aligned} \tag{67}$$

The character of the change in the PSD brought about in the multiplication process is best appreciated by first considering the limiting curves of $G_{e_m e_m}$ and of

$$G_{\tilde{e}_{out}\tilde{e}_{out}} = G_{CC} + G_{\theta\theta} + \tfrac{1}{2}G_{NN},$$

respectively, which are, term by term

$$\begin{aligned} G_{e_m e_m} \to {} & (K^2/4)(1 - \omega/n\omega_1)^{-2} + (K^2/4)(1 - \omega/n\omega_1)^{-2} \\ & + (B_2^2/4Q_c^2)(1 + Q_c/Q_T)(1 - \omega/n\omega_1)^{-2}, \end{aligned} \tag{68a}$$

$$\begin{aligned} G_{\tilde{e}_{out}\tilde{e}_{out}} \to {} & (K^2/4)(1 - \omega/\omega_1)^{-2} + (K^2/4)(1 - \omega/\omega_1)^{-2} \\ & + (B_2^2/4Q_c^2)(1 + Q_c/Q_T)(1 - \omega/\omega_1)^{-2}. \end{aligned} \tag{68b}$$

Whereas these curves are identical on a relative frequency scale, the peak densities of the noise terms increase with n^2 while the peak density of the carrier decreases with $1/n^2$ due to the multiplication. Nevertheless, the PSD of the carrier of the multiplied signal is equal to that of a signal derived from an oscillator operating at $\omega_1' = n\omega_1$ in every detail, provided A_1 and K^2 are equal.

For many practical applications it is of interest to know which conditions are potentially capable of providing the highest signal-to-noise ratio at a given frequency. The above equations are quite useful to answer a number of questions in this area. A few examples will be discussed below by comparing the N/S ratio of

two signals, one derived with the aid of an ideal n times multiplier from Oscillator I, which operates at ω_1, the other derived from Oscillator II which operates directly at $\omega_1' = n\omega_1$. Amplitude disturbances are assumed removed from both signals. Each oscillator is considered under two conditions I-1, I-2, and II-1, II-2, respectively, whereby 1 and 2 refer to

1) no additive noise $K^2 \gg B_2^2/2Q_c^2$
2) white additive noise $K^2 \ll B_2^2/2Q_c^2$.

Condition 1) occurs either when the oscillator is free running, or when an output filter of sufficiently high quality factor is used to suppress the additive noise to insignificant levels. Condition 2) can occur only in oscillators with large reduced quality factor $\overline{Q}$, and only over that range of frequencies away from the carrier where G_{NN} in (55b) is effectively constant (i.e., where the additive noise is essentially white).

The N/S is computed according to (50), whereby under Condition 1) only the limiting curves are considered. $H(F, f)$ is assumed to represent a rectangular window whose pass band extends from f_a to f_b and, with $F=0$, is wide enough for P_S to be given by $A_1^2/2$ to a good approximation. F is the mean frequency separation of the filter pass band from the carrier at $nf_1 = f_1'$, defined by

$$F^2 = (f_1' - f_a)(f_1' - f_b).$$

One finds for the N/S ratios of the high frequency signals under Condition 1)

$$\text{I-1} \qquad N_I/S_I = (K_I^2/A_{II}^2)(n^2 f_1^2/F^2)(f_b - f_a)$$

$$\text{II-1} \qquad N_{II}/S_{II} = (K_{II}^2/A_{III}^2)(f_1'^2/F^2)(f_b - f_a) \tag{69}$$

and under Condition 2)

$$\text{I-2} \qquad N_I/S_I = (2n^2 B_{2I}^2/A_{II}^2)(f_b - f_a)$$

$$\text{II-2} \qquad N_{II}/S_{II} = (2B_{2II}^2/A_{III}^2)(f_b - f_a) \tag{70}$$

whereby the parameters of the two oscillators are identified by the subscripts I and II, respectively.

Evaluating the ratios $(N_I/S_I)/(N_{II}/S_{II})$ permits estimation of the relative merits of deriving a desired signal from a lower frequency oscillator by multiplication or directly from an oscillator operating at the required frequency.

In many cases, the noise intensities in the two oscillators can be assumed to have about the same magnitude. The ratio K_I^2/K_{II}^2 is then determined primarily by the reduced quality factors $\overline{Q}_I$ and $\overline{Q}_{II}$ of the oscillator, i.e.,

$$k_I^2/k_{II}^2 \approx \overline{Q}_{II}^2/\overline{Q})$$

while the signal amplitudes are determined by the dissipated power, i.e.,

$$A_{II}^2/A_{III}^2 \approx P_I^2/P_{II}^2.$$

If, furthermore, B_2^2 is assumed to be about an order of magnitude smaller than the sum of all sources entering K^2, the relative merits can be estimated from the following table.

TABLE I

	I 1	I 2
II 1	$\overline{Q}_{II}^2 P_{II}/\overline{Q}_I^2 P_I$	$\frac{n^2F^2}{f_1'^2}\frac{P_{II}\overline{Q}_{II}^2}{P_I}$
II 2	$\frac{f_1'^2}{F^2}\frac{P_{II}}{P_I\overline{Q}_I^2}$	$n^2 P_{II}/P_I$

When the ratios shown in Table I are smaller than one, the Oscillator I-multiplier combination is to be preferred; when they are larger, the high frequency oscillator has the lower N/S ratio under the conditions stated. The upper left ratio shows that, if the additive noise is suppressed, the oscillator with the larger $\overline{Q}^2 P_S$ product will be superior, regardless of the multiplication ratio. If additive noise is present in both oscillators, however, the lower right ratio shows that regardless of its reduced quality factor, the lower frequency oscillator is almost invariably an inferior source. The upper right ratio indicates that a free running microwave oscillator can give a lower N/S ratio than a quartz crystal oscillator multiplier combination, if no high Q output filter is used in the crystal oscillator and n is large. The lower left ratio might be of interest when the signal derived from a quartz crystal oscillator under optimum conditions is compared to the output of a maser, where the additive noise is not readily suppressed.

Whereas these relations are valid only under a number of restrictive conditions, they do illustrate the very serious degradation of the signal properties caused by the additive noise from the source across the output of the oscillator stage. If this noise is eliminated, however, only the properties of the feedback network and the final frequency determine the N/S ratio of the signal, but not the multiplication factor.

Conclusions

A major result of this paper is the derivation of the explicit form of the output signal from a noise perturbed oscillator. It permits the complete evaluation of any desired performance aspect, such as the spectral properties of the signal and of the signal phase, as well as the short term frequency stability, and it clarifies the interrelations between them. Although a specific oscillator model was chosen to formulate the basic differential equation, the results apply to all oscillators whose frequency does not depend on signal amplitude in first order. They are expressed in terms of a few general parameters all of which, except one, can be determined from the impulse response function and the noise output of the passive feedback network alone. Only the parameter γ, which depends upon the magnitude of the instantaneous nonlinearity in the circuit, requires a knowledge of the active device characteristics.

The expression for the output signal was derived by means of a novel first-order perturbation technique which, in contrast to earlier work, does not require

smoothing of the instantaneous nonlinearity. Consideration of the instantaneous nonlinearities reveals a parametric pumping action to occur within the oscillator loop which, however, affects only the low-pass equivalent of the disturbed system, but not the band-pass equivalent. Two dissimilar performance aspects are thus created and the oscillator appears to act simultaneously like a linear AGC oscillator and like a high Q passive tuned circuit. The investigation of the significance of this effect in oscillators with time variable circuit parameters and in frequency pulling, pushing, and entrainment phenomena is believed to be a very fruitful area for further research.

Appendix

Solution of the Perturbation Equation

A. The General Solution

The basic equation (3) for the noise perturbed oscillator contains no approximations or restrictive conditions, other than that the circuit elements are independent of time. The perturbation equation (9)

$$\ddot{u} + \frac{L_1}{LC_2}\left(\frac{R_3C_2}{L_1} - \frac{df(e_{g0})}{de_{g0}}\right)\dot{u} + \left(\omega_0^2 - \frac{L_1}{LC_2}\frac{d}{dt}\frac{df(e_{g0})}{de_{g0}}\right)u = F(t) \tag{71}$$

results from it when $u(t)$ in (8) is assumed small, so that higher order terms are negligible. A small $u(t)$ can, regardless of its time dependence, cause only small variations in e_g. Hence, (9) already contains the basic premise of the method of the slowly varying amplitude and phase [26], and this method cannot be applied again to impose additional restrictions on $u(t)$.

In the general case e_{g0}, the steady-state signal in the unperturbed oscillator can be represented as

$$e_{g0} = \sum_n [a_n \cos n(\omega_1 t + \varphi) + b_n \sin n(\omega_1 t + \varphi)], \tag{72}$$

and with it we find

$$\frac{df(e_{g0})}{de_{g0}} = \sum_n [\alpha_n \cos n(\omega_1 t + \varphi) + \beta_n \sin n(\omega_1 t + \varphi)]$$

where the α_n and β_n are (nonlinear) aggregates of the constants a_n and b_n in (72) and the coefficients of the various powers of e_{g0} when $f(e_{g0})$ in (1) is developed into a Taylor series. When $u(t)$ is then assumed to have the form

$$u(t) = \sum_n [x_n(t) \cos n(\omega_1 t + \varphi) + y_n(t) \sin n(\omega_1 t + \varphi)], \tag{73}$$

(71) can be written as

$$\sum_{n=1}^{\infty} [\ddot{x}_n + g_n(t)] \cos n(\omega_1 t + \varphi) + \sum_{n=1}^{\infty} [\ddot{y}_n + h_n(t)] \sin n(\omega_1 t + \varphi) = F(t) \tag{74}$$

whereby the

$$g_n(t) = g_n(\dot{x}_1, x_1, \dot{x}_2, x_2, \cdots; \dot{y}_1, y_1, \dot{y}_2, y_2, \cdots)$$
$$h_n(t) = h_n(\dot{x}_1, x_1, \dot{x}_2, x_2, \cdots; \dot{y}_1, y_1, \dot{y}_2, y_2, \cdots) \tag{75}$$

are linear functions in the $\dot{x}_q$, x_r, $\dot{y}_s$, y_t with constant coefficients.

It is noted that only identities have been used in transforming (9) into (74). We can furthermore introduce the identity

$$F(t) = F(t)\cdot 1 = F_c(t) \cos(\omega_1 t + \varphi) + F_s \sin(\omega_1 t + \varphi)$$
$$F_c(t) = F(t) \cos(\omega_1 t + \varphi)$$
$$F_s(t) = F(t) \sin(\omega_1 t + \varphi) \tag{76}$$

into (74) and we find (71) to be identical with

$$[\ddot{x}_1 + g_1(t) - F_c] \cos(\omega_1 t + \varphi) + [\ddot{y}_1 + h_1(t) - F_s] \sin(\omega_1 t + \varphi) + \sum_{n=2}^{\infty} [\ddot{x}_n + g_n(t)] \cos n(\omega_1 t + \varphi) + \sum_{n=2}^{\infty} [\ddot{y}_n + h_n(t)] \sin n(\omega_1 t + \varphi) = 0. \tag{77}$$

Equation (77) is of the form

$$\sum_{n=0}^{\infty} G_n(t)\psi_n(t) = 0 \qquad [G_0(t) = 0] \tag{78}$$

whereby $\{\psi_n(t)\}$ is a complete set of orthogonal functions in the interval $[t, t+2\pi/\omega_1]$. Solving (71) is thus reduced to determining the functions $G_n(t)$ such that (78) is satisfied for all t.

The trivial solution of (78) is

$$G_n(t) = 0 \qquad n = 1, 2, 3, \cdots. \tag{79}$$

Hence, if a set of functions $\{{}^{(0)}x_n\}$ and $\{{}^{(0)}y_n\}$ is found which satisfies the simultaneous equations

$$\begin{aligned} \ddot{x}_1 + g_1(t) &= F_c(t) & \ddot{y}_1 + h_1(t) &= F_s(t) \\ \ddot{x}_2 + g_2(t) &= 0 & \ddot{y}_2 + h_2(t) &= 0 \\ \ddot{x}_3 + g_3(t) &= 0 & \ddot{y}_3 + h_3(t) &= 0 \\ &\vdots & &\vdots \end{aligned} \tag{80}$$

it will, inserted into (73), define a $u(t)$ which satisfies (9). The condition (79) is thus clearly sufficient to define a solution $u(t)$ of (9).

Aside from (79), however, there exists a countably infinite number of other conditions which satisfy (78). These are obtained by demanding that the terms in the sum (78) cancel in pairs, triplets, quadruplets and so on. For example, (78) is satisfied if

$$G_1(t)\psi_1(t) + G_2(t)\psi_2(t) = 0, \; G_n(t) = 0, \; n = 3, 4, \cdots. \tag{81}$$

This condition is met, of course, when (79) holds; but it is also met when $G_1(t) = \lambda_1\psi_2(t)$, $G_2(t) = -\lambda_1\psi_1(t)$ with λ_1

an arbitrary constant. This leads to

$$\begin{aligned} \ddot{x}_1 + g_1(t) &= F_c + \lambda_1 \sin(\omega_1 t + \varphi), \\ \ddot{y}_1 + h_1(t) &= F_s - \lambda_1 \cos(\omega_1 t + \varphi) \\ \ddot{x}_2 + g_2(t) &= 0 \\ \ddot{y}_2 + h_2(t) &= 0. \\ &\vdots \end{aligned} \tag{82}$$

The set of functions which satisfies these equations can be written as $\{{}^{(0)}x_n + \lambda_1{}^{(1)}x_n\}\{{}^{(0)}y_n + \lambda_1{}^{(1)}y_n\}$ whereby the ${}^{(0)}x_n$, ${}^{(0)}y_n$ are the solutions of (80) and the $\lambda_1{}^{(1)}x_n$, $\lambda_1{}^{(1)}y_n$ are the particular integrals for the sinusoidal forcing terms in (82). Proceeding in this manner, it can be shown that any possible choice of the $G_n(t)$, other than (79), which satisfies (78), leads to solutions $\{x_n\}$ and $\{y_n\}$, which consist of the integrals of (80) plus the particular integrals in response to some (in every case well defined) combination of the ψ_n as forcing terms, multiplied by an arbitrary constant λ_j. In all cases, it is necessary, therefore, that the equations (80) be satisfied.

Considering all possible choices of the $G_n(t)$, the solution of (77) can be written as

$$\begin{aligned} \{x_n\} &= \left\{{}^{(0)}x_n + \sum_{j=1}^{\infty} \lambda_j {}^{(j)}x_n\right\} \\ \{y_n\} &= \left\{{}^{(0)}y_n + \sum_{j=1}^{\infty} \lambda_j {}^{(j)}y_n\right\}. \end{aligned} \tag{83}$$

When the development leading to (82) is traced backwards and (82) is written in the form (74), it is found that the left-hand side of (74) remains unaltered, while the right-hand side becomes

$$F(t) + \lambda_1[\sin(\omega_1 t + \varphi)\cos(\omega_1 t + \varphi) - \cos(\omega_1 t + \varphi)\sin(\omega_1 t + \varphi)] = F(t) + \lambda_1 0.$$

This result is true for all possible choices of the $G_n(t)$ other than (79). Hence, when the ${}^{(j)}x_n$, ${}^{(j)}y_n$ are inserted into (73), the resulting ${}^{(j)}u(t)$ is but the complementary function of (9), which is, of course, contained already in ${}^{(0)}u(t)$. All solutions of (77) which are not based on (79) are redundant. Therefore, all λ_j in (83) can be set equal to zero without affecting the solutions in any way.

It follows that the equations (80) are necessary and sufficient to determine, with (73), a solution $u(t)$ of (71), and that this solution is the general solution of that equation.

B. An Approximate Solution

The equation (77) was obtained by neglecting higher order terms in $u(t)$. Its solution will thus only give the first-order approximation to the disturbances of the oscillator signal. When the perturbing forces are thermal and shot noise, knowledge of this first-order approximation is adequate, beyond any reasonable doubt, for all practical purposes. However, to determine it fully requires solving the infinite set of simultaneous linear equations (80), a task which clearly calls for an iteration procedure.

The zero-order solution is found by setting all $\dot{x}_n$, $\dot{x}_n$; $\dot{y}_n$, y_n for $n \geq 2$ equal to zero; that is, by assuming $u(t)$ to be given by (14):

$$u(t) = x_1 \cos(\omega_1 t + \varphi) + y_1 \sin(\omega_1 t + \varphi). \tag{84}$$

Then x_1 and y_1 are to be determined from the first two equations in (80); that is, from the equations which result when the coefficients of the fundamental frequency terms in (77) are required to be zero for all t. These equations are of the form

$$\begin{aligned} \ddot{x}_1 + c_1\dot{x}_1 + c_2 x_1 + c_3\dot{y}_1 + c_4 y_1 &= F_c \\ \ddot{y}_1 + d_1\dot{x}_1 + d_2 x_2 + d_3\dot{y}_1 + d_4 y_1 &= F_s \end{aligned} \tag{85}$$

whereby the c_k and d_k are constants which depend upon the parameters of the nonlinear current voltage characteristic of the active device and upon the harmonic amplitudes of e_{g0}. The latter, of course, must be determined from the homogeneous part of (3).

The evaluation of these constants becomes very simple if it is justified to approximate e_{g0} by its fundamental component, as in (13). It then is also adequate to approximate the Taylor series $f(e_{g0})$ by (16), since the second power in e_{g0} contributes no fundamental component to (77), and the fourth and higher order terms are usually small enough to be neglected in nearly harmonic oscillators.

With these approximations the equations (85) become

$$\begin{aligned} \ddot{x}_1 + 2\omega_1\dot{y}_1 + 2\omega_1\gamma\dot{x}_1 &= F_c \\ \ddot{y}_1 - 2\omega_1\dot{x}_1 - 2\omega_1{}^2\gamma x_1 &= F_s \end{aligned} \tag{86}$$

which are most easily verified when (84) and (76) are inserted into the Mathieu equation (21). And, to determine $u(t)$ to within these approximations, it remains to solve (86) when $F(t)$ is given by (5) with (12).

The impulse response functions of the system (86) can be found in the following manner. With

$$i_s = 0, \quad v_1 = v_2 = 0, \quad v_3 = a_{3k}\delta(t - t_k) \tag{87}$$

F_c and F_s in (76) become

$$\begin{aligned} F_c &= (\omega_1{}^2/\overline{Q})a_{3k}\delta(t - t_k)\cos(\omega_1 t + \varphi) \\ F_s &= (\omega_1{}^2/\overline{Q})a_{3k}\delta(t - t_k)\sin(\omega_1 t + \varphi). \end{aligned} \tag{88}$$

Since the initial conditions for x_1 and y_1 and their first derivatives are zero when, at $t = t_k$ the oscillator is assumed to be in that state which would result had all previous impulses been in the infinite past, the Laplace transform of (86) is

$$\begin{aligned} (p^2 + 2\omega_1\gamma p)X_{13} + 2\omega_1 p Y_{13} &= L_c e^{-pt_k} \\ -2\omega_1(p + \omega_1\gamma)X_{13} - p^2 Y_{13} &= L_s e^{-pt_k} \end{aligned} \tag{89a}$$

whereby

$$\begin{aligned} L_c &= (\omega_1{}^2/\overline{Q})a_{3k}\cos(\omega_1 t_k + \varphi) \\ L_s &= (\omega_1{}^2/\overline{Q})a_{3k}\sin(\omega_1 t_k + \varphi). \end{aligned} \tag{89b}$$

The characteristic equation of (89a)

$$p[p^2(p + 2\omega_1\gamma) + 4\omega_1{}^2(p + \omega_1\gamma)] = 0 \tag{90a}$$

can be approximated by

$$p(p + \omega_1\gamma)(p^2 + \omega_1\gamma p + 4\omega_1^2) = 0 \qquad (90b)$$

when $\gamma^2 \ll 4$.

The inverse transforms of the solutions X_{13} and Y_{13} of (89a) are then, except for terms with γ^2 as a factor

$$\begin{aligned} x_{13k} &= -\frac{L_s}{2\omega_1} e^{-\omega_1\gamma(t-t_k)} \\ &+ \frac{1}{2\omega_1} e^{-(\omega_1\gamma/2)(t-t_k)}[L_c \sin 2\omega_1(t-t_k) + L_s \cos 2\omega_1(t-t_k)] \\ &+ \frac{\gamma}{2\omega_1} e^{-(\omega_1\gamma/2)(t-t_k)}\left[L_c \cos 2\omega_1(t-t_k) - \frac{L_s}{2}\sin 2\omega_1(t-t_k)\right]. \\ Y_{13k} &= \frac{L_c}{2\omega_1} \\ &+ \frac{1}{2\omega_1} e^{-\omega_1\gamma/2(t-t_k)}[L_s \sin 2\omega_1(t-t_k) \\ &- L_c \cos 2\omega_1(t-t_k)] \\ &- \frac{\gamma}{4\omega_1} e^{-\omega_1\gamma/2(t-t_k)}\left[L_s \cos 2\omega_1(t-t_k) \right. \\ &\left. + \frac{L_s}{2}\sin 2\omega_1(t-t_k)\right] \end{aligned} \qquad (91)$$

It is noted that approximating e_{g0} by (13) and $f(e_{g0})$ by (16) is a matter of convenience only and is not necessary to solve (85). Likewise, approximating (90a) by (90b), that is, assuming $\gamma^2 \ll 4$, results in a simplification of the expressions (91), but (86) can, of course, be solved for arbitrary values of γ. Since the parameter γ determines the effective bandwidth of the oscillator, it is realized that the narrow band approximation need not be introduced at any point to solve (71). It is also realized, however, that (13) will be a rather poor approximation of e_{g0} if γ is not very small. Hence, even assuming $\gamma \ll 1$ is thoroughly consistent with the approximation (13), and the terms with γ as a factor in (91) need not be considered further.

When (91), without the γ terms, is inserted into (84), one finds as the approximate solution of (21) for a single impulse from v_3,

$$\begin{aligned} u(t) &= (\omega_1/2\overline{Q})a_{3k}\cos(\omega_1 t_k + \varphi)\sin(\omega_1 t + \varphi) \\ &- (\omega_1/2\overline{Q})a_{3k}\sin(\omega_1 t_k + \varphi)e^{-\omega_1\gamma(t-t_k)}\cos(\omega_1 t + \varphi) \\ &+ (\omega_1/2\overline{Q})a_{3k}e^{-(\omega_1\gamma/2)(t-t_k)}\sin\omega_1(t-t_k). \end{aligned} \qquad (92)$$

When the expressions corresponding to (92) are derived for a single impulse from each of the other noise generators contributing in $F(t)$ and the results are added, the relations (24) shown in the text are obtained. The Laplace transforms of derivatives of delta functions required thereby are readily found according to established techniques [36].

C. Comparison with Prior Art Results

Previous investigations of the effects of noise in oscillators have led to results the essence of which is equivalent to the statement that the perturbation $u(t)$ is given by

$$\begin{aligned} u(t) &= (\omega_1/\overline{Q})a_{3k}\cos(\omega_1 t_k + \varphi)\sin(\omega_1 t + \varphi) \\ &- (\omega_1/\overline{Q})a_{3k}\sin(\omega_1 t_k + \varphi)e^{-\omega_1\gamma(t-t_k)}\cos(\omega_1 t + \varphi) \end{aligned} \qquad (93)$$

instead of by (92). The most conspicuous difference is the absence in (93) of the third term in (92), which is recognized as the response function of a tuned circuit of relative bandwidth γ, excited by one half of the strength of the original impulse. The impulse response (93) corresponding to the prior art agrees in form with the first two terms in (92), but is twice as strong.

Considering the discussions of the oscillator perturbations in the Sections II-D through III-C of the text it is noted that (92) will assume the form (93) when the instantaneous nonlinearity in the oscillator is zero (i.e., $\gamma = 0$), and when the amplitude disturbances are corrected by a delayed action mechanism (i.e., AGC). In that case, however, $\omega_1\gamma$ in (93) must be interpreted to mean the reciprocal of τ_a, the time required for the AGC mechanism to correct a given amplitude disturbance. (When the time constant T of the AGC loop is short, τ_a is almost directly proportional to the AGC loop gain and is nearly independent of T, a fact that follows from Golay's equations [15].)

The techniques used in the literature for the analytic treatment of noise perturbed nonlinear oscillators differ substantially [8]–[13]. But upon closer study one can find that, indeed, all include, at various stages of the development, the assumption that there are no instantaneous nonlinearities in the circuit. The present results, thus, are not in conflict with those derived earlier; they extend them to the more realistic case where the presence of instantaneous nonlinearities in the circuit is admitted.

The most commonly used procedure employs the averaging principle [14], which involves smoothing of the instantaneous nonlinearity [12]. Its application in effect reduces a differential equation of the type

$$\ddot{e} + \omega_0\left(\frac{1}{Q} - \alpha + \beta_1 e^2\right)\dot{e} + \omega_0^2 e = \omega_0^2 E(t), \qquad (94)$$

which describes rather generally the voltage e in a van der Pol [20] oscillator under the influence of a disturbing noise voltage $E(t)$ to [7]

$$\ddot{e} + \omega_0\left(\frac{1}{Q} - \alpha + \beta A^2\right)\dot{e} + \omega_0^2 e = \omega_0^2 E(t). \qquad (95)$$

A, thereby, is the amplitude of the oscillator voltage which is apparently always approximated at the very beginning by $e = AV\cos\omega t$; (95), however, no longer describes the behavior of a truly nonlinear oscillator. It refers to an oscillator with a linear active element whose effective transconductance $(\alpha - \beta A^2)$ is controlled by an amplitude sensitive mechanism with a small, but finite, time constant. The smoothing process cannot physically be accomplished within the nonlinear active element; rather, the signal from a linear oscillator must be rectified externally and the information obtained used to steer the gain of the device.

The perturbation equation corresponding to (95) with $e = e_0 + u$ is not (21), but

$$\ddot{u} + \omega_0^2 u = \omega_0^2 E(t), \qquad (96)$$

plus an additional equation, similar to that used by

Golay [15], which describes the behavior of the amplitude disturbance under the influence of the delayed action mechanism.

The Mathieu equation (21) and, hence, ultimately the solution (92) is obtained when the parametric pumping action [13] that occurs in the oscillator because the $\beta e^2 = \frac{1}{2}\beta A^2(1+\cos 2\omega t)$ term in (94) is properly represented in the perturbation equation. Conversely, when the perturbations are assumed to be given by (93), the existence of an instantaneous nonlinearity in the oscillator is ignored.

The difference between the present results and those of the earlier work can also be illustrated by tracing the development of the equations (86) back to the form (74), that is, to

$$(\ddot{x}_1 + 2\omega_1\dot{y}_1 + 2\omega_1\gamma\dot{x}_1)\cos(\omega_1 t + \varphi) + (\ddot{y}_1 - 2\omega_1\dot{x}_1 - 2\omega_1^2\gamma x_1)\sin(\omega_1 t + \varphi) = F(t) \quad (97)$$

which, with

$$u_a = x_1(t)\cos(\omega_1 t + \varphi)$$
$$u_\vartheta = y_1(t)\sin(\omega_1 t + \varphi) \quad (98)$$

can be rewritten as

$$\ddot{u}_a + 2\omega_1\gamma\dot{u}_a + \omega_1^2 u_a + \ddot{u}_\vartheta + \omega_1^2 u_\vartheta = F(t). \quad (99)$$

The earlier results (93) are obtained when (99) is assumed to be identical [11] with

$$\ddot{u}_a + 2\omega_1\gamma\dot{u}_a + \omega_1^2 u_a = F_c\cos(\omega_1 t + \varphi)$$
$$\ddot{u}_\vartheta + \omega_1^2 u_\vartheta = F_s\sin(\omega_1 t + \varphi). \quad (100)$$

While any pair of solutions u_a and u_ϑ of (100) satisfies (99), the converse is true only when γ in (99) is zero, that is, when the oscillator contains no instantaneous nonlinearity. Replacing (99) by (100) thus involves smoothing of the instantaneous nonlinearity. The first equation in (100) is then equivalent to the additional equation mentioned above in connection with (96). It represents the action of the AGC mechanism, that is, $\omega_1\gamma$ in (100) is to be considered the reciprocal of τ_a as explained before.

According to the discussions in Section III-C, the solutions (92) and (93) are stochastically equivalent when $\gamma \leq 2\omega_1 K^2/A_1^2$. Whereas the response of the L circuit satisfies the equations (100) for any value of γ, the fact that the solution of (99) representing the response of the B circuit does not satisfy (100) becomes increasingly more important as γ exceeds $2\omega_1 K^2/A_1^2$, and the solutions (92) must be used to describe the oscillator behavior.

References

[1] A. Spälti, "Der Einfluss des thermischen Widerstandsrauschens and des Schrotteffektes auf die Störmodulation von Oscillatoren," *Bulletin des Schweizerischen Electrotechnischen Vereins*, vol. 39, pp. 419–427, June 1948.

[2] R. M. Lerner, "The effects of noise on the frequency stability of a linear oscillator," *Proc. Nat. Electr. Conf.*, vol. 7, pp. 275–280, 1951.

[3] W. A. Edson, *Vacuum Tube Oscillators*. New York: Wiley, 1953, p. 367 ff.

[4] J. P. Gordon, H. J. Zeiger, and C. H. Townes, "The maser, a new type of microwave amplifier, frequency standard and spectrometer," *Phys. Rev.*, vol. 99, pp. 1264–1274, 1955.

[5] W. A. Edson, "Noise in oscillators," *Proc. IRE*, vol. 48, pp. 1454–1466, August 1960.

[6] A. Blaquiére and P. Grivet, "Comments on normalized equations of the regenerative oscillator—noise, phase-locking and pulling," *Proc. IEEE (Correspondence)*, vol. 53, pp. 518–519, May 1965.

[7] P. Grivet and A. Blaquiére, "Masers and classical oscillators," *Proc. of the Symposium on Optical Masers*, p. 72, 1963.

[8] I. L. Berstein, "On fluctuations in the neighborhood of periodic motion of an auto-oscillating system," *Doklady Akad. Nauk.*, vol. 20, p. 11, 1938.

[9] P. I. Kuznetsov, R. L. Stratonovich, and V. I. Tikhonov, "The effects of electrical fluctuations on a vacuum tube oscillator," *J. Exp. Theor. Phys. USSR*, vol. 1, pp. 510–519, 1955; in Russian, vol. 28, pp. 509–523, 1955.

[10] S. M. Rytov, "Fluctuations in oscillating systems of the Thomson type," *Soviet Physics*, vol. 2, pp. 217–235, 1956; in Russian, *J. Exp. Theor. Phys. USSR*, vol. 29, pp. 304–333, 1955.

[11] A. Blaquiére, "Effet du Bruit de Fond sur la Fréquence des Auto-Oscillateurs à Lampes," *Ann. Radio Elect.*, vol. 8, pp. 36–80, January 1953. "Spectre de Puissance d'un Oscillateur Non-Linéaire Perturbé par le Bruit," *Ann. Radio Elect.*, vol. 8, pp. 153–179, August 1953.

[12] P. Grivet and A. Blaquiére, "Nonlinear effects of noise in electronic clocks," *Proc. IEEE*, vol. 51, pp. 1606–1614, November 1963.

[13] J. A. Mullen, "Background noise in nonlinear oscillators," *Proc. IRE*, vol. 48, pp. 1467–1473, August 1960.

[14] N. Kryloff and N. Bogoliuboff, *Introduction to Nonlinear Mechanics*. Princeton, N. J.: Princeton University Press, 1949, pp. 12, 14, 27, 28.

[15] M. J. E. Golay, "Normalized equations of the regenerative oscillator—noise, phase-locking, and pulling," *Proc. IEEE*, vol. 52, pp. 1311–1330, November 1964.

[16] E. J. Baghdady, R. N. Lincoln, and B. D. Nelin, "Short term frequency stability: characterization, theory, and measurement," *Proc. IEEE*, vol. 53, pp. 704–722, July 1965.

[17] E. T. Jaynes and F. W. Cummings, "Comparison of quantum and semiclassical radiation theories with application to the beam maser," *Proc. IEEE*, vol. 51, pp. 89–109, January 1963.

[18] W. E. Lamb, Jr., "Theory of an optical maser," *Phys. Rev.*, vol. 134, pp. A1429–A1450, 1966.

[19] Y. P. Yu, "Application of network theorems to transient analysis," *J. Franklin Inst.*, vol. 248, pp. 381–398, 1949.

[20] B. Van der Pol, "The non-linear theory of electrical oscillations," *Proc. IRE*, vol. 22, pp. 1051–1086, September 1936.

[21] P. Le Corbeiller, "Two-stroke oscillators," *IRE Trans. on Circuit Theory*, vol. CT-7, pp. 387–398, December 1960.

[22] W. J. Cunningham, *Introduction to Non-Linear Analysis*. New York: McGraw-Hill, 1958.

[23] J. Groszkovski, *Frequency of Self-Oscillations*. New York: Macmillan, 1964, pp. 186, 293.

[24] N. W. McLachlan, *Ordinary Non-Linear Differential Equations*. Oxford: Clarendon Press, 1950, p. 110.

[25] H. J. Reich, *Functional Circuits and Oscillators*. Princeton, N. J.: van Nostrand, 1961, pp. 327, 353.

[26] N. W. McLachlan, *Theory and Application of Mathieu Functions*. Oxford: Clarendon Press, 1947.

[27] S. Chandrasekhar, "Stochastic problems in physics and astronomy," in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed. New York: Dover, 1954, p. 22.

[28] Ming Chen Wang and G. E. Uhlenbeck, "Theory of the Brownian motion," in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed. New York: Dover, 1956, p. 122.

[29] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–332, 1944, and vol. 24, pp. 46–156, 1945, eq. (2.1-5).

[30] J. S. Bendat, *Principles and Applications of Random Noise Theory*. New York: Wiley, 1958, p. 79.

[31] N. Bloembergen, "Wave propagation in nonlinear electromagnetic media," *Proc. IEEE*, vol. 51, pp. 124–131, January 1963.

[32] W. R. Bennett, "Methods of solving noise problems," *Proc. IRE*, vol. 44, pp. 609–638, May 1956.

[33] W. A. Edson, "Intermittent behavior in oscillators," *Bell Sys. Tech. J.*, vol. 24, pp. 1–22, 1945.

[34] E. R. Chenette, "Low noise transistor amplifiers," *Solid State Design*, vol. 5, pp. 27–30, February 1964.

[35] The importance of auxiliary filtering in improving the spectrum was already pointed out by Edson in [5] above.

[36] W. B. Davenport, Jr. and W. L. Root, "*An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill, 1958.

A Simple Model of Feedback Oscillator Noise Spectrum

Introduction

This letter contains brief thoughts on the following points.

1) The relationships among four commonly used spectral descriptions of oscillator short-term stability or noise behavior.
2) A heuristic derivation, presented without formal proof, of the expected spectrum of a feedback oscillator in terms of known oscillator parameters.
3) Some experimental results which illustrate the validity of the simple model.
4) Comments on the effect of nonlinearity, specific spectral requirements for several applications, choice of resonator frequency and active element, and expected spectrum characteristics of several oscillator types.

Spectral Models of Phase Variations

Consider a stable oscillator whose measurable output can be expressed as

$$v(t) = A \cos [\omega_0 t + \phi(t)].$$

It is common to treat $\phi(t)$ as a zero-mean stationary random process describing deviations of the phase from the ideal. The frequency domain information about phase or frequency variations is contained in the "power" spectral density $S_\phi(\omega_m)$ of the phase $\phi(t)$ or, alternatively, in the "power" spectral density $S_{\dot\phi}(\omega_m)$ of the frequency $\dot\phi$. By analogy to modulation theory, we use ω_m to mean the modulation, video, baseband, or offset frequency associated with the noise-like variations in $\phi(t)$. The units of $S_\phi(\omega_m)$ are radians^2/cps bandwidth or dB relative to 1 radians^2/cps BW; $S_{\dot\phi}(\omega_m)$ is expressed in (radians/sec)^2 per c/s BW [1] [2]. The two are related by $S_{\dot\phi}(\omega_m) = \omega_m^2 S_\phi(\omega_m)$.

$S_{\dot\phi}(\omega_m)$ can also be expressed in terms of the equivalent rms frequency deviation Δf_{rms} in a given video bandwidth. Further, subject to the limitations that $\overline{\phi^2} \ll 1$ (small total modulation indes) and that AM$\ll$FM components, the normalized RF power spectrum $G(\omega - \omega_0)$ is identical to the two-sided spectrum of the phase $S_\phi(\omega_m)$; i.e., RF sidebands relative to the carrier are down by $S_\phi(\omega_m)$ expressed in decibels relative to 1 radian^2/BW.

Relation to Oscillator Internal Noise

A basic requirement on an oscillator noise model is that it show clearly the relationship of the spectrum of the phase $S_\phi(\omega_m)$ to the known or expected noise and signal levels and resonator characteristics of the oscillator. A simple picture can be constructed using a model of a linear feedback oscillator. Minor corrections to the results are necessary to account for nonlinear effects which must be present in a physical oscillator. Assume a single resonator feedback network of fractional bandwidth $2B/\omega_0 = 1/Q$, where Q is the operating, or loaded, quality factor. For small phase deviations at video rates which fall within the feedback half-bandwidth $\omega_0/2Q$, a phase error at the oscillator input due to noise or parameter variations results in a frequency error determined by the phase-frequency relationship of the feedback network, $\Delta\theta = 2Q\dot\phi/\omega_0$. Thus, for modulation rates *less* than the half-bandwidth of the feedback loop, the spectrum of the *frequency* $S_{\dot\phi}(\omega)$ is identical (with a scale factor) to the spectrum of the uncertainty of the oscillator input phase due to noise and parameter variations. This uncertainty will be denoted $\Delta\theta(t)$, and its two-sided power spectral density $S_{\Delta\theta}(\omega_m)$.

For modulation rates *large* compared to the feedback bandwidth, a series feedback loop is out of the circuit. At these modulation rates, the power spectral density of the output *phase* $S_\phi(\omega)$ is identical to the spectrum of the oscillator input phase uncertainty $S_{\Delta\theta}(\omega_m)$.

For a physical oscillator the spectrum $S_{\Delta\theta}(\omega_m)$ of the input phase uncertainty $\Delta\theta(t)$ is expected to have two principal components. One component is due to phase uncertainties resulting from additive white noise at frequencies around the oscillator frequency, as well as noise at other frequencies mixed into the pass band of interest by nonlinearities. The second component is due to parameter variations at video frequencies which affect the phase (such as variations in the phase shift of a transistor due to carrier density fluctuations in the base resistance). The additive noise component of $S_{\Delta\theta}(\omega_m)$ is identical to the spectral density of the noise voltage squared relative to the mean square signal voltage. For white additive noise, this component is flat with frequency. For a feedback oscillator with an effective noise figure F, the two-sided $S_{\Delta\theta}(\omega) = 2\text{FKT}/P_S$; P_S is the signal level at oscillator active element input.

The video spectrum of parameter variations is found typically to have a power spectral density varying inversely with frequency (a $1/\omega_m$ or $1/f$ spectrum). The total power spectral density of oscillator input phase errors is of the form $S_{\Delta\theta}(\omega_m) = \alpha/\omega_m + \beta$, where α is a constant determined by the level of $1/f$ variations and β is $= 2\text{FKT}/P_S$ for two-sided spectra.

To find $S_\phi(\omega_m)$ or $S_{\dot\phi}(\omega_m)$, we use the fact that

$$\text{for} \quad \omega_m < \frac{\omega_0}{2Q} \qquad S_\phi(\omega) = \left[\frac{\omega_0}{2Q}\right]^2 S_{\Delta\theta}(\omega_m)$$

$$\omega_m > \frac{\omega_0}{2Q} \qquad S_{\dot\phi}(\omega) = S_{\Delta\theta}(\omega_m).$$

A suitable composite expression is

$$S_\phi(\omega_m) = S_{\Delta\theta}\left[1 + \left(\frac{\omega_0}{2Q\omega_m}\right)^2\right].$$

This yields an asymptotic model for $S_\phi(\omega)$ shown on log-log scales in Fig. 1.

The model can be summarized as follows.

$S_\phi(\omega_m)$ decreases with ω_m
- at 9 dB/octave up to the point where $1/f$ effects no longer predominate.
- at 6 dB/octave from that point up to the feedback loop half-bandwidth.
- at 0 dB/octave above that frequency up to a limit imposed by subsequent filtering.

$S_{\dot\phi}(\omega)$ decreases at 3 dB/octave up to the first breakpoint, is flat with frequency up to the feedback baseband bandwidth, and increases at 6 dB octave above that point.

The case where $1/f$ effects predominate only for frequencies small compared with the feedback loop bandwidth is shown here as an

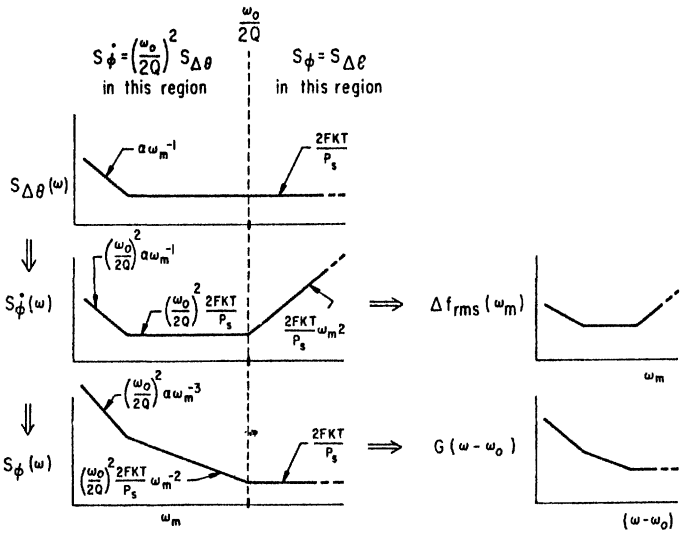


Fig. 1. Derivation of Oscillator Spectra. The logical sequence leading from oscillator parameters to spectrum characteristics is presented here. The power spectra of output phase or frequency are derived from the spectrum of input phase uncertainties and from the oscillator feedback bandwidth. The calculable constants of the oscillator are FKT, P_s, and $\omega_0/2Q$; the $1/f$ constant α is not accurately predictable but can be inferred from data. The amplitude spectrum of frequency deviation and the RF spectrum can be derived as shown, subject to limitations discussed in the text.

Manuscript received December 10, 1965; revised December 29, 1965

Reprinted from *Proc. of IEEE*, pp. 329-330, February 1966.

example. For a high-Q oscillator, $1/f$ effects in $S_{\Delta\theta}$ can predominate out to a modulation rate exceeding $\omega_0/2Q$; in this case there is no 6 dB per octave region in $S_\phi(\omega)$. A similar spectrum results where large additive noise in following amplifier stages or measuring equipment obscures the oscillator internal noise, except at very low modulation frequencies.

Note that there is a portion of the curve $S_\phi(\omega_m)$ which is proportional to $1/\omega_m^2$, leading to a $1/\omega_m$ or $1/f$ variation for rms phase deviation. This is often confused with the true $1/f$ effects associated with parameter variations leading to the $1/f$ portion of the curve for $S_{\Delta\theta}(\omega_m)$ and $S_\phi(\omega_m)$. These two are not the same thing; "$1/f$" refers to a power spectral density rather than an amplitude spectrum.

In practice, the measurable $S_\phi(\omega_m)$ is always modified by subsequent bandlimiting filtering and by additive noise contributed by following amplifiers. It is conceivable that, for a two-terminal oscillator, the filtering action of the resonator eliminates the additive phase noise component for $\omega_m > (\omega_0/2Q)$.

Experimental Verification

Measurements were taken on a stable microwave signal source[1] designed to have a spectral purity limited only by the oscillator, which was a 100 M/s crystal oscillator. This unit employs two large-jump step recovery diode multipliers with amplification between them. The data are presented in Fig. 2 in comparison with a model derived from the following constants:

Feedback bandwidth	= 16 kc/s
P_S	= −4 dBm
F	= 9 dB
KT	= −174 dBm in 1 c/s BW
Multiplication ratio	= 100 = 40 dB
$N^2 2FKT/P_S$	= +40+3+9−174+4 = −118 dB.

This leads to an asymptotic value for $S_\phi(\omega_m)$ of −118 dB relative to 1 radian²/BW in 1 c/s bandwidth, i.e., a carrier-to-sideband ratio of 118 dB. The "$1/f$" region (9 dB/octave) constant α is estimated for best data fit.

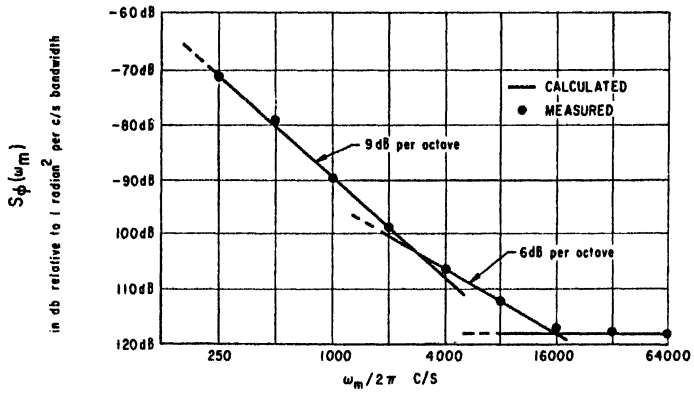


Fig. 2. $S_\phi(\omega_m)$ for Stable Microwave Signal Source. The data presented here is the average of two independent measurements which were in excellent agreement. These measurements were made at X Band on the multiplied output of a 100 Mc/s voltage controlled crystal oscillator having a 16 kc/s feedback half-bandwidth. Since this bandwidth can be reduced by a considerable factor without exceeding the present state of the art, the data is not intended to represent ultimate attainable levels, but rather serves as an illustrative example. The $1/f$ constant is chosen for best data fit. Slopes and other calculated parameters are derived from known oscillator characteristics.

Nonlinear Effects

The data was based on an estimated transistor noise figure of 9 dB. This was taken high to account for nonlinear mixing of noise at third harmonic and higher frequencies which is mixed into the pass band of interest by second harmonic periodic parameter variations caused by the nonlinearity. The excellent fit of the data implies that this degradation of effective noise figure may well be an adequate description of the effect of nonlinearity.

Video Frequency Range of Interest

A number of applications which have been dealt with in this issue of the Proceedings may be summarized in terms of the video frequency range of interest. Space systems and Doppler radar applications are of particular interest to the author. For these two, interest lies in the range of a few c/s up to 100 kc/s. Space applications typically concentrate on the range where, for a crystal oscillator, $S_\phi(\omega_m)$ is proportional to $S_{\Delta\theta}(\omega_m)$ [3], while Doppler radar applications place additional emphasis on the region above the oscillator feedback loop bandwidth [4]. Both applications typically require microwave systems which employ multiplication from the oscillator frequency.

Choice of Oscillator Frequency for Crystal Oscillator-Multiplier

It is of interest to inspect the effect of oscillator frequency upon the output spectrum of an oscillator-multiplier system having a fixed output frequency. Two assumptions which aid the calculation are a) constant oscillator input signal-to-noise ratio, and b) resonator Q varying inversely with the oscillator frequency ω_0. Under these assumptions a comparison of two oscillator frequencies yields the following results.

1) For $\omega_m < (\omega_0/2Q)$, of the lower frequency oscillator, the multiplied output $S_\phi(\omega_m)$ is identical for either choice.
2) For $\omega_m \gg (\omega_0/2Q)$, the output $S_\phi(\omega_m)$ varies as the square of the multiplication ratio (i.e., inversely as the square of the oscillator frequency).

This can be verified by a simple graphical construction.

Choice of Active Element in a Transistor Oscillator

It is apparent that $1/f$ variations and nonlinearity can have significant deleterious effects on the attainable low levels of $S_\phi(\omega_m)$. In the light of suggestions by O. Mueller that microthermal effects [6] contribute to $1/f$ noise in transistors, it is suggested that AGC oscillators using large area transistors having high power capabilities may provide simultaneous improvements in $1/f$ level and in nonlinear effects.

Spectrum Characteristics of Microwave Solid State Sources

The spectrum model given here allows simple prediction of spectrum shape and level for microwave sources of the types discussed by Johnson et al [5]. Comparison with their data shows good agreement —their measurements for crystal oscillator units extend to $\omega_m \gg (\omega_0/2Q)$, while microwave oscillators are characterized by Q factors such that, for the measurements cited, $\omega_m < (\omega_0/2Q)$.

Acknowledgment

The author is pleased to acknowledge helpful discussions with members of IEEE Subcommittee 14.7, of which this letter may be considered a brief summary. Prepublication access to all of the papers contained in this issue is also freely acknowledged. The influence of L. S. Cutler, J. A. Mullen, and W. L. O. Smith has been of special value in the preparation of this correspondence.

D. B. Leeson
Applied Technology, Inc.
Palo Alto, Calif.

References

[1] L. S. Cutler and C. L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," *this issue*, page 136.
[2] E. J. Baghdady, R. N. Lincoln, and B. D. Nelin, "Short-term frequency stability: characterization, theory, and measurement" *Proc. IEEE*, vol. 53, pp. 704–722, July 1965.
[3] R. L. Sydnor, J. J. Caldwell, and B. E. Rose, "Frequency stability requirements for space communications and tracking systems," *this issue*, page 231.
[4] D. B. Leeson and G. F. Johnson, "Short-term stability for a Doppler radar: requirements, measurements, and techniques," *this issue*, page 244.
[5] S. L. Johnson, B. H. Smith, and D. A. Calder, "Noise spectrum characteristics of low-noise microwave tubes and solid-state devices," *this issue*, page 258.
[6] O. Mueller, "Thermal feedback and $1/f$-flicker noise in semiconductor devices," *1965 Internat'l Solid State Circuits Conference, Digest*, p. 68 ff.

9.5 Gc/s Solid State Local Oscillator PN 31-007191, manufactured by Applied Technology, Inc., Palo Alto, Calif. Measurements are average of values measured by the author and D. J. Healey, III, Westinghouse Corp., Baltimore, Md., using Spectra Electronics SE-200 and Westinghouse proprietary noise test sets.

Noise in Relaxation Oscillators

ASAD A. ABIDI AND ROBERT G. MEYER, FELLOW, IEEE

Abstract—**The timing jitter in relaxation oscillators is analyzed. This jitter is described by a single normalized equation whose solution allows prediction of noise in practical oscillators. The theory is confirmed by measurements on practical oscillators and is used to develop a prototype low noise oscillator with a measured jitter of 1.5 ppm rms.**

I. INTRODUCTION

LOW noise in the output of an oscillator is important in many applications. The noise produced in the active and passive components of the oscillator circuit adds random perturbations to the amplitude and phase of the oscillatory waveform at its output. These perturbations then set a limit on the sensitivity of such systems as receivers, detectors, and data transmission links whose performance relies on the precise periodicity of an oscillation.

A number of papers [1]–[6] were published over the past several decades in which theories were developed for the prediction of noise in high-Q LC and crystal oscillators which are widely used in high-frequency receivers. Noise in these circuits is filtered into a narrow bandwidth by the high-Q frequency-selective elements. This fact allows a relatively simple analysis of the noise in the oscillation, the results of which show that the signal to noise ratio of the oscillation varies inversely with Q.

Relaxation oscillators are an important class of oscillators used in applications such as voltage-controllable frequency and waveform generation. In contrast to LC oscillators, they require only one energy storage element, and rely on the nonlinear characteristics of the circuit rather than on a frequency-selective element to define an oscillatory waveform. These circuits have recently become common because they are easy to fabricate as monolithic integrated circuits.

Due to their broad-band nature, these oscillators often suffer from large random fluctuations in the period of their output waveforms, termed the *timing jitter*, or simply, *jitter* in the oscillator. In an application such as an FM demodulator, the relaxation oscillator in a phase-locked loop will be limited in its dynamic range, and hence sensitivity, due to this jitter. There are no systematic studies in the literature, however, on either measurements of jitter in relaxation oscillators, or on an analysis of how noise voltages and currents in the components of the oscillator randomly modulate the periodic waveform. Such an analysis was perhaps discouraged by the nonlinear fashion in which the oscillator operates, as shall become evident in Section III below.

Despite this state of affairs, circuits designers have successfully used methods based on qualitative reasoning to reduce the jitter in relaxation oscillators. The purpose of this study has been to develop an explicit background for these methods. By analyzing the switching of such oscillators in the presence of noise, circuit methods are developed to reduce the timing jitter.

The results of this study have been verified experimentally, and a prototype low jitter oscillator was built with jitter less than 2 parts per million, nearly an order of magnitude better than most commonly available circuits.

II. OSCILLATOR TOPOLOGIES AND DEFINITION OF JITTER

One of the most popular relaxation oscillator circuits is the emitter-coupled multivibrator [7] with a floating timing capacitor shown in Fig. 1, which uses bipolar transistors as the active devices. Transistors $Q1$ and $Q2$ alternately switch on and off, and the timing capacitor C is charged and discharged via current sources I. Transistors $Q3$ and $Q4$ are level shifting emitter followers, and diodes $D1$ and $D2$ define the voltage swings at the collectors of $Q1$ and $Q2$. A triangle wave is obtained across the capacitor and square waves at the collectors of $Q1$ and $Q2$.

This circuit is sometimes modified for greater stability against temperature drifts, and other types of active devices are used, but in essence it is always equivalent to Fig. 1. The oscillator operates by sensing the capacitor voltage and reversing the current through it when this voltage exceeds a predetermined threshold.

Another common relaxation oscillator is shown in Fig. 2(a), which uses a grounded timing capacitor. The charging current is reversed by the Schmitt trigger output, whose two input thresholds determine the peak-to-peak amplitude of the triangle wave across the capacitor. The block diagram of this circuit is shown in Fig. 2(b). As the ground in an oscillator is defined only with respect to a load, the circuit of Fig. 1 is also represented by the block diagram of Fig. 2(b).

Manuscript received November 16, 1982; revised April 14, 1983. This research work was supported by the U.S. Army Research Office under Grant DAAG29-80-K-0067.

A. A. Abidi is with the Bell Laboratories, Murray Hill, NJ 07974.
R. G. Meyer is with the Electronics Research Laboratory, University of California, Berkeley, CA 94720.

Reprinted from *IEEE Journal of Solid-State Circuits*, vol. SC-18, pp. 794-802, December 1983.

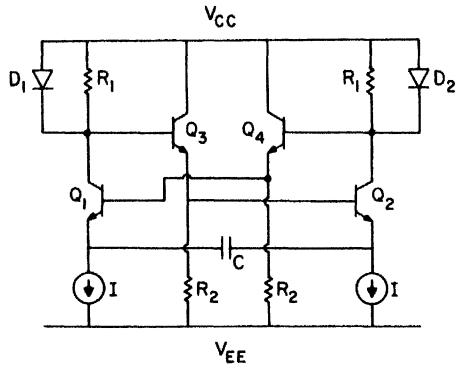


Fig. 1. Cross-coupled relaxation oscillator with floating timing capacitor.

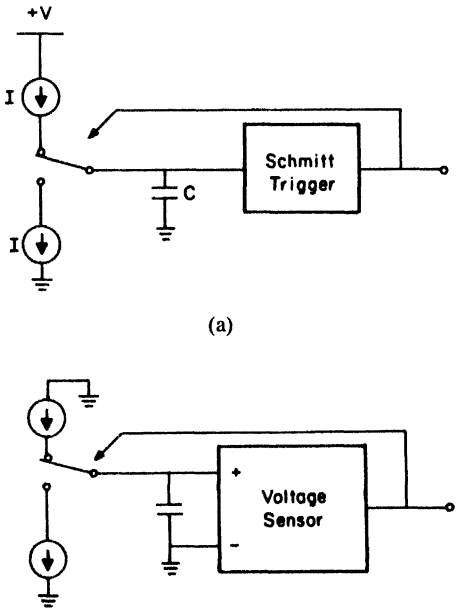


Fig. 2. (a) Relaxation oscillator with grounded timing capacitor. (b) Topological equivalent of oscillator.

The voltage sensor in such oscillators is a bistable circuit. When the capacitor voltage crosses one of the two trigger points at its input, the sensor changes state. The sensor has a vanishingly small gain, while the capacitor voltage is between these trigger points; but as a trigger point is approached, the operating points of the active devices in the sensor change in such a way that it becomes an amplifier of varying gain. The small-signal gain of the circuit is determined by an internal positive feedback loop, and becomes unboundedly large at the trigger point, causing the sensor to switch regeneratively and change the direction of the capacitor current.

In contrast to the linear voltage waveform on the capacitor (Fig. 3), the currents in the active devices of the sensor circuit are quite nonlinear because of this regeneration. The slope of these currents increases as the trigger point is approached, as shown in (Fig. 4), so that noise in the circuit is amplified and randomly modulates the time at which the circuit switches. Thus, the noisy current of Fig. 5 produces the randomly pulsewidth modulated waveform of

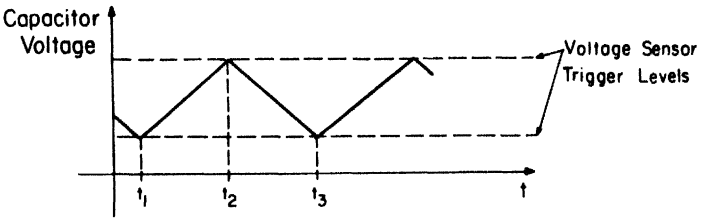


Fig. 4. Typical waveform of the current in a switching device in a relaxation oscillator.

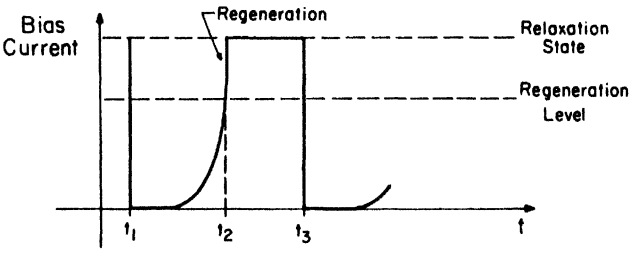


Fig. 4. Typical waveform of the current in a switching device in a relaxation oscillator.

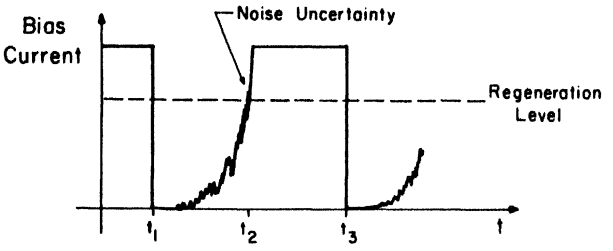


Fig. 5. Actual waveform (including noise) of the current in a switching device in a relaxation oscillator.

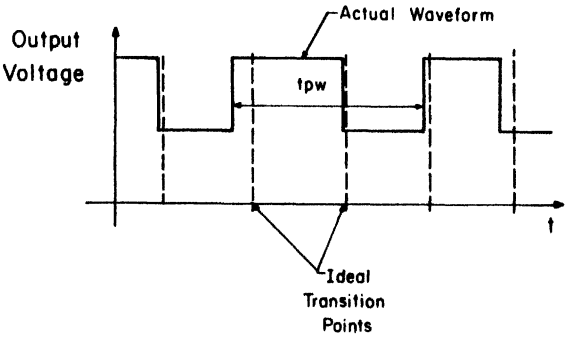


Fig. 6. Output voltage waveform of a relaxation oscillator showing the effect of noise.

Fig. 6. The timing jitter may be defined in terms of the mean μ_t and standard deviation σ_t of the pulsewidth as

$$\text{jitter} = \sigma_t/\mu_t \tag{1}$$

which is usually expressed in parts per million.

To determine the noise in FM demodulation, it is more desirable to know the frequency spectrum of an oscillation with jitter. However, the problem is more clearly stated, and solved, in the time domain, and the spectrum of the jitter should, in principle, be obtainable from a Fourier transformation.

III. The Process of Jitter Production

The switching of a floating capacitor oscillator in the presence of noise is now analyzed by examining the circuit of Fig. 7 as it approaches regeneration. The device and parasitic capacitance are assumed to be negligibly small. Suppose $Q2$ conducts a small current I_1, while $Q1$ carries the larger current $2I_0 - I_1$. The current flowing through the timing capacitor produces a negative-going ramp at V_4, causing an increase in $V_{BE}(Q_2)$ and thus in the current through Q_2. A single stationary noise source I_n is assumed to be present in the circuit, as shown in Fig. 7. The equations describing the circuit are

$$V_{BE_1} = V_T \log_e \frac{2I_0 - I_1}{I_S} \tag{2}$$

$$V_{BE_2} = V_T \log_e \frac{I_1}{I_S} \tag{3}$$

$$V_C = (2I_0 - I_1)R + V_{BE_2} - (I_1 - I_n)R - V_{BE_1} \tag{4}$$

$$\frac{dV_C}{dt} = \frac{I_0 - I_1}{C} \tag{5}$$

where $V_T = kT/q$ and I_S is the reverse leakage current at the base of the transistor.

Substituting (2) and (3) into (4), and applying (5), the differential equation describing the circuit is

$$\left(\frac{V_T}{2I_0 - I_1} + \frac{V_T}{I_1} - 2R\right)\frac{dI_1}{dt} + \frac{R\,dI_n}{dt} = \frac{I_0 - I_1}{C}. \tag{6}$$

This may be rewritten as

$$\frac{dI_1}{dt} = \left(\frac{1}{\dfrac{V_T}{2I_0 - I_1} + \dfrac{V_T}{I_1} - 2R}\right)\left(\frac{I_0 - I_1}{C} - \frac{R\,dI_n}{dt}\right) \tag{7}$$

where the right-hand side consists of two terms, one due to the autonomous dynamics of the circuit and one due to noise. As I_1 increases, the denominator of the right-hand side diminishes until it becomes zero when

$$I_1 = I_R \approx \frac{V_T}{2R} \tag{8}$$

where $I_1 \ll I_0$; I_1 must satisfy this inequality for the circuit to oscillate. I_R is defined as the *regeneration threshold* of the circuit: upon exceeding it, and in the absence of any device or stray capacitance, I_1 would change at an infinite rate until one of the circuit voltages limits. Accordingly, the circuit is said to be in the *relaxation mode* while $0 < I_1 < I_R$, and in *regeneration* when $I_R < I_1 < 2I_0$. (I_1 may be the current through either $Q1$ or $Q2$, depending on the particular half cycle of oscillation.)

Suppose that at some time $t = t_A$ the circuit is in relaxation so that the current $I_1 = I_A$, and that it builds up the threshold of regeneration $I_1 = I_R$ at time $t = t_2$. Equation

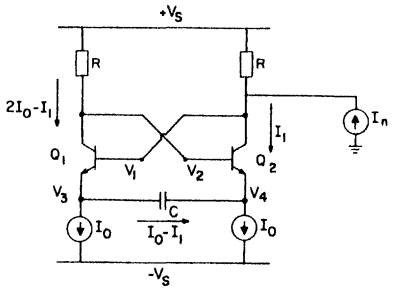


Fig. 7. Generalized equivalent circuit of a relaxation oscillator for noise analysis.

(6) can then be integrated from t_A to t_2 as follows:

$$\int_{I_A}^{I_R}\left(\frac{V_T}{2I_0 - I_1} + \frac{V_T}{I_1} - 2R\right)dI_1 = \int_{t_A}^{t_2}\left(\frac{I_0 - I_1}{C}\right)dt - R\{I_n(t_2) - I_n(t_A)\} \tag{9}$$

so that

$$V_K = \frac{I_0}{C}(t_2 - t_A) - \int_{t_A}^{t_2}\frac{I_1}{C}dt - R\{I_n(t_2) - I_n(t_A)\} \tag{10}$$

where V_K, the left-hand side of (9), is a constant which depends only on the choice of initial current I_A, I_0, and, from (8), on V_T and R.

The influence of the noise current on the instant of switching is now evident from (10): random fluctuations in the value of $I_n(t_2)$ must induce corresponding fluctuations in t_2 so that the sum of the terms on the right-hand side of (10) remains equal to the constant V_K. More precisely, if $t_A = 0$ and $t_2 = T$ (the half-period of the oscillation), then

$$F(I_n(T), T) \overset{\text{def}}{=} \frac{I_0}{C}T - \int_0^T \frac{I_1(t)}{C}dt - R\{I_n(T) - I_n(0)\} = \text{constant} \tag{11}$$

and thus

$$\delta F = \left.\frac{\partial F}{\partial I_n}\right|_T \delta I_n(T) + \frac{\partial F(I_n, T)}{\partial T}\delta T = 0 \tag{12}$$

which implies

$$-R\delta I_n(T) + \left\{\frac{I_0}{C} - \frac{I_1(T)}{C}\right\}\delta T = 0 \tag{13}$$

so that

$$\delta T = \frac{R}{\left\{\dfrac{I_0 - I_1(T)}{C}\right\}}\delta I_n(T). \tag{14}$$

By definition, $I_1(T) = I_R$ so

$$\delta T = \frac{R}{\left\{\dfrac{I_0 - I_R}{C}\right\}}\delta I_n(T). \tag{15}$$

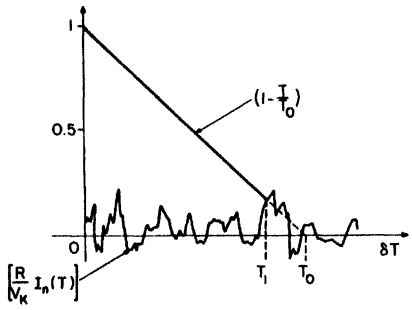


Fig. 8. Graphical interpretation of (19).

The main result of this paper lies in interpreting this equation as follows: the *variation* in switching times δT is equal to the variation in the times of first intersection of a *ramp* $(I_0 - I_R/C)T$ with the noise waveform $RI_n(T)$, as shown in Fig. 8. Note that the x axis in this figure is T, and *not* "real" time t. Thus, while the current waveform is nonlinear, as in Fig. 5, and while the time T, at which it starts to regenerate is given by the integral equation (10), *variations* in T appear as if they were produced by the first crossing of a noisy threshold by a linear ramp.

The statistics of T are contained in (10) in terms of the distribution of $I_n(t)$. This is the well-known equation for the first-crossings' distribution of a deterministic waveform with noise [8] and does not have a convenient closed form solution.

Nevertheless, a useful empirical result has been obtained for the distribution of T which allows the jitter in relaxation oscillators to be predicted quite accurately. This result is based on the following observations.

1) The linearized equation (15) describes completely the *deviations* in T; therefore, it contains information on the statistics of T.

2) It is plausible that the jitter will be directly proportional to the rms noise current for a fixed slope of the timing waveform.

3) If the dominant noise power in the oscillator lies at high frequencies, it acts to reduce the mean period of oscillation. This is evident from Fig. 8, where the ramp will almost always cross a positive-going peak of the noise.

4) If low-frequency noise is dominant, the resulting jitter will be greater than it would be if the same noise power were concentrated at higher frequencies. Again, Fig. 8 shows this, because if noise varies slowly compared to the ramp rate, the first-crossing will occur both above and below the x axis.

These observations can be quantitatively summarized as

$$\sigma(\delta T) = \sigma(T) = \alpha \frac{RC}{I_0 - I_R} \sigma(I_n) \tag{16}$$

where σ denotes the standard deviation of its argument and α is a constant of proportionality which by 1) and 2) above depends on whether the noise power is contained at high or at low frequencies. Equation (15) changes to (16) in going from the variation in T in one switching event to the standard deviation of an ensemble of such events.

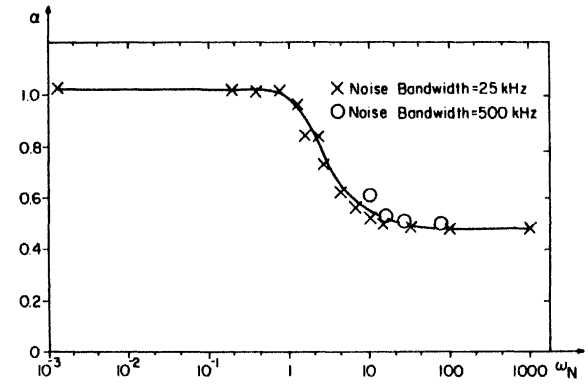


Fig. 9. Measured values of parameter α versus normalized noise bandwidth ω_N.

The constant α varies with the relative slope of the ramp to the rms slope of the noise. For white noise which has been low-pass filtered, the rms slope can be defined relative to the ramp rate by ω_N, where

$$\omega_N \overset{\text{def}}{=} 2\pi \times \frac{\text{rms noise voltage} \times \text{noise bandwidth}}{\text{voltage ramp rate}} \tag{17}$$

and where the rms noise voltage is responsible for modulating the first-crossing instant of the voltage ramp of Fig. 8.

The dependence of α on ω_N for low-pass filtered white noise is shown in Fig. 9. It was obtained from measurements on different oscillator circuits, and at varying noise levels, as described in Section V of this paper. This dependence should be the same for white noise in any relaxation oscillator. For $\omega_N \ll 1$, α approaches unity, as 4) above suggests, because the deviation in oscillator period faithfully follows the meanderings of the noise waveform. For $\omega_N \gg 1$, α is asymptotic to about 0.5 because only the positive peaks of noise determine the first-crossing, in accordance with 3).

If the effect of all the noise sources in an oscillator is represented by the single source I_n, then using the appropriate α, the jitter in its output can be predicted by (16). For example, in Fig. 7 devices $Q1$ and $Q2$, while forward-biased, act as voltage followers for the various noise voltage sources in the circuit, so that I_n is the rms sum of these noise voltages divided by the node resistance at the collector. This is true for all noise sources except the noise current flowing through the timing capacitor which is integrated into a voltage by the capacitor. As shown in Appendix I, its contribution to the jitter is usually negligible.

IV. Interpretation and Generalization of Results

We emphasize that the linear result of (16), which describes the variations in the nonlinear waveform of Fig. 5, is not merely an outcome of an incremental analysis of the problem, which would go as follows. As the loop approaches regeneration, the dc incremental gain increases and the small signal bandwidth due to the device and

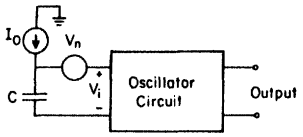


Fig. 10. Representation of noise in the relaxation oscillator by an equivalent input generator V_n.

parasitic capacitance decreases in inverse proportion, so that the incremental gain is infinite and the bandwidth zero at the onset of regeneration. Having entered the regeneration regime, the effect of the capacitance is to limit the maximum rate of change of the circuit waveforms. Thinking in terms of signal-to-noise ratio, the incremental device current is the signal which must compete with the amplified noise in the circuit to determine the time of switching. From the considerations of gain and bandwidth above, the rms value of white noise would become infinite at the regeneration threshold, so reducing the "signal"-to-noise ratio to zero. This implies infinite jitter, which is obviously not the case in reality. Such an approach demonstrates the inadequacy of thinking of this problem in terms of small signals.

A complete, large scale analysis shows that jitter production is better understood by thinking of the oscillator as the simplified threshold circuit of Fig. 10. The equivalent noise at the input of the circuit adds to the linear voltage waveform on the timing capacitor to produce an uncertainty in the time of regeneration. It is important to realize that Fig. 10 represents the oscillator only for an incrementally small time before the onset of regeneration.

This result is independent of the type of the oscillator circuit, and of the nature of the active devices used in it.

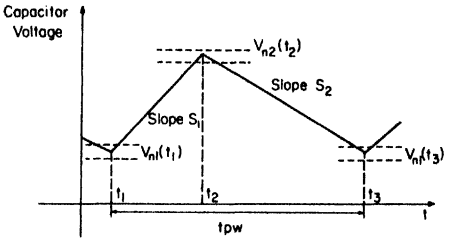


Fig. 11. Capacitor voltage waveform in a relaxation oscillator.

the measurement of the latter very difficult. By placing the oscillator in a phase-locked loop with a crystal-derived reference frequency, and by designing the loop filter with a cutoff well below the oscillation frequency, the thermal drift can be compensated, while the cycle-to-cycle jitter produced by noise frequencies above this cutoff remains unaffected. Appropriate precautions must be taken against fluctuations in the power supply, and against ground loop noise.

The effect of a noise voltage V_n over a complete cycle of oscillation is shown in Fig. 11, where the timing capacitor waveform is assumed to be asymmetrical for generality. As the device current I_1 in Fig. 7 always equals I_R at switching, the limits of the capacitor voltage have to fluctuate in response to the noise. Therefore, as the capacitor voltage is a continuous waveform, the fluctuations at times t_1, t_2, and t_3 all contribute to δT. Thus,

$$\delta T = \frac{V_n(t_1)}{S_1} + \frac{V_n(t_2)}{S_1} + \frac{V_n(t_2)}{S_2} + \frac{V_n(t_3)}{S_2}. \quad (19)$$

The standard deviation of the random variable δT is

$$\sigma(\delta T) = \alpha \times \left\{ \text{mean value of} \left\{ \frac{V_n^2(t_1)}{S_1^2} + V_n^2(t_2)\left(\frac{1}{S_1} + \frac{1}{S_2} \right)^2 + \frac{V_n^2(t_3)}{S_2^2} \right\} \right\}^{1/2} \quad (20)$$

The analysis of the grounded capacitor oscillator in Appendix II, and further generalizations given elsewhere [10] show that the jitter in any relaxation oscillator is given by the following expression:

$$\text{jitter} = \frac{\text{rms noise voltage in series with timing waveform}}{\text{slope of waveform at triggering point}} \times \text{a constant}. \quad (18)$$

V. Experimental Results

To verify the formulas for phase jitter developed above, and also to develop low noise oscillator circuits, it is necessary to be able to measure jitter with a resolution of about 1 ppm. This entails obtaining the distribution of pulsewidths from an accurate pulsewidth meter while the oscillator runs at a low frequency; the jitter is then the standard deviation of this distribution. However, the short-term thermal drift of the oscillation frequency can overwhelm the variations in cycle-to-cycle jitter, making

where $V_n(t_1)$, $V_n(t_2)$, and $V_n(t_3)$ are statistically independent values of the noise voltage $V_n(t)$, and α is the constant of proportionality defined in (16). When $S_1 = S_2 = S$ in magnitude,

$$\sigma(\delta T) = \alpha\sqrt{6}\,\frac{V_n(\text{rms})}{S} \quad (21)$$

where $\sigma(\delta T)$ is the jitter.

Measurements were made by dominating the oscillator's internal noise sources by an externally injected low-pass filtered white noise current. Two different oscillator circuits, described below, were used to obtain the experimental data. Histograms for the distributions of pulsewidths for injected noise with $\omega_N = 0.2$ and 1000 are shown in Fig. 12, where this range of ω_N was obtained by adjusting both the power and bandwidth of the injected noise. These histograms are unimodal and approximately symmetric, and they fit the normalized Gaussian function to within experimental error. Such experiments were also used to obtain the curve of α versus ω_N of Fig. 9, with the jitter defined as the standard deviation of these histograms.

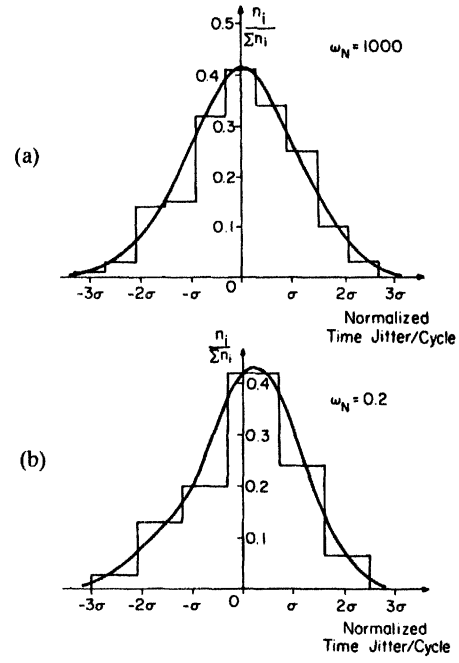


Fig. 12. Measured distributions of time jitter per cycle in the AD537 oscillator with (a) $\omega_N = 1000$ and (b) $\omega_N = 0.2$; fractional number of samples of single pulses versus their pulse widths.

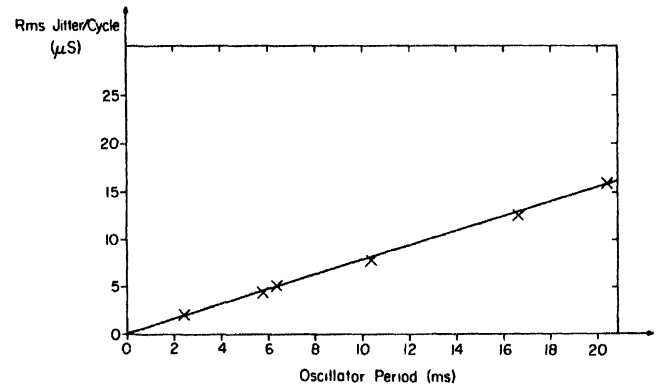


Fig. 13. Measured jitter per cycle versus oscillator period with noise injected into the AD537.

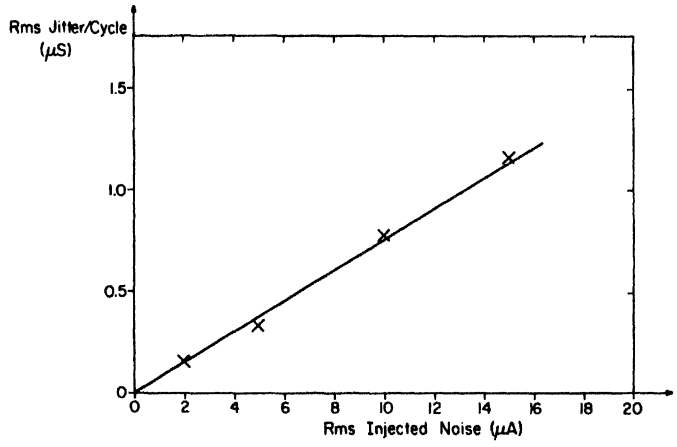


Fig. 14. Measured jitter per cycle versus injected noise amplitude in a grounded capacitor oscillator.

To verify the general result of (18), the following experiments were performed:

1) For a fixed injected rms noise, the jitter was measured as a function of the period of oscillation; the results are shown in Fig. 13.

2) For a fixed period of oscillation, the jitter was measured for a varying rms input noise; the results are plotted in Fig. 14.

The straight line fit of the data confirmed (18). In both 1) and 2), the range of the independent variable was restricted such that $\omega_N \gg 1$, thus ensuring that α was at its lower asymptotic value of 0.48, so that variations in α did not confound the measurements. The data of Fig. 13 were obtained from the AD 537 [9] floating capacitor oscillator, with a noise voltage applied in series with the voltage reference; Fig. 14 was obtained from measurements on a discrete component grounded capacitor oscillator [10], with a noise current injected into the Schmitt trigger. The reduction in mean period of oscillation predicted by 3) in Section III was also observed.

The most important application of this theory is in determining the jitter due to the inherent noise sources in an oscillator circuit. However, it is essential to know the bandwidth that applies to these noise sources in determining their total noise power. The small-signal bandwidth changes with the approach to the regeneration threshold when the circuit starts to behave like an integrator, as discussed in Section IV. While a detailed analysis of this is beyond the scope of this paper, the situation can be examined qualitatively. Consider the circuit described by (7) in the relaxation regime, and with all devices in the active region, when, say, $I_1 = 0.1 \times I_R$. If the spectrum of the superposition of all the noise sources on I_1 is considered as the "output" noise variable, the noise bandwidth of the circuit can be defined as the 3 dB down frequency of this spectrum. The equivalent input noise source of Fig. 10 is then this superposed noise in I_1 referred to a voltage source in series with the capacitor.

As I_1 approaches I_R, the circuit acts more like an integrator, so that fluctuations in the switching instant T are proportional to the fluctuations in the initial conditions of the integrator. The latter are simply produced by the noise in the circuit when the approach towards regeneration is started, which may roughly be defined as the time when $I_1 = 0.1 \times I_R$.

The spectral density of some of the noise sources in the circuit will depend on I_1, but usually they make a small contribution to the total noise, and their value at $0.1 \times I_R$ is a good approximation.

The noise bandwidth of the oscillator can be measured experimentally in two ways. The first relies on the availability of a white noise source with an adjustable output filter whose cutoff extends beyond the bandwidth to be measured. In response to a constant injected noise power, with the filter cutoff being progressively increased, the jitter will drop by 3 dB at the noise bandwidth of the oscillator. Alternatively, if the noise source has a fixed cutoff frequency, it can be used to modulate the amplitude of a carrier frequency, which is then injected into the oscillator. The random modulation will produce jitter, and

as the carrier frequency is increased, the jitter will reduce by 3 dB at the noise bandwidth. This relies on the fact that while the jitter is determined by the amplitude fluctuations (Fig. 8), the frequency components of the injected noise are concentrated around the carrier frequency, and increase with it.

As an example, the inherent jitter of the AD 537 VCO can be predicted. The noise bandwidth was measured to be 16 MHz using the second method described above. This 16 MHz value gives a fair agreement with SPICE noise simulations of the circuit biased at $I_1 = 0.1 \times I_R$, despite the fact that the exact values of the device capacitances were not available. The noise spectral density was simulated to be 1.3×10^{-8} V/$\sqrt{\text{Hz}}$ using SPICE, and was primarily due to the current sources and the base resistances of the level shift and switching transistors. Thus,

$$V_n = 1.3 \times 10^{-8} \times \sqrt{16 \times 10^6} = 52\ \mu\text{V rms}$$

At an oscillation frequency of 1 kHz, the slope of the timing capacitor ramp was 2.6×10^3 V/s, giving $\omega_N = 2.01$ and thus $\alpha = 0.8$ from Fig. 9. The predicted jitter is

$$\sigma(T) = \sqrt{6} \times 0.8 \times \frac{52 \times 10^{-6}}{2.6 \times 10^3}$$
$$= 39 \text{ ns}$$
$$= 39 \text{ ppm.}$$

The measured jitter from several samples had a mean value of 35 ppm, an excellent agreement in view of the approximate values of the active device noise models.

VI. A Low Jitter Oscillator Circuit

Many applications require even lower values of jitter than that of the AD 537 which is one of the lowest jitter monolithic VCO's widely available. The theory developed in this paper allows oscillators to be designed to a specified noise performance; as an example, a circuit with 1 ppm jitter was designed.

In the block diagram of Fig. 10, suppose that the timing voltage on the capacitor is a triangle wave with a peak-to-peak value V and period T. The slope of the ramp is

$$S = 2V/T \tag{22}$$

so the fractional jitter is

$$J = \frac{\alpha\sqrt{6}}{2} \frac{V_n}{V} \tag{23}$$

where V_n is the rms noise voltage. Thus, to obtain a small jitter, it is necessary to reduce V_n and increase V within the constraints of the circuit. V is limited by the power supply of the circuit, and V_n depends on both the characteristics of the active devices and the circuit topology. In the AD 537 topology (Fig. 15), for example, many devices additively contribute to the total noise in series with the timing capacitor because the functions of regeneration and threshold voltage detection are combined into one circuit.

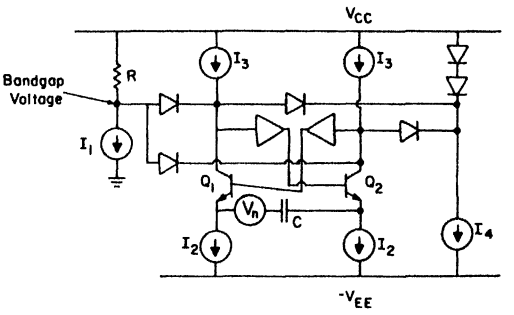


Fig. 15. Simplified schematic of the AD537.

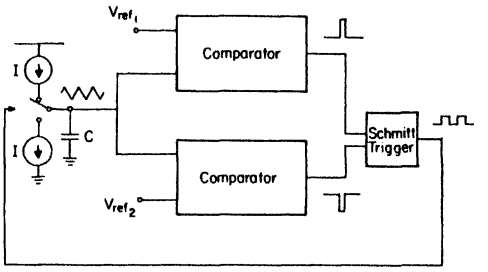


Fig. 16. Low-noise grounded capacitor oscillator topology.

In the grounded capacitor circuit [Fig. 2(a)], however, the current switch is separate from the regenerative Schmitt trigger; their noise contributions can thus be minimized independently.

Fig. 16 shows a variation of this topology where the Schmitt trigger is driven by fast pulses produced by high gain comparators following the timing capacitor. The result (18) shows that the contribution of the noise in the Schmitt trigger to the total jitter is inversely proportional to the slope of the waveform which drives it, and so is very small. Instead, the input noise of the comparators, which are driven by a slow ramp, determines the jitter. This scheme was used because very low input noise comparators are easily available, although there is no fundamental reason why a Schmitt trigger of comparable noise could not be designed. The complete schematic of the discrete component oscillator circuit is shown in Fig. 17. The differential comparators have a gain of 100 and an equivalent input noise of 6.3 μV rms over a noise bandwidth of 75 kHz. Using ± 6 V power supplies, the timing capacitor waveform was 8.8 V peak-to-peak, so (23) predicted the jitter to be 0.9 ppm rms. At an oscillation frequency of 1 kHz, the cycle-to-cycle jitter was measured to be 1.5 ppm rms; the additional jitter probably came from inadequate decoupling in the circuit.

An oscillator working from a single 5 V power supply with a jitter of about 1 ppm would be desirable in many systems applications. Such a circuit has been developed [12] on the basis of the results above, and the jitter measured to be less than 1 ppm at a 1 kHz oscillation frequency.

VII. Conclusions

Noise in relaxation oscillators can be described by a single normalized equation, which allows the jitter in any

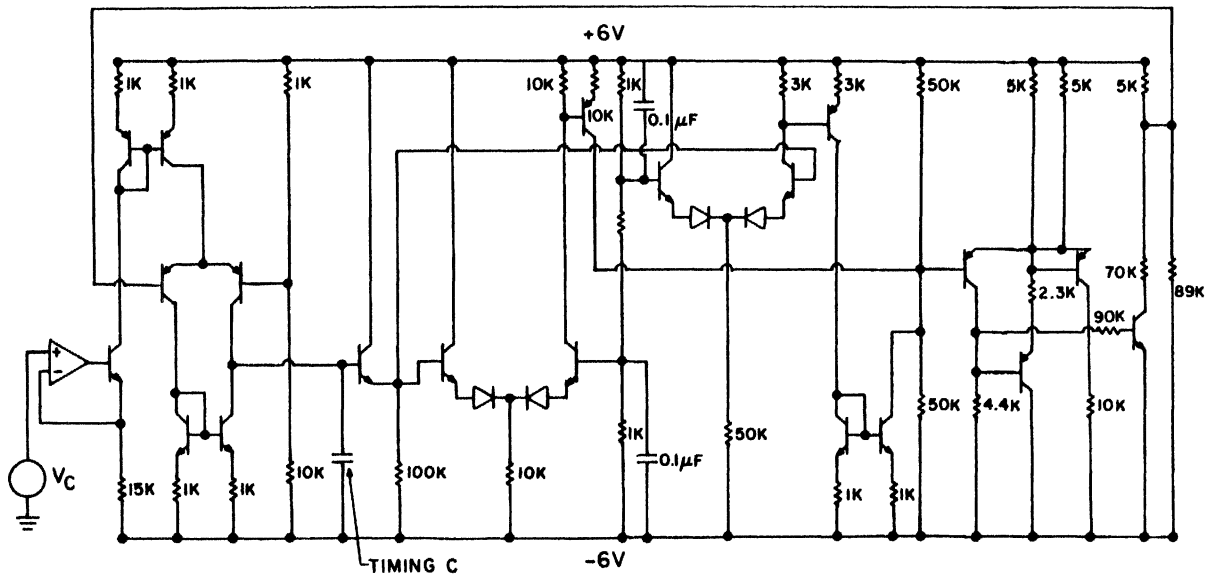


Fig. 17. Complete schematic for the low noise grounded-capacitor oscillator.

such oscillator to be predicted. The importance of this equation is that it linearizes the nonlinear regenerative waveforms in the oscillator. Further, it suggests those circuit topologies which promise low jitter, as demonstrated by an experimental prototype, whose jitter was measured to be 1.5 ppm at 1 kHz.

Appendix I

Effect of Noise Current Through the Timing Capacitor

In the schematics of Fig. 1 and Fig. 2 it can be seen that noise in the current sources appears directly in series the timing capacitor. In practice, these current sources are almost always active sources whose noise can be represented by a band-limited output noise current generator $I_{nc}(t)$. The capacitor then acts as a low-pass filter and the contribution V_{nc} from I_{nc} to the equivalent input noise voltage V_n of the oscillator (see Fig. 10) is

$$V_{nc}(t_r) = \frac{1}{C}\int_0^{t_r} I_{nc}(t)\,dt \qquad \text{(A1)}$$

where t_r is the time during which I_{nc} charges C. If I_{nc} is subject to a single-pole frequency roll off with bandwidth B and flat spectral density $S_{nc}(f)$, then it can be shown that [11]

$$(V_{nc})(\text{rms}) = \frac{1}{C}\sqrt{S_{nc}(f)} \times \sqrt{t_r} \qquad \text{(A2)}$$

where it is assumed that $t_r \gg 1/B$.

The rms noise contribution in V_n due to I_{nc} as given in (A2) can be compared with the contribution V_{nn} from I_n at the collector

$$(V_{nn})(\text{rms}) = (I_n)(\text{rms})R.$$

The relative importance of these two terms is now examined. Taking some typical values, if 1 mA current sources produce the shot noise,

$$S_{nc}(f) = 3.2\times10^{-22}\ \text{A}^2/\text{Hz}$$

and if $t_r = 1$ ms and $C = 0.5$ μF, then from (A2)

$$(V_{nc})(\text{rms}) = 1.1\times10^{-6}\ \text{V}.$$

In comparison, when I_n is the shot noise in 1 mA of dc current, and it is band-limited to 16 MHz by the circuit capacitances, then $R = 500\ \Omega$ gives

$$(V_{nn})(\text{rms}) = \sqrt{3.2\times10^{-22}\times16\times10^{6}}\times500$$
$$= 36\times10^{-6}\ \text{V}.$$

Thus $V_{nc} \ll V_{nn}$, and the difference is even larger when additional contributions to V_{nn} are considered.

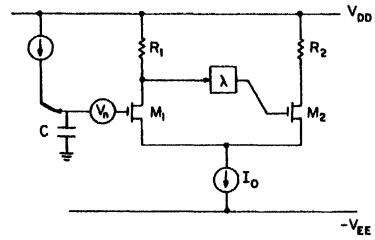


Fig. 18. Grounded capacitor oscillator's Schmitt trigger using MOS transistors.

Appendix II

Noise in an MOS Grounded Capacitor Oscillator

A simplified circuit of a grounded capacitor oscillator consisting of only the timing capacitor and the Schmitt trigger is shown in Fig. 18, where the latter uses MOS active devices with square law characteristics. The current I through device $M1$ is driven by the capacitor voltage and

makes the Schmitt trigger approach its regeneration threshold. The noise in the circuit is represented by the equivalent source V_n, and adds to the timing capacitor ramp in the manner of Fig. 10; it could equally well have been represented as a noise current source within the Schmitt circuit.

The circuit equations are as follows:

$$\frac{I_0 t}{C} + V_n = V_{GS_1} + V_S = V_T + \left(\frac{I}{\beta}\right)^{1/2} + V_S \qquad \text{(B1)}$$

$$\lambda(V_{DD} - IR_1) = V_{GS_2} + V_S = V_T + \left(\frac{I_0 - I}{\beta}\right)^{1/2} + V_S. \qquad \text{(B2)}$$

Subtracting (B2) from (B1)

$$\frac{I_0 t}{C} + V_n - \lambda V_{DD} + \lambda I R_1 = \left(\frac{I}{\beta}\right)^{1/2} - \left(\frac{I_0 - I}{\beta}\right)^{1/2} \qquad \text{(B3)}$$

and differentiating (B3),

$$\frac{I_0}{C} + \frac{dV_n}{dT} + \lambda R_1 \frac{dI}{dt} = \frac{1}{2\beta}\left\{\left(\frac{\beta}{I}\right)^{1/2} + \left(\frac{\beta}{I_0 - I}\right)^{1/2}\right\}\frac{dI}{dt} \qquad \text{(B4)}$$

rewriting which

$$\frac{dI}{dt} = \frac{\frac{I_0}{C} + \frac{dV_n}{dt}}{\frac{1}{2\beta}\left\{\left(\frac{\beta}{I}\right)^{1/2} + \left(\frac{\beta}{I_0 - I}\right)^{1/2}\right\} - \lambda R_1}. \qquad \text{(B5)}$$

The regeneration threshold $I_R = 1/4\beta\lambda^2 R_1^2$ is the current at which the denominator of the right-hand side becomes zero. If this happens at $t = T$ relative to some time $t = 0$, then integrating (B5) from 0 to T gives

$$V_K = \frac{I_0 T}{C} + \{V_n(T) - V_n(0)\} \qquad \text{(B6)}$$

where V_K is a constant voltage, depending only on the circuit parameters. This is exactly the same equation as (10), and gives the same result for jitter as (16).

References

[1] J. L. Stewart, "Frequency modulation noise in oscillators," *Proc. IRE*, vol. 44, pp. 372–376, Mar. 1956.

[2] W. A. Edson, "Noise in oscillators," *Proc. IRE*, vol. 48, pp. 1454–1466, Aug. 1960.

[3] J. A. Mullen, "Background noise in nonlinear oscillators," *Proc. IRE*, vol. 48, pp. 1467–1473, Aug. 1960.

[4] M. G. E. Golay, "Monochromaticity and noise in a regenerative electrical oscillator," *Proc. IRE*, vol. 48, pp. 1473–1477, Aug 1960.

[5] P. Grivet and A. Blaquiere, "Nonlinear effects of noise in electronic clocks," *Proc. IEEE*, vol. 51, pp. 1606–1614, Nov. 1963.

[6] J. Rutman, "Characterization of phase and frequency instabilities in precision frequency sources: Fifteen years of progress," *Proc. IEEE*, vol. 66, pp. 1048–1075, Sept. 1978.

[7] A. B. Grebene, "The monolithic phase-locked loop–A versatile building block," *IEEE Spectrum*, vol. 8, pp. 38–49, Mar. 1971.

[8] D. Middleton, *Statistical Communication Theory*. New York: McGraw-Hill, 1960.

[9] B. Gilbert, "A versatile monolithic voltage-to-frequency converter," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 852–864, Dec. 1976.

[10] A. A. Abidi, "Effects of random and periodic excitations on relaxation oscillators," Univ. California, Berkeley, Memo UCB/ERL M81/80, 1981.

[11] E. Parzen, *Stochastic Processes*. San Francisco: Holden-Day, 1962.

[12] T. P. Liu and R. G. Meyer, private communication.

Analysis of Timing Jitter in CMOS Ring Oscillators

Todd C. Weigandt, Beomsup Kim*, Paul R. Gray

Department of Electrical Engineering and Computer Science
University of California, Berkeley

* Department of Electrical Engineering
Korea Advanced Institute of Science and Technology

Abstract

In this paper the effects of thermal noise in transistors on timing jitter in CMOS ring-oscillators composed of source-coupled differential resistively-loaded delay cells is investigated. The relationship between delay element design parameters and the inherent thermal noise-induced jitter of the generated waveform are analyzed. These results are compared with simulated results from a Monte-carlo analysis with good agreement. The analysis shows that timing jitter is inversely proportional to the square root of the total capacitance at the output of each inverter, and inversely proportional to the gate-source bias voltage above threshold of the source-coupled devices in the balanced state. Furthermore, these dependencies imply an inverse relationship between jitter and power consumption for an oscillator with fixed output period. Phase noise and timing jitter performance are predicted to improve at a rate of 10 dB per decade increase in power consumption.[1]

I. Introduction

Ring oscillators are widely used in phase-locked-loops (PLL) for clock and data recovery, frequency synthesis, clock synchronization in microprocessors, and many applications which require multi-phase sampling [1] [2]. In many such applications, clock signals are generated to drive mixers or sampling circuits in which the random variation of the sampling instant, or jitter, is a critical performance parameter. In some applications the frequency domain equivalent of jitter, called phase noise, is important. A block diagram of a typical PLL using a ring-oscillator for multi-phase clock generation is shown in figure 1. Jitter requirements in typical applications range from on the order of 100 picoseconds r.m.s. down to less than 5 picoseconds in very high-speed communications receivers, for example.

Jitter can arise from many sources, including inadvertent injection of signals from other parts of the circuit through the power supply. However, interfering sources like these can often be minimized by the use of circuit techniques such as differential implementations. In a fully optimized design the main source of timing jitter is the inherent thermal and/or shot noise of the active and passive devices that make up the inverter cell. 1/f noise is usually not of practical importance since it is rejected by the PLL loop filter, and does not effect the stage-to-stage delay in a DLL. Therefore minimizing the impacts of thermal and shot noise in the basic inverter cells becomes the key to attaining low timing jitter.

1. Research supported by NSF, ARPA, and the California MICRO Program

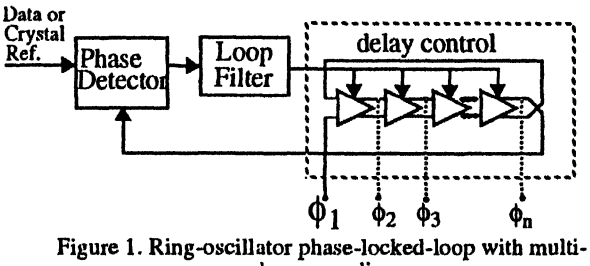


Figure 1. Ring-oscillator phase-locked-loop with multi-phase sampling

This paper attempts to determine analytically and through simulation the relationship between the design parameters of the inverter cell used in the ring-oscillator and the resulting noise-induced jitter. The class of circuits analyzed is source-coupled differential delay cells with resistive loads, implemented in CMOS technology, where the loads are realized by PMOS transistors in the triode region (figure 2). This particular implementation has proven useful in practical applications because of its high speed and rejection of supply noise [1]. In this paper we will first consider jitter for the individual delay stages in a ring-oscillator, and then look at the implications for design of the overall ring-oscillator phase-locked-loop.

II. First Order Timing Jitter Analysis

The period of a ring-oscillator is determined by the number of stages in the ring and the delay for each stage. Accompanying each cycle of oscillation is a random timing error due to noise. The goal of this section is to determine the contribution of thermal noise sources in an ECL type inverter circuit, like that shown in figure 2, to the timing jitter of the ring-oscillator.

In this analysis, each inverter stage in a ring-oscillator is assumed to contribute a nominal time delay, t_d, and a timing error, $\Delta\tau$, to each cycle of oscillation. The timing error has a mean of zero and a variance denoted by $\Delta\tau_1^2$. To first order, the delay per stage is measured from the time when the outputs begin switching to the time when the differential output reaches zero, as illustrated in figure 3. With this assumption the nominal delay per stage is given by

$$t_d \cong V_{PP}\left(\frac{C_L}{I_{SS}}\right) \tag{1}$$

where I_{SS} / C_L is the output slew rate and V_{PP} is one half the full differential output swing. The load capacitance, C_L, is the total capacitance at the output of each inverter.

The random component of the timing delay is estimated using the first crossing approximation ([3]), illustrated in figure 4. Here, the simplifying assumption is made that the next stage begins switch-

Reprinted from *Proc. of ISCAS*, June 1994.

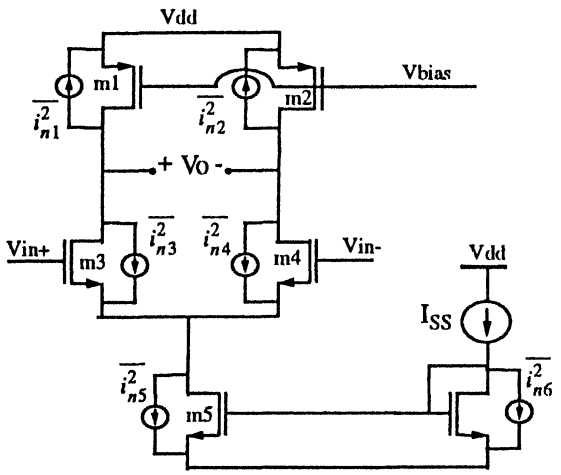


Figure 2. Differential delay cell with noise sources

ing when the differential output voltage crosses zero, and the error in the actual time of crossing is the timing error passed on to future stages in the delay chain. Figure 4 shows that an error voltage at the nominal time of crossing shifts the actual time by an amount proportional to the voltage error divided by the slew rate of the output. Using this approximation, the timing error variance is given by :

$$\overline{\Delta\tau_1^2} \cong \overline{\Delta v_n^2} \times \left(\frac{C_L}{I_{SS}}\right)^2 \qquad (2)$$

The voltage noise variance in equation (2) is the sum of contributions of each of the thermal noise sources in figure 2. The contribution of these noise sources to the differential output voltage is actually time varying in nature since they change as the circuit switches. In this section the simplifying assumption is made that the voltage noise variance will be the same as if the circuit where in equilibrium. In this case traditional noise analysis techniques [4] apply and the output referred voltage noise can be determined by integrating the noise spectral density over the bandwidth of the low-pass filter formed by the load resistor and the gate capacitance of the next stage. If this result is combined with (2) and (1), the r.m.s. timing jitter error for one stage normalized to the time delay per stage can be shown to be :

$$\frac{\Delta\tau_{1rms}}{t_d} \cong \frac{\Delta v_{rms}}{V_{PP}} = \sqrt{\frac{2kT}{C_L}} \bullet \left(\sqrt{1+\frac{2}{3}a_v}\right) \bullet \frac{1}{V_{PP}} \qquad (3)$$

Interestingly, the ratio of the timing error to the time delay per stage is just given by the ratio of the r.m.s. voltage noise to the voltage swing, V_{PP}. The voltage noise has the familiar *kT/C* dependence, and is proportional to another term called the noise contribution factor, ξ. In this case $\xi = \sqrt{1 + (2/3)\,a_v}$ where a_v is the small-signal gain of the inverter. The NMOS noise contribution is given by the second term in this expression and is proportional to the gain since, for a fixed output bandwidth, higher gain implies higher transconductance and hence a larger noise contribution. The PMOS contribution is the first term which in this case is just one.

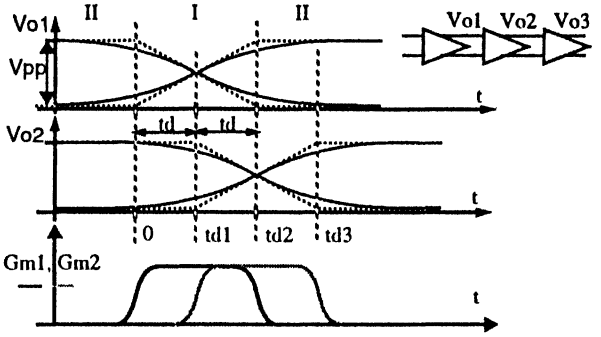
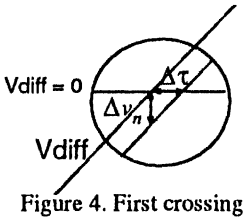


Figure 3. Output waveforms for CMOS inverter chain



$$\overline{\Delta\tau^2} = \frac{\overline{\Delta v_n^2}}{(\text{Slew Rate})^2}$$

Figure 4. First crossing approximation for timing jitter

III . Second Order Analysis

The first order analysis neglects many important contributions to noise. A more thorough analysis must consider the time varying nature of the noise sources, the effects of the tail current noise sources, and interactions between stages.

Time varying noise sources

The assumption that the voltage noise variance is the same as its equilibrium value is not valid for the NMOS differential pair transistors since each side switches from fully on to fully off, during which the transconductance, and hence the noise contribution changes dramatically. Furthermore the tail current noise, although rejected by the circuit when balanced, contributes to the output voltage noise during other parts of the switching transient.

To simplify the analysis we break up the noise contributions into two piecewise constant regions of operation, as shown in figure 3. The tail current noise seen at the output is assumed zero while in balanced mode (I), and fully on during the unbalanced mode (II). The NMOS differential pair noise source contribution is approximately zero for the unbalanced mode[1], and is approximated as constant for the balanced region of operation. The contribution of the triode-region PMOS noise sources is nearly constant for both. To find the voltage noise at the output as a function the current noise sources, analysis is carried out in the time domain using autocorrelation functions and convolution. The result is a new noise contribution factor which captures the time dependence of the output voltage noise.

$$\xi = \sqrt{1 + \frac{2}{3}a_v\left(1 - e^{-t/\tau}\right) + \frac{2\sqrt{2}}{3}a_v e^{-t/\tau}} \qquad (4)$$

This equation shows that as the circuit begins switching (t=0), the diff. pair noise contribution (second term) rises exponentially from

1. In the unbalanced mode, one side of the differential pair is off, and the other side's contribution is reduced by emitter degeneration.

zero to its equilibrium value considered previously. The third component of the noise contribution factor is due to the tail current noise source and decays exponentially from its equilibrium value in the unbalanced mode (II) towards zero when switching begins. It can be shown that the time constant τ is approximately equal to the time delay of the stage, in which case the exponentials in this expression reduce to constants at the time of interest, t_d. This means that the noise contribution factor is relatively insensitive to most design parameters except gain.

Inter-stage Interaction

Figure 4 shows that for a typical CMOS inverter chain the switching times of adjacent stages overlap and there are times when more than one stage is in the active region of amplification. In this case it is not sufficient to consider the noise contribution of a single inverter alone since noise from one inverter may be amplified and filtered by the next stage, contributing to the jitter in the subsequent stages in that manner. A better model is to consider two successive stages, and determine the voltage noise at the output of the second stage directly from the thermal noise current sources in the first stage.

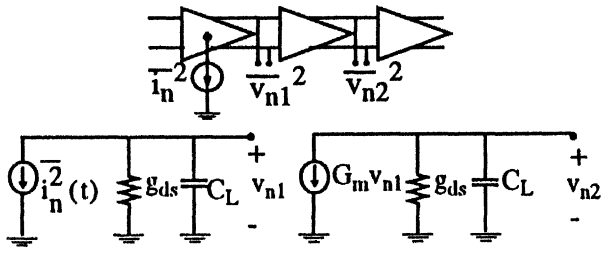


Figure 5. Extended circuit model for inter-stage interaction

Analysis for this case yields a slightly different noise contribution factor ξ than before, and an increase in the voltage noise variance by a factor of 1/2 $(a_v)^2$. With some re-arrangement, the new normalized timing jitter expression can be shown to be :

$$\frac{\Delta\tau_{1\,rms}}{t_d} = \sqrt{\frac{kT}{C_L}} \bullet \frac{1}{(V_{GS}-V_T)} \bullet \xi \qquad (5)$$

This means the normalized r.m.s. timing jitter is actually given by the ratio of the *kT/C* noise level to the gate bias voltage above threshold for the balanced state, (V_{GS} - V_T). If the second term is brought under the radical, then it is apparent that this fundamental timing error can also be expressed as the ratio of the thermal noise energy level to the electrical energy stored on the gate capacitance of the next stage.

IV. Simulations

A monte-carlo approach to transient noise analysis was taken to simulate the jitter performance of ring-oscillators in SPICE. This approach includes the effects of time-varying transconductances and inter-stage interaction. Figure 6 shows that, as expected, the normalized timing jitter improves with the square root of C_L. In the graph, C_L is scaled by changing the gate width. The gate width and current are scaled proportionally so as to keep (V_{GS}-V_T) constant and fix the delay per stage. Since the static power consumption is proportional to I_{SS}, jitter improves with the square root of power consumption, as indicated by the top axes.

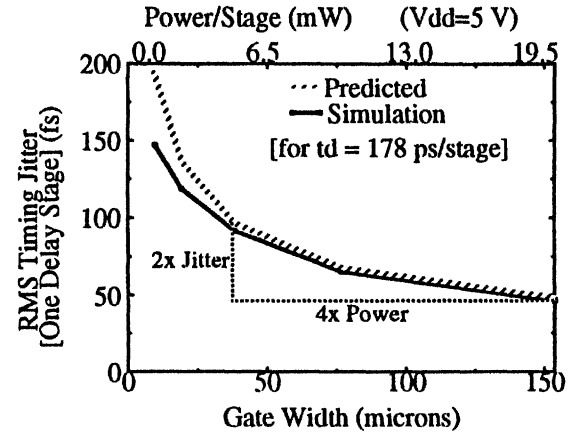


Figure 6. RMS timing jitter versus inverter size / power per stage

V. Design Implications

When designing a ring-oscillator, the parameter of interest is the jitter per cycle of oscillation. The analysis to this point has investigated the intrinsic jitter per delay stage, and we now extend these results to consider the jitter of the overall ring-oscillator. The jitter per cycle of oscillation can be used to determine the total PLL jitter for a ring-oscillator configured in a phase-locked-loop, and can also be used to predict the oscillator's phase noise spectrum.

Cycle-to-cycle jitter

Suppose the goal is to design a ring-oscillator with a fixed period, T_0, and minimal jitter. For an N-stage configuration the period of the oscillator is given by $2N \times t_d$, and the total jitter variance for once cycle of oscillation is given by $2N \times \Delta\tau_1^2$, provided noise sources in successive stages are independent. Using the results of the last section, the jitter per cycle of oscillation, or cycle-to-cycle jitter, can be shown to be

$$\overline{\Delta\tau_N^2} = \overline{\Delta\tau_1}^2 \times \frac{T_0}{t_d} = \frac{kT}{I_{SS}} \frac{a_v \xi^2}{(V_{GS}-V_T)} \times T_0 \qquad (6)$$

where the substitution $2N = T_0 / t_d$ is used so that the jitter can be expresses as function of T_0, rather than N.

To design for low jitter, $(V_{GS}-V_{TN})$ should chosen as large as possible. The inverter gain term, a_v, is the result of inter-stage amplification consideration. For designs where this is a factor (more true of CMOS than bipolar), this implies that a for a fixed delay and fixed current, the jitter improves with lower gain per stage. Inverter gain must be kept greater than one, however, for oscillation to occur. The noise contribution factor, ξ, is a weak function of most design parameters except gain. For many CMOS designs, a_v is kept in the range of 1.5-3, and ξ ranges from 1.3 to 1.9.

The main result of equation (7) is that with everything else fixed, the timing jitter variance improves linearly with an increase in supply current. Since power consumption depends on the quiescent current level, this implies, at least for the class of circuits considered here, a direct trade-off between power consumption and timing jitter.

Interestingly, the implications of equation (7) to *first order* do not change with changes in supply voltage, technology scaling, and configuration. If $(V_{GS}-V_{TN})$ is proportional to the supply voltage, then for a constant jitter, decreasing the supply voltage requires increasing the supply current by the same amount. This means that the power consumption stays the same. Scaling of the gate length gives access to higher speeds, but equation (7) shows that for a fixed T_0, the jitter is proportional to the current itself, and does not depend directly on the gate length. Velocity saturation effects have not been neglected, to first order, either, since no form for the current equation has been assumed. Another interesting result, is that the jitter variance does not depend on the exact configuration of the oscillator itself. Each of the configurations in figure 7, for instance, have the same period if inverters with the same $(V_{GS} - V_T)$ and I_{SS} are used; but by equation (7), they also have the same jitter variance. Power is minimized, in this case, by using the configuration with as few delay stages as necessary.

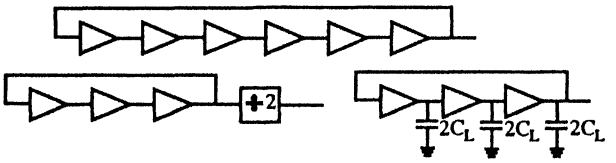


Figure 7. Multiple oscillator configurations with same period

Equation (7) shows that cycle-to-cycle jitter variance is proportional to the period, T_0, itself. A better figure of merit, however, is the jitter normalized to the period of oscillation. The r.m.s. jitter as a percentage of the output period $(\Delta\tau_{Nrms}/T_0)$ actually varies as $1/\sqrt{T_0}$. This implies that higher frequency oscillators will have poorer jitter for the same power consumption.

Overall ring-oscillator PLL Jitter

In a ring-oscillator the variance of the timing error relative to a fixed reference transition, grows with each successive period of oscillation unless the oscillator is configured in a PLL. Analysis in [5] [6] shows that the total r.m.s. jitter when locked in a PLL will be α times the cycle-to-cycle jitter, where α is a multiplying factor which is inversely proportional to the bandwidth of the PLL. A wider bandwidth PLL corrects timing errors more quickly, resulting in a smaller overall jitter and an earlier roll-off point in figure 8. The minimum practical value of α is limited by clock feed-through, and other PLL design issues. α is typically in the range of 10-100. For a delay-locked-loop [4], jitter is not accumulated between periods, and α is effectively equal to one.

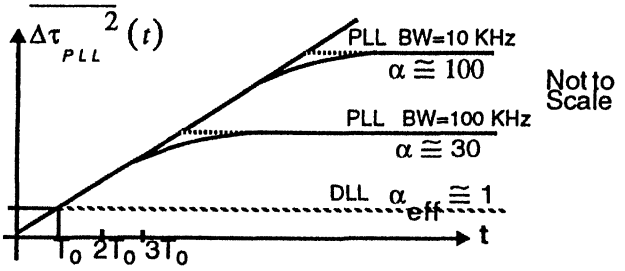


Figure 8. Jitter variance vs. time for reference transition at t=0

Ring-oscillator phase noise

Phase noise is an important figure of merit for oscillators used in RF applications. An ideal oscillator spectrum is an impulse in the frequency domain at the carrier frequency f_0. A real oscillator has its energy spread over a narrow bandwidth around f_0, and phase noise is a measure of the noise power in a 1-Hz bandwidth at a frequency f_m, offset from the carrier, relative to the total power of the oscillator. Phase noise can be determined from ring-oscillator timing jitter in a few ways. One way is to note that the accumulated phase error is a Wiener process, in which case its power spectrum can be shown to be a lorentzian. The other is to relate the spectral density of the normalized frequency fluctuations to the instantaneous error in the period of oscillation, $(\Delta\tau / T_0)$, and then use the relationships in [4] to arrive at the spectral density of phase fluctuations. With some re-arrangement of terms, the phase noise spectrum can be shown to have the following form :

$$S_\Phi(f_m) \cong \frac{f_0}{f_m^2}\left(\frac{\Delta\tau_{rms}}{T_0}\right)^2 = \left(\frac{f_0}{f_m}\right)^2 \bullet \frac{kTa_v\xi^2}{I_{SS}(V_{GS}-V_T)} \quad (7)$$

The phase noise is related to the ratio of the offset frequency to the oscillator frequency, and falls off at a rate of *20 dB /* decade for higher offsets. For a given frequency offset from carrier, the phase noise improves with higher inverter cell supply currents. Therefore phase noise is expected to improve with power consumption at a rate of 10 *dB /* decade.

VI. Conclusions

This paper has analyzed the relationship between design parameters of ECL type inverter cells and the resulting thermal-noise-induced jitter. The jitter per stage was shown to depend on the ratio of the *kT/C* noise level to the $(V_{GS}-V_T)$ bias point. The cycle-to-cycle jitter of a ring oscillator was shown to improve with larger bias currents and the normalized jitter was proportional to $1/\sqrt{T_0}$, indicating inherently higher jitter for higher speeds. Finally, the overall jitter in a PLL and the phase-noise of a ring oscillator were determined from the cycle to cycle jitter, and phase noise was predicted to improve at a rate of 10 dB / decade increase in power consumption.

VII. References

[1] B. Kim, D.N. Helman, P.R. Gray, "A 30-MHz Hybrid Analog/Digital Clock Recovery Circuit in 2 mm CMOS," *IEEE Journal of Solid-State Circuits*, Vol. 25, No. 6, pp. 1385-1394, December 1990.

[2] M. Johnson, E.L. Hudson, "A Variable Delay Line PLL for CPU-coprocessor Synchronization," *IEEE Journal of Solid-State Circuits*, Vol. 23, No. 5, pp. 1218-1223, October 1988.

[3] A. A. Abidi, R. G. Meyer, "Noise in Relaxation Oscillators," *IEEE Journal of Solid-State Circuits*, Vol. 18, No. 6, pp. 794-802, December 1983.

[4] P. R. Gray and R. G. Meyer, "Analysis and Design of Analog Integrated Circuits", John Wiley & Sons, NY, 1984

[5] B. Kim, T. C. Weigandt, P. R. Gray, "PLL/DLL System Noise Analysis for Low Jitter Clock Synthesizer Design," *ISCAS '94 Proceedings*, June 1994.

[6] J.A. McNeill, "A 200 mW, 155 MHz, Phase-Locked-Loop with Low Jitter VCO", *ISCAS '94 Proceedings*, June 1994.

Analysis, Modeling, and Simulation of Phase Noise in Monolithic Voltage-Controlled Oscillators

Behzad Razavi
AT&T Bell Laboratories, Holmdel, NJ07733

Abstract

In this paper, the phase noise of monolithic voltage-controlled oscillators is formulated with the aid of a linearized model. A new definition of Q is introduced and three mechanisms leading to phase noise are identified. A simulation technique using sinusoidal noise components is also described.

I. Introduction

Low-noise voltage-controlled oscillators (VCOs) are an integral part of high-performance phase-locked systems such as frequency synthesizers used in wireless tranceivers. While most present implementations of RF VCOs employ external inductors to achieve a low phase noise, the trend towards large-scale integration and low cost mandates monolithic solutions. For example, ring oscillators have been proposed as a suitable candidate [1], but their phase noise is generally known to be "high."

This paper describes an approach to analyzing, modeling, and simulating the phase noise of monolithic VCOs, with particular attention to CMOS ring oscillators. Following an analysis of a general oscillatory system and a new definition of the quality factor, Q, we employ a linearized model of ring oscillators to predict the phase noise with reasonable accuracy.

II. General Oscillatory System

Consider the linear feedback system depicted in Fig. 1. The system oscillates at $\omega = \omega_0$ if the transfer function

$$\frac{Y}{X}(j\omega) = \frac{H(j\omega)}{1 + H(j\omega)} \tag{1}$$

goes to infinity at this frequency, e.g., $H(j\omega_0) = -1$. (We call ω_0 the "carrier frequency.") The phase noise

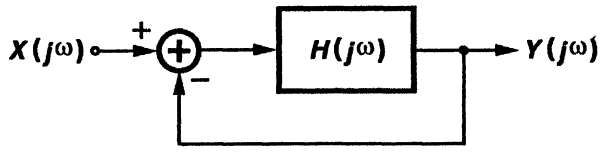


Fig. 1. General oscillatory system.

observed at the output is a function of: 1) sources of noise in the circuit, and 2) how much the feedback system rejects (or amplifies) various noise components. Modeling each source of noise as an input, $X(j\omega)$, to the system, we first quantify the latter effect for frequencies close to ω_0. If $\omega = \omega_0 + \Delta\omega$, then $H(j\omega) \approx H(j\omega_0) + \Delta\omega dH/d\omega$ and the noise transfer function is

$$\frac{Y}{X}[j(\omega_0 + \Delta\omega)] = \frac{H(j\omega_0) + \Delta\omega\dfrac{dH}{d\omega}}{1 + H(j\omega_0) + \Delta\omega\dfrac{dH}{d\omega}}. \tag{2}$$

Since $H(j\omega_0) = -1$ and for most practical cases $|\Delta\omega dH/d\omega| \ll 1$, (2) reduces to

$$\frac{Y}{X}[j(\omega_0 + \Delta\omega)] \approx \frac{-1}{\Delta\omega\dfrac{dH}{d\omega}}. \tag{3}$$

This equation indicates that a noise component at $\omega = \omega_0 + \Delta\omega$ is multiplied by $-(\Delta\omega dH/d\omega)^{-1}$ when it appears at the output of the oscillator. In other words, the noise power spectral density is shaped by

$$\left|\frac{Y}{X}[j(\omega_0 + \Delta\omega)]\right|^2 = \frac{1}{(\Delta\omega)^2\left|\dfrac{dH}{d\omega}\right|^2}. \tag{4}$$

This is illustrated in Fig. 2. As we will see later, (4) assumes a simple form for ring oscillators.

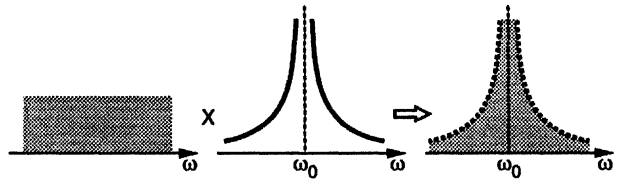


Fig. 2. Noise shaping in oscillators.

To gain more insight, let $H(j\omega) = A(\omega)\exp[j\Phi(\omega)]$. Thus, (4) can be written as

$$\left|\frac{Y}{X}[j(\omega_0 + \Delta\omega)]\right|^2 = \frac{1}{(\Delta\omega)^2[(\dfrac{dA}{d\omega})^2 + (\dfrac{d\Phi}{d\omega})^2]}. \tag{5}$$

Reprinted from *Proc. CICC*, pp. 323-326, May 1995.

We define the open-loop Q as

$$Q = \frac{\omega_0}{2}\sqrt{(\frac{dA}{d\omega})^2 + (\frac{d\Phi}{d\omega})^2}. \quad (6)$$

The *open-loop Q* is a measure of how much the *closed-loop* system opposes variations in the frequency, as is better seen when (5) and (6) are combined:

$$\left|\frac{Y}{X}[j(\omega_0 + \Delta\omega)]\right|^2 = \frac{1}{4Q^2}(\frac{\omega_0}{\Delta\omega})^2, \quad (7)$$

a familiar form previously derived for simple LC oscillators [2]. It is interesting to note that in an LC tank at resonance, $dA/d\omega = 0$ and (6) reduces to the conventional definition of Q: $\omega_0(d\Phi/d\omega)/2$. As will be seen later, in ring oscillators $dA/d\omega$ and $d\Phi/d\omega$ are of the same order and only the more general definition proposed in (6) can be used.

III. Linearized Model of CMOS VCOs

Submicron CMOS technologies have demonstrated potential for RF phase-locked systems [3]. Fig. 3 shows a fully differential 3-stage ring oscillator suitable for such applications. To calculate the phase noise, we model the signal path in the VCO with a linearized (single-ended) circuit as in Fig. 4. Here, R and C represent the output resistance and the load capacitance of each stage, respectively, ($R \approx 1/g_{m3} = 1/g_{m4}$), and G_mR is the gain required for steady oscillations. The noise of each differential pair and its load devices is modeled as current sources I_{n1}-I_{n3}, injected onto nodes 1-3, respectively.

Before calculating the noise transfer function, we note that the circuit of Fig. 4 oscillates if, at ω_0, each stage has unity voltage gain and 120° of phase shift. Thus, $\omega_0 = \sqrt{3}/(RC)$, and $G_mR = 2$, and the open-loop transfer function is given by

$$H(j\omega) = \frac{-8}{(1 + j\sqrt{3}\frac{\omega}{\omega_0})^3}. \quad (8)$$

Therefore, $|dA/d\omega| = 9/(4\omega_0)$ and $|d\Phi/d\omega| = 3\sqrt{3}/(4\omega_0)$. It follows from (5) that if a noise current I_{n1} is injected onto node 1 in the oscillator of Fig. 4, then its power spectrum is shaped by

$$\left|\frac{V_1}{I_{n1}}[j(\omega_0 + \Delta\omega)]\right|^2 = \frac{R^2}{27}(\frac{\omega_0}{\Delta\omega})^2. \quad (9)$$

This equation is the key to predicting various phase noise components in the ring oscillator.

IV. Additive and Multiplicative Noise

Modeling the ring oscillator of Fig. 3 with the linearized circuit of Fig. 4 entails a number of issues. While the stages in Fig. 3 turn off for part of the period, the linearized model exhibits no such behavior. Furthermore, the dependence of the delay upon the tail current I_{SS} is not reflected in Fig. 4. In order to incorporate these effects, we identify three types of phase noise.

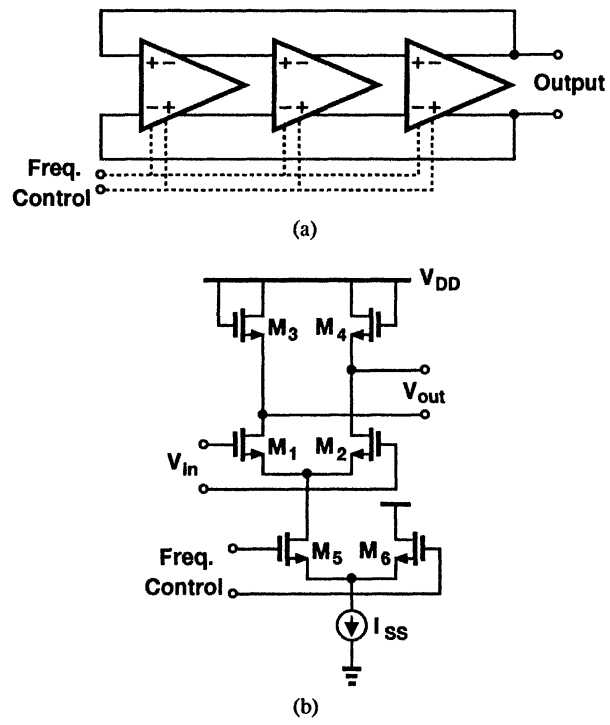


Fig. 3. CMOS VCO. (a) Block diagram, (b) implementation of one stage.

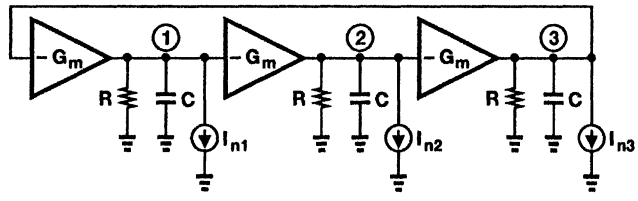


Fig. 4. Linearized model of the VCO.

A. Additive Noise

Additive noise consists of components that are directly added to the output as shown in Fig. 2 and formulated by (4) and (9).

To calculate the additive phase noise in Fig. 4 with the aid of (9), we note that for $\omega \approx \omega_0$ the voltage gain in each stage is close to unity, and the total output phase noise power density due to I_{n1}-I_{n3} is

$$|V_{1tot}[j(\omega_0 + \Delta\omega)]|^2 = \frac{R^2}{9}(\frac{\omega_0}{\Delta\omega})^2\overline{I_n^2}, \quad (10)$$

where it is assumed $\overline{I_{n1}^2} = \overline{I_{n2}^2} = \overline{I_{n3}^2} = \overline{I_n^2}$. For the differential stage of Fig. 3(b), the noise current per unit

bandwidth is equal to $\overline{I_n^2} = 8kT(g_{m1} + g_{m3})/3 \approx 8kT/R$. Thus,

$$|V_{1tot}[j(\omega_0 + \Delta\omega)]|^2 = 8kT\frac{R}{9}(\frac{\omega_0}{\Delta\omega})^2. \qquad (11)$$

Additive phase noise is predicted by the linearized model with high accuracy if the stages in the ring oscillator turn off for only a small portion of the period. In a 3-stage CMOS oscillator designed for the 900-MHz range, the differential pairs turn off for less than 10% of the period. Furthermore, at zero-crossing points —where most of the phase noise is generated— all stages are on. Therefore, the linearized model emulates the CMOS oscillator with reasonable accuracy. A simple measure of this accuracy is the error in the oscillation frequency of the model with respect to that of the actual circuit. This error remains below 10% for a 3-stage ring and 20% for a 4-stage ring.

Since additive noise is shaped according to (11), its effect is significant only for components close to the carrier frequency.

B. High-Frequency Multiplicative Noise

The nonlinearity in the differential stages of Fig. 3, especially as they turn off, causes noise components to be multiplied by the carrier (and by each other). If the input/output characteristic of each stage is expressed as $V_{out} = \alpha_1 V_{in} + \alpha_2 V_{in}^2 + \alpha_3 V_{in}^3$, then for an input consisting of the carrier and a noise component, e.g., $V_{in}(t) = A_0 \cos\omega_0 t + A_n \cos\omega_n t$, the output exhibits the following important components:

$V_{out1}(t) \propto \alpha_2 A_0 A_n \cos(\omega_0 \pm \omega_n)t$,
$V_{out2}(t) \propto \alpha_3 A_0 A_n^2 \cos(\omega_0 - 2\omega_n)t$,
$V_{out3}(t) \propto \alpha_3 A_0^2 A_n \cos(2\omega_0 - \omega_n)t$.

Note that $V_{out1}(t)$ appears in band if ω_n is small, i.e., if it is a *low-frequency* component, but in a fully differential configuration, $V_{out1}(t) = 0$ because $\alpha_2 = 0$. Also, $V_{out2}(t)$ is negligible because $A_n \ll A_0$, leaving $V_{out3}(t)$ as the only significant cross-product.

Simulations indicate that the feedback in the oscillator yields approximately equal magnitudes for $V_{out3}(t)$ and the original component at ω_n. Thus, the nonlinearity folds all the noise components below ω_0 to the region above and vice versa, effectively doubling the noise power predicted by (11). Such components are significant if they are close to ω_0 and are herein called high-frequency multiplicative noise. This phenomenon is illustrated in Fig. 5.

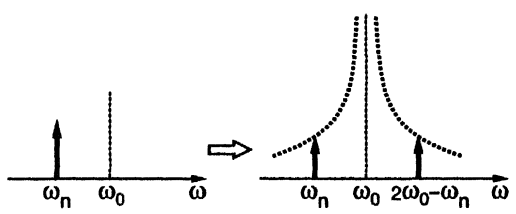


Fig. 5. High-frequency multiplicative noise.

C. Low-Frequency Multiplicative Noise

Since the frequency of oscillation in Fig. 3 is a function of the tail current in each differential pair, noise components in this current modulate the frequency, thereby contributing phase noise. Depicted in Fig. 6, this effect can be significant because, in CMOS oscillators, ω_0 must be adjustable by approximately ±20% to compensate for process variations, thus making the frequency quite sensitive to noise in the tail current.

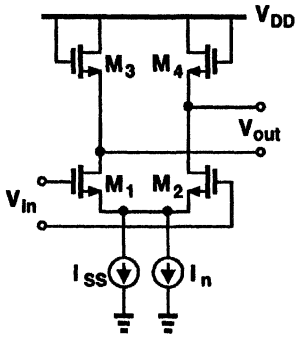


Fig. 6. Carrier modulation by tail noise current.

To quantify this phenomenon, we note that since variations in the tail current modulate the impedance of M_3 and M_4, the resistor R in the linearized circuit can be modeled as the sum of a constant term and a small current-dependent term. It can be proved that if the noise current $I_n = I_{n0}\cos\omega_n t$, then two current components described by $\frac{\sqrt{3}}{4} I_{n0}\cos(\omega_0 \pm \omega_n)t$ appear in the signal path and hence are multiplied by the transfer function in (9). Thus,

$$|V_n|^2 = \frac{R^2}{48}(\frac{\omega_0}{\Delta\omega})^2 |I_n|^2, \qquad (12)$$

where $\Delta\omega = \omega_n$. This mechanism is illustrated in Fig. 7.

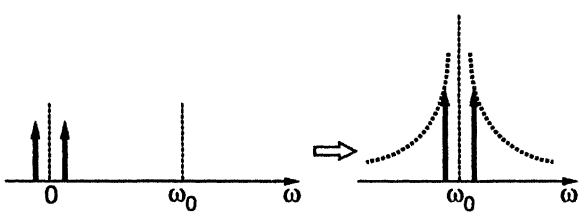


Fig. 7. Low-frequency multiplicative noise.

It is seen that modulation of the carrier brings the low frequency noise components of the tail current to the band around ω_0. Thus, flicker noise in I_n becomes particularly important.

In the differential stage of Fig. 3(b), two sources of low-frequency multiplicative noise can be identified: noise in

I_{SS} and noise in M_5 and M_6. For comparable device size, these two sources are of the same order and must be both taken into account.

V. Simulation

To simulate the phase noise in the oscillator and verify the accuracy of the above derivations, we use a small sinusoidal "noise" current that is injected onto different nodes of the circuit. This approach is justified by the fact that random noise can be expressed as a Fourier series of sinusoids with random phase [4].

Designed to operate at 970 MHz, each oscillator is simulated in the time domain for 2 μsec with 30-psec steps and the resulting output waveform is processed by Matlab to obtain the spectrum. Shown in Fig. 8 is the output spectrum of the linearized model in response to a sinusoidal current with 2-nA amplitude at 980 MHz. The vertical axis represents $10 \log V_{rms}^2$. The observed magnitude of the 980-MHz component is in exact agreement with (10).

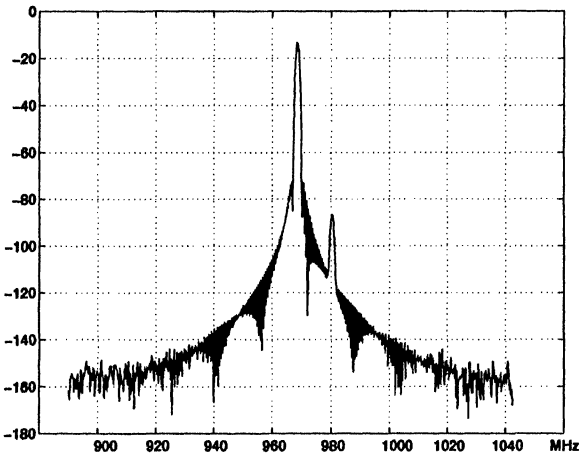


Fig. 8. Simulated spectrum of linearized model.

A similar test on the CMOS oscillator of Fig. 3 yields the spectrum in Fig. 9. Note that the component at 980 MHz has approximately the same magnitude as that in Fig. 8, indicating that the linearized model is indeed an accurate representation. As explained in Section IV, the 960-MHz component originates from third-order mixing of the carrier and the 980-MHz component and essentially doubles the phase noise.

Low-frequency multiplicative noise is simulated by modulating the tail current of one stage in the CMOS oscillator with a 2-nA 10-MHz sinusoid. The resulting sideband magnitudes, shown in Fig. 10, closely agree with (12).

VI. Conclusion

Analysis of a general oscillatory system leads to a linearized model of ring oscillators that predicts the phase noise with reasonable accuracy. Quantified in this paper are three mechanisms, namely, additive noise, high-frequency multiplicative noise, and low-frequency multiplicative noise, that contribute to the phase noise of oscillators. Simulations confirm the accuracy of the model and the derivations.

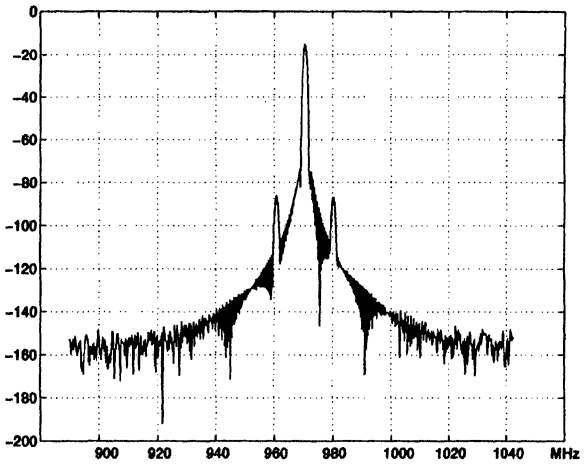


Fig. 9. Simulated spectrum of CMOS VCO.

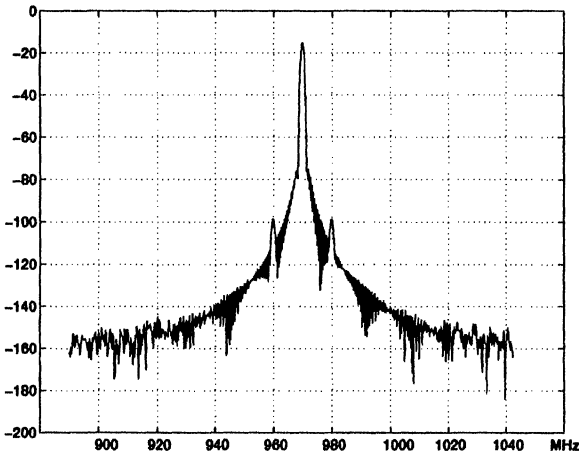


Fig. 10. Simulated spectrum of CMOS VCO with tail current noise.

References

[1] A. A. Abidi, "Radio-Frequency Integrated Circuits for Portable Communications," *Proc. CICC*, pp. 151-158, May 1994.

[2] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. IEEE*, pp. 329-330, Feb. 1966.

[3] B. Razavi, "A 3-GHz 25-mW CMOS Phase-Locked Loop," *VLSI Circuits Symp. Dig. Tech. Papers*, pp. 131-132, June 1994.

[4] S. O. Rice, "Mathematical Analysis of Random Noise," *Bell System Tech. J.*, pp. 282-332, July 1944, and pp. 46-156, Jan. 1945.