Simultaneous Adaptive Estimation and Decision Algorithm for Carrier Modulated Data Transmission Systems

HISASHI KOBAYASHI, MEMBER, IEEE

Abstract—**The problems of sequence decision, sample timing, and carrier phase recovery in a class of linear modulation data transmission systems are treated from the viewpoint of multiparameter estimation theory. The structure of the maximum-likelihood estimator is first obtained, and a decision-directed receiver is then derived. These receivers are different from the conventional one in that the carrier phase is extracted from the signal components themselves in an adaptive fashion.**

The structure of this adaptive demodulator and detector is then extended to the case in which the channel characteristic is unknown, and the algorithm for adjusting the carrier phase and sample instant is discussed in combination with that of adaptive equalization.

I. Introduction

THE ERROR in carrier phase acquisition along with sample timing error and intersymbol interference is one of the major obstacles of high-speed digital data-transmission systems [1]. In the conventional pulse-amplitude modulation-vestigial sideband (PAM-VSB) (or -single sideband (-SSB)) systems, the phase of the demodulator is controlled by a phase-locked loop (PLL) tracking the phase of a reference carrier (sometimes called a pilot carrier), which is added in quadrature with the modulating carrier. The following are the difficulties associated with this conventional technique. 1) The phase jitter of the pilot carrier, caused by data components near the carrier frequency, contributes to error in the phase estimate and hence results in additional distortion of the demodulated signal. 2) The phase characteristic of the transmission media tends to be highly nonlinear near the cutoff frequency region,[1] at which the pilot carrier is located for the purpose of efficient bandwidth utilization. Thus the amount of phase shift (averaged over the signal spectrum band), which the modulated signal components receive, is different from the phase shift of the pilot carrier. In other words, the phase estimate based on the carrier is not the optimum value to be used for demodulation. This error, too, results in distortion of the demodulated waveform (sometimes called phase-intercept distortion) [3].

In conventional PLL systems, the modulated signal, which contains much more energy than the pilot carrier, as far as the acquisition of carrier frequency and phase is concerned, is in effect regarded as an interfering component or noise. Such a treatment may be quite natural in analog communication systems in which a signal process is essentially a random process, usually a Gaussian process [4]. In digital data communication systems, on the other hand, it would be more natural to regard the signal as a sequence of known (at least partially) waveforms modulated by a random sequence.

In this paper, the problem of sequence decision, demodulation, sampling, and equalization for a class of linear modulation systems [1] is treated from the viewpoint of multiparameter estimation. It will be clear that any deviation of the receiver from its optimum characteristic would be reflected in additional distortion of the demodulated signal. Hence the error-control signal should be obtained from distorted output. It is shown that the sampled values of the demodulated signal provide a sufficient statistic for controlling the carrier phase and sample timing of the receiver. It is shown that the maximum-likelihood receiver (MLR) is realized by modifying the conventional receiver to the one of recursive type. The structure of an adaptive demodulator and sampler is then derived where the control signal is obtained from the decision output sequences, rather than from the continuous waveform as in the traditional PLL or in the Costas loop [1], [3]–[5]. The convergence problem of this decision-directed receiver (DDR) is discussed by applying Robbins–Monro methods [6]–[8].

The structure of this adaptive demodulator and sampler is further extended to the case in which the channel characteristic is unknown. This adaptive algorithm for adjusting the carrier phase and sample instant is discussed in combination with that of adaptive equalization which has been extensively studied by Lucky [1], [9], [10] and others [11]–[13]. The joint equalization, carrier acquisition, and timing recovery for PAM-VSB (or -SSB) systems have been studied independently by Chang [14], [15] and by the present author [16]. This paper is intended to provide a unified treatment for a class of linear modulation systems extending the earlier work.

Paper approved by the Data Communication Systems Committee of the IEEE Communication Technology Group for publication without oral presentation. Manuscript received June 22, 1970; revised September 7, 1970.

The author is with the IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y. 10598.

[1] This is typically the case in telephone channels. See, for example, [1, p. 36] and [2].

II. Maximum-Likelihood Receiver

Let $\{a_n\}$ be the data sequence and ω_0 the carrier frequency. If we assume that ω_0 is larger than the modulating signal spectrum bandwidth, a concise representation of such a signal is given in terms of the complex envelope [17]–[20] which we denote by $s(t)$. The functional form of $s(t)$ depends on the modulation scheme. For example,

PAM-DSB

$$s(t) = \{\sum_n a_n f(t - nT - \tau)\} \exp(j\phi) \tag{1a}$$

PAM-VSB (or -SSB)

$$s(t) = [\sum_n a_n\{f(t - nT - \tau) - j\check{f}(t - nT - \tau)\}] \exp(j\phi) \tag{1b}$$

quadrature amplitude modulation (QAM)

$$s(t) = \{\sum_n (a_{1,n} + ja_{2,n}) f(t - nT - \tau)\} \exp(j\phi) \tag{1c}$$

digital PM

$$s(t) = \{\sum_n \exp(ja_n) f(t - nT - \tau)\} \exp(j\phi) \tag{1d}$$

differential PM

$$s(t) = \{\sum_n \exp(jb_n) f(t - nT - \tau)\} \exp(j\phi),$$

$$b_n = b_{n-1} + a_n \bmod 2\pi \tag{1e}$$

PAM-PM

$$s(t) = \{\sum_n a_n \exp(jb_n) f(t - nT - \tau)\} \exp(j\phi). \tag{1f}$$

Here ϕ and τ represent the carrier phase and delay time, respectively, which are unknown to the receiver. The function $f(t)$ in (1a) and (1b) represents the baseband signal element, and $\check{f}(t)$ of (1b) is a linear transformation of $f(t)$ and is equal to the Hilbert transform of $f(t)$ in the case of SSB modulation.[2]

Signal forms of (1) have a common representation

$$s(t) = \{\sum_n c_n f(t - nT - \tau)\} \exp(j\phi) \tag{2}$$

where $\{c_n\}$ is a real or complex number relating to the information sequence $\{a_n\}$. For example, $c_n = a_{1,n} + ja_{2,n}$ in QAM and $c_n = c^{ja_n}$ in digital PM. Similarly, $f(t)$ is a real or complex function representing a signal element, e.g., $f(t)$ is complex for VSB (or SSB). The actual waveform to be transmitted is given from the definition of the complex envelope by

$$\mathrm{Re}\{s(t) \exp(j\omega_0 t)\} \tag{3}$$

and the receiver input is denoted by $\mathrm{Re}\{x(t) \exp(j\omega_0 t)\}$ where

$$x(t) = r(t) + n(t). \tag{4}$$

Here, $r(t)$, the complex envelope of the received signal, takes the form

$$r(t) = \{\sum_n c_n g(t - nT - \tau)\} \exp(j\phi) \tag{5}$$

where

$$g(t) = f(t) \otimes h_b(t). \tag{6}$$

In (6), $h_b(t)$ is the impulse response of the baseband equivalent of a given channel $h(t)$ and is in general a complex function [21], and $\otimes$ represents convolution. That is, Fourier transforms of $h(t)$ and $h_b(t)$ are related by

$$H_b(\omega) = \begin{cases} H(\omega + \omega_0), & |\omega| < W \\ 0, & \text{elsewhere} \end{cases} \tag{7}$$

where W corresponds to the bandwidth of a low-pass filter which follows the demodulator. W can be chosen equal to or greater than W_f, the bandwidth of the signal element $f(t)$ [21].

As shown by (5), the received signal is a known waveform with unknown parameters ϕ, τ, and $\{c_n\}$. We assume that the additive noise is a Gaussian process with zero mean and covariance function $N_0 K(t,t')$. If the noise is stationary, $K(t,t') = K(t - t')$, and if it is white, $K(t,t') = \delta(t - t')$. Let $P_{r+n}(x)$ and $P_n(x)$ be the probability measures under the hypothesis that the signal is present and absent, respectively.[3] Then the likelihood-ratio function is given by

$$L(x \mid \{c_n\},\tau,\phi) = \frac{dP_{r+n}(x)}{dP_n(x)} = \exp\left\{\frac{1}{2N_0}([x,r]_K + [r,x]_K - [r,r]_K)\right\} \tag{8}$$

where the inner product $[x,r]_K$ is defined by

$$[x,r]_K = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^*(t) K^{-1}(t - t') r(t')\, dt\, dt' \tag{9}$$

where $K^{-1}(t - t')$ is the function[4] which satisfies

$$\int_{-\infty}^{\infty} K^{-1}(t - t') K(t' - t'')\, dt' = \delta(t - t''). \tag{10}$$

where $\tilde{\delta}(t)$ is any function whose Fourier transform is unity in the frequency range $|w| < W$.

[2] The representation of an SSB signal in terms of Hilbert transform is widely used [1], [19]. It is rather straightforward to extend this concept to represent a VSB signal [1].

[3] Although the problem is not a detection problem, it is convenient to introduce these two hypotheses in order to define the likelihood-ratio function of (8).

[4] The inner product of (9) can be well defined using the reproducing kernel Hilbert space method, even if $K^{-1}(t,t')$ does not exist. See [22]–[24].

By substituting (5) and (6) into (9), we obtain

$$[x,r]_K = \sum_n z_n^* c_n \triangleq \mathbf{z}^{*\prime}\mathbf{c} \tag{11}$$

where the complex-valued vectors $\mathbf{c}$ and $\mathbf{z}$ represent the sequences $\{c_n\}$ and $\{z_n\}$ which may be of an infinite length. The asterisk and prime mean complex conjugate and transpose, respectively. The random variable z_n is a linear observable of $x(t)$ defined by

$$\begin{aligned} z_n = z_n(\tau,\phi) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t)K^{-1}(t-t') \\ &\quad \cdot g^*(t'-nT-\tau)\exp(-j\phi)\,dt\,dt' \\ &= [\{x(t)\exp(-j\phi)\} \otimes q^*(-t)]_{t=nT+\tau} \end{aligned} \tag{12}$$

where $q(t)$ is the solution of the integral equation

$$\int_{-\infty}^{\infty} K(t-t')q(t')\,dt' = g(t). \tag{13}$$

In (12), the real and imaginary parts of $x(t)\exp(-j\phi)$ correspond to the so-called in-phase and quadrature components of the demodulator output. The function $q^*(-t)$ represents a complex filter (Appendix I) matched to the signal $g(t)$, and thus z_n is the sampled value of the matched filter output. Similarly, we obtain

$$[r,r]_K = \sum_n \sum_{n'} c_n^* R_{nn'} c_{n'} = \mathbf{c}^{*\prime}\mathbf{R}\mathbf{c} \tag{14}$$

where $\mathbf{R}$ represents a complex-valued Hermitian (or self-adjoint) matrix whose components $R_{nn'}$ are given by

$$\begin{aligned} R_{nn'} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g^*(t-nT)K^{-1}(t-t')g(t'-n'T)\,dt\,dt' \\ &= \int_{-\infty}^{\infty} g^*(t-nT)q(t-n'T)\,dt = R((n-n')T) \end{aligned} \tag{15}$$

i.e., sampled values of the matched filter response of the signal element $g(t)$. From (8), (11), and (14), the log-likelihood function is obtained as

$$\begin{aligned} \ln L(x \mid \mathbf{c},\tau,\phi) &= \frac{1}{2N_0}\{\mathbf{z}^{*\prime}\mathbf{c} + \mathbf{c}^{*\prime}\mathbf{z} - \mathbf{c}^{*\prime}\mathbf{R}\mathbf{c}\} \\ &= \frac{1}{2N_0}\mathbf{z}^{*\prime}\mathbf{R}^{-1}\mathbf{z} \\ &\quad - \frac{1}{2N_0}(\mathbf{c}-\mathbf{R}^{-1}\mathbf{z})^{*\prime}\mathbf{R}(\mathbf{c}-\mathbf{R}^{-1}\mathbf{z}). \end{aligned} \tag{16}$$

The last expression shows that sampled values of the matched filter output provide a sufficient statistic for extracting all the parameters.

The maximum-likelihood estimates (MLE) of parameters $\mathbf{c}$, τ, and ϕ are those values which jointly maximize the expression of (16). First we demonstrate that the MLEs are asymptotically (i.e., for a high SNR or for a large size of data) unbiased. Under the asymptotic assumption we can neglect the noise component of the matched filter output z_n, obtaining

$$\begin{aligned} z_n &\cong [\{\sum_m c_m g(t-mT-\tau)\exp(j\phi)\} \\ &\quad \cdot \exp(-j\hat{\phi}) \otimes q^*(-t)]_{t=nT+\hat{\tau}} \\ &= \sum_m R((n-m)T+\hat{\tau}-\tau)\exp[-j(\hat{\phi}-\phi)]c_m \end{aligned} \tag{17}$$

where $\hat{\phi}$ and $\hat{\tau}$ are the estimates of ϕ and τ used in the receiver. Then applying Taylor's series expansion to $R((n-m)T+\hat{\tau}-\tau)\exp[j(\hat{\phi}-\phi)]$ around $\hat{\tau}=\tau$ and $\hat{\phi}=\phi$, we obtain

$$\begin{aligned} \mathbf{z}^{*\prime}\mathbf{c} + \mathbf{c}^{*\prime}\mathbf{z} &= 2\,\mathrm{Re}\Bigg[\Bigg\{\sum_m\sum_n c_n^* R((n-m)T)c_m \\ &\quad + (\hat{\tau}-\tau)\sum_m\sum_n c_n^*\dot{R}((n-m)T)c_m \\ &\quad + \frac{(\hat{\tau}-\tau)^2}{2}\sum_m\sum_n c_n^*\ddot{R}((n-m)T)c_m \\ &\quad + \text{higher order terms}\Bigg\}\exp[-j(\hat{\phi}-\phi)]\Bigg] \\ &\cong 2\Bigg\{\mathbf{c}^{*\prime}\mathbf{R}\mathbf{c} - \frac{(\hat{\tau}-\tau)^2}{2}\mathbf{c}^{*\prime}(-\ddot{\mathbf{R}})\mathbf{c} \\ &\quad - \frac{(\hat{\phi}-\phi)^2}{2}\mathbf{c}^{*\prime}\mathbf{R}\mathbf{c}\Bigg\} \end{aligned} \tag{18}$$

where $\ddot{\mathbf{R}}$ represents a matrix whose (n,m) component is

$$\ddot{R}((n-m)T) = \left[\frac{d^2}{dt^2}R(t)\right]_{t=(n-m)T}.$$

Note that $\ddot{\mathbf{R}}$ is a negative-definite matrix, since $\ddot{R}(t-t')$ is a negative-definite function (Appendix II). In deriving (18), we used the property that

$$\dot{R}((n-m)T) = \left[\frac{d}{dt}R(t)\right]_{t=(n-m)T}$$

is skew symmetric and hence its quadratic form is always zero:

$$\mathbf{c}^{*\prime}\dot{\mathbf{R}}\mathbf{c} = \sum_m\sum_n c_n^*\dot{R}((n-m)T)c_m = 0. \tag{19}$$

On defining quantities σ_τ^2 and σ_ϕ^2 by

$$\sigma_\tau^2 = \frac{N_0}{\mathbf{c}^{*\prime}(-\ddot{\mathbf{R}})\mathbf{c}} \tag{20}$$

$$\sigma_\phi^2 = \frac{N_0}{\mathbf{c}^{*\prime}\mathbf{R}\mathbf{c}} \tag{21}$$

we can write the likelihood function in the following form:

$$L(x \mid \mathbf{c},\hat{\tau},\hat{\phi}) \cong C \cdot \exp\left\{-\frac{(\hat{\tau}-\tau)^2}{2\sigma_\tau^2} - \frac{(\hat{\phi}-\phi)^2}{2\sigma_\phi^2}\right\} \quad (22)$$

where C is a positive constant. Thus we see that the MLEs of ϕ and τ are asymptotically normally distributed with the true values as their means and with variances σ_ϕ^2 and σ_τ^2, respectively. It is known that σ_τ^2 and σ_ϕ^2 of (20) and (21) correspond to Cramer–Rao lower bounds, that is, the MLEs are asymptotically efficient [25]–[27].

For given estimates $\hat{\tau}$ and $\hat{\phi}$, the log-likelihood ratio of (16) is maximized by choosing $\hat{\mathbf{c}}$ such that

$$J = (\hat{\mathbf{c}} - \mathbf{R}^{-1}z)^{*\prime}\mathbf{R}(\hat{\mathbf{c}} - \mathbf{R}^{-1}\mathbf{z}) \quad (23)$$

is minimized. Thus we see that the maximum-likelihood decision rule is equivalent to a generalized minimum-distance decision rule. If the information sequence $\mathbf{c}$ is independent from digit to digit, the decision rule is realized by the bit-by-bit detection method in which the sequence $\mathbf{R}^{-1}\mathbf{z}$ is passed into a decision box (e.g., an amplitude-threshold detector if $\mathbf{c}$ represents an amplitude sequence). If $\mathbf{c}$ is a correlated sequence like outputs of a convolutional encoder [28], or a correlative level encoder [29] (or a partial response channel), then the maximum-likelihood decision rule takes a form quite different from the conventional bit-by-bit detection method [30]–[34].

Multiplication by the matrix $\mathbf{R}^{-1}$ corresponds to the (two-dimensional) equalization which removes intersymbol interference completely when the demodulator phase $\hat{\phi}$ and the sample timing $\hat{\tau}$ are exact. Thus the sequence $\mathbf{R}^{-1}\mathbf{z}$ is obtained by passing the received waveform $x(t)$ into the demodulator, the matched filter, the sampler, and then into the equalizer. This optimum receiver structure has been derived by Tufts [35], who reached this configuration based on the criterion that the output noise power be minimized with zero intersymbol interference. The solution under this criterion is in fact the minimum variance unbiased estimate of the sequence $\mathbf{c}$ and is known to agree with the MLE when the noise is Gaussian [25]–[27]. A further relationship between the matrix $\mathbf{R}^{-1}$ and the equalizer is summarized in Appendix III.

Let us denote the equalizer output by $\check{\mathbf{c}}$, i.e.,

$$\check{\mathbf{c}} = \mathbf{R}^{-1}\mathbf{z}. \quad (24)$$

Note that $\check{\mathbf{c}}$ is not the MLE, since we know that information sequence takes on only discrete values, whereas $\check{\mathbf{c}}$ is an analog value.[5] Equation (16) indicates that $\check{\mathbf{c}}$ is normally distributed,[6] and the variance is calculated by applying Taylor's series expansion as follows:

$$\check{\mathbf{c}} = \mathbf{R}^{-1}\mathbf{z} = \mathbf{R}^{-1}\{\mathbf{R} + (\hat{\tau}-\tau)\dot{\mathbf{R}} + \cdots\}\mathbf{c}\exp[-j(\hat{\phi}-\phi)]$$
$$\cong \mathbf{c} + (\hat{\tau}-\tau)\mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{c} - j(\hat{\phi}-\phi)\mathbf{R}^{-1}\mathbf{c}. \quad (25)$$

The second and third terms represent intersymbol interference terms introduced when the demodulator phase and sample timing are not correct. The error variance of $\check{c}_n$ is then obtained from (16) and (25):

$$E[\mid \check{c}_n - c_n \mid^2] = N_0[\mathbf{R}^{-1}]_{n,n} + \sigma_\tau^2 \mid [\mathbf{R}^{-1}\dot{\mathbf{R}}\mathbf{c}]_n \mid^2 + \sigma_\phi^2 \mid [\mathbf{R}^{-1}\mathbf{c}]_n \mid^2 \quad (26)$$

where $[\mathbf{R}^{-1}]_{n,n}$ denotes the (n,n) element of the matrix $\mathbf{R}^{-1}$ and $[\mathbf{a}]_n$ means the nth component of vector $\mathbf{a}$. The right side of (26) could have been derived from the Cramer–Rao lower bound extended to a complex-valued random variable [38]. It should be remarked here that the development from (8) to (26) is analogous to radar parameter estimation problems [38]–[41]. The MLE has also been applied to the estimation of the time of arrival and signal waveforms in array processing [42], [43].

Since we know that the log-likelihood function (16) is a concave (at least locally) function of parameters $\mathbf{c}$, τ, ϕ, we can apply the gradient technique [44], [57] (or the steepest ascent method), if the initial estimates are within the convergence region. We will defer the convergence problem till the end of Section III. Taking the derivatives of the log-likelihood function with respect to $\hat{\phi}$ and $\hat{\tau}$, we obtain

$$\frac{\partial \ln L}{\partial \hat{\phi}} = 2\,\mathrm{Re}\left\{\frac{\partial \mathbf{z}^{*\prime}}{\partial \hat{\phi}}\hat{\mathbf{c}}\right\} \quad (27)$$

$$\frac{\partial \ln L}{\partial \hat{\tau}} = 2\,\mathrm{Re}\{\dot{\mathbf{z}}^{*\prime}\hat{\mathbf{c}}\} \quad (28)$$

where the nth components of sequences $\partial\mathbf{z}/\partial\phi$ and $\dot{\mathbf{z}}$ are defined by

$$\frac{\partial z_n}{\partial \hat{\phi}} = [\{-jx(t)\exp(-j\hat{\phi})\} \otimes q^*(-t)]_{t=nT+\hat{\tau}} = -jz_n \quad (29)$$

$$\dot{z}_n = \frac{\partial z_n}{\partial \hat{\tau}} = \frac{\partial}{\partial \hat{\tau}}[\{x(t)\exp(-j\hat{\phi})\} \otimes q^*(-t)]_{t=nT+\hat{\tau}}$$
$$= \left[\frac{d}{dt}\{x(t)\exp(-j\hat{\phi}) \otimes q^*(-t)\}\right]_{t=nT+\hat{\tau}} \quad (30)$$

which is the sampled value of differentiated waveform of the matched filter output. From (27)–(30), we obtain the iterative estimation formulas

$$\hat{\phi}_{i+1} = \hat{\phi}_i - \alpha_i\,\mathrm{Im}\{\hat{\mathbf{c}}_i^{*\prime}\mathbf{z}_i\}, \qquad \alpha_i > 0 \quad (31)$$

$$\hat{\tau}_{i+1} = \hat{\tau}_i + \beta_i\,\mathrm{Re}\{\hat{\mathbf{c}}_i^{*\prime}\dot{\mathbf{z}}_i\}, \qquad \beta_i > 0 \quad (32)$$

[5] In other words, we take into account *a priori* probability of random variable $\mathbf{c}$. We could have included this in our formulation by multiplying (8) by the prior probability of $\mathbf{c}$ (and those of τ and ϕ, if available). Such an estimate is called *the unconditional maximum-likelihood estimate* [27].

[6] Complex-valued Gaussian random variable and vectors are discussed in [17], [36], and [37].

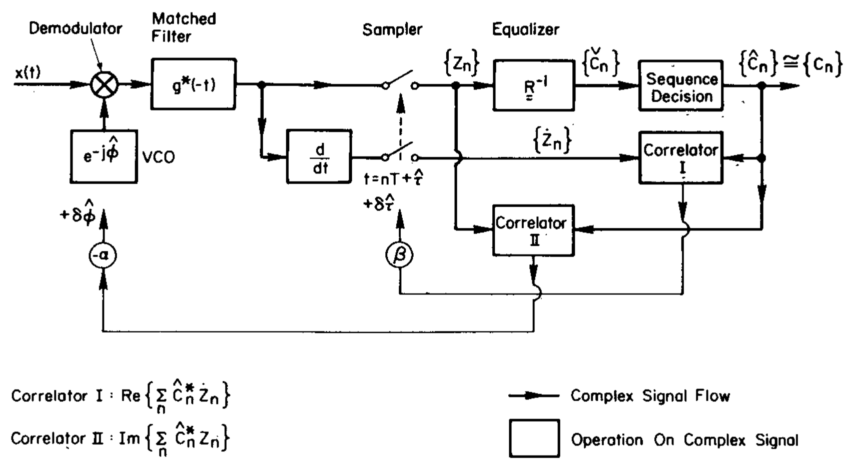

Fig. 1. Maximum-likelihood receiver and decision-directed receiver. Heavy lines represent complex signal flows or operations on complex signals.

where $\mathbf{z}_i = \mathbf{z}(\hat{\phi}_i,\hat{\tau}_i)$, and $\hat{\mathbf{c}}_i$ is the estimate of the information sequence at the ith iteration, i.e., it is the legitimate sequence $\hat{\mathbf{c}}$ that minimizes

$$J = (\hat{\mathbf{c}} - \mathbf{R}^{-1}\mathbf{z}_i)^{*\prime}\mathbf{R}(\hat{\mathbf{c}} - \mathbf{R}^{-1}\mathbf{z}_i). \tag{33}$$

The structure of the maximum-likelihood receiver (MLR) is given in Fig. 1, where the estimated sequences converge to the MLEs $\tilde{\tau}$, $\tilde{\phi}$, and $\tilde{\mathbf{c}}$:

$$\lim_{i\to\infty} \hat{\tau}_i = \tilde{\tau} \tag{34}$$

$$\lim_{i\to\infty} \hat{\phi}_i = \tilde{\phi} \tag{35}$$

$$\lim_{i\to\infty} \hat{\mathbf{c}}_i = \tilde{\mathbf{c}}. \tag{36}$$

The optimum gain sequence α_i and β_i in (31) and (32) are chosen in such a way that the next approximation gives the point $(\hat{\tau}_{i+1},\hat{\phi}_{i+1})$ that maximizes the log-likelihood function (16) over all points on the line of action of the gradient passing through $(\hat{\tau}_i,\hat{\phi}_i)$. The optimum values of α_i and β_i are derived in Appendix IV, when $(\hat{\tau}_i,\hat{\phi}_i)$ are close[7] to their true values and the SNR is high. These values are

$$\alpha_i = \frac{1}{\mathbf{c}^{*\prime}\mathbf{R}\mathbf{c}} = \frac{\sigma_\phi^2}{N_0} \tag{37}$$

$$\beta_i = \frac{1}{\mathbf{c}^{*\prime}(-\ddot{R})\mathbf{c}} = \frac{\sigma_\tau^2}{N_0} \tag{38}$$

where σ_ϕ^2 and σ_τ^2 are Cramer–Rao bounds of $\hat{\phi}$ and $\hat{\tau}$. (See (20) and (21).)

III. Decision-Directed Receiver

The MLR obtained in the preceding section is of a recursive type, that is, the received signal is processed repeatedly until the estimate of unknown parameters converge to their final values. Although this recursive algorithm provides us with the best estimate we can obtain, this is clearly an impractical scheme, since we have to store all the data received during the observation period. Moreover, computation must be performed for the whole data by applying the formulas (31)–(33) iteratively. We shall be able, however, to develop a more practical estimation algorithm by modifying the structure obtained before.

Let us divide the observation interval into sequential disjoint subintervals of N digit length. Then the observed data $x(t)$, $0 \le t < \infty$, is transformed into a sequence of random processes $x_1(t'),x_2(t'),\cdots,x_k(t')$, $0 \le t' \le NT$, such that

$$x_k(t') = \begin{cases} x(t + (k-1)NT), & 0 \le t' \le NT \\ 0, & \text{elsewhere} \end{cases} \tag{39}$$

where $k = 1,2,3,\cdots$. We assume that the processes $x_k(t')$, $k = 1,2,3\cdots$, are independently and identically distributed. Then we shall be able to apply to our problem the Robbins–Monro stochastic approximation method [6]. Sakrison [7], [8] discusses extensively applications of the stochastic approximation method to parameter estimation problems in various communication systems. His results are directly applicable to the present problem. Tong and Liu [45] applied the stochastic approximation method to the automatic equalization problem.

Let $L(x_k \mid \tau,\phi)$ be the likelihood function defined for the kth interval:

$$L(x_k \mid \tau,\phi) = \exp\left\{\frac{1}{2N_0}([x_k,r_k]_K + [r_k,x_k]_K - [r_k,r_k]_K)\right\}. \tag{40}$$

The true values of ϕ and τ will be those values $\hat{\phi}$ and $\hat{\tau}$ for which the expected value of the gradients of the log-likelihood function associated with the parameters $\hat{\phi}$ and $\hat{\tau}$ are zero:

$$E\left[\frac{\partial}{\partial\hat{\phi}} \ln L(x_k \mid \hat{\tau},\hat{\phi})\right] = 0 \tag{41}$$

[7] We mean here that the point $(\hat{\tau}_i,\hat{\phi}_i)$ belongs to the region in which Taylor's expansion of the log-likelihood function around (τ,ϕ) gives a good approximation.

and

$$E\left[\frac{\partial}{\partial \hat{\tau}} \ln L(x_k \mid \hat{\tau},\hat{\phi})\right] = 0 \tag{42}$$

where E means the expectation with respect to both the additive noise process $n_k(t)$ and the data sequence $\mathbf{c}_k$. The Robbins–Monro method is an algorithm that uses the sequential observations $\{x_k(t),\ k = 1,2,\cdots\}$ to iteratively estimate the values of ϕ and τ which satisfy (41) and (42). As will be shown later, the Robbins–Monro method requires less computation than the MLE method discussed in Section II. Furthermore, the Robbins–Monro method can yield asymptotically efficient estimates, so that for a large number of observations the error variance is as small as can be obtained by the MLE method.

On applying the Robbins–Monro method, we generate the sequence of estimates $\{\hat{\phi}_k\}$ and $\{\hat{\tau}_k\}$ by the following recursive relation:

$$\hat{\phi}_{k+1} = \hat{\phi}_k + \alpha_k \left[\frac{\partial}{\partial \hat{\phi}} \ln L(x_k \mid \hat{\tau}_k,\hat{\phi})\right]_{\hat{\phi}=\hat{\phi}_k}$$

$$= \hat{\phi}_k + \frac{\alpha_k}{N_0} \cdot \mathrm{Im}\,\{\mathbf{c}_k^{*\prime}\mathbf{z}_k\} \tag{43}$$

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \beta_k \left[\frac{\partial}{\partial \hat{\tau}} \ln L(x_k \mid \hat{\tau},\hat{\phi}_k)\right]_{\hat{\tau}=\hat{\tau}_k}$$

$$= \hat{\tau}_k + \frac{\beta_k}{N_0} \cdot \mathrm{Re}\,\{\mathbf{c}_k^{*\prime}\dot{\mathbf{z}}_k\} \tag{44}$$

where $\mathbf{z}_k$ is a vector whose elements are

$$z_{k,n} = [\{x_k(t) \exp(-i\hat{\phi}_k)\} \otimes q^*(-t)]_{t=nT+\hat{\tau}_k}. \tag{45}$$

Note that (43) and (44) are similar to (31) and (32), with the exception that in the present recursive formula the sequentially observed data $\{x_k(t)\}$ is used only once, whereas in (31) and (32), the data $x(t)$ observed over the entire period is processed repeatedly.

The second terms of (43) and (44) include the information sequence $\mathbf{c}_k$, not their estimate $\hat{\mathbf{c}}_k$. This condition is clearly met in the initial training period during which some sequence of known pattern is sent. The condition will be practically satisfied in a DDR as far as the majority of decision outputs are correct and estimates of τ and ϕ are improved as the number of iterations increases. Strictly speaking, however, there is the possibility of a "run away" in the decision-directed approach. This occurs when the detector makes a series of errors resulting in a degradation of parameter estimates which, in turn, results in a further performance deterioration. A DDR for synchronous detection was investigated by Proakis *et al.* [46]. Their study showed that the run away, though acknowledged as a possibility in theory, was not observed in simulations. Another successful application of the DDR approach is the adaptive channel equalization done by Lucky [1], [10] and by others [12], [13]. Analysis of a DDR is quite difficult because of the dependence introduced in each iteration. Davisson and Schwartz discuss the run-away problem in estimating unknown prior probabilities in a binary detection problem [47] and in estimating unknown transition probabilities of a Markov sequence [48]. The run-away problem in the DDR will be further discussed at the end of this section.

We now make the following assumptions.

Assumption 1) The processes $\{x_k(t),\ k = 1,2,\cdots\}$ are independently and identically distributed, that is, the additive noise is stationary and independent among the subintervals, and the information sequence is also stationary with zero mean and with correlation function

$$E[c_{k,i}c_{k,j}^*] = S_0\Phi_{i-j}, \qquad \text{for all } k. \tag{46}$$

Assumption 2) There exist constants $0 < C_\phi \leq C_\phi' < \infty$ and $0 < C_\tau \leq C_\tau' < \infty$, such that

$$-C_\phi(\hat{\phi} - \phi)^2 \geq E\left[\frac{\partial}{\partial \hat{\phi}} \ln L(x_k \mid \hat{\tau},\hat{\phi})\right]\cdot(\hat{\phi} - \phi) \geq -C_\phi'(\hat{\phi} - \phi)^2 \tag{47}$$

$$-C_\tau(\hat{\tau} - \tau)^2 \geq E\left[\frac{\partial}{\partial \hat{\tau}} \ln L(x_k \mid \hat{\tau},\hat{\phi})\right]\cdot(\hat{\tau} - \tau) \geq -C_\tau'(\hat{\tau} - \tau)^2 \tag{48}$$

or equivalently

$$C_\phi(\hat{\phi} - \phi)^2 \leq \frac{S_0}{N_0}(\hat{\phi} - \phi) \cdot \mathrm{Im}\,\{\mathrm{tr}\,[\mathbf{R}(\hat{\tau} - \tau)\cdot\mathbf{\Phi}] \exp[j(\hat{\phi} - \phi)]\} \leq C_\phi'(\hat{\phi} - \phi)^2 \tag{49}$$

$$C_\tau(\hat{\tau} - \tau)^2 \leq \frac{S_0}{N_0}(\hat{\tau} - \tau) \cdot \mathrm{Re}\,\{\mathrm{tr}\,[\dot{\mathbf{R}}(\hat{\tau} - \tau)\cdot\mathbf{\Phi}] \exp[j(\hat{\phi} - \phi)]\} \leq C_\tau'(\hat{\tau} - \tau)^2 \tag{50}$$

for any $\hat{\phi}$ and $\hat{\tau}$ in the convergence region. In other words, the convergence regions $\hat{\phi}$ and $\hat{\tau}$ are determined by (49) and (50).

Assumption 3) For all values of $\hat{\tau}$ and $\hat{\phi}$ in the convergence region, we have

$$\mathrm{var}\left\{\frac{\partial}{\partial \hat{\phi}} \ln L(x_k \mid \hat{\tau},\hat{\phi})\right\} = \frac{S_0}{N_0}\,\mathrm{tr}\,[\mathbf{R}\cdot\mathbf{\Phi}] \triangleq d_\phi^2 < \infty \tag{51}$$

$$\mathrm{var}\left\{\frac{\partial}{\partial \tau} \ln L(x_k \mid \hat{\tau},\hat{\phi})\right\} = \frac{S_0}{N_0}\,\mathrm{tr}\,[-\ddot{\mathbf{R}}\cdot\mathbf{\Phi}] \triangleq d_\tau^2 < \infty. \tag{52}$$

Assumption 4) The gain sequences $\{\alpha_k\}$ and $\{\beta_k\}$ are positive, monotonically decreasing, and satisfy

$$\sum_{k=1}^{\infty} \alpha_k = \infty, \qquad \sum_{k=1}^{\infty} \beta_k = \infty \tag{53}$$

and

$$\sum_{k=1}^{\infty} \alpha_k^2 < \infty, \qquad \sum_{k=1}^{\infty} \beta_k^2 < \infty. \tag{54}$$

Then it can be shown (Appendix V) that the sequence of estimates $\{\hat{\phi}_k\}$ and $\{\hat{\tau}_k\}$ converge to their true values in mean-square:

$$\lim_{k\to\infty} E[(\hat{\phi}_k - \phi)^2] = 0 \tag{55}$$

$$\lim_{k\to\infty} E[(\hat{\tau}_k - \tau)^2] = 0. \tag{56}$$

The choice of the gain sequences affect both the convergence rate and the error variance. Sequences which satisfy (53) and (54) are

$$\alpha_k = \frac{A}{k^\delta}, \qquad \frac{1}{2} < \delta \le 1 \tag{57}$$

$$\beta_k = \frac{B}{k^\epsilon}, \qquad \frac{1}{2} < \epsilon \le 1. \tag{58}$$

From the results established by Chung [49] we know that sequences

$$\alpha_k = \frac{A}{k}, \qquad \beta_k = \frac{B}{k} \tag{59}$$

give the most rapid convergence of $\hat{\phi}_k$ and $\hat{\tau}_k$ to ϕ and τ, respectively. For this choice of α_k and β_k, the asymptotic distribution of $k^{1/2}(\hat{\phi}_k - \phi)$ and $k^{1/2}(\hat{\tau}_k - \phi)$ are both Gaussian with zero mean and variances [50]

$$\frac{A^2 d_\phi^2}{2Ad_\phi^2 - 1}, \qquad \frac{B^2 d_\tau^2}{2Bd_\tau^2 - 1}. \tag{60}$$

Therefore the asymptotic variances of (60) are minimized by choosing

$$A = \frac{1}{d_\phi^2}, \qquad B = \frac{1}{d_\tau^2} \tag{61}$$

and hence

$$\lim_{k\to\infty} E[k^{1/2}(\hat{\phi}_k - \phi)]^2 = d_\phi^2 \tag{62}$$

$$\lim_{k\to\infty} E[k^{1/2}(\hat{\tau}_k - \tau)]^2 = d_\tau^2. \tag{63}$$

The right-hand sides of (62) and (63) are equal to the attainable lower bound on the mean-square error of unconditional estimates [38]. Therefore we can conclude that the Robbins–Monro method can yield the asymptotically efficient estimate so that, for a sufficiently long period of observation, the error variance is as small as can be obtained by any other method.

Before closing this section, an important question associated with the MLR and the DDR should be answered; namely, a possible convergence to a local maximum rather than to the true values (τ,ϕ). If the information sequence **c** is known (for example, during the initial training period), the convergence to a local maximum seems hardly a problem. The log-likelihood function of (16) can be written, in the absence of noise, as

$$\begin{aligned} 2N_0 \ln L(x \mid \mathbf{c},\hat{\tau},\hat{\phi}) &= 2\,\mathrm{Re}\left\{\sum_m \sum_n c_n^* R((n-m)T + \hat{\tau} - \tau) c_m \exp[j(\hat{\phi} - \phi)]\right\} \\ &\quad - \sum_m \sum_n c_n^* R((n-m)T) c_m. \end{aligned} \tag{64}$$

If the sequence **c** is of a sufficient length and is independent, i.e.,

$$E[c_n^* c_m] = S_0 \delta_{n-m,0} \tag{65}$$

then (64) is approximated by

$$2N_0 \ln L(x \mid \mathbf{c},\hat{\tau},\hat{\phi}) = KS_0[2\,\mathrm{Re}\,\{R(\hat{\tau} - \tau) \exp[j(\hat{\phi} - \phi)]\} - R(0)] \tag{66}$$

where K is the size of the sequence **c**. For a given modulation scheme, signal waveform, and channel characteristic, (66) is a function of $\hat{\tau} - \tau$ and $\hat{\phi} - \phi$. We see from (66) that the convergence to a local maximum should not occur if the initial error $(\hat{\tau} - \tau)$ is within the main lobe of the function $R(t)$ observed at the matched filter output.

If the sequence **c** represents the information in a multiphase modulation system (see (1d)), i.e., if $\{c_n\}$ takes on one of L discrete phases, say,

$$l\Delta = l\frac{2\pi}{L}, \qquad l = 1,2,\cdots,L \tag{67}$$

then there arises a problem of ambiguity in the DDR. In other words, if **c** is some information sequence, then $\mathbf{c} \exp(jl\Delta)$ is another legitimate sequence. Therefore, once the phase of the demodulator falls in the convergence region of phase $\phi + l\Delta$, it will converge to $\phi + l\Delta$ rather than to the true value ϕ. This ambiguity problem will be solved by using the differential phase coding [1]. This coding is usually adopted together with the comparison detection method in which the phase of the preceding digit is used as the reference [51], [52] for the present digit. The binary antipodal signaling is equivalent to the phase modulation of $L = 2$, hence the ambiguity problem can be solved. However, in a multilevel amplitude-modulation system there exists no coding method to compensate for this possible ambiguity of the phase, unless some redundancy is introduced in the sequence **c**.

An alternative method is the use of a pilot carrier as in the conventional PAM-VSB (or -SSB) system. Let A be the amplitude of the pilot carrier which is added in quadrature with the modulated carrier. Then the complex envelope of the transmitted signal is

$$s_0(t) = s(t) + jA \exp(j\phi) \tag{68}$$

where $s(t)$ is defined by (2). Correspondingly, the received signal is

$$x(t) = r(t) + jB \exp(j\phi) + n(t) \qquad (69)$$

where $r(t)$ is given by (5), and B is a complex number given by

$$B = A \int_{-\infty}^{\infty} h_b(t)\, dt. \qquad (70)$$

Then the log-likelihood ratio function is given by the following

$$\begin{aligned} 2N_0 \ln L(x \mid \mathbf{c},\hat{\tau},\hat{\phi}) &= 2\,\mathrm{Re}\,\{[r,x]_K\} - [r,r]_K + 2\,\mathrm{Re}\,[jB \exp(j\hat{\phi}),x]_K \\ &\quad - 2\,\mathrm{Re}\,\{[jB \exp(j\hat{\phi}),r]_K\} - \frac{|B|^2}{k_0}. \end{aligned} \qquad (71)$$

The first two terms are the same as (8), and the last two terms are not related to the received input $x(t)$. A new term to be added to the receiver is then

$$\mathrm{Re}\,[jB \exp(j\hat{\phi}),x]_K = \mathrm{Im}\left\{\frac{B^*}{k_0}\int_{-\infty}^{\infty} x(t) \exp(-j\hat{\phi})\, dt\right\} \qquad (72)$$

where the constant k_0 in (71) and (72) is

$$\begin{aligned} k_0 &= \frac{1}{\mathfrak{F}\{K^{-1}(t)\}_{\omega=0}} \\ &= \mathfrak{F}\{K(t)\}_{\omega=0} > 0 \end{aligned} \qquad (73)$$

which is the zero-frequency component of the Fourier transform of $K(t)$. The operation represented by (72) should, in reality, be achieved by a filter with some bandwidth (e.g., DLL) rather than an ideal integrator in order to track a possible phase jitter or a frequency shift of the pilot carrier. Notice that the multiplicative factor B^* is required to compensate for the phase shift which the pilot carrier receives in the transmission medium. If both the noise process and the sequence $\mathbf{c}$ have zero mean, we can use the following approximation:

$$2\,\mathrm{Im}\left\{\frac{B^*}{k_0}\int_{-\infty}^{\infty} x(t) \exp(-j\hat{\phi})\, dt\right\} \cong \frac{2|B|^2 T_0}{k_0} \cos(\hat{\phi} - \phi) \qquad (74)$$

where T_0 is the integration period. Thus the ambiguity of the phase can be avoided by resorting to the pilot carrier. From (22) and (74), the variance of the phase estimate $\hat{\phi}$ is improved to

$$\begin{aligned} \mathrm{var}\,\{\hat{\phi}\} &= \frac{1}{1/\sigma_\phi^2 + |B|^2 T_0/k_0 N_0} \\ &= \frac{N_0}{\mathbf{c}^{*\prime}\mathbf{R}\mathbf{c} + |B|^2 T_0/k_0} \end{aligned} \qquad (75)$$

in the case of the MLR. Similarly, the variance of $\hat{\phi}_k$ in the DDR is

$$\begin{aligned} E[k^{1/2}(\hat{\phi} - \phi)]^2 &= \frac{1}{1/d^2 + |B|^2 NT/k_0 \cdot N_0} \\ &= \frac{N_0}{S_0\,\mathrm{tr}\,(\mathbf{R}\mathbf{\Phi}) + |B|^2 NT/k_0} \end{aligned} \qquad (76)$$

where N is the number of terms in each subinterval (see (39)).

The analysis in this section has been performed only for the case where the observation interval is divided into sequential disjoint subintervals. This is solely because the stochastic approximation method can be successfully applied under this assumption. In practice, however, one may want to update $\hat{\tau}$ and $\hat{\phi}$ at the digit rate, i.e., at every T seconds. Such an algorithm may be explained as a discrete version of a PLL system. Performance analysis of a discrete PLL seems difficult, since the Fokker–Planck equation [4], [53] cannot be well defined in this case.

IV. Adaptive Receivers for Unknown Channels

In the previous section it has been assumed that the channel response is known and time invariant. Matched filters are used at the demodulator output, obtaining the maximum signal-to-noise ratio. Equalization filters are represented by the $N \times N$ matrix $\mathbf{R}^{-1}$ (which becomes a stationary filter as N approaches infinity). In actual situations, however, the channel characteristic is not exactly known; moreover, it may be slowly changing with time. In this section, an adaptive algorithm for adjusting the carrier phase and sample instant will be discussed in conjunction with that of adaptive equalization. Adaptive equalization has been extensively studied by Lucky [1], [10] and by others [11]–[13]. In our proposed scheme the demodulator, sampler, and equalizer work interactively to seek the joint optimization.

Since the channel response function $h_b(t)$ (see (7)) is not known exactly, the signal waveform $g(t)$ and the matched filter $q^*(-t)$ cannot be specified. Therefore, we are no longer able to define the likelihood ratio function in a meaningful way. Let the matched filter $q^*(-t)$ be replaced by some low-pass filter, and let the sampled output be denoted again by z_n. The equalizer $\{w_i\}$ defined by (94) is no longer valid, since the unit signal response function observed at the low-pass filter is not equal to $R(t)$. Thus we must change the equalizer to an adaptive one. The criterion we choose here is minimization of the distance function

$$J = (\mathbf{c} - \mathbf{W}\mathbf{z})^{*\prime}\mathbf{P}(\mathbf{c} - \mathbf{W}\mathbf{z}) \qquad (77)$$

where $\mathbf{W}$ is a Toeplitz matrix whose (i,j) component is equal to w_{j-i}, and $\mathbf{P}$ is some positive definite matrix. The vector $\mathbf{W}\mathbf{z}$ represents the equalizer output sequence. Note that (77) is similar to (23), or the second term of (16).

Let the equalizer be of a finite size, i.e., $w_i = 0$, for $|i| > M$, and let us choose the simplest cost matrix

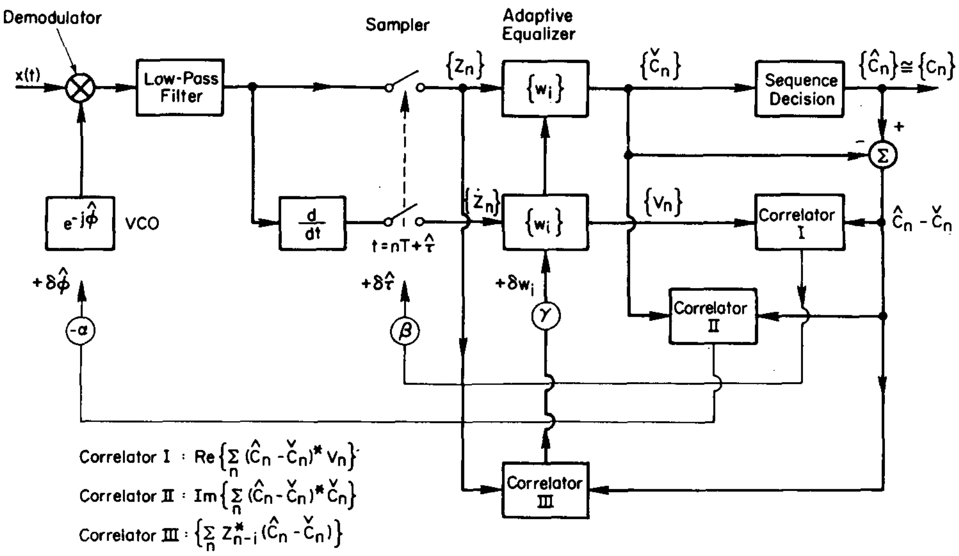

Fig. 2. Adaptive receiver.

$\mathbf{P} = \mathbf{I}$, the identity matrix. Then the performance index J becomes the sum of mean-squared errors:

$$J = \sum_n | c_n - \sum_{i=-M}^{M} w_i z_{n-i} |^2. \tag{78}$$

In order to minimize J, its gradient with respect to $\hat{\phi}$, $\hat{\tau}$, and w_i must be obtained:

$$\frac{\partial J}{\partial \hat{\phi}} = 2 \operatorname{Re} [\sum_n \{ (c_n - \sum_{i=-M}^{M} w_i z_{n-i})^* (\sum_{i'=-M}^{M} j w_{i'} z_{n-i'}) \}]$$
$$= 2 \operatorname{Im} [\sum_n (c_n - \check{c}_n)^* \check{c}_n] = 2 \operatorname{Im} [\sum_n c_n^* \check{c}_n] \tag{79}$$

where

$$\check{c}_n = \sum_{i=-M}^{M} w_i z_{n-i} \tag{80}$$

is the equalizer output (analog voltage), whereas c_n is the true sequence which takes on discrete values. In the DDR $\{c_n\}$ is substituted by the decision output $\{\hat{c}_n\}$, where we assume the majority of the decisions are correct. Similarly, the gradient of J with respect to $\hat{\tau}$ is

$$\frac{\partial J}{\partial \hat{\tau}} = 2 \operatorname{Re} [\sum_n \{ (c_n - \sum_{i=-M}^{M} w_i z_{n-i})^* (- \sum_{i'=-M}^{M} w_{i'} \dot{z}_{n-i'}) \}]$$
$$= -2 \operatorname{Re} [\sum_n (c_n - \check{c}_n)^* v_n] \tag{81}$$

where

$$v_n = \sum_{i=-M}^{M} w_i \dot{z}_{n-i} \tag{82}$$

which is obtained by passing the sequence $\{\dot{z}_n\}$ into the same filter as the equalizer.

The tap gain $\{w_i\}$ of the equalizer is also adjusted according to the gradients

$$\frac{\partial J}{\partial \operatorname{Re} \{w_i\}} + j \frac{\partial J}{\partial \operatorname{Im} \{w_i\}} = 2 \frac{\partial J}{\partial w_i^*} = \sum_n \{ z_{n-i}^* (c_n - \check{c}_n) \} \tag{83}$$

which is obtained by cross correlating the equalizer output sequence with the equalizer input. The structure of the adaptive receiver derived above is diagrammatically shown in Fig. 2, where γ is the gain to be multiplied with the gradient of (83). Here again the demodulator and filter take a complex form as explained in Appendix I.

V. Conclusions and Remarks

The problems of sequence decision, demodulation, and sample timing in carrier-modulated data-transmission systems have been treated from the viewpoint of multiparameter estimation theory. A general structure of the MLR was obtained for a class of linear modulation systems such as PAM-DSB, PAM-VSB (or -SSB), QAM, digital PM, differential PM, and PAM-PM. Asymptotic expressions for the variances of the MLE have been derived.

A DDR has been derived based on the MLR obtained. Convergence and asymptotic efficiency of this sequential-type receiver have been shown by applying the Robbins–Monro stochastic approximation method.

The structure of the receivers obtained in the present paper is different from the conventional one in the method of extracting the carrier phase. It has been shown that the carrier can be estimated from the modulated signal components themselves. This new structure suggests the use of a pilot carrier only for the purpose of solving the phase ambiguity, if needed. It is expected that distortions of demodulator output obtained by the conventional PLL technique will be greatly reduced. It has been shown analytically also that sampled values of the demodulator output provide a sufficient statistic for controlling the receiver. Thus the algorithm is suitable for digital implementation.

The structure of the DDR is then extended to the case in which the channel characteristic is unknown. The algorithm for adjusting the carrier phase and sample timing is discussed in combination with that of adaptive equalization. Equalizers thus obtained eliminate both intersymbol and interchannel interferences. In the proposed scheme, the demodulation, sampling, equalization, and sequence decision algorithms work interactively together to seek the operating point of the joint optimization.

Throughout this paper, iterative estimation formulas were derived from the method of the steepest descent. The analysis can be extended to different algorithms such as the conjugate gradient method [54]. Applications of the conjugate gradient method to equalization and to other linear filter synthesis problems under the minimum mean-square-error criterion have been discussed by Devieux and Pickholtz [55] and by the present author [56]. It is known that convergence rates of the steepest descent method and the conjugate gradient method are geometric when the optimum gain coefficients are chosen [12], [56], [57].

Appendix I

Complex-Envelope and Complex-Filter Representation

The complex-envelope representation of a narrow-band signal has been used in the literature [17]–[20]. In this appendix we simply state the relation between equations and the corresponding structures, which may help the reader to apply our results stated in a general form to the modulation system he specifically designs.

Demodulation

$$x(t)\exp(-j\hat{\phi}). \tag{84}$$

$x(t)$ is the complex envelope of the receiver input; thus the actual input is written as $\mathrm{Re}\{x(t)\exp(j\omega_0 t)\}$. The term $x(t)\exp(-j\hat{\phi})$ contains the real part (or inphase component) and the imaginary part (or quadrature component). The block-diagram representations are given in Fig. 3.

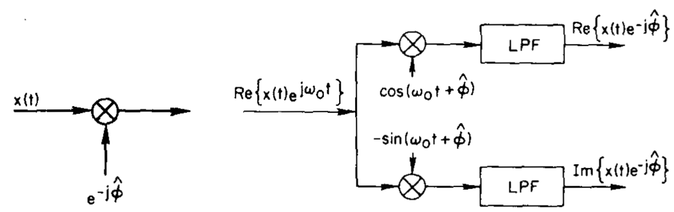

Fig. 3. Demodulator: complex representation and its implementation.

Linear Filtering

$$z(t) = y(t) \otimes h(t). \tag{85}$$

Let $y(t)$ be the input to a linear filter $h(t)$ and let $z(t)$ be the corresponding output. They are all complex functions. Letting $y(t)$ and $h(t)$ be decomposed into real and imaginary parts,

$$y(t) = y^{(1)}(t) + jy^{(2)}(t)$$
$$h(t) = h^{(1)}(t) + jh^{(2)}(t)$$

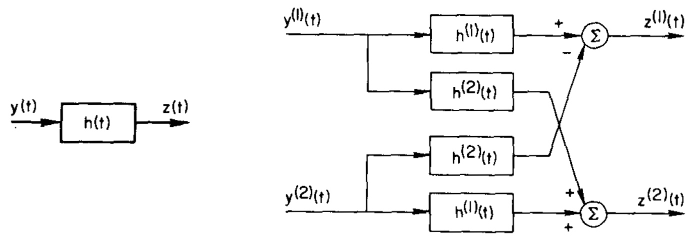

Fig. 4. Linear filter $h(t) = h^{(1)}(t) + jh^{(2)}(t)$.

we have the following matrix representation:

$$\begin{bmatrix} z^{(1)}(t) \\ z^{(2)}(t) \end{bmatrix} = \begin{bmatrix} h^{(1)}(t) & -h^{(2)}(t) \\ h^{(2)}(t) & h^{(1)}(t) \end{bmatrix} \otimes \begin{bmatrix} y^{(1)}(t) \\ y^{(2)}(t) \end{bmatrix}.$$

We thus obtain Fig. 4.

Appendix II

Negative Definiteness of $\ddot{R}(t - t')$

From the definition of (15)

$$R(t-t') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g^*(u-t)K^{-1}(u-u') \cdot g(u'-t')\,du\,du'. \tag{86}$$

Since the covariance function $K(u-u')$ is a positive-definite function, so is its inverse kernel $K^{-1}(t-t')$. Then it immediately follows that $R(t-t')$ is positive definite.

The second derivative $\ddot{R}(t-t')$ is obtained by differentiating $R(t-t')$ with respect to t and $(-t')$, i.e.,

$$\ddot{R}(t-t') = -\frac{d^2}{dt\,dt'}R(t-t') = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dot{g}^*(u-t)K^{-1}(u,u') \cdot \dot{g}(u'-t')\,du\,du'. \tag{87}$$

It is then clear that $\ddot{R}(t-t')$ is negative definite:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^*(t)\ddot{R}(t-t')x(t')\,dt\,dt' = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^*(u)K^{-1}(u-u')y(u')\,du\,du' < 0 \tag{88}$$

for any function $x(t)$, where

$$y(u) = \int_{-\infty}^{\infty} \dot{g}(u-t)x(t)\, dt. \tag{89}$$

Appendix III

Matrix $\mathbf{R}^{-1}$ and Equalizer

If the dimension of vector $\mathbf{c}$ is N, then the matrix $\mathbf{R}$ is an $N \times N$ positive-definite Hermitian matrix (see (15)). Furthermore, $\mathbf{R}$ is a Toeplitz matrix [58]. When N is finite, the operator $\mathbf{R}^{-1}$ can be regarded as a nonstationary linear filter since $\mathbf{R}^{-1}$ is, in general, not a Toeplitz form. However, if N is sufficiently large, that is, if the autocorrelation function $R(t)$ of (15) is virtually zero for $|t| > NT$ (where T is the bit interval), then $\mathbf{R}^{-1}$ can be approximated by a stationary filter such as a transversal filter. This asymptotic relation may be best understood by applying the z-transform method.

Let $R(z)$ be the z transform of the sampled autocorrelation function $R(kT)$, $k = 0,\pm 1,\pm 2,\cdots$, i.e.,

$$R(z) = \sum_{k=-\infty}^{\infty} R(kT)z^{-k}. \tag{90}$$

Then the z transform of the equalizer output $\{\check{c}_k\}$ is

$$\check{C}(z) = \sum_{k=-\infty}^{\infty} \check{c}_k z^{-k} = \frac{Z(z;\hat{\phi},\hat{\tau})}{R(z)} \tag{91}$$

where[8]

$$Z(z;\hat{\phi},\hat{\tau}) = \sum_{k=-\infty}^{\infty} z_k(\hat{\phi},\hat{\tau})z^{-k}. \tag{92}$$

The equalizer $1/R(z)$ is realized by a transversal filter with tap gains $\{w_i\}$ which are in general complex numbers,

$$\frac{1}{R(z)} = \sum_{i=-\infty}^{\infty} w_i z^{-i}. \tag{93}$$

The coefficients w_i are obtained from (93) as

$$w_i = \frac{1}{2\pi j}\oint_{|z|=1} \frac{z^{i-1}}{R(z)}\, dz = \int_{-1/2}^{1/2} \frac{\exp(j2\pi i\lambda)}{R(\exp(j2\pi\lambda))}\, d\lambda \tag{94}$$

where $j = (-1)^{1/2}$. Furthermore, the entries of the $N \times N$ matrix $\mathbf{R}^{-1}$ and the set of coefficients $\{w_i\}$ possess the following asymptotic relation:

$$w_i = \lim_{N\to\infty} [\mathbf{R}^{-1}]_{k,k+i}, \qquad \text{for all } k. \tag{95}$$

Then the first term of (26) is asymptotically equal to $N_0 w_0$, where w_0 is the center tap gain of the equalizer. From this last result we see that the center tap gain is a positive real number. This could have been derived from (94).

[8] The z is the z-transform variable, and Z is used to denote the z transform of the random sequence of (12).

Appendix IV

Optimum Gain Sequences α_i and β_i

Define the following performance index $I(\hat{\phi},\hat{\tau},\hat{\mathbf{c}})$:

$$I(\hat{\phi},\hat{\tau},\hat{\mathbf{c}}) = \mathbf{z}^{*\prime}\hat{\mathbf{c}} + \hat{\mathbf{c}}^{*\prime}\mathbf{z} - \hat{\mathbf{c}}^{*\prime}\mathbf{R}\hat{\mathbf{c}} \tag{96}$$

which is proportional to the log-likelihood function of (16). The performance index after the $(i+1)$th iteration is given by

$$I_{i+1} = \mathbf{z}_{i+1}^{*\prime}\hat{\mathbf{c}}_{i+1} + \hat{\mathbf{c}}_{i+1}^{*\prime}\mathbf{z}_{i+1} - \hat{\mathbf{c}}_{i+1}^{*\prime}\mathbf{R}\hat{\mathbf{c}}_{i+1}. \tag{97}$$

Substituting (31) and (32) into (97) and expanding around $\alpha_i = 0$ and $\beta_i = 0$, we obtain the following formula under a high SNR condition, approximating the performance index by a parabola:

$$I_{i+1} = I_i + [\operatorname{Im}\{\mathbf{z}_i^{*\prime}\mathbf{c}\}]\{2\alpha_i - \alpha_i^2\mathbf{c}^{*\prime}\mathbf{R}\mathbf{c}\} + [\operatorname{Re}\{\dot{\mathbf{z}}^{*\prime}\mathbf{c}\}]\{2\beta_i - \beta_i^2\mathbf{c}^{*\prime}(-\ddot{R})\mathbf{c}\}. \tag{98}$$

Here we replaced $\hat{\mathbf{c}}_i$ by $\mathbf{c}$, since within the convergence region we can assume that the maximum-likelihood decision output sequence should be close to the true sequence after some iterations. Then the optimum choice of α_i and β_i of (37) and (38) immediately follows.

Appendix V

Proof of Convergence for (55) and (56)

The essence of the procedure employed in the following proof is found in Sakrison [7]. On defining the quantities p_k and q_k by

$$p_k = \frac{1}{N_0}\operatorname{Im}\{\mathbf{c}_k^{*\prime}\mathbf{z}_k\} \tag{99}$$

$$q_k = \frac{1}{N_0}\operatorname{Re}\{\mathbf{c}_k^{*\prime}\dot{\mathbf{z}}_k\} \tag{100}$$

(43) and (44) become

$$\hat{\phi}_{k+1} = \hat{\phi}_k + \alpha_k p_k \tag{101}$$

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \beta_k q_k. \tag{102}$$

On subtracting true values from both sides of (101) and (102) and squaring, we have

$$(\hat{\phi}_{k+1} - \phi)^2 = (\hat{\phi}_k - \phi)^2 + 2\alpha_k p_k(\hat{\phi}_k - \phi) + \alpha_k^2 p_k^2 \tag{103}$$

$$(\hat{\tau}_{k+1} - \tau)^2 = (\hat{\tau}_k - \tau)^2 + 2\beta_k q_k(\hat{\tau}_k - \tau) + \beta_k^2 q_k^2. \tag{104}$$

Because of Assumption 1) of Section III, the kth estimates $\hat{\phi}_k$ and $\hat{\tau}_k$ are random variables which depend on $x_1(t),x_2(t),\cdots,x_{k-1}(t)$, but not on $x_k(t)$. Therefore, we obtain the expectation of p_k and p_k^2 conditioned on $\hat{\tau}_k$

and $\hat{\phi}_k$:

$$E[p_k \mid \hat{\tau}_k, \hat{\phi}_k] = \frac{1}{N_0} E[\mathrm{Im}\,\{\mathbf{c}_k{}^{*\prime}\mathbf{z}_k(\hat{\tau}_k,\hat{\phi}_k)\}]$$

$$= \frac{1}{N_0}\mathrm{Im}\,\{\mathrm{tr}\,[\mathbf{R}(\hat{\tau}-\tau)\Phi]\exp\,[j(\hat{\phi}_k-\phi)]\}$$

$$\triangleq l(\hat{\tau}_k - \tau, \hat{\phi}_k - \phi) \tag{105}$$

$$E[p_k{}^2 \mid \hat{\tau}_k, \hat{\phi}_k] = l^2(\hat{\tau}_k - \tau, \hat{\phi}_k - \phi) + \mathrm{var}\,\{p_k\}. \tag{106}$$

By substituting (105) and (106) into (103), we obtain

$$E[(\hat{\phi}_{k+1} - \phi)^2 \mid \hat{\tau}_k, \hat{\phi}_k]$$
$$= (\hat{\phi}_k - \phi)^2 + 2\alpha_k(\hat{\phi}_k - \phi)l(\hat{\tau}_k - \tau, \hat{\phi}_k - \phi)$$
$$+ \alpha_k{}^2\{l^2(\hat{\tau}_k - \tau, \hat{\phi}_k - \phi) + \mathrm{var}\,p_k\}$$
$$\leq (\phi_k - \phi)^2[1 - 2\alpha_k C_\phi + \alpha_k{}^2 C_\phi{}'^2] + \alpha_k{}^2 d_\phi{}^2. \tag{107}$$

The last expression was obtained by use of Assumptions 2) and 3). Then taking the expected values of both sides of (107) with respect to $\hat{\tau}_k$ and $\hat{\phi}_k$, and using the notation

$$b_k = E[(\hat{\phi}_k - \phi)^2] \tag{108}$$

we obtain

$$b_{k+1} \leq b_k(1 - 2\alpha_k C_\phi + \alpha_k{}^2 C_\phi{}'^2) + \alpha_k{}^2 d_\phi{}^2. \tag{109}$$

The rest of the proof is the same as that of Sakrison [7, eqs. (17)–(28)], and we can show

$$\lim_{k\to\infty} b_k = 0. \tag{110}$$

Thus the sequence $\{\hat{\phi}_k\}$ converges to ϕ in the mean square. The proof for the convergence of $\{\hat{\tau}_k\}$ is exactly the same.

Acknowledgment

The author is grateful to anonymous reviewers for their valuable suggestions and criticisms given to the original version of the present paper.

References

[1] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication.* New York: McGraw-Hill, 1968.
[2] A. A. Alexander, R. M. Gryb, and D. W. Nast, "Capabilities of the telephone network for data transmission," *Bell Syst. Tech. J.*, vol. 39, May 1960, p. 431.
[3] W. R. Bennet and J. R. Davey, *Data Transmission.* New York: McGraw-Hill, 1965.
[4] A. J. Viterbi, *Principles of Coherent Communication.* New York: McGraw-Hill, 1966.
[5] J. P. Costas, "Synchronous communications," *Proc. IRE*, vol. 44, Dec. 1956, pp. 1713–1718.
[6] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, 1951, pp. 400–407.
[7] D. J. Sakrison, "Stochastic approximation: a recursive method for solving regression problems," in *Advances in Communication Systems*, vol. 2, A. V. Balakrishnan, Ed. New York: Academic Press, 1966.
[8] D. J. Sakrison, R. J. McAulay, V. C. Tyree, and J. H. Yuen, "An adaptive receiver implementation for the Gaussian scatter channel," *IEEE Trans. Commun. Technol.*, vol. COM-17, Dec. 1969, pp. 640–648.
[9] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, Apr. 1965, p. 547.
[10] ——, "Techniques for adaptive equalization of digital communication systems," *Bell Syst. Tech. J.*, vol. 45, Feb. 1966, p. 255.
[11] J. R. O'Neil and B. R. Saltzberg, "An automatic equalizer for coherent quadrature carrier data transmission systems," presented at 1966 IEEE Int. Conf. Communications, Philadelphia, Pa.
[12] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Tech. J.*, vol. 48, Jan. 1969, p. 55.
[13] M. J. DiToro, "Communication in time-frequency spread media using adaptive equalization," *Proc. IEEE*, vol. 56, Oct. 1968, pp. 1653–1679.
[14] R. W. Chang, "Joint automatic equalization for data communication," in *1969 Proc. ACM Symp. Problems in the Optimization of Data Communication Systems.*
[15] ——, "Joint equalization, carrier acquisition, and timing recovery for data communication," in *Proc. 1970 Int. Conf. Communications.*
[16] H. Kobayashi, "Iterative and sequential methods for recovering sample-timing and carrier phase in data transmission systems," IBM Res. Rep. RC-2491, May 1969.
[17] C. W. Helstrom, *Statistical Theory of Signal Detection,* (revised ed.). New York: Pergamon, 1968.
[18] R. Arens, "Complex processes for envelopes of normal noise," *IRE Trans. Inform. Theory*, vol. IT-3, Sept. 1957, pp. 204–207.
[19] H. B. Voelcker, "Demodulation of single-sideband signals via envelope detection," *IEEE Trans. Commun. Technol.*, vol. COM-14, Feb. 1966, pp. 22–30.
[20] ——, "Toward a unified theory of modulation—pt. I: phase-envelope relationship," *Proc. IEEE*, vol. 54, Mar. 1966, pp. 340–353.
——, "Toward a unified theory of modulation—pt. II: zero manipulation," *ibid.*, May 1966, pp. 735–755.
[21] J. M. Wozencraft and I. R. Jacobs, *Principles of Communication Engineering.* New York: Wiley, 1965, pp. 492–508.
[22] E. Parzen, "Extraction and detection problems and reproducing kernel Hilbert space," *J. SIAM*, vol. 1, 1962, p. 35.
[23] T. Kailath, "Some application of reproducing kernel Hilbert space," presented at 3rd Annual Allerton Conf. Circuit and System Theory, University of Illinois, Urbana, 1965.
[24] J. Capon, "Hilbert space methods for detection theory and pattern recognition," *IEEE Trans. Inform. Theory*, vol. IT-11, Apr. 1965, pp. 247–259.
[25] C. R. Rao, "Minimum variance and the estimation of several parameters," *Proc. Cambridge Phil. Soc.*, vol. 43, 1947, p. 280.
[26] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, vol. 2: Inference and Relationship.* New York: Hafner, 1961.
[27] D. Middleton, *An Introduction to Statistical Communication Theory.* New York: McGraw-Hill, 1960.
[28] R. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[29] A. Lender, "Correlative level coding for binary-data transmission," *IEEE Spectrum*, vol. 3, Feb. 1966, pp. 104–115.
[30] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, Apr. 1967, pp. 260–269.
[31] H. Kobayashi, "Application of probabilistic decoding to digital magnetic recording systems," *IBM J. Res. Develop.*, Jan. 1971, pp. 69–74.
[32] ——, "Correlative level coding and the maximum likelihood decoding," presented at 1970 Canadian IEEE Symp. Communication.
[33] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," to be published.
[34] J. K. Omura, "Optimal receiver design for convolutional codes and channels with memory via control theoretic concepts," to be published.
[35] D. W. Tufts, "Nyquist's problem—the joint optimization of transmitter and receiver in pulse amplitude modulation," *Proc. IEEE*, vol. 53, Mar. 1965, pp. 248–259.
[36] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals From Noise*, R. A. Silverman, Transl. Englewood Cliffs, N. J.: Prentice-Hall, 1962.
[37] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes; Sample Function Properties and Their Application.* New York: Wiley, 1967.
[38] H. Kobayashi and J. B. Thomas, "The generalized Cramer–Rao inequality and its application to parameter estimation," in *Proc. 1st Asilomar Conf. Circuits and Systems*, 1967, pp. 1–10.
[39] D. Slepian, "Estimation of signal parameters in the presence of noise," *IRE Trans. Inform. Theory*, vol. PGIT-3, Mar. 1954, pp. 68–89.

[40] E. J. Kelly, I. S. Reed, and W. L. Root, "The detection of radar echoes in noise, pt. II," *J. SIAM*, vol. 8, 1960, p. 481.
[41] C. E. Cook and M. Bernfeld, *Radar Signals; an Introduction to Theory and Application*. New York: Academic Press, 1967.
[42] F. C. Schweppe, "Sensor-array data processing for multiple-signal sources," *IEEE Trans. Inform. Theory*, vol. IT-14, Mar. 1968, pp. 294–305.
[43] H. Kobayashi and P. D. Welch, "The detection and estimation of two simultaneous seismic events," in *Proc. Symp. Computer Processing in Communication*. Brooklyn, N. Y.: Polytechnic Press, 1969, pp. 757–777.
[44] A. V. Balakrishnan, "Filtering and prediction theory," in *Communication Theory*. New York: McGraw-Hill, 1968, ch. 3.
[45] P. S. Tong and B. Liu, "Automatic time domain equalization in the presence of noise," in *Proc. 1967 Nat. Elec. Conf.*, vol. 23.
[46] J. G. Proakis, P. R. Drouilhet, and R. Price, "Performance of coherent detection systems using decision-directed channel measurement," *IEEE Trans. Commun. Syst.* vol. CS-12, Mar. 1964, pp. 56–63.
[47] L. D. Davisson and S. C. Schwartz, "Analysis of a decision-directed receiver with unknown priors," *IEEE Trans. Inform. Theory*, vol. IT-16, May 1970, pp. 270–276.
[48] S. C. Schwartz and L. D. Davisson, "Analysis of a decision-directed receiver for a Markov sequence with unknown transition probabilities," presented at the Symp. Information Theory, Dubna, USSR, June 19–25, 1969.
[49] K. L. Chung, "On a stochastic approximation method," *Ann. Math. Stat.*, 1954, pp. 463–483.
[50] J. Sachs, "Asymptotic distribution of stochastic approximation procedures," *Ann. Math. Stat.*, 1958, pp. 370–405.
[51] C. R. Cahn, "Performance of digital phase-modulation communication systems," *IRE Trans. Commun. Syst.*, vol. CS-7, May 1959, pp. 3–6.
[52] J. Salz and B. R. Saltzberg, "Double error rates in differentially coherent phase systems," *IEEE Trans. Commun. Syst.*, vol. CS-12, June 1964, pp. 202–205.
[53] W. C. Lindsey and R. C. Weber, "On the theory of automatic phase control," in *Stochastic Optimization and Control*. New York: Wiley, 1968.
[54] M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Stand.*, vol. 49, Dec. 1952, p. 409.
[55] C. Devieux and R. Pickholtz, "Adaptive equalization with a second order gradient algorithm," in *Proc. Symp. Computer Processing in Communication*. Brooklyn, N. Y.: Polytechnic Press, 1969, pp. 665–681.
[56] H. Kobayashi, "Iterative synthesis methods for a seismic array processor," *IEEE Trans. Geosci. Electron.*, vol. GE-8, July 1970, pp. 169–178.
[57] A. A. Goldstein, *Constructive Real Analysis*. New York: Harper and Row, 1967.
[58] U. Grenander and G. Szego, *Toeplitz Form and Their Applications*. Berkeley, Calif.: Univ. of California Press, 1958.

Hisashi Kobayashi (S'66–M'68) was born in Tokyo, Japan, on June 13, 1938. He received the B.S. and M.S. degrees from the University of Tokyo in 1961 and 1963, respectively, and the M.A. and Ph.D. degrees from Princeton University, Princeton, N. J., in 1967, all in electrical engineering.

In 1960 he was awarded RCA David Sarnoff Scholarship and in 1965 the Orson Desaix Munn Fellowship. From 1963 to 1965 he worked at the Toshiba Electric Company, Tokyo, where he was engaged in research and development of advanced radar systems. In 1967 he joined the IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y., as a Research Staff Member. At present he is Manager of the analytic studies of system performance group in the Computer Science Department. He has also been involved in seismic signal processing, data communication theory, digital magnetic recording theory, image data compaction, and computer-system performance evaluation. From 1969 to 1970 he was an Acting Assistant Professor at the Department of System Science, University of California, Los Angeles, where he taught courses in probability, information theory, and signal detection theory.

Dr. Kobayashi is a member of Eta Kappa Nu, the Institute of Electronics and Communication Engineers of Japan, and Association for Computing Machinery.