THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 9, November 1978
Printed in U.S.A.

Optimum Reception of Digital Data Signals in the Presence of Timing-Phase Hits

By D. D. FALCONER and R. D. GITLIN

(Manuscript received March 29, 1978)

Protection switching of digital radio channels results in a timing-phase discontinuity and occasional long error bursts in the demodulated data signal. Using an idealized mathematical model, we derive maximum likelihood receivers which rapidly track such delay hits, whether or not a timing-pilot tone is used. When the receiver is at a different physical location from the switch, the tracking algorithm must also sense the occurrence of a switch. A dual-mode, data-directed structure is revealed as being optimum; a narrowband tracking loop is used for steady-state operation, while a wideband tracking loop provides rapid recovery from the timing transient. An error-sensing nonlinearity, which incorporates hysteresis, inhibits erroneous noise-induced mode transitions. Oversampling of the demodulated data signal rapidly establishes a coarsely quantized, optimum sampling phase and permits the dual-mode tracking loop to operate in a data-directed manner. Data-directed operation permits greater loop bandwidths, since the data energy is not perceived as noise. Simulation of a digital data transmission system employing a dual-mode, data-directed, and coarse-quantized timing loop has demonstrated dramatic reduction in the length of error bursts following a protection switch. For example, at the data rate of 1.544 Mb/s, a conventional phase-locked loop with a 100-Hz bandwidth, when displaced a half-symbol interval by a delay hit, would typically sustain an error burst 15,000 bits in duration. When such a delay hit stresses the dual-mode timing loop, simulation has indicated error bursts on the order of 15 bits in duration.

I. INTRODUCTION

Some channels used for data transmission exhibit occasional abrupt changes in their absolute delay. For example, during severe fading, line-of-sight microwave facilities commonly switch signals from their regularly assigned channel to a protection channel which, owing to different filtering, cable lengths, etc., may impart a different delay. Such a switch, unknown to the receiver, changes the best phase with which a synchronous receiver should sample the incoming signal during each symbol interval. Until this new optimal timing phase is acquired by the receiver, data errors may proliferate if the delay change is a significant fraction of the symbol interval. The length of the succession of errors will be essentially inversely proportional to the bandwidth of the timing-recovery loop or filter.

The object of this investigation is to determine and analyze signal processing structures which rapidly respond to a sudden change in timing phase (a delay hit) while also providing accurate steady-state timing information when the protection channel is not required. Based upon an idealized channel model, we determine the maximum likelihood (optimum) receiver, and practically motivated approximations are made to provide realizable signal processors. The proposed receivers mediate the inherent conflict between using a narrowband timing recovery loop for steady-state operation, so that accurate and stable timing can be derived from the noisy received signal, and using a wideband loop to follow a timing-phase transient. As might be expected, the derived tracking loop is of the dual-mode variety; i.e., it automatically senses the state of the system (i.e., transient or steady state) and adjusts its structure accordingly. The exact form of the loop depends on the detailed manner in which the disturbances are modeled; yet it is demonstrated that the essential features of the signal processors are quite robust and have significant intuitive appeal.

In this study, our development is for an arbitrary protection-switched data communication system; however, specific simulation results and special emphasis are given to a 4-input level, Class-IV, partial response signaling format, such as used in the DUV system.[1]

In Section II, we describe the system model—principally the statistical mechanism for generating a "delay hit." The optimum receiver is described in Section III, and various suboptimum realizable structures are developed in Section IV. Digital implementation of these techniques in the partial response system is reported, via simulation, in Sections V and VI.

II. SYSTEM MODEL FOR TIMING-RECOVERY PROBLEM

In this section, we propose and develop a mathematical model for the data transmission system under consideration. Any attempt to exactly model the end-to-end data channel will be exceedingly tedious and probably fruitless, since the FM demodulator and phase-hit mechanism of most microwave facilities are highly nonlinear effects. Our approach is to isolate the major manifestations of a timing-phase discontinuity and background noise via a simple model, and then to apply maximum likelihood detection to obtain useful receivers. The validity of this approach is measured by simulation of the receiver over a real channel. We begin by writing the transmitted baseband data signal[2] as

$$s(t) = \sum_n a_n h(t - nT) + \rho_c \sin \frac{\pi}{T} t, \tag{1}$$

where $\{a_n\}$ is a sequence of independent multilevel symbols, $h(t)$ is a band-limited transmitted pulse, $1/T$ is the symbol rate, and ρ_c is the parameter which indicates the power in the pilot tone located at $1/2T$ Hz. The purpose of the pilot tone is to aid in providing the receiver timing phase and frequency. It will be assumed that the end-to-end pulse shaping used in the system is such that the desired sampling instants are $t = nT$. Whenever the pulse $h(t)$ possesses more than the minimum Nyquist bandwidth, it is convenient to rewrite (1) as

$$s(t) = \sum_n [a_n + \rho(-1)^n] h(t - nT), \tag{2}$$

where it is recognized that $\Sigma_n (-1)^n h(t - nT)$ is periodic with period $2T$, and ρ_c is the product of ρ and the energy in the pulse at $1/2T$ Hz. Since $h(t)$ is customarily band-limited to less than $1/T$ Hz, only the fundamental component of the signal $\Sigma_n (-1)^n h(t - nT)$ will be transmitted through the filter $h(t)$, thus the sinusoid may be represented by the alternating (dotting pattern) series. Indeed, in practice, a dotting pattern is frequently used to generate the tone. The transmitted signal may be rewritten as

$$s(t) = \sum_n c_n h(t - nT), \tag{3}$$

where

$$c_n = a_n + \rho(-1)^n. \tag{4}$$

Recall that partial response signals can be generated by either digital filtering of the independent data symbols, $\{a_n\}$, or by the use of special non-Nyquist pulse shapes. The receiver structures derived in the sequel will be discussed for both Nyquist and partial-response shaping. We now turn to the specific idealizations we will make to model the transmission path.

In the absence of any transmission distortion, the received baseband signal, $r(t)$, can be modeled as

$$r(t) = \mathcal{J}[s(t)] + \nu(t), \tag{5}$$

where $\mathcal{J}[s(t)]$ is the time-jittered signal and where the additive noise $\nu(t)$ will be taken as white Gaussian with spectral density N_0. For the purpose of analytical tractability, any instability in the timing phase will be modeled by representing the received signal as

$$r(t) = \sum_n c_n h(t - nT - \Delta_n) + \nu(t), \tag{6}$$

where Δ_n is a random process whose characteristics will be described below. In the above model, which is shown in Fig. 1, the phase of the pilot tone is presumed to be jittered at the discrete instants, $\{nT\}$, in the same manner as the phase of the data signal. This is accurate when the pilot tone is generated via the dotting pattern method, and the timing instabilities arise solely in the *transmitter* clock. Any timing-phase jitter, $\Delta(t)$, that occurs during transmission should properly be modeled by $r(t) = s(t - \Delta(t)) + \nu(t)$. However, this leads to analytical difficulty in characterizing the statistical nature of the random process $\Delta(t)$, as well as having to contend with jitter-induced amplitude modulation of the received signal. With this caveat in mind, we lump all sources of timing-phase jitter into the model given by (6). Of course, the utility of the above model will be measured by the performance of the derived receivers in the real-world environment.

The standard approach to the tracking of a slowly varying timing phase, and thus the timing frequency, is to use a narrowband filter centered about $1/2T$ Hz to extract the transmitted pilot tone. Of course, the bandwidth of this filter must be quite narrow to attenuate the in-band data-plus-noise energy. Extremely small effective bandwidths are

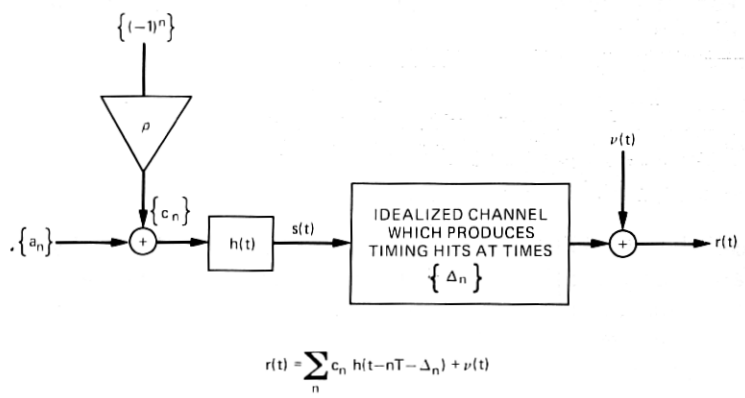

Fig. 1—Idealized model of timing hits in a digital data transmission system.

achieved in practice by following a narrowband zonal filter with a phase-locked loop (PLL) whose voltage-controlled oscillator (VCO) is tuned to $1/2T$ Hz. However, the narrow bandwidth of the PLL will preclude rapid tracking of any sudden change in the timing phase.

We digress momentarily to recall that the conventional envelope-derived timing scheme,[3,4] which does not utilize a pilot tone, will not provide timing information as the bandwidth of the system decreases to $1/2T$ Hz. It should be pointed out, however, that other nonquadratic techniques not requiring a pilot tone will provide a tone at the symbol rate for such minimum bandwidth systems; in particular, Saltzberg[5] has shown that the average of $\operatorname{sgn}[s(t)s(t-T)]$ provides a tone at $1/T$ Hz, and it is apparent that quartic[6] and similar operations will also provide the desired tone.

Returning to the formulation of our system model, we let the dynamic evolution of the timing jitter be given by the difference equation

$$\Delta_{n+1} = \Delta_n + w_n + \alpha_n v_n, \tag{7}$$

where $\{w_n\}$ and $\{v_n\}$ are sequences of mutually and self-independent Gaussian random variables with variances σ^2 and μ^2 respectively, and where $\mu^2 \gg \sigma^2$. The variable α_n is governed by

$$\alpha_n = \begin{cases} 0, & \text{with probability } 1 - p_0 \\ 1, & \text{with probability } p_0 \end{cases}, \tag{8}$$

where $0 \leq p_0 \ll 1$. Note that $\{\Delta_n\}$ is a Markov sequence where the mutually independent sequences $\{w_n\}$ and $\{v_n\}$ model the steady-state and the transient (delay-hit) modes, respectively. The initial value Δ_0 will be assumed to be uniformly distributed on $(0,T)$. Clearly, most of the time there are no delay hits; i.e., $\alpha_n = 0$, and the timing phase wanders about the correct value. Thus, p_0 is the probability that a timing discontinuity (which would follow a protection switch) occurs during a symbol interval. A typical sample path, or realization, of $\{\Delta_n\}$ is shown in Fig. 2, where the relative frequency of delay hits is determined by p_0. This simple two-mode model for the timing phase will be used to derive the optimum and various suboptimum data detectors, where an integral component of these detectors will be the timing-recovery loop. The steady-state jitter, w_n, is incorporated so that the timing loop will continually adjust the receiver's timing phase; note that this mechanism allows the receiver to presume nominal knowledge of the timing frequency, and any inaccuracy or drift in this quantity will be compensated for by the timing-phase tracking system.

With $\{\Delta_n\}$ specified by (7), the joint probability density function (pdf) of the $\{\Delta_n\}$ sequence is given by

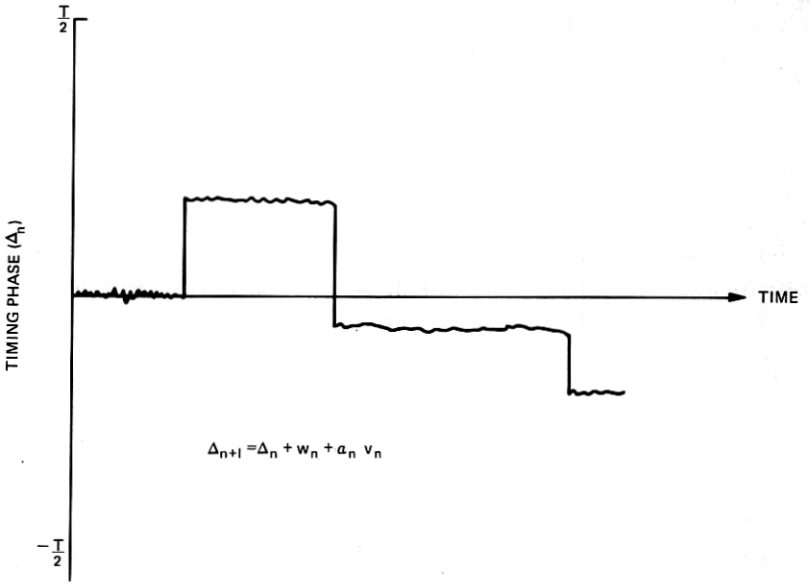

Fig. 2—A typical timing-phase trajectory ($\alpha_n = 1$ signifies that a timing hit has occurred).

$$
\begin{aligned}
p(\boldsymbol{\Delta}) \equiv p(\Delta_0, \Delta_1 \cdots, \Delta_n) &= p(\Delta_0) p(\Delta_1, \Delta_2, \cdots, \Delta_n \mid \Delta_0) \\
&= p(\Delta_0) p(\Delta_1 \mid \Delta_0) p(\Delta_2, \cdots, \Delta_n \mid \Delta_1, \Delta_0) \\
&= p(\Delta_0) p(\Delta_1 \mid \Delta_0) p(\Delta_2 \mid \Delta_0, \Delta_1) p(\Delta_3, \cdots, \Delta_n \mid \Delta_0, \Delta_1, \Delta_2) \\
&= p(\Delta_0) p(\Delta_1 \mid \Delta_0) p(\Delta_2 \mid \Delta_1) p(\Delta_3 \mid \Delta_2) \cdots p(\Delta_n \mid \Delta_{n-1}) \\
&= p(\Delta_0) \prod_{i=1}^{n} p(\Delta_i \mid \Delta_{i-1}), \qquad (9)
\end{aligned}
$$

where $\boldsymbol{\Delta} \equiv (\Delta_0, \Delta_1, \cdots, \Delta_n)$.

Using (7) and (8), the conditional density is given by the mixture

$$
\begin{aligned}
p(\Delta_i \mid \Delta_{i-1}) = \frac{(1 - p_0)}{\sqrt{2\pi}\,\sigma} e^{-(\Delta_i - \Delta_{i-1})^2/2\sigma^2} \\
+ \frac{p_0}{\sqrt{2\pi}\,(\mu^2 + \sigma^2)^{1/2}} e^{-(\Delta_i - \Delta_{i-1})^2/2(\mu^2 + \sigma^2)}, \qquad (10)
\end{aligned}
$$

and thus the joint pdf is given by (9) and (10) where Δ_0 is distributed uniformly over $(0,T)$.

Now that the system model has been specified, we turn to our professed goal of deriving optimum and suboptimum receivers.

III. OPTIMUM RECEPTION OF TIME-JITTERED PAM DATA SIGNALS

It is well known that the optimum (minimum probability of error) data-sequence detector *maximizes* the *a-posteriori* (MAP) probability density of the received signal with respect to the data sequence.* Thus the MAP receiver will supply the end-user with a sequence of decisions which maximize the probability density function $p[\{\hat{a}_m\}|r(t), 0 \leq t \leq \mathcal{T}]$, where the observation interval is $(0,\mathcal{T})$. By virtue of (4) and the properties of MAP receivers, we may estimate $\hat{a}_m$ via $\hat{a}_m = \hat{c}_m - \rho(-1)^m$, i.e., the estimates of $\{c_m\}$ and $\{a_m\}$ are related as above. Since all the data sequences $\{a_m\}$ are equiprobable, the relevant probability density can be obtained by averaging over the jittered timing phases $\{\Delta_i\}$, i.e., the MAP density is proportional to

$$p[r(t)|\{\hat{a}_m\}, 0 \leq t \leq \mathcal{T}] = \int p[r(t)|\{\hat{a}_m\}, \{\hat{\Delta}_m\}, 0 \leq t \leq \mathcal{T}]p[\hat{\Delta}]d\hat{\Delta}, \quad (11)$$

where the conditional density in the integrand is given by the standard formula for the probability density functional of a known signal in white Gaussian noise,

$$p[r(t)|\{\hat{a}_m\}, \{\hat{\Delta}_m\}, 0 \leq t \leq \mathcal{T}] = k \exp\left\{-\frac{1}{2N_0}\int_0^{\mathcal{T}} [r(t) - \sum_m \hat{c}_m h(t - mT - \hat{\Delta}_m)]^2 dt\right\}. \quad (12)$$

In the above equation, k is a constant independent of both $\{a_m\}$ *and* $\{\Delta_m\}$. The maximization of (11) with respect to $\{\hat{c}_m\}$, or equivalently $\{\hat{a}_m\}$, can be facilitated by writing (11) as

$$p[r(t)|\{\hat{a}_m\}, 0 \leq t \leq \mathcal{T}] = k \int d\hat{\Delta} \exp\left\{-\frac{1}{2N_0}\left[\int_0^{\mathcal{T}} [r(t) - \sum_m \hat{c}_m h(t - mT - \hat{\Delta}_m)]^2 dt - 2N_0 \ln p(\hat{\Delta})\right]\right\}. \quad (13)$$

It can be shown that, in a high signal-to-noise-ratio environment [i.e., as $N_0 \to 0$], the above integral with respect to $\hat{\Delta}$ can be replaced by the maximum value of the integrand, i.e., as $N_0 \to 0$

$$p[r(t)|\{\hat{a}_m\}, 0 \leq t \leq \mathcal{T}] \sim \exp\left\{-\frac{1}{2N_0}\left[\int_0^{\mathcal{T}} [r(t) - \sum_m \hat{c}_m h(t - mT - \hat{\Delta}_m^*)]^2 dt - 2N_0 \ln p(\hat{\Delta}^*)\right]\right\}, \quad (14)$$

where $\{\hat{\Delta}_n^*\}$ is the maximizing sequence. Thus, under the asymptotic

* The bit-optimum detector has been shown to be asymptotically approximated (at high signal-to-noise ratio) in performance by the optimum sequence detector (Ref. 7).

condition described above, the optimum receiver computes the *joint* minimum* of

$$\Lambda[r(t)|\{\hat{a}_m\}, \{\hat{\Delta}_m\}, 0 \le t \le \mathcal{T}] \equiv -\ln p[r(t)|\{\hat{a}_m\}, \{\hat{\Delta}_m\}, 0 \le t \le \mathcal{T}] - 2N_0 \ln p(\hat{\Delta}) \quad (15a)$$

$$= \int_o^{\mathcal{T}} \left[r(t) - \sum_m \hat{c}_m h(t - mT - \hat{\Delta}_m) \right]^2 dt - 2N_0 \left[\sum_{m=0}^{\infty} \ln p(\hat{\Delta}_m | \hat{\Delta}_{m-1}) + \ln p(\Delta_0) \right] \quad (15b)$$

$$\equiv \Lambda_1[r(t)|\{\hat{a}_m\}, \{\hat{\Delta}_m\}] + \Lambda_2[\{\hat{\Delta}_m\}], \quad (15c)$$

where $\Lambda_1[\cdot]$ and $\Lambda_2[\cdot]$ are defined in the obvious manner from (15b). For convenience in notation, we drop the $\hat{a}_m$ and $\hat{\Delta}_m$ symbols in favor of a_m and Δ_m whenever there is no possibility of confusion. We also adopt the notational shorthand

$$\ell[\{a_m\}, \{\Delta_m\}] \equiv \Lambda[r(t)|\{a_m\}, \{\Delta_m\}, 0 \le t \le \mathcal{T}], \quad (16a)$$

$$\equiv \ell_1[\{a_m\}, \{\Delta_m\}] + \ell_2[\{\Delta_m\}], \quad (16b)$$

where the ℓ_i corresponds to the appropriate Λ_i $(i = 1, 2)$ in (15c). Thus, our task is to jointly minimize $\ell[\{a_m\}, \{\Delta_m\}]$ with respect to the discrete-valued variables $\{a_m\}$ and the continuous-range variables $\{\Delta_m\}$. Since $\int_0^{\mathcal{T}} r^2(t)dt$ is independent of $\{c_m\}$ and $\{\Delta_m\}$, the relevant portion of $\ell_1[\{a_m\}, \{\Delta_m\}]$ is given by

$$\ell_1[\{a_m\}, \{\Delta_m\}] = -2 \sum_m c_m z(mT + \Delta_m) + \sum_m \sum_k c_m c_k g((m-k)T + \Delta_k - \Delta_m), \quad (17)$$

where the matched-filter output $z(t)$ is given by

$$z(t) = \int_{-\infty}^{\infty} h(t' - t) r(t') dt' \quad (18)$$

and

$$g(t) = \int_{-\infty}^{\infty} h(t') h(t + t') dt' \quad (19)$$

is the channel correlation function. Thus the sufficient statistics are the set of matched-filter output samples $\{z(mT + \Delta_m)\}$, where the sampling phases $\{\Delta_m\}$ are still to be determined. The other component of the likelihood is given by

* It should be clear from (15b) and (10) that, in the absence of a noise or timing-phase hit, $\hat{\Delta}_n^* \to \Delta_n$ and $\hat{a}_n \to a_n$; i.e., the estimates tend to the true parameter values.

$$\ell_2[\{\Delta_m\}] = -2N_0 \ln p(\Delta_0) - 2N_0 \sum_{m=1} \ln \left\{ \frac{(1-p_0)}{\sqrt{2\pi}\,\sigma} e^{-(\Delta_m - \Delta_{m-1})^2/2\sigma^2} + \frac{p_0}{\sqrt{2\pi}\,(\mu^2+\sigma^2)^{1/2}} e^{-(\Delta_m-\Delta_{m-1})^2/2(\mu^2+\sigma^2)} \right\}, \quad (20)$$

and the optimum receiver minimizes $\ell = \ell_1 + \ell_2$, where ℓ_1 and ℓ_2 are given by (17) and (20), respectively, with respect to $\{a_m\}$ and $\{\Delta_m\}$. Optimization with respect to $\{\Delta_m\}$ is via differentiation and gives

$$(\Delta_1 - \Delta_0)G[\Delta_1 - \Delta_0] + \frac{\partial}{\partial \Delta_0} \ell_1[\{a_m\}, \{\Delta_m\}] = 0 \quad (21)$$

and

$$(\Delta_{k+1} - \Delta_k)G[\Delta_{k+1} - \Delta_k] - (\Delta_k - \Delta_{k-1})G[\Delta_k - \Delta_{k-1}] + \frac{\partial}{\partial \Delta_k} \ell_1[\{a_m\}, \{\Delta_m\}] = 0, \qquad k = 1, 2, \cdots \quad (22)$$

where the function $G[\cdot]$ is defined by*

$$G[x] = \left[\frac{\frac{(1-p_0)}{\sigma(\sigma^2/2N_0)} \exp\{-x^2/2\sigma^2\} + \frac{p_0}{\mu(\mu^2/2N_0)} \exp\{-x^2/2\mu^2\}}{\frac{1-p_0}{\sigma} \exp\{-x^2/2\sigma^2\} + \frac{p_0}{\mu} \exp\{-x^2/2\mu^2\}} \right]. \quad (23)$$

As we see in the next section, the nature of $G[\ ]$ will impart a dual-mode character to the various tracking loops described in the sequel. It is convenient to define the weighted differential-epoch

$$\eta_k \equiv (\Delta_k - \Delta_{k-1})G[\Delta_k - \Delta_{k-1}], \quad (24)$$

and (21) and (22) can thus be written as

$$\eta_{k+1} = \eta_k - \frac{\partial \ell_1}{\partial \Delta_k}[\{a_m\}, \{\Delta_m\}] \qquad k = 0, 1, 2, \cdots, \quad (25)$$

where $\eta_0 = 0$.

Several difficulties associated with the "iteration" prescribed by (25) preclude incorporation in a realistic detector: (*i*) as already mentioned, Δ_0 is unknown, (*ii*) as it stands, the optimization over Δ_k is for a *given* set of $\{a_m\}$, (*iii*) from (17) it is clear that $\partial \ell_1/\partial \Delta_k$ depends on all the $\{a_m\}$ and $\{\Delta_m\}$, and (*iv*) optimization of (17) with respect to the $\{a_m\}$ requires a Viterbi-related dynamic programming algorithm.[8] (The state size is

* We have assumed that $\mu \gg \sigma$ so that $\mu^2 + \sigma^2 \approx \mu^2$. The detailed nature of the function $G[\cdot]$ is discussed in the next section.

somewhat ambiguous, since the presumably finite memory of $g(t)$ is enhanced by the arbitrarily large size of $\Delta_k - \Delta_m$.)

Because of the above factors, the level of complexity associated with the optimum receiver is prohibitive, and thus even for our simplified model we must resort to suboptimum reception. In a sense, this is not surprising since the maximum likelihood receiver has at its disposal the entire observation record, and it is only in special cases that the optimum procedure can be implemented in a sequential manner.

IV. SUBOPTIMUM RECEPTION

In this section, we indicate several reasonable receivers suggested by the optimum receiver of the previous section.

4.1 Data-directed receiver

Our approach to deriving a useful suboptimum receiver is to remove, via approximation, the difficulties associated with implementing the optimum receiver—the principal simplification we will make is to take a decision-directed approach. We begin by noting that (25) would be a practical and realizable recursion if: (*i*) $\partial \ell_1 / \partial \hat{\Delta}_k$ depended only on $\hat{\Delta}_k$ and $\hat{a}_k$, and (*ii*) the optimum value of $\hat{a}_k$ depends only on $z(kT - \hat{\Delta}_k)$ and $\hat{\Delta}_k$. With these desiderata in mind, we note from (17) that

$$\frac{\partial \ell_1}{\partial \hat{\Delta}_k}[\{\hat{a}_m\}, \{\hat{\Delta}_m\}] = -2c_k \dot{z}(kT + \hat{\Delta}_k) - \hat{c}_k \times \sum_{n \neq k} \hat{c}_n [\dot{g}((k-n)T - \hat{\Delta}_k + \hat{\Delta}_n) - \dot{g}((n-k)T + \hat{\Delta}_n - \hat{\Delta}_k)] \quad (26)$$

where the "dot" indicates the time derivative. The first concession we make to realizability is to neglect the second term in (26). Note that if a tone is not transmitted and the data levels are uncorrelated, then the expected value of this term is zero. We realize, of course, that this term would make a contribution whenever a timing-phase hit occurs; however, we are relying on the $\ell_2[\cdot]$ component of the likelihood to provide the dominant indication of this event. With this approximation, (25) reduces to

$$\eta_{k+1} = \eta_k + 2\hat{c}_k \dot{z}(kT + \Delta_k), \quad (27)$$

and from (24) we have

$$\eta_{k+1} = (\Delta_{k+1} - \Delta_k) G[(\Delta_{k+1} - \Delta_k)].$$

The value of Δ_{k+1} may be generated from η_{k+1} and Δ_k via the inverse relation

$$\Delta_{k+1} = \Delta_k + F^{-1}[\eta_{k+1}], \quad (28)$$

where if $\eta = xG[x] \equiv F[x]$, then $F^{-1}[\ ]$ is defined by

$$x \equiv F^{-1}[\eta]. \quad (29)$$

Thus (27) and (28) provide a second-order system of iterative equations for generating the desired estimates of $\{\Delta_k\}$ provided: (*i*) that either the sequence $\{c_k\}$ is known or reliable decisions are available, and (*ii*) that the initial phase estimate Δ_0 is known. These equations may be viewed as a second-order, discrete-time phase-locked loop (PLL) with a nonlinearity $F^{-1}[\ ]$ necessitated by the dual-mode nature of the timing phase.

The function $F^{-1}[x]$ is plotted in Fig. 3 for $p_0 \approx 0$ and $\sigma^2 \ll \mu^2$. Note that this function is odd, exhibits hysteresis, and is multivalued over a certain range, and as shown in Fig. 3, $F^{-1}[\eta]$ can be approximated by the two straight-line segments

$$F^{-1}(\eta) = \begin{cases} \dfrac{\sigma^2}{2N_0}\eta, & |\eta| \le \eta^{(1)} \approx \sqrt{2\sigma}\sqrt{\log_e\left(\dfrac{p_0}{1-p_0}\dfrac{\mu}{\sigma}\right)} \\ \dfrac{\mu^2}{2N_0}\eta, & |\eta| \ge \eta^{(2)} \approx \sqrt{2\sigma}\sqrt{\log_e\left(\dfrac{p_0}{1-p_0}\dfrac{\mu^3}{\sigma^3}\right)} \\ \dfrac{\sigma^2}{2N_0}\eta \text{ or } \dfrac{\mu^2}{2N_0}\eta, & \eta^{(1)} \le |\eta| \le \eta^{(2)}. \end{cases} \tag{30}$$

With regard to the recursion (28), the parameters σ^2/N_0 and μ^2/N_0 can

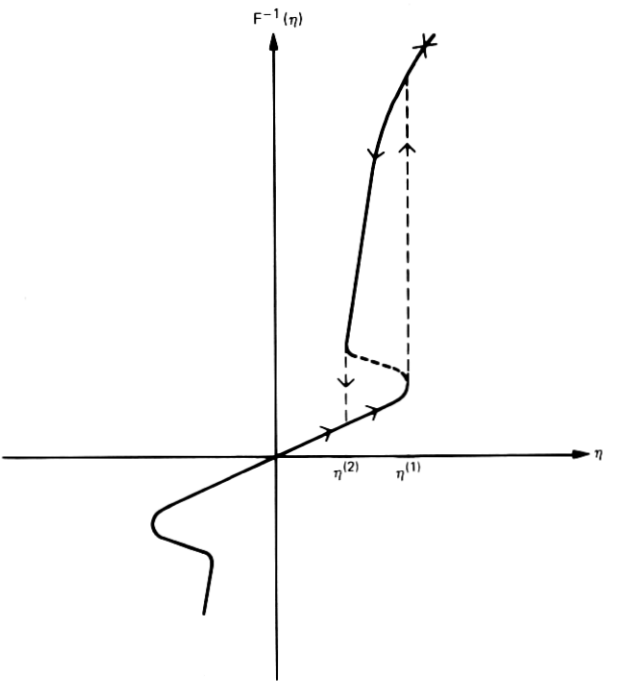

Fig. 3—The nonlinearity $F^{-1}[\eta]$ of (30).

be interpreted as a small and a large "step-size," respectively. The magnitude of the step-size depends on the "old" η_k and the "processed" observation $\hat{c}_k \dot{z}(kT + \Delta_k)$, and a typical trajectory of step-sizes is indicated by the arrows in Fig. 3.

Returning to (17), we rewrite this expression as

$$\ell_1[\{a_m\}, \{\Delta_m\}] = -2 \sum_m c_m z(mT + \Delta_m) + \sum_m c_m^2 g_0 + \sum_{m \neq k} \sum_k c_m c_k g((m-k)T + \Delta_k - \Delta_m). \quad (31)$$

If the system pulse shape is Nyquist $[g(n-k)T = g_0 \delta_{n-k}]$, then the third summation will be close to zero when the $\{\Delta_n\}$ are approximately equal over the duration of $g(t)$. With this approximation in mind, we neglect the cross term $(m \neq k)$ and complete the square to obtain

$$\ell_1[\{a_m\}, \{\Delta_m\}] \cong \sum_m \left\{ g_0 \left(c_m - \frac{z_m}{g_0} \right)^2 - \frac{z_m^2}{g_0} \right\}$$

$$= \sum_m \left\{ g_0 \left(a_m - \left[\frac{z_m}{g_0} - \rho(-1)^m \right] \right)^2 - \frac{z_m^2}{g_0} \right\},$$

and thus

$$\hat{a}_m = Q\left[\frac{z_m}{g_0} - \rho(-1)^m \right], \quad (32a)$$

where $Q[\ ]$ is a function which quantizes its argument to the nearest symbol level. Observe that, with the assumptions we have made, the optimum value of $\hat{a}_m$ depends only on $z_m \equiv z(mT + \hat{\Delta}_m)$ and is in fact the symbol level closest to $(z_m/g_0) - \rho(-1)^m$. The receiver sketched in Fig. 4 implements (27), (28), and (32a).

If the system pulse shape is of the partial response type,[2] then a modified procedure is called for. We illustrate this technique when the received pulse $g'(t)$ is a Class IV partial response pulse and the shaping

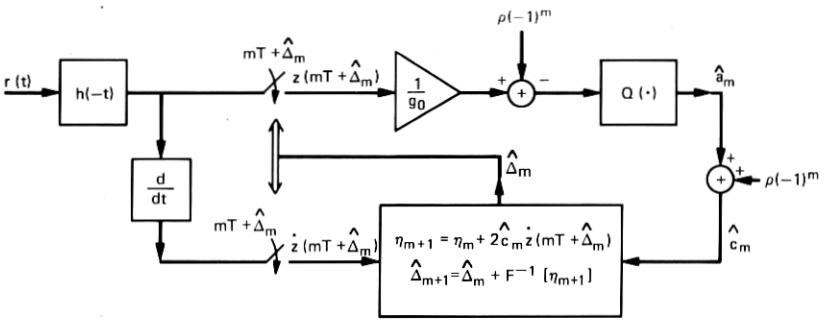

Fig. 4—Data-directed timing loop and receiver.

is split between the transmitter and receiver, the overall* characteristic being $j2T \sin \omega T$. In this case, $g_k' = g_0'(\delta_k - \delta_{k-2})$ and, again neglecting the delay differences $\Delta_k - \Delta_m$, the likelihood becomes

$$\ell_1[\{a_m\}, \{\Delta_m\}] = -2 \sum_m c_m z(mT + \Delta_m) + g_0' \left(\sum_m c_m^2 + \sum_m c_m c_{m-2} \right). \tag{32b}$$

Because of the coupling between c_m and c_{m-2}, the optimization procedure to determine the $\{\hat{c}_m\}$, from (32b), requires the use of dynamic programming (the Viterbi algorithm[8]). While the Viterbi algorithm (VA) can be implemented in a rather straightforward manner for Class IV partial response systems, the decoding delay in the VA makes tracking of the timing phase rather unwieldy, and consequently practical receivers would probably employ a suboptimum technique which directly examines the output of the receiving filter,

$$z(kT) = \sum_n c_n g'(kT - nT - \Delta_n) + \nu(kT). \tag{32c}$$

In the above equation, the samples are obtained from the output of the receiver filter and, neglecting timing jitter, we have

$$z(kT) = c_k - c_{k-2} + \nu_k = d_k + \nu_k, \tag{32d}$$

where $\{d_k\} = \{c_k - c_{k-2}\}$ is the dependent or correlated data sequence. For example, if the input data symbols c_k assume the values $\pm 1, \pm 3$, then d_k would be one of the seven output values $0, \pm 2, \pm 4, \pm 6$. Practical detectors would quantize z_k to one of the seven allowed output values, and the desired data $\{\hat{c}_k\}$ is recovered from the relation $\hat{c}_k = \hat{d}_k - \hat{d}_{k-2}$, where the data are typically precoded[2] to prevent an erroneous decision from propagating. Note that the partial response waveform can be written either as (32c) or as $\Sigma_n d_n' g(t - nT - \Delta_n)$, where d_n' are the correlated output levels and $g(t)$ is the minimum-bandwidth symmetric Nyquist pulse, $\sin(\pi t/T)/(\pi t/T)$. If we adopt this latter representation, then our maximum likelihood development can proceed as before—the only additional approximation being that, while the various sequences of $\{d_n'\}$ are not all equally likely, we have implicitly taken them to be equiprobable. Thus, an approximation to the optimum receiver shown in Fig. 5 would be to quantize z_k, using (32c), to the nearest output level and to use the corresponding $\hat{d}_n$ in (27) and (28).

Returning to Fig. 4, we recall that (27) to (29) have the appearance of a second-order, discrete-time, phase-locked loop with a bi-variable step-size. The choice of step-size is dictated by the current value of η_k which provides a measurement of the "jump" $\Delta_k - \Delta_{k-1}$. A large value of η_{k+1} is indicative of a large jump, while a small value of η_{k+1} reassures

* Note that, in this case, $g'(t)$ is antisymmetric and the receiver filter is not matched to the transmitter filter.

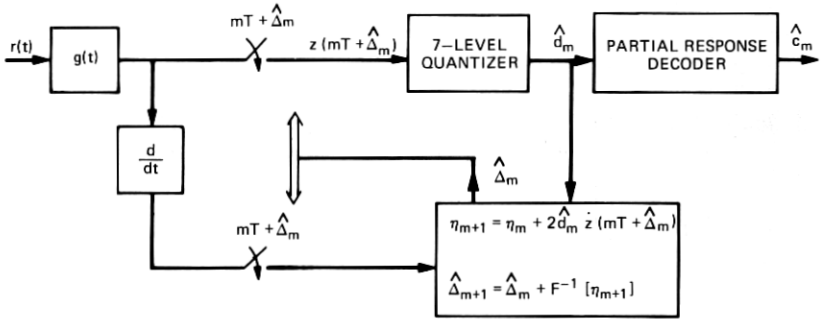

Fig. 5—Data-directed timing loop and receiver for partial-response signaling.

the tracking loop that its estimate of Δ_k is close to the correct value. The nonlinearity shown in Fig. 3 is interpreted as providing hysteresis, since via (30) we know that in the range $\eta^{(1)} \leq |\eta| \leq \eta^{(2)}$

$$\Delta_{k+1} = \Delta_k + \beta_{k+1}\eta_{k+1}, \tag{33}$$

where

$$\beta_{k+1} \approx \begin{cases} \sigma^2/N_0, & \text{if } |\eta_{k+1}| < \eta^{(2)} \text{ and } |\eta_k| < \eta^{(2)} \\ \mu^2/N_0, & \text{if } |\eta_{k+1}| > \eta^{(2)} \text{ or if } |\eta_{k+1}| > \eta^{(1)} \text{ and } |\eta_k| > \eta^{(2)}. \end{cases}$$

The bi-variable step-size appearing in the tracking loop of Fig. 6 has the effect of adaptively varying the loop's bandwidth and thus accelerating recovery from a delay jump. The omission of the quadratic term appearing in (31) may prolong this recovery by several symbol intervals, but this is a small price to pay for the resulting simplicity in implementation.* The receiver shown in Fig. 6 quantizes the filtered and sampled sequence with the aid of a decision-directed phase-locked tracking loop

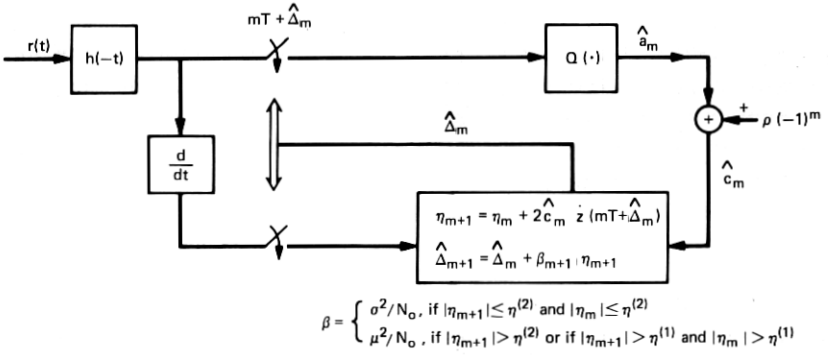

Fig. 6—Simplified data-directed receiver.

* Omission of the quadratic term is tantamount to neglecting the *amplitude* transient, caused by the delay jump, which propagates through the channel and receiver filters. In other words, if one accepts the model given by (6), then any amplitude transients are implicitly neglected.

which provides the sequence of sampling phases, given knowledge of the maximum-likelihood, initial-phase estimate, Δ_0. As described in a later section, the algorithm can be suitably modified to incorporate an estimate of Δ_0. The above decision-directed timing loop is similar to that described by Gitlin and Salz,[9] but the novel aspects here are the bivariable step-size and the hysteresis associated with the tracking loop nonlinearity.

4.2 Modifications to the decision-directed receiver

A drawback of the receiver described in Section 4.1 is the possibility of a relatively long error burst following a timing-phase jump. Suppose such a delay jump causes a large deviation from the optimum sampling time; in a bandlimited system with a narrow eye-opening,[2] this results in a large amount of intersymbol interference and consequently a high probability of error in the next symbol interval. The resulting incorrect decision, used in the decision-directed timing recovery loop, may move the estimated sampling phase in the wrong direction, further increasing the intersymbol interference. This type of effect is called *runaway* and is possible in nearly all decision-directed parameter tracking systems. Runaway is of particular concern in Class IV partial-response systems since the eye is open only for a small fraction of the symbol interval.[10] The possibility of runaway is further enhanced by our neglecting the quadratic cross-term appearing in (31).

To diminish the possibility of error proliferation due to delay jumps in a bandwidth-limited system, we propose the *coarse-quantized* timing recovery system shown in Fig. 7. The incoming signal is sampled at times $\{mT + \Delta_m + (iT/M),\ 0 \le i \le M - 1\}$, where M is some integer, i.e., the receiver samples are taken at the rate M/T instead of $1/T$ samples/s. The sampling phase is still controlled by a *single* tracking loop and, as before, after suitable filtering, each sample is quantized to the nearest data level. The number of samples M is chosen large enough that T/M is less than the width of the eye-opening corresponding to the pulse $g(t)$. Thus for a system with an open eye,[2] in the absence of noise, at least one of the sampling phases $\{\Delta_m + (iT/M)\}$ will result in a correct output* decision. Expressed mathematically, for at least one integer "i," the maximum possible interference,

$$\max_{\{c_m\}} \sum_{m \neq 0} \left| c_m g\left(mT + \Delta_m + \frac{iT}{M}\right) \right|,$$

is less than the minimum distance between two possible received signal levels.

The idea behind the increased sampling rate (which might be readily

* For a four-input level, Class IV partial response system, recall that the output sequence is chosen from one of seven levels.

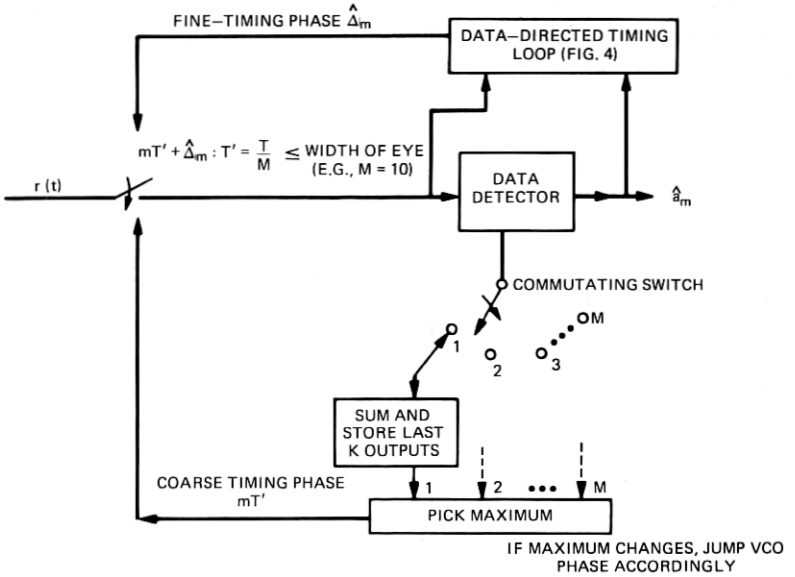

Fig. 7—Coarse-quantized timing recovery system.

available in a digital receiver) is that one of the M possible coarsely quantized sampling phases results in an open eye, and therefore in the absence of noise, supplies a correct sequence $\{c_n\}$ suitable for updating the decision-directed phase-tracking loop. The particular (coarse) timing phase chosen to supply the decision sequence used by the tracking loop is determined by reformulating the problem as picking the static timing epoch $\{iT/M\}$ which maximizes the *a posteriori* likelihood over the recent past. This strategy is mechanized from (32a) by computing a running likelihood

$$\ell_k^{(i)}\left[\{c_m^{(i)}, \{\Delta_m\}\}\right] = g_0 \sum_{j=k-K}^{k} \left[c_j^{(i)} - \frac{1}{g_0} z\left(jT + \Delta_j + \frac{iT}{M}\right)\right]^2 - \frac{1}{g_0} \sum_{j=k-K}^{k} z^2\left(jT + \Delta_j + \frac{iT}{M}\right), \quad (34)$$

where K is some suitably chosen number, $\{c_j^{(i)}\}$ are the decisions† corresponding to the sampling phase $iT/M + \Delta_j$, and the decisions are obtained by quantizing the appropriate output sample. The coarse-timing phase, i^*T/M, is chosen if $\ell_k^{(i^*)} \leq \ell_k^{(i)}$ for all $i \neq i^*$. In practice, one might use a small number of sampling epochs; e.g., three, which would bracket the correct phase. The sum in (34) is truncated to run over a finite span of duration KT seconds so that the effects of ancient delay jumps do not affect the *current sampling epoch*. Once the best coarse timing phase

† The Δ_j are supplied by the phase-locked loop driven by the decisions corresponding to the current most likely coarse-quantized timing phase.

is determined, the bi-variable step-size PLL previously described is used to determine the exact sampling phase. Whenever a new sampling phase becomes the most likely, the current estimate of the timing phase is incremented by the appropriate amount.† It should be pointed out that additive noise following a delay jump may prolong the recovery somewhat. This effect is difficult to assess analytically, and a similar statement can be made concerning the size of M (the number of timing epochs) and K (the memory of the running likelihood). The sensitivity of system performance to these parameters is best determined experimentally.

The realization shown in Fig. 7, which incorporates the coarse-quantized timing recovery scheme, has the related mechanization depicted in Fig. 8, which is specialized to a Class IV partial-response system. Here the quantized seven-level outputs are computed for each sampling phase, and each sequence is monitored for partial-response violations. The coarse-quantized sampling phase used to control the timing tracking loop is chosen as the phase which has the fewest associated partial response violations. Again, the actual logic which dictates when and how switches to a new timing phase are accomplished is probably best determined by an experimental and/or simulation study of the actual system.

4.3 A refined loop which estimates Δ_0

As it stands, the bi-modal phase-locked loop described by (27) and (28) is initialized from a random or arbitrary initial condition, Δ_0. A consequence of this initialization is that in either mode (delay hit/no delay hit) the loop exhibits a double integration (or direct second difference) structure. We now show that, when a constant step-size is used, the algorithm is potentially unstable.‡ We begin by recalling that in either mode the tracking loop is governed by equations of the form

$$\eta_{k+1} = \eta_k + c_k \dot{z}(kT + \Delta_k) \tag{27}$$

$$\Delta_{k+1} = \Delta_k + \beta_k \eta_{k+1}, \tag{33}$$

where β_k is a positive nonincreasing sequence.

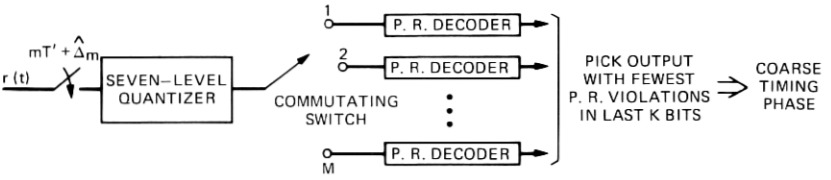

Fig. 8—Coarse-quantized timing recovery for partial-response system.

† For example, suppose $\ell_k^{(i)}$ was the maximum and the new maximum is $\ell_k^{(i^*)}$, then the timing phase should be incremented by $(i^* - i)\ T/M$ s.

‡ This instability can be regarded as a manifestation of the sensitivity of the system equations to the initial unknown phase; i.e., the effect of a wrong choice of this phase propagates endlessly. This is a consequence of viewing the system of simultaneous equations for the timing-phase estimate, (27) and (28), as a recursion with an arbitrary initial condition.

For the purpose of this discussion, we consider a fixed but unknown delay, Δ, and examine the average values of the above equations, i.e.,

$$\bar{\eta}_{k+1} = \bar{\eta}_k + \dot{g}(\Delta_k - \Delta) \tag{35a}$$

$$\bar{\Delta}_{k+1} = \bar{\Delta}_k + \beta_k \bar{\eta}_{k+1}, \tag{35b}$$

where the overbar denotes expectation. We can combine the above equations to obtain the recursion

$$\bar{\Delta}_{k+1} - \left(1 + \frac{\beta_k}{\beta_{k-1}}\right) \bar{\Delta}_k + \frac{\beta_k}{\beta_{k-1}} \bar{\Delta}_{k-1} = \beta_k \dot{g}(\bar{\Delta}_k - \Delta), \tag{36}$$

and if we denote the tracking error by

$$\epsilon_k = \overline{\Delta_k - \Delta}, \tag{37}$$

then we have

$$\epsilon_{k+1} - \left(1 + \frac{\beta_k}{\beta_{k-1}}\right) \epsilon_k + \frac{\beta_k}{\beta_{k-1}} \epsilon_{k-1} \approx \beta_k [\dot{g}(0) + \epsilon_k \ddot{g}(0)]$$

$$= \beta_k \ddot{g}(0) \epsilon_k, \tag{38}$$

where we have used a Taylor Series expansion which is valid for small ϵ_k. Note that the solution to the above time-varying difference equation will decay when the product of the "instantaneous roots," β_k/β_{k-1} is less than unity. However, if we were to use a constant step-size, i.e., $\beta_k = \beta_{k-1} = \beta$, and if $\ddot{g}(0) \approx 0$, then the solutions are of the form $\epsilon_k = \epsilon_0 \sin k\theta$; i.e., the error does not decay to zero but oscillates as soon as the error penetrates the linear region (clearly, this is an unacceptable situation). The existence of oscillations can be deduced directly from (27) and (28). Note from (28) that Δ_k is the accumulated sum of the past errors. When a phase-hit occurs, this sum will become large and the loop will enter the large step-size mode. In order that the loop ultimately converge to the correct phase, it is clear that the accumulator will have to "see" many terms opposite in sign to the original accumulants, i.e., the loop can oscillate.

In the light of the above discussion, we now derive an estimate of Δ_0 and indicate how this estimate may be incorporated into the existing timing loop to produce a stable loop. We first let

$$\Delta'_m \equiv \Delta_m - \Delta_0, \tag{39}$$

and in terms of $\{\Delta'_m\}$ we have from (20)

$$\ell_2[\{\Delta'_m\}] = -N_0 \sum_{m=1}^{\infty} \log \left\{ \frac{(1 - p_0)}{\sqrt{2\pi}\,\sigma} \exp\{-(\Delta'_m - \Delta'_{m-1})/2\sigma^2\} \right.$$

$$\left. + \frac{p_0}{\sqrt{2\pi}\,\mu} \exp\{-(\Delta'_m - \Delta'_{m-1})^2/2\mu^2\} \right\}, \tag{40}$$

where we note that $\Delta_0' = 0$; it is our intention to estimate $\{\Delta_n'\}$ and $\{\Delta_0\}$ separately and then to combine these quantities to construct $\{\hat{\Delta}_n\}$.

Proceeding as before, we define

$$\eta_k' \equiv (\Delta_k' - \Delta_{k-1}')G[\Delta_k' - \Delta_{k-1}'], \tag{41}$$

and taking the derivative of the likelihood with respect to Δ_k' gives

$$\begin{aligned}\eta_{k+1}' &= \eta_k' - \frac{\partial \ell_1[\{a_m\}, \{\Delta_m'\}, \Delta_0]}{\partial \hat{\Delta}_k'} \\ &= \eta_k' + 2c_k\dot{z}(kT + \Delta_k' + \Delta_0).\end{aligned} \tag{42}$$

Inverting (41) gives

$$\Delta_{k+1}' = \Delta_k' + F^{-1}[\eta_{k+1}'], \tag{43}$$

where $F^{-1}[\ ]$ has been previously defined and where (42) and (43) are initialized with $\Delta_0' = 0$.

An estimate of Δ_0 will be obtained by applying stochastic approximation theory,[11] and using as the increment the derivative of the current term in the likelihood, $\ell_1[\{a_m\}, \{\Delta_m'\}, \Delta_0]$, with respect to Δ_0.

The resulting stochastic approximation algorithm for the estimate of Δ_0 is

$$\hat{\Delta}_{0,k+1} = \hat{\Delta}_{0,k} + \gamma_k c_k \dot{z}(kT + \hat{\Delta}_k' + \hat{\Delta}_{0,k}), \qquad k = 0, 1, 2, \cdots, \tag{44}$$

where γ_k is a positive step-size sequence. Since the estimate of Δ_k is the sum of the component estimates, i.e.,

$$\hat{\Delta}_k \equiv \hat{\Delta}_k' + \hat{\Delta}_{0,k}, \tag{45}$$

adding (43) and (44) gives the structure shown in Fig. 9, which implements the recursions:

$$\hat{\Delta}_{k+1} = \hat{\Delta}_k + F^{-1}[\hat{\eta}_{k+1}] + \gamma_k \hat{c}_k \dot{z}(kT + \hat{\Delta}_k) \tag{46}$$

$$\hat{\eta}_{k+1}' = \hat{\eta}_k' + 2\hat{c}_k \dot{z}(kT + \hat{\Delta}_k). \tag{47}$$

In implementing (46) and (47), the step-size γ_k would probably be switched to a larger step-size whenever the $F^{-1}[\cdot]$ function indicates that a mode switch is taking place—this can be thought of as reinitializing the estimate of Δ_0. Contrasting (46) and (47) with (27) and (28), we see that, when the state of the system is such that $F^{-1}[\eta] = \beta\eta$, the latter system can be written as the second-order difference equation

$$\Delta_{k+1} - 2\Delta_k + \Delta_{k-1} = \beta c_k \dot{z}(kT + \Delta_k), \tag{48}$$

while the former system is equivalent to

$$\begin{aligned}\Delta_{k-1} - 2\Delta_k + \Delta_{k-1} = (\gamma_k + \beta)c_k\dot{z}(kT + \Delta_k) \\ - \gamma_{k-1}c_{k-1}\dot{z}(kT - T + \Delta_{k-1}).\end{aligned} \tag{49}$$

The effect of the direct feeding of the input, $\gamma_k c_{k-1} \dot{z}(kT - T + \Delta_{k-1})$, to the second summer in Fig. 9 can be seen by considering the evolution of the average phase error, (37). It is clear from (49) that with $\gamma_k = \gamma$ the modified tracking-loop structure can provide roots within the unit circle and hence eliminate the possibility of oscillations.

V. APPLICATION IN A SIMULATED CLASS IV PARTIAL-RESPONSE SYSTEM

The application of dual-mode, decision-directed timing recovery and coarse-quantized timing recovery to a data communication system subject to additive noise and occasional delay jumps was tested by means of a computer simulation of a digital version of the baseband data transmission system shown in Fig. 10. Transmission through radio channels was modeled by the addition of additive white Gaussian noise to the signal and the insertion of abrupt delay changes. Since the simulated system is not an exact replica of the idealized equations used for analysis, the actual receiver differed in some small details from the structure previously derived.

A seven-level, Class IV, partial-response waveform was generated in sampled form with sampling rate $10/T$ (10 times the symbol rate), and channel and receiver signal processing were also carried out digitally at this sampling rate.

Because the simulation was carried out in nonreal time on a digital computer, exact realization of time delays of other than multiples of $T/10$ was not possible. Moreover, the actual sampling phase of the oversampled input was not under the receiver's control. Instead, arbitrary channel and receiver sampling delays were approximated by linear interpolation. The samples at the output of the receiver filter were denoted $\{z(mT + iT'); m = 0, 1, 2, \cdots, \infty; i = 0, 1, \cdots, 9\}$.

The index i denotes a timing phase, quantized to a multiple of $T' = T/10$. The output sampled at $mT + iT' + \hat{\Delta}_m$ was taken to be, by linear interpolation,

$$z(mT + iT' + \hat{\Delta}_m) = \left(1 - \frac{\hat{\Delta}_m}{T'}\right) z(mT + iT') + \frac{\hat{\Delta}_m}{T'} z(mT + (i+1)T'). \quad (50)$$

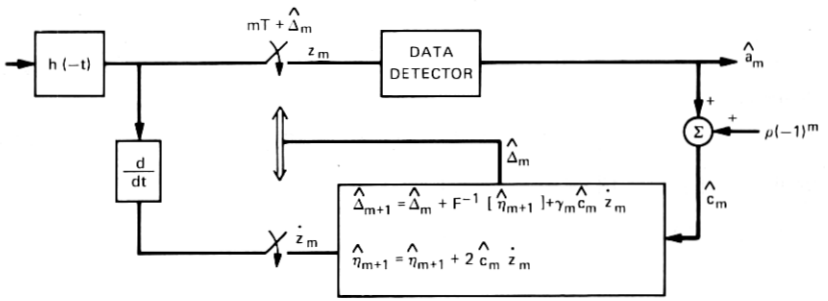

Fig. 9—Modified data-directed, fine-tuned, timing loop.

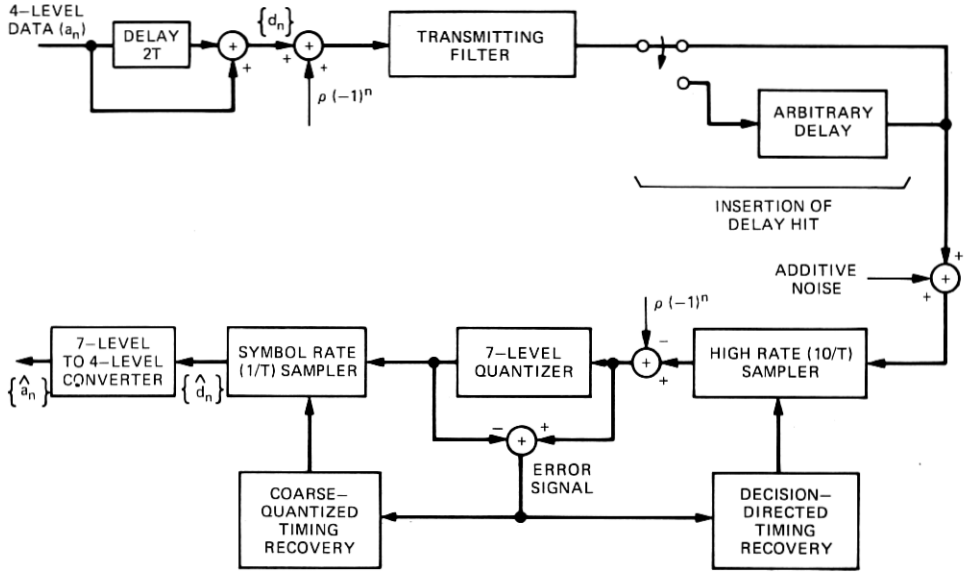

Fig. 10—Simulation of Class IV partial response system with delay hits.

The quantity $\hat{\Delta}_m$ ($0 \le \hat{\Delta}_m \le T'$) is the receiver's estimate of the sampling phase in the mth symbol interval, mod T'.

The 10-fold oversampling also permits the application of the coarse-quantized timing recovery method described in Section 4.2, with $M = 10$. Each sample† $z(mT + iT' + \hat{\Delta}_m)$ is quantized into $\hat{c}_m^{(i)} = \hat{d}_m^{(i)} + \rho(-1)^m$ where $\hat{d}_m^{(i)}$ is one of the seven levels $\hat{d}_m^{(i)} = 0, \pm 2, \pm 4, \pm 6$, and an error $e_m^{(i)}$ is formed as:

$$e_m^{(i)} \equiv z(mT + iT' + \hat{\Delta}_m) - \hat{c}_m^{(i)}. \tag{51}$$

For each integer i from 0 to 9, the sum of squared errors over the past K symbol intervals was formed

$$\ell_k^{(i)} = \sum_{m=k-K}^{k} e_m^{(i)2}.$$

The values of K used in the simulations were 20 and 40.

That integer $i = i^*$ which minimized $\ell_k^{(i)}$ was taken to be the current most likely coarse-quantized sampling phase. Whenever $i = i^*$ (once per symbol interval) the current decision $\hat{c}_m^{(i^*)} - \rho(-1)^m = \hat{d}_m^{(i^*)}$ is passed on as the receiver's decision on d_m. The program records the occurrence of errors (discrepancies between $\hat{d}_m^{(i^*)}$ and d_m). Note that the above definition of $\ell_k^{(i)}$ differs from that proposed in Section 4.2 in the omission of the sum of squares of z-samples. A further modification was the inhibition of a change in i^* when $\ell_k^{(i)}$ is greater than 90 percent of $\ell_k^{(i^*)}$. This "dead zone" modification reduced the occurrence of switches back and forth between two values of i for which the values of $\ell_k^{(i)}$ are nearly equal.

We remark that abrupt changes in the intervals between receiver output samples should not be passed on to the data recipient. The receiver's output samples would in practice enter an elastic buffer and be clocked out under the control of a very narrowband phase-locked loop.

The value of $\hat{\Delta}_m$ used in the interpolative sampling procedure was obtained by a digital implementation of the decision-directed, second-order timing recovery algorithm described in Section 4.3. Instead of using the correction term $\hat{c}_m^{(i^*)}\, \dot{z}(mT + \hat{\Delta}_m + i^*T')$ in the loop, we use an approximation to the negative of the derivative (gradient) of the squared error $e_m^{(i^*)2}$ with respect to $\hat{\Delta}_m$; i.e.,

$$\delta_m \equiv -e_m^{(i^*)}\, [z(mT + (i^* + 1)T') - z(mT + i^*T')]. \tag{52}$$

When the loop error is zero, the modified correction term (52) guarantees that no adjustment will be made, while the original correction term only provides this condition on the average.

† The overall impulse response was scaled so that the ideal sampled outputs in the absence of noise are 0, ±2, ±4, ±6.

The average value of the correction term δ_m defined by (52) can be computed from equations (51), (32c), and (32d) for an ideal Class IV partial-response system for which

$$g'(t) = \frac{\sin\left(\frac{\pi t}{T}\right)}{\left(\frac{\pi t}{T}\right)} - \frac{\sin\left(\frac{\pi(t-2T)}{T}\right)}{\left(\frac{\pi(t-2T)}{T}\right)}. \tag{53}$$

For small timing errors $(\Delta_m - \hat{\Delta}_m - i^*T')$ between the true phase Δ_m and the estimated phase $\hat{\Delta}_m + i^*T'$ (that is, neglecting quadratic and higher order terms in $(\Delta_m - \hat{\Delta}_m - i^*T')/T'$), the linearized average value of δ_m is

$$\langle \delta_m \rangle \approx 0.278 \left(\frac{\Delta_m - \hat{\Delta}_m - i^*T'}{T'}\right) \tag{54}$$

when a timing tone is not transmitted ($\rho = 0$). The corresponding linearized average value of the correction term $\langle \hat{d}_m^{(i^*)} \dot{z}(mT + \hat{\Delta}_m + i^*T') \rangle$ can similarly be shown to equal this same quantity. Note that the correction term δ_m given by (52) arises from an attempt to minimize the mean-squared error of a linear interpolation scheme applied to a digital receiver whose actual input sampling phase is not under its control. Thus, we have established a connection between this simple linear interpolation scheme with a mean-squared-error optimality criterion and the decision-directed, timing-phase recovery scheme dictated by minimum error probability considerations.

As prescribed in Section IV, a second-order decision-directed, sampling-phase, updating algorithm including a direct correction term and a cumulative correction term is used. The following equations summarize the simulated receiver's operation:

(*i*) *Interpolative sampling*:

$$z(mT + iT' + \hat{\Delta}_m) \equiv \left(1 - \frac{\hat{\Delta}_m}{T'}\right) z(mT + iT') + \frac{\hat{\Delta}_m}{T'} z(mT + (i+1)T').$$

(*ii*) *Quantization*:

$$\hat{d}_m^{(i)} = \text{quantization of } [z(mT + iT' + \hat{\Delta}_m) - \rho(-1)^m]$$

$$c_m^{(i)} = \hat{d}_m^{(i)} + \rho(-1)^m.$$

(*iii*) *Error*:

$$e_m^{(i)} \equiv z(mT + iT' + \hat{\Delta}_m) - \hat{c}_m^{(i)}.$$

(*iv*) *Coarse-quantized timing recovery*:

i^* = value of i which minimizes

$$\ell_k^{(i)} = \sum_{m=k-K}^{k} e_m^{(i)2}.$$

(v) *Decision-directed fine timing recovery:*

$$\frac{\hat{\Delta}_{m+1}}{T'} = \frac{\hat{\Delta}_m}{T'} + \gamma_m \delta_m + \beta_m \eta_{m+1},$$

where

$$\delta_m \equiv e_m^{(i^*)} [z(mT + (i^* + 1)T') - z(mT + i^*T')]$$

and

$$\eta_{m+1} \equiv (1 - \alpha)\eta_m + \delta_m,$$

with

$$\hat{\Delta}_0 = \eta_0 = 0.$$

Since in the simulated and real systems, the timing transient does not instantaneously affect the received signal, the recursion defining η_{m+1} introduces a small amount of "leakage," represented by $\alpha = 0.0005$. Suitable values of the second-order loop parameters γ_m and β_m in the narrowband and wideband modes were established by loop-bandwidth considerations and observations of the transient response of $\hat{\Delta}_m$ to simulated delay jumps. The parameter values picked for the narrowband mode were

$$\gamma_m = 0.005$$

and

$$\beta_m = 3.42 \times 10^{-6}.$$

Assuming the linearized average correction term of (54), we have a discrete-time linear model of the second-order fine-timing recovery loop shown in Fig. 11. This loop's bandwidth, for the above values of γ_m and β_m and $1/T = 772$ kHz, is readily shown to be 240 Hz.

The wideband mode is initiated first whenever the value of i^* is changed by the coarse-quantized algorithm, or second whenever the following recursively generated quantity exceeds a threshold:

$$S_{m+1} = 0.99S_m + \delta_m. \tag{55}$$

The quantity S_m is a weighted average of past correction terms, in contrast to the *cumulative sum* of all past correction terms envisaged in (47).

With the initiation of the wideband mode, γ_m is set to 1 and β_m to 3.42×10^{-4}. Thereafter,

$$\gamma_m = \max(0.005, 1/L)$$

$$\beta_m = 3.42 \times 10^{-4}/L \quad \text{up to} \quad L = 200,$$

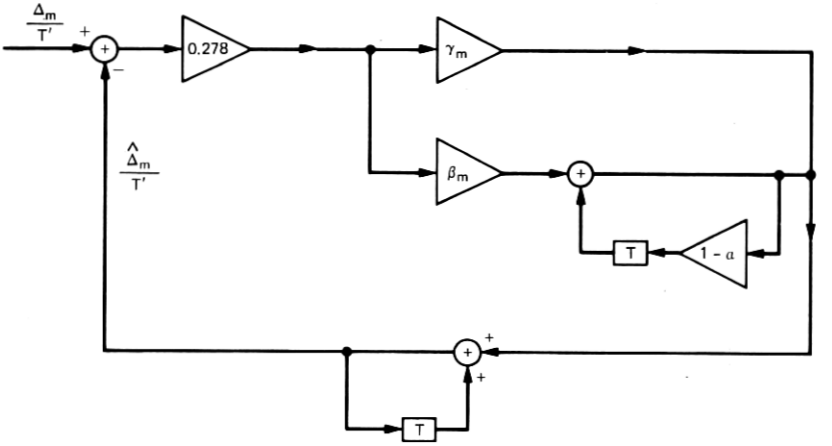

Fig. 11—Discrete time linearized, second-order loop model.

where L is the number of symbol intervals elapsed since initiation of the wideband mode. Thus, the initiation of the wideband mode restarts a stochastic approximation algorithm, with step-size decreasing toward a fixed minimum value. The duration of the wideband mode is 200 symbol intervals, after which the narrowband mode resumes.

VI. SIMULATION RESULTS

The channel model and receiver structure described in Section V were simulated with a 24-dB signal-to-noise ratio and with the insertion of occasional delay hits.

The value of K (the number of past squared errors stored by the coarse-quantized timing recovery algorithm) was set to either 20 or 40. Transmission both with and without a −18-dB ($\rho = 0.554$) $1/2T$ tone was simulated. The results are summarized in Fig. 12, which displays the observed average number of symbol errors (errors in $\hat{d}_m^{(i^*)}$) vs the delay hit expressed as a fraction of a symbol interval.

Each average plotted in Fig. 12 is only over five delay hits of the same magnitude, and thus the curves display considerable variability. Nevertheless, it is clear that a receiver employing two-mode decision-directed and coarse-quantized timing recovery can tolerate delay hits of up to almost half a symbol interval, while sustaining error bursts on the order of a dozen or less, rather than several thousand, which might be expected in a conventional tracking loop with a bandwidth on the order of 100 Hz. Greater delay hits unavoidably cause the deletion or repetition of data.

The number of errors sustained roughly doubled as K was doubled from 20 to 40. This is understandable, since the delay in detecting a phase change, by the coarse-quantized timing recovery system, is proportional to K. The risk of "false-alarm switching" decreases with K, and therefore

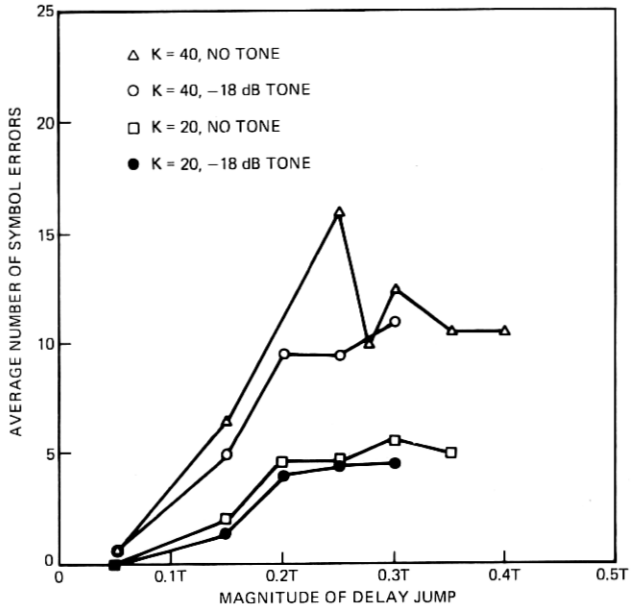

Fig. 12—Average number of errors produced by delay jumps in the simulated system employing two-mode, decision-directed and coarse-quantized timing recovery.

a relatively large value of K such as 40 may be worth the price of, say, a dozen extra errors sustained per delay jump. On the other hand, decreasing K will increase the number of errors due to "false-alarm" switching. The optimum value of K can best be determined by experience with a real system.

It is interesting to note from the curve that the presence or absence of a transmitted timing tone 18 dB below the data signal does not make a dramatic difference in the robustness of the system against delay hits. It therefore appears safe to omit the tone in a system employing decision-directed and coarse-quantized timing recovery. We note that the simulated system displayed rapid start-up characteristics in the coarse-quantized, decision-directed mode. The timing phase, correct to within $T/20$, was acquired in 20 to 25 symbol intervals in the absence of a transmitted tone. The transmission of a −18-dB tone unaccountably delayed timing-phase acquisition during start-up.

Figure 13 shows the evolution of the receiver's sampling-phase estimate following a delay jump of $-1.5T'$. The horizontal coordinate is the number of elapsed symbol intervals. The dotted curves show the phase estimate $i*T' + \hat{\Delta}_m$ (quantized by the limited resolution of the computer plotting routine). The x's at height 8.5 indicate the occurrence of symbol errors. In this example, the errors occur after the delay hit but before initiation of a coarse-quantized timing jump.

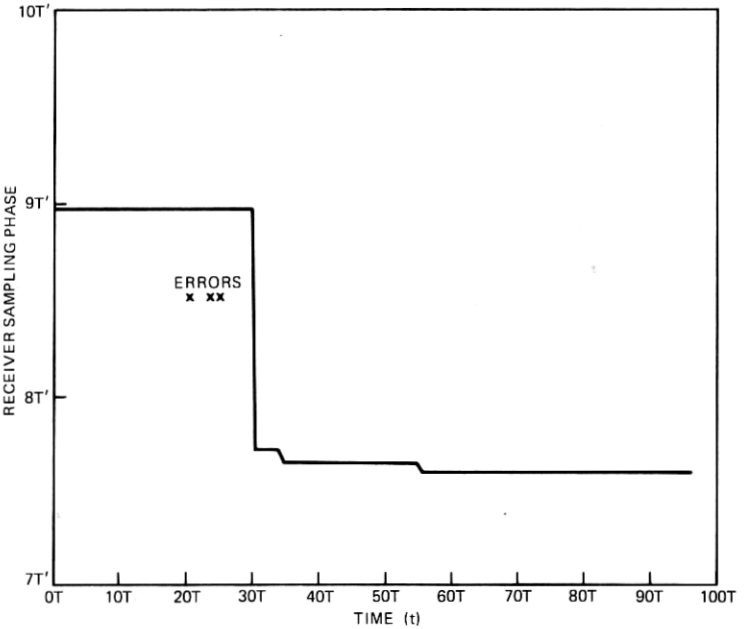

Fig. 13—Response of system to delay jump applied at $t = 0$.

VII. CONCLUSIONS

The technique of data-directed, coarse-quantized, dual-mode timing recovery has been derived and applied to the rapid acquisition of timing phase in systems subject to delay hits. In a simulated system, typical error-burst lengths, following a timing discontinuity of up to a half symbol interval, have been reduced to a dozen or so—two orders of magnitude less than that expected with a conventional phase-locked tracking system. Furthermore, the derivation and simulations have demonstrated the viability of these timing-recovery techniques in the absence of a transmitted pilot tone.

REFERENCES

1. K. L. Seastrand and L. L. Sheets, "Digital Transmission Over Analog Microwave Radio Systems," Proc. IEEE 1972 Int. Conf. Comm. (ICC'72), Philadelphia.
2. R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication*, New York: McGraw-Hill, 1968.
3. W. R. Bennett, "Statistics of Regenerative Digital Transmission," B.S.T.J., *37*, No. 6 (November 1958), pp. 1501–1542.
4. R. D. Gitlin and J. F. Hayes, "Timing Recovery and Scramblers in Data Transmission," B.S.T.J., *54*, No. 3 (March 1975), pp. 569–593.
5. B. R. Saltzberg, unpublished work.
6. J. E. Mazo, "Jitter Comparison of Tones Generated by Squaring and by Fourth-Power Circuits," B.S.T.J., *57*, No. 5 (May-June 1978), pp. 1489–1498.
7. G. D. Forney, Jr., "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory, *IT-18* (May 1972), pp. 363–378.
8. G. D. Forney, Jr., "The Viterbi Algorithm," Proc. IEEE, *61*, No. 3, pp. 268–278.
9. R. D. Gitlin and J. Salz, "Timing Recovery in PAM Systems," B.S.T.J., *50*, No. 5 (May–June 1971), pp. 1645–1669.

10. J. Steel and B. M. Smith, "The Effects of Equalization, Timing and Carrier Phase on the Eye Patterns of Class-4 Partial Response Data Signals," IEEE Trans. Comm., February 1975.
11. D. J. Sakrison, "Stochastic Approximation: A Recursive Method for Solving Regression Problems," in *Advances in Communications*, Vol. 2, A. V. Balakrishran, ed., New York: Academic Press, 1966.