Least-squares storage-channel identification

by J. M. Cioffi

Pulse (dibit) and step (transition) responses for magnetic-storage channels are important for detection-circuitry design and for comparison of various media, heads, and other channel components. This paper presents a least-squares procedure that can be used to identify the dibit and transition responses from measurements of the read-head response to any known data sequence written on the medium. The method yields significantly higher-quality estimates for the dibit and step shapes than does determining these same characteristics by measuring the average response to isolated transition or by performing a Discrete Fourier Transform (DFT) on the response to a pseudorandom data pattern. The new method can be implemented off line but also can be made sufficiently efficient to be implemented with a microprocessor for use in self-optimizing (adaptive) channel detection circuitry.

1. Introduction

Storage-channel identification is the measurement and/or computation of the characteristics of the read-back channel in a data storage device, such as a magnetic disk, magnetic tape, or optical disk. The identified characteristics are most often the channel's response to a step input (the "transition" response) or to a pulse (the "dibit" response). These characteristics are important for many purposes, such as the design of the detection circuitry (especially for equalizers and for maximum-likelihood detectors), for determining the maximum data density of the device, and for comparing various media, heads, and other channel components.

This paper presents a least-squares procedure for identification of the linear time-invariant filter that most closely approximates the desired step or pulse responses. The storage device is excited with a known data sequence, and, later, the read-head response to the known sequence is measured (or digitized) at regular intervals. The resulting measurements are then processed via the least-squares procedure to determine the step and/or pulse responses.

The resultant estimates of these responses are of significantly higher resolution (higher quality) than those produced by previous procedures, such as measuring the average response to isolated transitions (or isolated dibits) or computing the Discrete Fourier Transform (DFT) of the response to some known (usually pseudorandom) data pattern. Furthermore, the new method, although based on a linear model of the channel as presented here, can indicate the average accuracy of the linear model over any data pattern, thus indicating the presence of potential nonlinearities in the responses, unlike the aforementioned methods. The degree of agreement between the linear model and measurements can be useful in determining the data rates at which various data detection methods do and do not apply.

Section 2 defines in more detail the quantities used in channel identification and the least-squares procedure, and it compares the quality of estimates of the new and previous procedures. Section 3 studies some details of the solution and displays the results of the new procedure for several measurements taken from actual storage devices, including magnetic disks with thin-film heads, tape systems with magnetoresistive heads, and optical disks. Section 4 is a brief conclusion. Appendix A extends the channel identification procedure to apply at any digitizer sampling rate (an integer ratio of the sampling to data rates is assumed in the main body of the paper). Appendix B discusses streamlining of the least-squares procedure for possible use with adaptive detection methods, while Appendix C discusses the detection of nonlinearities.

2. Storage-channel identification methods

This section mathematically defines and analyzes the quantities and procedures used in storage-channel identification. **Figures 1(a)** and **1(b)** summarize the definitions used throughout this section.

- *Variable definitions*

The read-back channel and associated identification parameters are illustrated in Figures 1(a) and 1(b). The continuous read-head output signal, $d(t)$, can be modeled in one of two ways [1]:

$$d(t) = \sum_k x_k h(t - kT) + u(t), \tag{1a}$$

$$d(t) = \sum_k s_k h_s(t - kT) + u(t), \tag{1b}$$

where $h(t)$ and $h_s(t)$ are the unknown linear time-invariant pulse and step responses, respectively, and $u(t)$ denotes an uncorrelated, additive, zero-mean noise,† x_k takes on the values ± 1 (or $+1$ and 0 for some optical storage systems), corresponding to 1's and 0's, respectively, in the stored data sequence at time kT, $1/T$ is the data rate, and k is an integer. In Equation (1b), s_k can take on the values ± 2 or 0 (± 1 or 0 for optical) according to the relation

$$s_k = x_k - x_{k-1}. \tag{2}$$

Likewise, one determines for a linear channel

$$h(t) = h_s(t) - h_s(t - T). \tag{3}$$

It is a property of the method presented that the estimates also obey Equation (3); however, it is sometimes informative to separately identify $h(t)$ and $h_s(t)$, rather than identify only one and compute the other from it. It is assumed that $d(t)$ is digitized at some rate T_d, such that

$$T_d \triangleq \frac{T}{p}, \tag{4}$$

where p is an integer (≥ 1) oversampling factor. This restriction is relaxed to a rational fraction in Appendix A. The sampled read-head output is then, with $t = mT_d$ in (1),

$$d(mT_d) = \sum_k x_k h(mT_d - kT) + u(mT_d)$$
$$= \sum_k x_k h[(m - kp)T_d] + u(mT_d) \tag{5a}$$

† Even though the assumption that the noise is additive may not be completely true in practice, our objective is to find the values for the parameters in such a model that most closely approximate the measured responses, and deviations from such a model appear in the final results of the method in this paper.

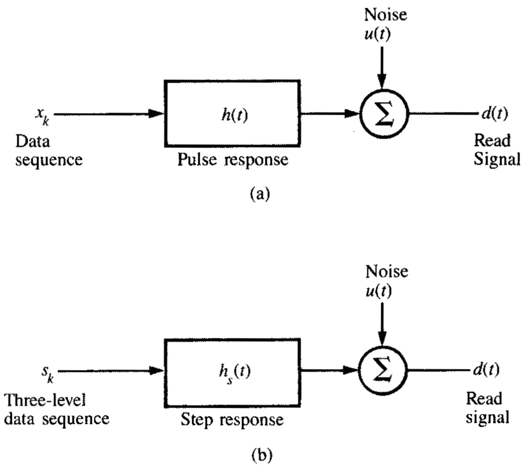

Figure 1

Summary of storage quantity definitions (a) for pulse responses and (b) for step responses.

or‡

$$d(mT_d) = \sum_k s_k h_s(mT_d - kT) + u(mT_d)$$
$$= \sum_k s_k h_s[(m - kp)T_d] + u(mT_d). \tag{5b}$$

The channel is estimated by

$$\hat{d}(mT_d) \triangleq \sum_k x_k w(mT_d - kT), \tag{6}$$

where $w(t)$ is a linear filter response whose sampled values at times mT_d are to be computed via the channel identification procedure [ideally $w(t) = h(t)$]. Likewise, for the step response, the estimate is

$$\hat{d}_s(mT_d) \triangleq \sum_k s_k w_s(mT_d - kpT_d). \tag{7}$$

We also define an error signal

$$\varepsilon(mT_d) \triangleq d(mT_d) - \hat{d}(mT_d). \tag{8}$$

As an example, note that, if x_k or s_k is a sequence corresponding to an isolated pulse or transition input, (5a) and (5b) reduce to

$$d(mT_d) = h(mT_d) + u(mT_d) \tag{9a}$$

or

$$d_s(mT_d) = h_s(mT_d) + u(mT_d), \tag{9b}$$

respectively, the desired pulse shapes in noise. Then, $w(mT_d)$

‡ The reader may note that (5a) and (5b) are equivalent to p subchannels, each at spacing T; this observation is exploited to reduce computation in the new procedure in Section 3.

and $w_s(mT_d)$ can be estimated by the averages

$$w(mT_d) = \frac{1}{n} \sum_{k=1}^{n} d(mT_d;k), \tag{10a}$$

$$w_s(mT_d) = \frac{1}{n} \sum_{k=1}^{n} d_s(mT_d;k), \tag{10b}$$

where the index k denotes the kth experiment. That is, one measures the response n times and averages, which is the basis for the aforementioned isolated step and dibit identification methods. Some deficiencies of the estimates identified via such isolated step or pulse methods are discussed later. Equations (9) and (10) were given only to verify the utility of the definitions in (1)–(8). We now proceed with a discussion of the least-squares channel-identification procedure.

• *The application of least squares*

In the least-squares identification procedure, a known data pattern is written on the storage device. The $w(mT_d)$ are chosen to minimize

$$\xi_l = \sum_{m=1}^{l} \varepsilon(mT_d)^2, \tag{11}$$

where $\varepsilon(mT_d)$ is given in (8). If we denote $W_{M,l}$ by the $M \times 1$ column vector

$$W_{M,l} \triangleq \begin{bmatrix} w_l(0) \\ \cdot \\ w_l[(M-1)T_d] \end{bmatrix}, \tag{12}$$

then the solution to (11) is conveniently written [2]

$$W_{M,l} = \Big(\sum_{m=1}^{l} X_{M,m}X'_{M,m}\Big)^{-1} \Big(\sum_{m=1}^{l} X_{M,m}d(mT_d)\Big), \tag{13}$$

where $'$ denotes transpose, and

$$X_{M,m} \triangleq \begin{bmatrix} x_m \\ \cdot \\ x_{m-M+1} \end{bmatrix} \tag{14}$$

for $p = 1$. There are $p - 1$ zeros between entries in (14) if $p > 1$. We have further assumed that M is large enough to span the nonzero extent of the pulse (step) response in intervals of sampling periods or $MT_d = NT$ data periods containing p samples each, $M = Np$. Equation (13) can be rewritten

$$W_{M,l} = R_{M,l}^{-1}P_{M,l}, \tag{15}$$

where

$$R_{M,l} \triangleq \frac{1}{l} \sum_{m=1}^{l} X_{M,m}X'_{M,m},$$

$$P_{M,l} \triangleq \frac{1}{l} \sum_{m=1}^{l} X_{M,m}d(mT_d). \tag{16}$$

A similar expression holds for the step response, with x's replaced by s's and w's replaced by w_s's in the solution. Note that $M \times M$ matrix inversion is explicit in (13); however, because of the special structure in this problem, no matrix need ever be inverted directly. For more details, see Section 3 and especially [2].

• *A performance measure*

The mean of $W_{M,l}$ can be easily determined as

$$E[W_{M,l}] = H_M = \begin{bmatrix} h(0) \\ \cdot \\ \cdot \\ h[(M-1)T_d] \end{bmatrix}, \tag{17}$$

the desired solution, when the above least-squares method is used. The Norm Tap Deviation is a mean-square measure of statistically how far the estimated $W_{M,l}$ is from $H_{M,l}$ and is also easily computed, if $u(t)$ is white (spectrally flat over the frequency range of interest), as

$$\theta_{M,l} = E[\| W_{M,l} - H_{M,l}\|^2] = \frac{1}{l} \operatorname{trace}(R_{M,l}^{-1})\sigma_u^2, \tag{18}$$

where

$$\sigma_u^2 \triangleq E[u(KT_d)^2]. \tag{19}$$

We show in the next few sections that both the isolated transition (or dibit) and DFT methods are special cases of the general least-squares method with very special restrictions on the input sequence and on M and l. Thus, we are able to use (18) as a performance indicator for those methods as well.

• *Isolated transition example and analysis of resolution*

As an example, once again consider an isolated dibit; then $X_{M,m}$ has only one nonnegative entry per column and (13) reduces to (using generalized inverses, see [3])

$$W_{M,l} = \begin{bmatrix} d(MT_d) \\ \cdot \\ d(T_d) \end{bmatrix}. \tag{20a}$$

A string of n "isolated" (far enough apart) dibits occurring within a large data record (length l) has a least-squares solution,

$$W_{M,l} = \frac{1}{n} \sum_{k=0}^{n-1} \begin{bmatrix} d(kMT_d + T_d) \\ \cdot \\ d(kMT_d + MT_d) \end{bmatrix}, \tag{20b}$$

that is exactly the same as the isolated pulse solution in (10a). The least-squares identification procedure is more general in that the input need not be an isolated transition or dibit.

Equation (18) allows us to compare the quality of the least-squares estimates of $H_{M,l}$ ($W_{M,l}$) for different input sequences. Note that, for a string of n isolated (MT_d apart, so $l = Mn$) inputs, one determines for white zero-mean u

$$E\| W_{M,l} - H_{M,l}\|^2 = \frac{M}{n}\sigma_u^2. \tag{21}$$

Pseudorandom sequences are generally desirable [4, 5] for channel inputs because of their broadband spectral response. An identity for $R_{M,l}$ can easily be determined, if l = length of the pseudorandom sequence,§ as (see [6–8])

$$R_{M,l} = \frac{1}{l}[(l+1)I_M - 1_M 1_M'], \tag{22a}$$

where 1_M is an $M \times 1$ vector of M ones. One can also easily show that

$$R_{M,l}^{-1} = \frac{l}{l+1}\left(I_M + \frac{1}{l-M+1} 1_M 1_M'\right). \tag{22b}$$

Thus, (18) becomes, for a pseudorandom sequence of length M repeated n times,

$$\theta_{M,l} = \frac{1}{nM+1}\left\{\frac{(n-1)M^2+2M}{(n-1)M+1}\right\}\sigma_u^2 \rightarrow \frac{1}{n}\sigma_u^2. \tag{23}$$

For $n = 1$, there is an improvement of $(M+1)/2$ with respect to (21). As n increases to a large value, there is an improvement by a factor of M in estimate quality, or equivalently, M more digitized outputs from isolated dibits must be processed in the isolated dibit identification schemes to get the same resolution estimates as those produced by least squares with a pseudorandom length-M input. For oversampling ($p > 1$), the comparison favors the pseudorandom input by the same amount. Heuristically, when using pseudorandom or "scrambled" data in channel identification, the input is more spectrally "rich" and all frequencies are more equally weighted than when a single pulse is used. The resulting flat nature of the spectrum results in the inverse autocorrelation matrix being close to an identity which makes $\theta_{M,l}$ in (18) smaller (better). When x_k has a flat spectrum, s_k does not have a flat spectrum, but a similar slightly more complex argument can be given to justify the least-squares improvements.

In practice, it may not be difficult to average the extra data for the isolated input method. However, there is another very practical advantage of using more random data, as was first noted by C. M. Melas [9]. This is that in the isolated transition or isolated dibit methods, the AGC (Automatic Gain Control) must be removed from the channel to prevent the sudden change in energy associated with the isolated input from suddenly varying the gain parameter of the AGC. Then, the identified pulse characteristics will not include the effect of the AGC. This effect can commonly be more than a simple gain factor and is determined by the bandwidth and tracking rate of the AGC.

- *Comparison with frequency-domain methods*

Another more recent method used in storage-channel identification is [4, 5, 10] to compute the DFT of the response to some prescribed pattern written on the media. In order to invert the DFT to get a time-domain estimate of the pulse response, one must first divide the measured DFT by the DFT, *including phase*, of the input before the inverse DFT, which [10] also observes. Using this last restriction, one can also generalize the methods of [4, 5, 10] to estimate the channel response for any inputs, including the ±2, 0 normally associated with identification of the step (transition) response.

Nevertheless, with the division by input spectra, the frequency-domain method is the same as the time-domain least-squares method of this paper if $M = l$, and as we shall see, the case $M = l$ gives very poor estimate quality. In the case that $u(mT_d)$ is white and Gaussian, the least-squares method (see [11]) achieves the famed Cramer-Rao bound for a fixed l and M; that is, no other estimator has higher resolution for the given data. If the assumption on $u(t)$ is just white (not also necessarily Gaussian), then the least-squares estimator is a Best Linear Unbiased Estimator (BLUE) [3].

Theoretically, the difference between the DFT technique and the time-domain least-squares method can be quantified via the following analysis. It is usually wise to pick $M < l$ so as to introduce more noise averaging, or equivalently, to make the Cramer-Rao bound lower for fewer parameters. Generally speaking, in any estimation scheme, we desire $l > M$ to get good quality estimates. Nevertheless, picking M too small can introduce extraneous harmonic distortion in the estimated step response. The time-domain least-squares method can be rewritten as that $W_{M,l}$ that minimizes [2]

$$\xi_{M,l} = \underline{\varepsilon}'_{M,l}\underline{\varepsilon}_{M,l} = \|\underline{\varepsilon}_{M,l}\|^2, \tag{24}$$

where

$$\underline{\varepsilon}_{M,l} = \underline{d}_{l,l} - \underline{X}_{M,l,l}W_{M,l} \tag{25}$$

and

$$\underline{d}_{l,k} = \begin{bmatrix} d(kT_d) \\ \cdot \\ d[(k-l+1)T_d] \end{bmatrix}; \qquad \underline{x}_{l,k} = \begin{bmatrix} x_k \\ \cdot \\ x_{k-l+1} \end{bmatrix}, \tag{26a}$$

where $p - 1$ zeros can be inserted between nonzero entries in $\underline{x}_{l,k}$ and

$$\underline{X}_{M,l,k} = [\underline{x}_{l,k}, \underline{x}_{l,k-1}, \cdots, \underline{x}_{l,k-M+1}]. \tag{26b}$$

The DFT-based method is a special case of a linear $M \times l$ transformation on $\varepsilon_{M,l}$, that is, let

$$\underline{E}_{M,l} = \phi\underline{\varepsilon}_{M,l}, \tag{27}$$

where ϕ is an $M \times l$ (possibly complex) matrix representing the linear transformation. Then

$$\underline{E}^*_{M,l}\underline{E}_{M,l} = \underline{\varepsilon}^*_{M,l}\phi^*\phi\underline{\varepsilon}_{M,l}, \tag{28}$$

where * denotes conjugate transpose. If ϕ is a unitary transformation ($\phi^*\phi = I$), then

§ Even when the output is oversampled, we show later that the only autocorrelation matrix of interest is at the data rate; thus all of the analysis here is also valid for $p > 1$.

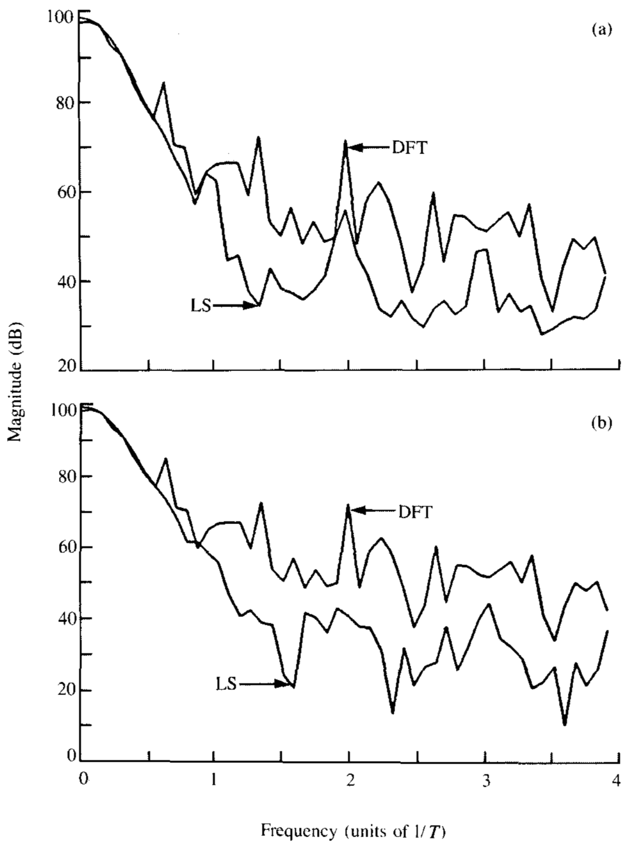

Figure 2

Comparison of DFT and least squares for (a) 8-bit and (b) 16-bit periods.

$$\|\underline{E}_{M,l}\|^2 = \|\underline{\varepsilon}_{M,l}\|^2, \tag{29}$$

and the minimized $\underline{\varepsilon}_{M,l}$ is obtained by

$$\underline{\varepsilon}_{M,l} = \phi^* \underline{E}_{M,l} . \tag{30}$$

In the DFT methods of [4, 5], the matrix ϕ is chosen, *under the very special assumptions that $M = l$ and the input is periodic (pseudorandom) of length $l = M$, as*

$$\phi = \phi_M = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & 1 & \cdot & 1 \\ 1 & e^{-j\omega_1 T} & \cdot & e^{-j\omega_1(M-1)T} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & e^{-j\omega_{M-1}T} & \cdot & e^{-j\omega_{M-1}(M-1)T} \end{bmatrix}, \tag{31}$$

where

$$\omega_i = i\frac{2\pi}{lT} \qquad i = 0, \cdots, M-1. \tag{32}$$

ϕ_M can easily be shown to be unitary [12], so the relation in (29) holds, apparently yielding the time-domain least-squares solution. ϕ_M^* is the inverse DFT in this case. However, in the time-domain method of this paper, M is much less than l to average the effects of noise and other nonideal effects.

Using our performance measure in (18) and (23) ($n = 1$, $l = M$) again, one determines the estimate quality as

$$\theta_{l,l} = \frac{2l}{l+1}\sigma_u^2 , \tag{33}$$

while the general formula for a pseudorandom sequence of length l with M parameters is

$$\theta_{M,l} = \frac{2M + Ml - M^2}{(l+1)(l-M+1)}\sigma_u^2 . \tag{34}$$

Substitution of $l = 10M$, a good practical rule of thumb, into (34) yields the advantage

$$\frac{\theta_{l,l}}{\theta_{M,l}} = \frac{2l(0.9l+1)}{0.09l^2 + 0.2l} = 20\,\frac{(0.9l+1)}{(0.9l+2)} . \tag{35}$$

Even for $l \cong 1000$, another reasonable number, the improvement in (35) is close to its limiting value of 20. This large improvement is typically evident when comparing the spectra of a pulse produced by the time-domain least squares and by the DFT method, as we have illustrated in **Figures 2(a)** and **2(b)**. Note from the level of "frequency ripple" in the DFT plot that the time-domain least squares is at least an order of magnitude improvement. Also note the lower "noise level" at higher frequencies with the least-squares identification procedure. It is also important to note that $l = M = 2^i - 1$ (i a positive integer) for a pseudorandom input, which, at least, requires special attention for efficient DFT implementation [12–14]. The reason for the two different lengths (M's) in Figures 3(a) and 3(b) is discussed later.

- *An averaged DFT identification scheme*

Here, we propose an averaged DFT method for the special case that $l = nM$, where n is an integer greater than 1, and the input sequence is periodic with period M. [The case of oversampling ($p > 1$) is identical for each of the subchannels (see Section 3).] There is a very special set of circumstances when the inverted matrix in (13) is Toeplitz and DFTs can be used. Generally, (13) is not Toeplitz and DFTs are not appropriate. This method is equivalent to least squares, as can be seen from the following. Define ϕ_l by

$$\underline{\phi}_l \triangleq \begin{bmatrix} \phi_M & 0 & \cdot & 0 \\ 0 & \phi_M & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \phi_M \end{bmatrix} . \tag{36}$$

Multiplication by ϕ is equivalent to n M-point DFTs performed on the n groups of M inputs. Note that $\underline{\phi}_l$ is unitary,

$$\underline{\phi}_l \underline{\phi}_l^* = I. \tag{37}$$

The least-squares estimates in the frequency domain are given for each frequency bin by

$$\nu(k) = \frac{\sum_{i=1}^{n} \delta(k, i)\chi^*(k)}{\sum_{i=1}^{n} \chi(k)\chi^*(k)} = \frac{1}{n} \frac{\sum_{i=1}^{n} \delta(k, i)}{\chi(k)}$$

$$k = 0, \cdots, M - 1, \quad (38)$$

where $\nu(k)$ and $\chi(k)$ are the M-point DFTs of $W_{M,l}$ and $X_{M,M}$, respectively. $\delta(k, i)$ is the M-point DFT of the time series d_k in the ith of the n groups. Equation (38) is really the average of n uses of the original DFT method, when a period-M input is recycled to fill l time periods. Then, some averaging will be introduced, in the optimal least-squares sense, into the DFT identification scheme. The method of (36) and (38), because of (37), is equivalent to an l-point least-squares time-domain procedure. Of course, an inverse DFT on the quantities in (38) must be performed to obtain the desired time-domain parameters, $W_{M,l}$. This method requires the unnecessary imposition of an integer ratio restriction on l and m, which is not required in the more general and straightforward time-domain least-squares solution (13).

• *A note on maximum-likelihood detection schemes*

The identified responses can be used in Maximum-Likelihood Sequence Detection (MLSD) [15, 16]. In this case, the Mean Square Error (MSE) is a more useful estimate of performance than (18). It is shown in [17] that (given a certain input sequence)

$$MSE = E[\varepsilon^2(mT_d)] = \sigma_u^2 \gamma_{M,l}, \quad (39)$$

where $\gamma_{M,l}$ is given by

$$\gamma_{M,l} = 1 - X'_{M,l} R_{M,l}^{-1} X_{M,l} \quad (40)$$

and $'$ denotes transpose. One also can show (see [17]) that

$$0 \le \gamma_{M,l} \le 1; \quad (41)$$

thus, the worst (because the desired value is σ_u^2) MSE after M measurements is

$$MSE_{\text{worst}} = 0, \quad (42)$$

which is exactly the value given by a length-M pseudorandom sequence. In fact, it is shown in [8] that choices for x_k other than length-M pseudorandom sequences can yield MSE between 0 and σ_u^2 after M data points, while still maintaining good (low) $E[\| W_{M,l} - H_{M,l} \|^2]$. Thus the length-$M$ pseudorandom sequence may not be the best training sequence if MLSD is used. Some data with statistics equivalent to what is expected in actual use would be the best choice for MLSD and other similar sampling detection schemes.

• *Signal-to-noise ratio estimation*

The SNR for the read-head response can be estimated (when $M \ll l$) by

$$SNR \cong \frac{(l - N) \| W_{M,l} \|^2}{p \cdot \xi_{M,l}}, \quad (43)$$

where $\| W_{M,l} \|^2/p$ is the signal power for the binary input to the pulse response, and $\xi_{M,l}/l - N$ is the noise power. However, one must ensure that data measured at the read-head output have NOT BEEN AVERAGED before digitizing to ensure a meaningful estimate in (43). Also, as Howell [4] has noted, that distortion in the measuring devices, particularly the nonlinearities in the CRT sweep rate if a storage scope is used, can add appreciable noise not inherent in the actual storage channel. Of course, such contamination would leave (37) as a measure of the mean-square distortion in the measuring procedure, rather than the desired channel noise + media noise + modeling mean-square errors. Even if measurements are carefully taken, (43) is usually more indicative of the levels of nonlinear mismatch to the model and can therefore be very useful in evaluating the potential success or failure of advanced detection schemes.

• *Determination of M*

We have previously assumed that the order M (number of identified parameters) was overestimated or known *a priori.* However, the best quality estimate for l data points is given by the so-called "Minimum Description Length" principle of [18], which jointly estimates M and the corresponding $W_{M,l}$ for l-points. The improvement in the general storage-channel identification problem is negligible if $l \ge 10M$. It is interesting to understand just what happens if M is chosen too small. Suppose $h(kT_d) \ne 0$ for $k < 0$, $k > M$. Then $u(kT_d)$ can be modeled as the sum of white noise and the distortion caused by the neglected terms in h. This second distortion term is just a linear filter acting on the pseudorandom pattern. When oversampled, the output of such a filter is the product of its transfer function and the transform of the oversampled pseudorandom pattern. The response of the oversampled pseudorandom pattern can easily be shown to be maximum at multiples of $1/T$, thus explaining why choosing M larger in Figure 2(b) than in Figure 2(a) caused the "harmonics" to disappear. Of course, picking M too large as in the DFT methods has a far more distorting effect on the output because of the lack of noise averaging. Generally speaking, conservative values for M and l are 15 bit periods and $l = 10M$, respectively.

• *Summary*

In this section, we have introduced the least-squares channel-identification procedure, compared its performance with other commonly used procedures, and found the least-squares method superior in the quality of estimates that it produces. We now turn to implementation/programming of this new procedure.

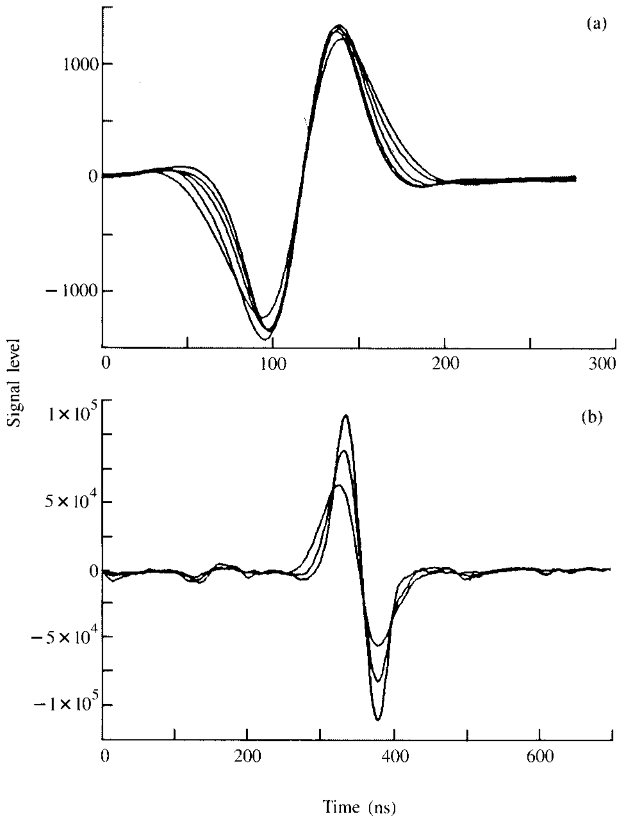

Figure 3

Pulse responses at 27 Mb/s for (a) thin-film medium and thin-film head and (b) particulate medium and thin-film head.

3. Efficient implementation of the off-line least-squares identification procedure

The time-domain least-squares solution is described using a matrix inverse in (13). This matrix can be large, requiring large storage and long processing time in an off-line computer program implementing the inversion. However, matrix inversion can be avoided to simplify the determination of $W_{M,l}$. This section describes several special features of the least-squares procedure that can be used to reduce considerably the computation and storage in an off-line implementation. Such simplifications could also become important if the characteristics of each particular storage device, and possibly at several different radii on each, were to be computed during the manufacturing process either for identifying defective devices or for optimization of the channel-detection circuitry for each particular unit. An efficient on-line procedure, similar to that of [8], is suggested in Appendix B.

• *Subchannels*

In most cases of practical interest, the oversampling factor p is greater than one. Then, one writes $mT_d = nT + iT_d$, where

$$n = \lfloor \left(\frac{m}{p}\right), \tag{44}$$

where $\lfloor$ denotes the "greatest integer less than," and i takes the values $0, \cdots, p-1$. Equation (5a) is rewritten [(5b) can be similarly rewritten]

$$d(nT + iT_d) = \sum_k x_k h[(n-k)T + iT_d] + u(nT + iT_d). \tag{45}$$

The index i has no effect upon the convolution operation, and the p phases of $d(mT_d)$ per sample period, $T = pT_d$, are described by

$${}^i d_n = \sum_k x_k \, {}^i h_{n-k} - {}^i u_n \qquad i = 0, \cdots, p-1, \tag{46}$$

where the ${}^i h_n$ are i independent "subchannels." With minor algebra, one can reduce the least-squares identification procedure to p subprocedures that can all be solved separately. The p solutions can be interspersed to obtain $W_{M,l} = W_{Np,l}$, where $N = M/p$ (we assume that p divides M or that M is picked slightly larger so that it does). Then, only one $N \times N$ matrix need be inverted (it is the same for all subchannels), rather than one $M \times M$ matrix, a considerable computational and storage saving. This matrix is the autocorrelation matrix of the input data alluded to in an earlier footnote (§). However, much greater savings are also available.

• *Use of fast algorithms*

The most efficient solution to the *general* least-squares identification problem appears in [2]. The DFT cannot be used in the general least-squares filtering problem because a Toeplitz structure must be imposed on (13) for its use. This solution requires about

$$\left(\frac{p+1}{p}\right) lN + 4.5N^2 + pN^2 \tag{47}$$

multiplications, divisions, and additions in comparison to $O(N^3)$ for straightforward matrix inversion. [$O(x)$ is a number that asymptotically rises no faster than in direct proportion to x.] The term $(lN/p) + 4.5N^2$ is the fixed cost of the equivalent of inverting the matrix $R_{N,l/p}$ (fixed because it is the same for each subchannel); the remaining term $pN^2 + lN$ is the additional cost, at $N^2 + lN/p$ per subchannel, for computing the equivalent of the product $R^{-1}_{N,l/p} \, {}^i P_{N,l/p} = {}^i W_{N,l/p}$ for each of the p subchannels. The storage requirements are about $6N + 2l$ locations for the algorithm in [2]. The cost reductions accrue to the shifted nature of $X_{N,k}$ with respect to $X_{N,k-1}$, or equivalently, that $R_{N,l/p}$ can be rewritten as a product of Toeplitz matrices,

$$R_{N,l/p} = \underline{X}'_{N,l/p,l} \underline{X}_{N,l/p,l}\,, \tag{48}$$

where $X_{N,l,k}$ is defined in (26b). For more details, see [2].

• *Choice of the input sequence*

Further computational and storage reductions are possible if the length-l sequence $x_{l,k}$ is chosen beforehand for all storage

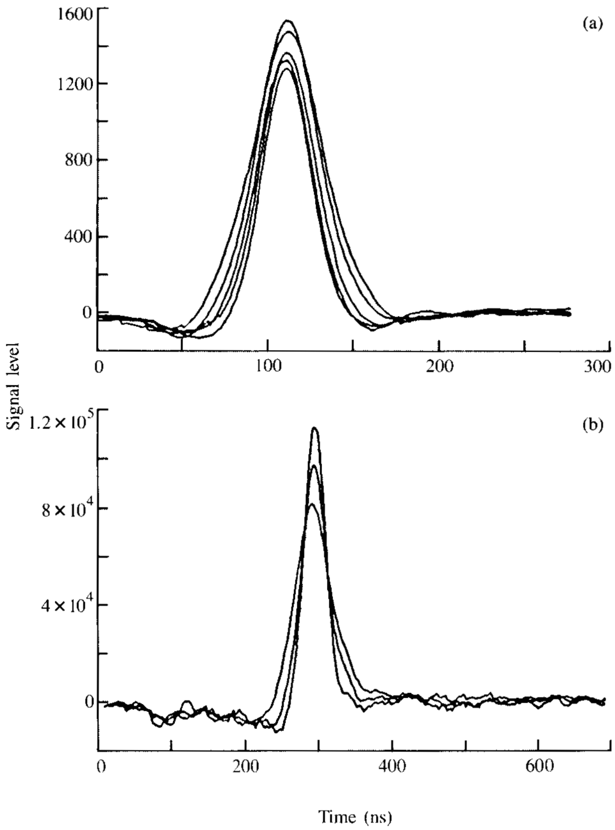

Figure 4

Step responses for (a) thin-film medium and thin-film head and (b) particulate medium and thin-film head.

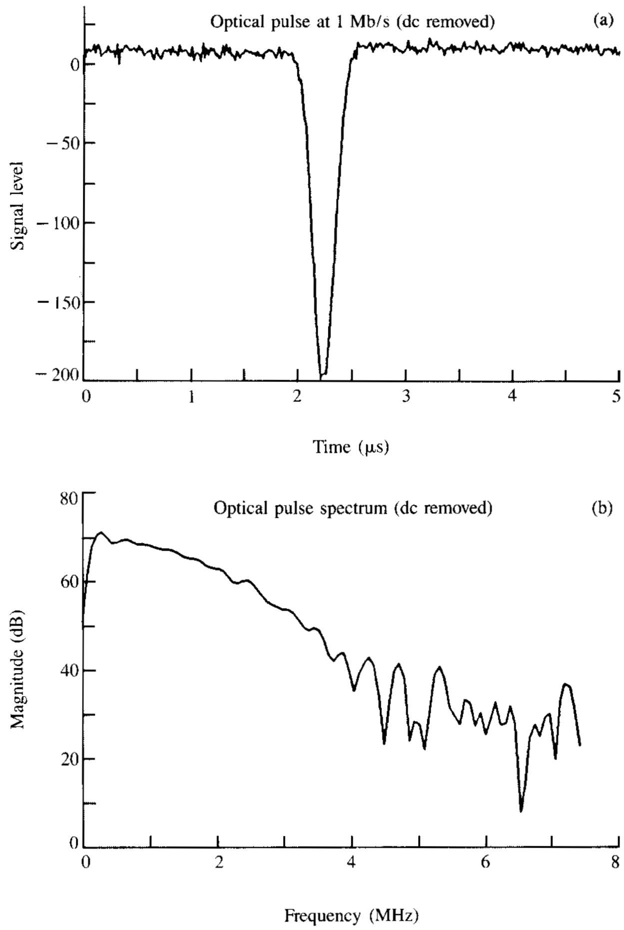

Figure 5

(a) Pulse response of optical medium at 1 Mb/s (dc removed) and (b) Fourier transform of pulse response.

channels to be identified. A currently popular choice is a 63-bit pseudorandom sequence. When the input data sequence is known beforehand, many of the quantities in the Fast (BFTF) algorithm of [2] can be precomputed and stored once, reducing computation to

$$pN^2 + lN \tag{49}$$

multiplications and additions (no divides) and storage (random access) to about

$$2N + l \tag{50}$$

locations. Neither these counts nor the counts in (47) and (49) can be matched by the DFT or other methods of comparable estimate quality for reasonable N (20 or less). Asymptotically, because of the $N \log_2 N$ computation in FFT implementations of the DFT, these FFT methods may have an advantage in terms of computational requirements, but N is never chosen that large in practice.

• *Experimental results*

To demonstrate the robustness of the new least-squares identification method, several channel pulse shapes are plotted in **Figures 3(a)** and **3(b)**, while the corresponding steps are plotted in **Figures 4(a)** and **4(b)**. These responses were obtained using the new procedure for a 63-bit pseudorandom sequence on digitized measurements of a thin-film disk/thin-film head channel [Figures 3(a) and 4(a)], and on a particulate disk/thin-film head channel [Figures 3(b) and 4(b)]. The measurements were taken at several different diameters on each device. The diameters for Figures 3(a) and 4(a) were 105, 120, 135, 150, and 165 mm, while those for 3(b) and 4(b) were 103, 136, and 172 mm. **Figures 5(a)** and **5(b)** show the pulse response and its spectrum, respectively, for an optical storage device. In **Figures 6(a)** and **6(b)**, we have plotted pulse and step responses for a magnetoresistive head in a magnetic-tape system; this time a 62-bit pattern corresponding to NRZI coding of two cycles of a 31-bit pseudorandom data pattern was used [10]. In **Figure 7**, the delay for the magnetoresistive head is plotted to illustrate the ability of the new least-squares identification

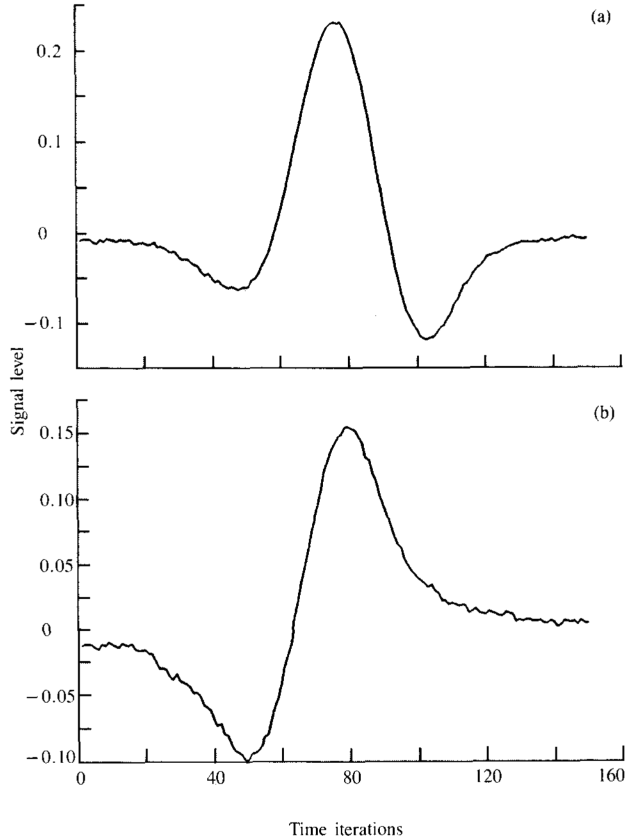

Figure 6

(a) Pulse response and (b) step response of magnetoresistive head.

procedure to capture that quantity as well. The dc level was removed from the desired signal for the optical device to facilitate inspection of the plots; the true optical channel is a baseband channel. The plots in Figures 3, 4, and 5 demonstrate the robust utility of the least-squares procedure.

4. Conclusions

In this paper, we have introduced a new least-squares storage-channel identification procedure. We have analyzed the procedure thoroughly and demonstrated via experiment its utility and its improvements over existing methods. Several methods for reducing the implementational cost of the procedure were also discussed. The procedure can become a uniform standard for identifying and comparing the channel characteristics of various storage media.

Acknowledgments

I would like to thank L. Barbosa, F. Dolivo, M. Haynes, K. Hense, T. Howell, R. Lynch, C. M. Melas, P. Seger, W. Schott, P. Siegel, H. Thapar, D. Turner, G. Ungerboeck, and J. E. Vaughn for their technical comments and/or data.

Appendix A: Arbitrary sampling rates

In this appendix, the sampling interval T_d is permitted to take the values

$$T_d = \frac{q}{p} T, \tag{A1}$$

where q and p are relatively prime positive integers such that $q < p$. Any arbitrary ratio of sampling to data rates can be as closely approximated as desired by the relation in (A1), as long as it is known, which implies some synchronization between digitizer and write clock. We also define a smaller time interval τ by

$$\tau = \frac{T_d}{q} = \frac{T}{p} \tag{A2}$$

or

$$qp\tau = pT_d = qT. \tag{A3}$$

The samples at rate T_d can be organized into successive disjoint sets of p members and of duration $pT_d = pq\tau$. Then any sampling instant mT_d can be rewritten as

$$mT_d = npq\tau + iq\tau = (np + i)T_d \qquad i = 0, \cdots, p - 1. \tag{A4}$$

The equivalent of (37) becomes

$$d[npq\tau + iq\tau] = \sum_k h(npq\tau - kp\tau + iq\tau)x_k + u(npq\tau + iq\tau). \tag{A5}$$

Note that, if p and q are relatively prime, as was assumed earlier, then h will be specified at intervals of τ in (A5), or equivalently at all time instants that are integer multiples of τ. At sample i within each group of p samples, only the values $h(kp\tau + iq\tau)$, where k is any integer, contribute to

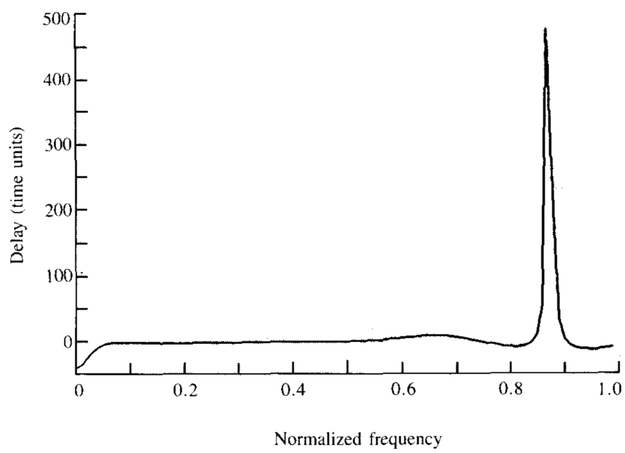

Figure 7

Delay characteristic for tape pulse response.

$d(npq\tau + iq\tau)$. Thus, d has p phases per group of p samples that can be independently modeled as

$$^{i}d_{np} = \sum_k x_k {}^{i}h_{nq-k} - {}^{i}u_{np} \qquad i = 0, \cdots, p-1, \tag{A6}$$

where, again,

$$^{i}h_n = h(nT + iT_d) \tag{A7}$$

and

$$^{i}d_n = d(nT_d + iT_d); \qquad {}^{i}u_n = u(nT + iT_d) \tag{A8}$$

for $i = 0, \cdots, p-1$. Each of the subchannels can be identified independently and the resultant responses overlaid (with delays of τ with respect to one another). The overall response can then be used directly or decimated to $p\tau$ (the data rate), $q\tau$ (the sampling rate), or any other integer divisor of the rate $1/\tau$. An important point to note is that there is a loss in resolution of a factor of approximately q for any fixed data length l with respect to the case where $T_d = T/p$. This last fact makes the alternative of resampling the data or phase-locking the ADC used to acquire the data (set $q = 1$) very attractive from a performance viewpoint.

Appendix B: On-line efficiency

It is possible to implement the least-squares storage-channel identification method in a sample-recursive manner. The procedure becomes a special case of the one considered previously by this author for echo cancelers in data transmission in [8]. The storage identification procedure could be performed on line, for example, to initialize, and possibly update (see [15, 16], a Maximum-Likelihood Sequence Detection Circuit.

A brief summary of the procedure is, where k is the recursive time index,

$$W_{M,k} = W_{M,k-1} + \varepsilon^{P}_{M,k} \cdot C_{M,k} , \tag{B1}$$

and where

$$\varepsilon^{P}_{M,k} = d(k) - W_{M,k-1} X_{M,k} . \tag{B2}$$

$C_{M,k}$ is an $M \times 1$ function of the input (presumably known or "training") data sequence and is given by

$$C_{M,k} = \Big(\sum_{m=0}^{k} X_{M,k} X'_{M,k} \Big)^{-1} X_{M,k} , \tag{B3}$$

and is presumably precomputed and stored prior to use. For more details on this procedure, and for an efficient recursive computation of $C_{M,k}$ when there is no prespecified training sequence, see [2, 8, 17, 19]. A final note is that, if the signal written just prior to the start of training is an erasure, then the prewindowed exact-initialization method of [8, 17] applies, rendering extremely low computational requirements; (B1) and (B3) simplify dramatically in that case.

Appendix C: Methods for nonlinear identification

The study of nonlinear identification of a data channel is an entire subject area in itself. For instance, one can refer to [20] and [21] for methods based on simplification of Volterra series under the constraints of a binary input. Here, a simple method suffices to verify the presence/absence of appreciable nonlinearities and to roughly quantify their magnitudes relative to the linear component of the channel response.

SNR measurement

Estimation of the SNR was discussed earlier. The minimized sum of squared errors, $\xi_{M,l}$, contains a component due to modeling error. If M is sufficiently large, most of this modeling error is due to nonlinearities. The size of the SNR is indicative of the level of nonlinearities. Generally speaking, SNRs well below those expected can be indicative of large modeling errors due to nonlinearities. Thus, one can use the size of the SNR as an indicator of nonlinearities, given that he has some prior experience with the particular media and head and knows what to expect in terms of a nominal SNR value. This type of procedure requires a very accurate phase-lock to the underlying data rate to ensure that nonlinearities are not artificially inserted by sampling-phase errors in the measurement process.

References

1. D. G. Messerschmitt, "A Study of Sampling Detectors for Magnetic Recording," University of California at Berkeley, private communication.
2. J. M. Cioffi, "The Block-Processing FTF Adaptive Algorithm," *IEEE Trans. Acoust., Speech, & Signal Proc.* **ASSP-34,** No. 1, 77–90 (February 1986).
3. C. L. Lawson and R. J. Hanson, *Solving Least-Squares Problems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974.
4. T. D. Howell, IBM Research Division, San Jose, CA, private communication.
5. G. Ungerboeck, IBM Research Division, Zurich, Switzerland, private communication.
6. J. Salz, "On the Start-Up Problem in Digital Echo Cancellers," *Bell Syst. Tech. J.* **6,** No. 2, Part 1, 1353–1364 (July–August 1983).
7. M. L. Honig, "Echo Cancellation of Voiceband Data Signals Using RLS and Stochastic-Gradient Algorithms," *IEEE Trans. Commun.* **COM-33,** No. 1, 65–73 (January 1985).
8. J. M. Cioffi and T. Kailath, "An Efficient, RLS, Data-Driven Echo Canceller for Fast Initialization of Full-Duplex Data Transmission," *IEEE Trans. Commun.* **COM-33,** No. 7, 601–611 (July 1985). See also *Proceedings of ICC'85*, June 1985, Chicago.
9. C. M. Melas, IBM Research Division, San Jose, CA, private communication.
10. M. K. Haynes, "Experimental Determination of the Loss and Phase Transfer Functions of a Magnetic Recording Channel," *IEEE Trans. Magnetics* **MAG-13**, No. 5, 1284–1286 (September 1977).
11. G. C. Goodwin and R. L. Payne, *Dynamic System Identification*, Academic Press, Inc., New York, 1977.
12. D. F. Elliot and K. Ramamohan Rao, *Fast Transforms: Algorithms, Analyses, Applications,* Academic Press, Inc., New York, 1982.
13. R. E. Blahut, *Fast Algorithms for Digital Signal Processing*, Addison-Wesley Publishing Co., Reading, MA, 1985.

14. H. J. Nussbaumer, *Fast Fourier Transforms and Convolution Algorithms*, Springer-Verlag, Berlin, 1981.
15. G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data-Transmission Systems," *IEEE Trans. Commun.* **COM-22,** No. 5, 624–636 (May 1974).
16. F. R. Magee and J. G. Proakis, "Adaptive Maximum-Likelihood Sequence Estimation for Digital Signaling in the Presence of Intersymbol Interference," *IEEE Trans. Info. Theory* **IT-19,** No. 1, 120–124 (January 1973).
17. J. M. Cioffi and T. Kailath, "Fast, Recursive-Least-Squares, Transversal Filters for Adaptive Filtering," *IEEE Trans. Acoust., Speech, & Signal Proc.* **ASSP-34,** No. 2, 304–337 (April 1984).
18. J. Rissanen, "Modeling by Shortest Data Description," *Automatica* **14,** 465–471 (1978).
19. J. M. Cioffi and T. Kailath, "Windowed FTF Adaptive Algorithms with Normalization," *IEEE Trans. Acoust., Speech, & Signal Proc.* **ASSP-33,** No. 3, 607–625 (June 1985).
20. O. Agazzi, D. G. Messerschmitt, and D. A Hodges, "Nonlinear Echo Cancellation of Data Signals," *IEEE Trans. Commun.* **COM-30,** No. 11, 2421–2433 (November 1982).
21. A. Gersho and E. Biglieri, "Adaptive Cancellation of Channel Nonlinearities for Data Transmission," *Proceedings of ICC'84*, Amsterdam, May 1984, pp. 1239–1242.

Received September 28, 1985; accepted for publication December 10, 1985

John M. Cioffi *Stanford University, Department of Electrical Engineering, Stanford, California 94305.*

Dr. Cioffi received the B.S. degree in electrical engineering from the University of Illinois, Urbana, in 1978 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 1979 and 1984, respectively. From 1979 to 1981, he worked in the Advanced Data Communications Department of Bell Laboratories in Holmdel, New Jersey, and from 1984 to 1986 in the signal processing and coding theory group of the IBM Research Division in San Jose, California. He is currently an assistant professor of electrical engineering at Stanford University. Dr. Cioffi is a member of Eta Kappa Nu, Phi Eta Sigma, Phi Kappa Phi, Sigma Xi, Tau Beta Pi, and the Institute of Electrical and Electronics Engineers. He is the associate editor for adaptive filtering of the *IEEE Transactions on Acoustics, Speech, and Signal Processing.*