Decision-Feedback Equalization for Pulse-Position Modulation

Andrew G. Klein, *Member, IEEE*, and Pierre Duhamel, *Fellow, IEEE*

***Abstract*—In this paper, we propose a minimum mean squared error (MMSE) decision feedback equalizer (DFE) for pulse position modulated (PPM) signals in the presence of intersymbol interference (ISI). While traditional uses of PPM may not have had ISI, PPM is increasingly being considered for use in situations where ISI is an issue, such as high-performance optical communication systems and ultrawideband communications. First, we review previous work on the subject which used the zero-forcing criterion under strict assumptions about the channel and equalizer lengths. Then, we derive a computationally efficient MMSE equalizer which removes these restrictions, and is suitable for use with training-based stochastic gradient–descent algorithms. Finally, we demonstrate the performance of the proposed equalizer with simulations.**

***Index Terms*—Decision feedback equalization (DFE), optical communication, pulse position modulation (PPM), ultrawideband.**

I. Introduction

PULSE-POSITION MODULATION (PPM) is a modulation format that has seen considerable attention over the past several decades, particularly in applications which require high energy efficiency. Traditional uses of PPM have been in situations with little or no intersymbol interference (ISI), and thus, until recently, there had been little motivation to explore equalization of such signals to the extent that equalization has been explored for linearly modulated signals. In the 1980s and early 1990s, much attention was given to M-ary PPM for its use on optical and nondirected infared channels (see, for example, [1] and references therein). More recently, PPM has been considered by the ultrawideband (UWB) community for so-called impulse radios [2]. ISI is known to be a problem in both applications, yet the only serious studies of ISI compensation for PPM were conducted a decade ago in the optical communications context [3], [4]. Most of the recent work on PPM in the UWB arena employs RAKE reception, which does nothing to combat ISI. Consequently, we take another look at equalization for PPM, which largely extends and overcomes many of the shortcomings of [3] and [4].

While the optimum PPM detector in ISI is the maximum-likelihood sequence estimator (MLSE) which was investigated in [5], its complexity is usually too high for practical implementation, and thus, suboptimal schemes are preferred in practice. The unconstrained-length zero-forcing (ZF) decision feedback equalizer (DFE) for PPM was derived in [3] under several simplifying assumptions: The channel was monic and minimum phase, the additive noise was ignored in the design of the equalizer (i.e., since it is a ZF equalizer), and the feedback portion of the equalizer was as long as the channel (and possibly infinite).

In this paper, we propose a minimum mean squared error (MMSE) DFE for PPM signals which removes the assumptions placed on the previously proposed equalizer. In addition, our proposed equalizer significantly outperforms the previous scheme [3] and is suitable for use with a least mean square (LMS)-based adaptive algorithm for determining the equalizer taps when training data is available. We show that the structure of the PPM signal has properties that contradict conventional wisdom about equalization for more traditional modulations like pulse amplitude modulation (PAM). For example, in PAM systems, it is well known that a finite-length linear equalizer cannot perfectly invert a finite-impulse response (FIR) channel unless oversampling is employed [6]. By exploiting the structure of the PPM signal, we show that critically sampled PPM systems enable perfect linear equalization—no oversampling is required.

In Section II, we present the system model and show the design equations for the proposed MMSE-DFE. Again, we remove several assumptions inherent in the previous work [3]—that is, we permit nonmonic and nonminimum phase channels, we account for the additive white Gaussian noise (AWGN) in the design of our equalizer, and we permit the lengths of the FIR feedforward and feedback portions of the equalizer to be design parameters. We show that a modification to the equalizer which exploits properties of the PPM signal permits perfect linear equalization in critically sampled PPM systems, which has the added benefit of introducing a slight computational savings. We also show how the MMSE equalizer taps can be determined using training-based estimation of the equalizer coefficients via the LMS algorithm. In Section III, we demonstrate the performance of our equalizer and compare our results to that of [3]. We show that our proposed structure significantly outperforms that of [3] and is typically more computationally efficient. Section IV concludes the paper and presents future avenues of research.

Throughout this paper, we use $^\top$ to denote matrix transpose, $\otimes$ to denote the Kronecker product, $\mathbf{1}_{m\times n}$ to denote the $m \times n$ matrix of all ones, and $\mathbf{0}_{m\times n}$ to denote the $m \times n$ matrix of all zeros. The unit vector consisting of a 1 in the ith location and zero everywhere else will be denoted $\boldsymbol{e}_i$, and the $M \times M$ identity matrix is denoted $\boldsymbol{I}_M$. For a vector $\boldsymbol{x}$, the function $\mathrm{diag}(\boldsymbol{x})$ is the square diagonal matrix with $\boldsymbol{x}$ along the diagonal.

Manuscript received September 22, 2006; revised March 9, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mounir Ghogho. The work of A. G. Klein was supported by the Scientific Chateaubriand Fellowship.

A. G. Klein is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: agk5@cornell.edu).

P. Duhamel is with the Laboratoire des Signaux et Systémes (LSS), SUPELEC, 91190 Gif-sur-Yvette, France (e-mail: pierre.duhamel@lss.supelec.fr).

Digital Object Identifier 10.1109/TSP.2007.899392

II. Equalization for PPM

A. Properties of the PPM Source Sequence

M-ary PPM is an orthogonal transmission scheme where a symbol consists of M *chips*, only one of which is nonzero. PPM can be thought of as a block coding scheme where information is conveyed by the location of the nonzero sample within the block of M chips. Thus, the symbol alphabet comprises the M columns of the identity matrix $\boldsymbol{I}_M$. We assume that adjacent symbols are sent with no guard time between them, and as in [3], we assume a discrete-time model where each chip is sampled once. We denote the symbol transmitted at time n by $\boldsymbol{x}[n] \in \{\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{M-1}\}$. While we assume that symbols are independent identically distributed (i.i.d.), the resulting chip rate sequence is certainly not, though it is cyclostationary with period M. Clearly, the PPM source is not zero-mean, and $\boldsymbol{\mu}_x \triangleq E[\boldsymbol{x}[n]] = (1/M)\mathbf{1}_{M\times 1}$. Since the vector random process $\boldsymbol{x}[n]$ can be seen as an i.i.d. selection of columns of an identity matrix, the autocorrelation and autocovariance matrices are

$$E\left[\boldsymbol{x}[n]\boldsymbol{x}^\top[m]\right] = \begin{cases} \frac{1}{M}\boldsymbol{I}_M, & m=n \\ \frac{1}{M^2}\mathbf{1}_{M\times M}, & m\neq n \end{cases} \tag{1}$$

$$E\left[(\boldsymbol{x}[n]-\boldsymbol{\mu}_x)(\boldsymbol{x}[m]-\boldsymbol{\mu}_x)^\top\right] = \begin{cases} \frac{1}{M}\boldsymbol{J}_M, & m=n \\ \mathbf{0}_{M\times M}, & m\neq n \end{cases} \tag{2}$$

where we define

$$\boldsymbol{J}_M \triangleq \boldsymbol{I}_M - \frac{1}{M}\mathbf{1}_{M\times M}. \tag{3}$$

We use the notation $\bar{\boldsymbol{x}}[n] \triangleq [\boldsymbol{x}^\top[n], \boldsymbol{x}^\top[n-1], \cdots]^\top$ to denote a stacked symbol stream. For a stacked symbol vector of length N, then, we can use (1) to write the $N \times N$ autocorrelation matrix as

$$\boldsymbol{\mathcal{R}}_N \triangleq E\left[\bar{\boldsymbol{x}}[n]\bar{\boldsymbol{x}}^\top[n]\right] = \frac{1}{M^2}\mathbf{1}_{N\times N} + \frac{1}{M}\left(\boldsymbol{I}_{\frac{N}{M}} \otimes \boldsymbol{J}_M\right). \tag{4}$$

B. System Model

The system model for the transmitter, channel, and equalizer/receiver is shown in Fig. 1. The i.i.d. M-ary PPM symbols $\boldsymbol{x}[n]$ are transmitted serially through a causal linear time-invariant FIR channel of length N_h with impulse response $\boldsymbol{h} = [h[0] \ldots h[N_h-1]]^\top$ and AWGN $\boldsymbol{w}[n]$ where each sample has variance σ_w^2 assumed to be uncorrelated with the data. A vector model for the length N_f received vector at time n is then

$$\bar{\boldsymbol{y}}[n] = \boldsymbol{\mathcal{H}}\bar{\boldsymbol{x}}[n] + \bar{\boldsymbol{w}}[n] \tag{5}$$

where $\bar{\boldsymbol{y}}[n] \in \mathbb{R}^{N_f}$ is the stacked vector of received symbols, $N_c \triangleq N_f + N_h - 1$ is the combined length of the channel and feedforward equalizer, $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{N_f\times N_c}$ is the Tœplitz channel convolution matrix defined as $[\boldsymbol{\mathcal{H}}]_{i,j} = h[j-i]$, $\bar{\boldsymbol{x}}[n] \in \mathbb{R}^{N_c}$ is the serialized vector of transmitted PPM symbols, and $\bar{\boldsymbol{w}}[n] \in \mathbb{R}^{N_f}$ is the AWGN. Note that, as in [3], our receiver employs coherent reception in the sense that it preserves the polarity of the received signal (as opposed to simple energy detection).

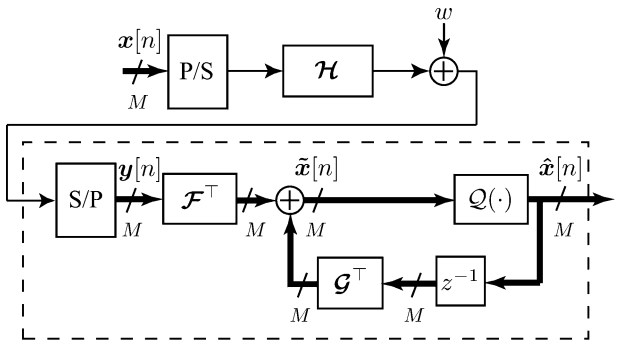

Fig. 1. Block equalizer system model.

To exploit the cyclostationary source statistics, we employ a symbol-rate block decision feedback equalizer (BDFE); a similar structure was previously proposed in [7] in the multiple-input–multiple-output (MIMO) context. Since this BDFE can alternatively be thought of as a single periodically time-varying chip rate DFE, it has equivalent computational complexity to a traditional single scalar DFE operating at the chip rate. The received signal is filtered with a block feedforward filter $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{N_f\times M}$, to which we add a contribution from a block feedback filter $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{N_g\times M}$. Note that the input to the feedback filter is the past decisions (i.e., the output of the decision device, delayed by one symbol). Thus, the input to the decision device can be expressed as

$$\tilde{\boldsymbol{x}}[n] = \boldsymbol{\mathcal{F}}^\top\bar{\boldsymbol{y}}[n] + \boldsymbol{\mathcal{G}}^\top\bar{\hat{\boldsymbol{x}}}[n-1] \tag{6}$$

where $\bar{\hat{\boldsymbol{x}}}[n] \in \mathbb{R}^{N_g}$ is the stacked vector of symbol estimates from the decision device.

The decision device that we employ here is the minimum Euclidean distance detector, which simply amounts to choosing the largest element of the length M received vector [8]. While we have chosen this decision device for its simplicity and low latency, we note that it is the optimal [i.e., maximum-likelihood (ML)] PPM decision device in the AWGN channel in the absence of ISI. A DFE requires a decision device with very low latency in the feedback path so that postcursor ISI can be subtracted; however, if desired, a more sophisticated decision device could make the ultimate decisions further down the receiver chain by using $\tilde{\boldsymbol{x}}[n]$ as input. We let $\mathcal{Q}(\cdot)$ denote the nonlinear decision-making operation, so that the decision device output can be written as

$$\hat{\boldsymbol{x}}[n] = \mathcal{Q}\left(\tilde{\boldsymbol{x}}[n]\right)$$

where the function $\mathcal{Q}(\cdot)$ is defined as

$$\mathcal{Q}\left(\tilde{\boldsymbol{x}}[n]\right) = \boldsymbol{e}_i \tag{7}$$

with i being the index of the largest element in $\tilde{\boldsymbol{x}}[n]$.

It is worth pointing out that the "BDFE-compare" structure proposed in [3] can be cast as a special case of our structure, with the following parameter choices: $N_f = M$, $\boldsymbol{\mathcal{F}} = \left([\boldsymbol{I}_M \mathbf{0}_{M\times(N_h-1)}]\boldsymbol{\mathcal{H}}^\top\right)^{-1}$, $N_g = N_h - 1$, and $\boldsymbol{\mathcal{G}} = -[\mathbf{0}_{(N_h-1)\times M}\boldsymbol{I}_{N_h-1}]\boldsymbol{\mathcal{H}}^\top\boldsymbol{\mathcal{F}}$. This is the (nonunique) ZF DFE, which also required the assumptions of a monic and minimum phase channel. Under the assumption of perfect feedback, the

equalizer will completely remove the ISI in this case, leaving only a filtered noise term. Several additional structures were considered in [3] which involve feedback of tentative chip decisions, but these structures are not considered in our paper.

C. Modified MMSE Equalizer

We now investigate a particular choice of equalizer coefficients that minimize the MSE. We *could* proceed in a straightforward manner by finding the choice of $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{G}}$ that minimizes the MSE between the equalizer output and the transmitted symbols

$$J_{\text{mse1}}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{G}}, \Delta) = E\left[\|\tilde{\boldsymbol{x}}[n] - \boldsymbol{x}[n-\Delta]\|_2^2\right]$$

for some combined channel/equalizer symbol delay Δ chosen by the system designer. However, this MSE criterion is stricter than it needs to be. Since the decision device is invariant to direct current (dc) offsets (i.e., adding some constant to all chips of a particular symbol), we are willing to accept an equalizer that introduces arbitrary dc offsets to a given symbol if it reduces the number of parameters in the equalizer. Because of this extra degree of freedom, we can reformulate the MSE criterion in a space of reduced dimension so that the equalizer operates only in the relevant subspace. To do this, we will consider a slight change to the equalizer structure in Fig. 1 and an alternate expression for MSE that enables us to exploit the structure of the PPM signal, thereby reaping a complexity reduction. First, we introduce three lemmas which serve to motivate our modified equalizer structure.

Lemma 1 [Properties of $\mathcal{Q}(\cdot)$]: The PPM minimum Euclidean distance detector function $\mathcal{Q}(\cdot)$ defined in (7) satisfies

$$\mathcal{Q}(\tilde{\boldsymbol{x}}[n]) = \mathcal{Q}(\boldsymbol{J}_M \tilde{\boldsymbol{x}}[n])$$

for $\boldsymbol{J}_M$ defined in (3) and any $\tilde{\boldsymbol{x}}[n] \in \mathbb{R}^M$.

Proof: See Appendix I. ∎

Lemma 2 (Properties of $\boldsymbol{J}_M$): The matrix $\boldsymbol{J}_M$ defined in (3) is idempotent, symmetric, positive semidefinite, and has rank $M-1$.

Proof: See Appendix II. ∎

Lemma 3 (Modified Cholesky Decomposition): For any square symmetric positive–semidefinite matrix $\boldsymbol{J}_M \in \mathbb{R}^{M\times M}$, there exists a rectangular upper triangular matrix $\boldsymbol{U}_M \in \mathbb{R}^{r\times M}$ with strictly positive diagonal elements such that

$$\boldsymbol{U}_M^\top \boldsymbol{U}_M = \boldsymbol{J}_M$$

where $r \le M$ is the rank of $\boldsymbol{J}_M$.

Proof: See Appendix III. ∎

Thus, according the Lemma 1, we could modify the original structure in Fig. 1 by inserting a block that premultiplies $\tilde{\boldsymbol{x}}[n]$ with $\boldsymbol{J}_M$ just before the decision device, and this change would have no effect on the output of the decision device. From Lemmas 2 and 3, $\boldsymbol{J}_M$ admits a factorization $\boldsymbol{U}_M^\top \boldsymbol{U}_M$, where $\boldsymbol{U}_M \in \mathbb{R}^{M-1\times M}$. Note that an explicit expression for $\boldsymbol{U}_M$ is provided in Appendix IV. Thus, we could alternatively modify Fig. 1 by absorbing $\boldsymbol{U}_M$ into the equalizer coefficients $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{G}}$, and then inserting a block that premultiplies $\tilde{\boldsymbol{x}}[n]$ with $\boldsymbol{U}_M^\top$ just before the decision device. This modified equalizer system model is shown in Fig. 2, where now the dimensions of the equalizers reduce to $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{N_f\times M-1}$ and $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{N_g\times M-1}$, the dimension of the equalizer output reduces to $\tilde{\boldsymbol{x}}[n] \in \mathbb{R}^{M-1}$, and the decision device output becomes $\hat{\boldsymbol{x}}[n] = \mathcal{Q}(\boldsymbol{U}_M^\top \tilde{\boldsymbol{x}}[n])$. Note that we incur a slight computational savings by reducing the number of equalizer taps by a factor of $M/(M-1)$.

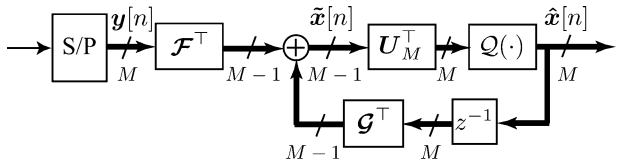

Fig. 2. Modified block equalizer system model.

It is worth considering the geometric interpretation of the projection via $\boldsymbol{J}_M$ and $\boldsymbol{U}_M$. As we noted previously, a PPM signal has nonzero mean. By simple coordinate translation of a PPM signal set (i.e., by subtracting the mean from each chip), we arrive at a new signal set which is the so-called transorthogonal or simplex set [8]. This is precisely the role of $\boldsymbol{J}_M$, which is apparent from the definition in (3); that is, $\boldsymbol{J}_M$ projects a PPM signal set into a corresponding simplex signal set via coordinate translation. Translation of the origin does not affect the Euclidean distance properties of the signal set, nor does premultiplication with $\boldsymbol{J}_M$ as we have seen from Lemma 1. It is also well known that the *dimensionality* of an M-ary simplex signal set is $M-1$ [8]; that is, through appropriate choice of coordinate system, an M-ary simplex signal set can be represented with just $M-1$ chips. Furthermore, an M-ary PPM signal set can be projected into an M-ary simplex set represented by $M-1$ chips, with identical Euclidean distance properties; this is the role of $\boldsymbol{U}_M$. Finally, we note that projection via $\boldsymbol{U}_M$ is invariant to dc offsets, so for some vector $\boldsymbol{x} \in \mathbb{R}^M$ and some scalar b

$$\boldsymbol{U}_M(\boldsymbol{x} + b\mathbf{1}_{M\times 1}) = \boldsymbol{U}_M \boldsymbol{x}$$

which can be proved by using the definition of $\boldsymbol{J}_M$ and showing that $\|\boldsymbol{U}_M \mathbf{1}_{M\times 1}\|_2^2 = 0$. Thus, by absorbing $\boldsymbol{U}_M$ into the equalizer coefficients, the equalizer output $\tilde{\boldsymbol{x}}[n] \in \mathbb{R}^{M-1}$ is of reduced dimension, is Euclidean distance preserving, and is invariant to dc shifts at the equalizer input. We reemphasize that this modification to the equalizer structure does not effect the input–output relationships as seen from the preceding lemmas.

D. Calculation of the MMSE Coefficients

We can now find the equalizer coefficients that minimize the MSE given by

$$J_{\text{mse2}}(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{G}}, \Delta) = E\left[\|\tilde{\boldsymbol{x}}[n] - \boldsymbol{U}_M \boldsymbol{x}[n-\Delta]\|_2^2\right] \qquad (8)$$

where we note that the appearance of $\boldsymbol{U}_M$ in the MSE equation serves to project the source signal into the space of reduced dimension so that it is compatible with the equalizer output $\tilde{\boldsymbol{x}}[n]$. Before proceeding, we make the following assumptions about the lengths of the channel and equalizers which serve to simplify the notation:

- N_c is a multiple of M;
- $N_g \le N_c - M(\Delta+1)$;

- N_g is a multiple of M.

By artificially appending zeros to $\boldsymbol{h}$, thus increasing N_c, the first two assumptions can be trivially satisfied without loss of generality for any set of system parameters. The third assumption is reasonable since it effectively requires feedback of whole symbols.

Next, we make several definitions to aid our development of the MMSE equalizer coefficients. Again, in addition to the equalizer lengths N_f and N_g, the MMSE equalizer accepts one other design parameter Δ, which represents the desired symbol delay through the channel/equalizer chain. By defining

$$\bar{\boldsymbol{E}}_\Delta \triangleq \begin{bmatrix} \boldsymbol{0}_{M\Delta\times M} \\ \boldsymbol{I}_M \\ \boldsymbol{0}_{N_c-M(\Delta+1)\times M} \end{bmatrix} \tag{9}$$

where $\boldsymbol{E}_\Delta \in \mathbb{R}^{N_c\times M}$, we can express the delayed symbol vector in terms of the source symbol stream as

$$\boldsymbol{x}[n-\Delta] = \boldsymbol{E}_\Delta^\top \bar{\boldsymbol{x}}[n]. \tag{10}$$

Substituting (5), (6), and (10) into (8), we can expand the MSE as

$$\begin{aligned} J_{\mathrm{mse2}}(\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{G}},\Delta) &= E\left[\left\| \boldsymbol{\mathcal{F}}^\top(\boldsymbol{\mathcal{H}}\bar{\boldsymbol{x}}[n]+\bar{\boldsymbol{w}}[n]) + \boldsymbol{\mathcal{G}}^\top\bar{\hat{\boldsymbol{x}}}[n-1] \right.\right. \\ &\qquad \left.\left. -\boldsymbol{U}_M\boldsymbol{E}_\Delta^\top\bar{\boldsymbol{x}}[n]\right\|_2^2\right] \\ &= \mathrm{tr}\Big(\boldsymbol{\mathcal{F}}^\top[\boldsymbol{\mathcal{H}}\boldsymbol{R}_{xx}\boldsymbol{\mathcal{H}}^\top + \boldsymbol{R}_{ww}]\boldsymbol{\mathcal{F}} \\ &\qquad + 2\boldsymbol{\mathcal{F}}^\top\boldsymbol{\mathcal{H}}\boldsymbol{R}_{x\hat{x}}\boldsymbol{\mathcal{G}} - 2\boldsymbol{\mathcal{F}}^\top\boldsymbol{\mathcal{H}}\boldsymbol{R}_{xx}\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top \\ &\qquad + \boldsymbol{\mathcal{G}}^\top\boldsymbol{R}_{\hat{x}\hat{x}}\boldsymbol{\mathcal{G}} - 2\boldsymbol{\mathcal{G}}^\top\boldsymbol{R}_{x\hat{x}}^\top\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top \\ &\qquad + \boldsymbol{U}_M\boldsymbol{E}_\Delta^\top\boldsymbol{R}_{xx}\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top\Big) \end{aligned}$$

where we define

$$\boldsymbol{R}_{xx} \triangleq E\left[\bar{\boldsymbol{x}}[n]\bar{\boldsymbol{x}}^\top[n]\right] \tag{11}$$

$$\boldsymbol{R}_{x\hat{x}} \triangleq E\left[\bar{\boldsymbol{x}}[n]\bar{\hat{\boldsymbol{x}}}^\top[n-1]\right] \tag{12}$$

$$\boldsymbol{R}_{\hat{x}\hat{x}} \triangleq E\left[\bar{\hat{\boldsymbol{x}}}[n-1]\bar{\hat{\boldsymbol{x}}}^\top[n-1]\right] \tag{13}$$

$$\boldsymbol{R}_{ww} \triangleq E\left[\bar{\boldsymbol{w}}[n]\bar{\boldsymbol{w}}^\top[n]\right]. \tag{14}$$

Under the AWGN assumption, we have $\boldsymbol{R}_{ww} = \sigma_w^2\boldsymbol{I}_{N_f}$. Under the assumption of correct feedback, we have that $\hat{\boldsymbol{x}}[n] = \boldsymbol{x}[n-\Delta]$ or, alternatively

$$\bar{\hat{\boldsymbol{x}}}[n-1] = \boldsymbol{\Sigma}_\Delta\bar{\boldsymbol{x}}[n] \tag{15}$$

where the matrix

$$\boldsymbol{\Sigma}_\Delta \triangleq \begin{bmatrix} \boldsymbol{0}_{N_g\times M(\Delta+1)} & \boldsymbol{I}_{N_g} & \boldsymbol{0}_{N_g\times N_c-M(\Delta+1)-N_g} \end{bmatrix} \tag{16}$$

selects the appropriate elements of $\bar{\boldsymbol{x}}[n]$ to construct $\bar{\hat{\boldsymbol{x}}}[n-1]$.

From (4) and (15), the correlation matrices can be rewritten as

$$\boldsymbol{R}_{xx} = \boldsymbol{\mathcal{R}}_{N_c} \tag{17}$$

$$\boldsymbol{R}_{x\hat{x}} = \boldsymbol{\mathcal{R}}_{N_c}\boldsymbol{\Sigma}_\Delta^\top \tag{18}$$

$$= \begin{bmatrix} \frac{1}{M^2}\boldsymbol{1}_{M(\Delta+1)\times N_g} \\ \boldsymbol{\mathcal{R}}_{N_g} \\ \frac{1}{M^2}\boldsymbol{1}_{N_c-N_g-M(\Delta+1)\times N_g} \end{bmatrix} \tag{19}$$

$$\boldsymbol{R}_{\hat{x}\hat{x}} = \boldsymbol{\Sigma}_\Delta\boldsymbol{\mathcal{R}}_{N_c}\boldsymbol{\Sigma}_\Delta^\top \tag{20}$$

$$= \boldsymbol{\mathcal{R}}_{N_g} \tag{21}$$

where we see that all the correlation matrices can be expressed in terms of appropriately sized $\boldsymbol{\mathcal{R}}_N$ defined in (4). Incidentally, we note that the assumption of correct feedback only enters the MMSE development in the definition of $\boldsymbol{R}_{x\hat{x}}$ in (18) or (19). To find the optimal feedforward and feedback filters, we set the gradient of the MSE to zero. First, we will find the feedback taps (in terms of the feedforward taps), and then we will find the feedforward taps. Taking the derivative with respect to $\boldsymbol{\mathcal{G}}$, setting to zero, and solving for $\boldsymbol{\mathcal{G}}$ gives

$$\begin{aligned} \frac{1}{2}\frac{\partial J_{\mathrm{mse2}}(\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{G}},\Delta)}{\partial\boldsymbol{\mathcal{G}}} &= \boldsymbol{R}_{x\hat{x}}^\top\left(\boldsymbol{\mathcal{H}}^\top\boldsymbol{\mathcal{F}} - \boldsymbol{E}_\Delta\boldsymbol{U}_M^\top\right) + \boldsymbol{R}_{\hat{x}\hat{x}}\boldsymbol{\mathcal{G}} \\ &= \boldsymbol{0} \\ \Longrightarrow \boldsymbol{R}_{\hat{x}\hat{x}}\boldsymbol{\mathcal{G}}^* &= \boldsymbol{R}_{x\hat{x}}^\top\left[\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top - \boldsymbol{\mathcal{H}}^\top\boldsymbol{\mathcal{F}}\right]. \end{aligned} \tag{22}$$

Solving this expression for $\boldsymbol{\mathcal{G}}^*$ (i.e., the choice of $\boldsymbol{\mathcal{G}}$ that minimizes the MSE) requires inversion of $\boldsymbol{R}_{\hat{x}\hat{x}}$, which we will show to have nullity $N_g/M - 1$. However, the linear system (22) is *consistent* since the right-hand side of the equation is in the column space of $\boldsymbol{R}_{\hat{x}\hat{x}}$. Thus, (22) is an underdetermined system with an infinite number of solutions. These points are made clear in the following two lemmas.

Lemma 4 [Singular Value Decomposition (SVD) of $\boldsymbol{\mathcal{R}}_N$]: For a length N stream of chips consisting of N/M symbols, the symmetric autocorrelation matrix $\boldsymbol{\mathcal{R}}_N$ defined in (4) has singular value decomposition given by $\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^\top$, where

$$\boldsymbol{V} = \left[\frac{1}{\sqrt{N}}\boldsymbol{1}_{N\times 1}\ \left(\boldsymbol{I}_{N/M}\otimes\boldsymbol{U}_M^\top\right)\ \frac{1}{\sqrt{M}}\left(\boldsymbol{U}_{N/M}^\top\otimes\boldsymbol{1}_{M\times 1}\right)\right]$$

and

$$\boldsymbol{D} = \mathrm{diag}([N/M^2, \underbrace{1/M,\ldots,1/M}_{N-N/M}, \underbrace{0,\ldots,0}_{N/M-1}]).$$

Proof: Recall that $\boldsymbol{U}_M \in \mathbb{R}^{M-1\times M}$ is the factorization of $\boldsymbol{U}_M^\top\boldsymbol{U}_M = \boldsymbol{I}_M - (1/M)\boldsymbol{1}_{M\times M}$ and $\boldsymbol{U}_{N/M}$ is defined identically with appropriate size. Clearly, $\boldsymbol{D}$ is diagonal. All that remains to be shown is that $\boldsymbol{V}$ is orthogonal and that $\boldsymbol{\mathcal{R}}_N = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^\top$, which follow from the definitions. ∎

Lemma 5 [On the Solvability of (22)]: The linear system

$$\boldsymbol{R}_{\hat{x}\hat{x}}\boldsymbol{x} = \boldsymbol{R}_{x\hat{x}}^\top\boldsymbol{b} \tag{23}$$

has an infinite number of solutions $\boldsymbol{x}$ for any $\boldsymbol{b}$.

Proof: Since $\boldsymbol{R}_{\hat{x}\hat{x}}$ is rank deficient, we only need to show that this system is consistent. By substituting (18) and (20), the system assumes the form of the normal equations, which are known to be consistent [9]. ∎

Consequently, we have flexibility in our choice of $\boldsymbol{\mathcal{G}}^*$, and we now consider two such choices. One possibility is to find a solution that attempts to reduce the effects of error propagation. As discussed in [10], large taps in the feedback portion of DFEs have a tendency to enhance the effects of error propagation. A sensible approach, then, might be to use the minimum Euclidean norm solution of $\boldsymbol{\mathcal{G}}^*$, which is found easily by multiplying both sides of (22) with the Moore–Penrose pseudoinverse of $\boldsymbol{R}_{\hat{x}\hat{x}}$. This solution may give a small performance improvement in environments with significant symbol errors. Alternatively, we can make another choice of $\boldsymbol{\mathcal{G}}^*$ which incurs a computational savings, by constraining entire rows of $\boldsymbol{\mathcal{G}}^*$ to be zero. This approach was explored in [11], where it was shown that $N_g/M-1$ of the N_g rows in $\boldsymbol{\mathcal{G}}^*$ can be set to zero.

We define $\boldsymbol{R}_{\hat{x}\hat{x}}^-$ to be the matrix 1-inverse[1] of $\boldsymbol{R}_{\hat{x}\hat{x}}$, chosen by the system designer to minimize the norm (in which case, it is the Moore–Penrose inverse), to zero rows in $\boldsymbol{\mathcal{G}}^*$, or perhaps to satisfy some other criterion. Thus, we write

$$\boldsymbol{\mathcal{G}}^* = \boldsymbol{R}_{\hat{x}\hat{x}}^- \boldsymbol{R}_{x\hat{x}}^\top \left[\boldsymbol{E}_\Delta \boldsymbol{U}_M^\top - \boldsymbol{\mathcal{H}}^\top \boldsymbol{\mathcal{F}}\right]. \tag{24}$$

Now, to calculate the feedforward taps, we proceed by taking the derivative of the MSE with respect to $\boldsymbol{\mathcal{F}}$, substituting $\boldsymbol{\mathcal{G}}^*$, and setting to zero to give

$$\begin{aligned}\frac{1}{2}\frac{\partial J_{\mathrm{mse2}}(\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{G}}^*,\Delta)}{\partial \boldsymbol{\mathcal{F}}} &= \boldsymbol{\mathcal{H}}\boldsymbol{R}_{xx}\boldsymbol{\mathcal{H}}^\top\boldsymbol{\mathcal{F}} + \boldsymbol{R}_{ww}\boldsymbol{\mathcal{F}} \\ &\quad + \boldsymbol{\mathcal{H}}\boldsymbol{R}_{x\hat{x}}\boldsymbol{\mathcal{G}}^* - \boldsymbol{\mathcal{H}}\boldsymbol{R}_{xx}\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top \\ &= \boldsymbol{0} \\ \Longrightarrow \boldsymbol{\mathcal{F}}^* &= (\boldsymbol{\mathcal{H}}\boldsymbol{\Psi}\boldsymbol{\mathcal{H}}^\top + \boldsymbol{R}_{ww})^{-1} \\ &\quad \cdot \boldsymbol{\mathcal{H}}\boldsymbol{\Psi}\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top \end{aligned} \tag{25}$$

where

$$\boldsymbol{\Psi} \triangleq \boldsymbol{R}_{xx} - \boldsymbol{R}_{x\hat{x}}\boldsymbol{R}_{\hat{x}\hat{x}}^-\boldsymbol{R}_{x\hat{x}}^\top.$$

In summary, then, the MMSE equalizer taps for the equalizer structure in Fig. 2 are found by first solving for the feedforward taps via (25) and then the feedback taps via (24).

E. Conditions for Perfect Feedforward Equalization

We now consider the conditions that permit a feedforward equalizer to perfectly remove the ISI. For more traditional modulations like PAM, it is well known that a finite-length linear equalizer cannot perfectly invert an FIR channel unless oversampling is employed. It is perhaps a bit surprising that, even without oversampling, all ISI can be removed in PPM systems with only a finite-length feedforward filter. We assume in this section that there is no feedback equalizer (or, alternatively, $N_g = 0$), no noise, and therefore, perfect equalization arises when the MSE is exactly zero. With no noise and no feedback section, the MSE reduces to

$$\begin{aligned}J_{\mathrm{mse2}}(\boldsymbol{\mathcal{F}},\Delta) = \mathrm{tr}\Big(&\boldsymbol{\mathcal{F}}^\top[\boldsymbol{\mathcal{H}}\boldsymbol{R}_{xx}\boldsymbol{\mathcal{H}}^\top]\boldsymbol{\mathcal{F}} \\ &+ \left[\boldsymbol{U}_M\boldsymbol{E}_\Delta^\top - 2\boldsymbol{\mathcal{F}}^\top\boldsymbol{\mathcal{H}}\right]\boldsymbol{R}_{xx}\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top\Big).\end{aligned}$$

From Lemma 4, the nullity of $\boldsymbol{R}_{xx} \in \mathbb{R}^{N_c\times N_c}$ is $N_c/M-1$, and from Lemma 3, there exists a decomposition $\boldsymbol{\Phi}_{xx}^\top\boldsymbol{\Phi}_{xx} = \boldsymbol{R}_{xx}$, where $\boldsymbol{\Phi}_{xx} \in \mathbb{R}^{N_c - N_c/M+1\times N_c}$. Letting

$$\boldsymbol{\mathcal{H}}' \triangleq \boldsymbol{\mathcal{H}}\boldsymbol{\Phi}_{xx}^\top$$

the MSE becomes

$$J_{\mathrm{mse2}}(\boldsymbol{\mathcal{F}},\Delta) = \left\|\boldsymbol{\mathcal{F}}^\top\boldsymbol{\mathcal{H}}' - \boldsymbol{U}_M\boldsymbol{E}_\Delta^\top\boldsymbol{\Phi}_{xx}^\top\right\|_{fro}^2$$

where $\|\cdot\|_{fro}$ denotes the Frobenius norm. The condition for perfect equalization is that the MSE is zero or

$$\boldsymbol{\mathcal{H}}'^\top\boldsymbol{\mathcal{F}} = \boldsymbol{\Phi}_{xx}\boldsymbol{E}_\Delta\boldsymbol{U}_M^\top. \tag{26}$$

For the existence of an exact solution of $\boldsymbol{\mathcal{F}}$ for this linear system, $\boldsymbol{\mathcal{H}}' \in \mathbb{R}^{N_f\times N_c - N_c/M+1}$ must be tall and full rank, leading to the following two conditions for perfect equalization.

1) *Equalizer length condition*: For $\boldsymbol{\mathcal{F}}$ to satisfy (26) for arbitrary Δ, $\boldsymbol{\mathcal{H}}'$ must be a tall matrix. Hence, it is required that

$$N_f > N_h(M-1). \tag{27}$$

2) *Channel disparity condition*: For $\boldsymbol{\mathcal{F}}$ to satisfy (26) for arbitrary Δ, $\boldsymbol{\mathcal{H}}'$ must have full column rank.

Thus, even in the absence of a feedback portion, our equalizer structure can succeed in perfectly equalizing the channel, so long as the length and disparity conditions are satisfied. While (27) can be used to provide the system designer some guidelines in selecting the length of the feedforward equalizer, it is by no means a necessary condition for adequate system performance, particularly when a feedback portion is in use. The channel disparity condition is similar to that required in fractionally spaced PAM systems [12], though finding an analogous interpretation that imposes conditions on "common subchannel roots" does not seem feasible. Nevertheless, we conjecture that the class of channels satisfying the disparity condition is quite large, and again, it is not a necessary condition for adequate system performance.

F. Adaptive Implementation via LMS

Up to now, we have assumed that the receiver has knowledge of the channel impulse response. Even with perfect knowledge of the channel impulse response, however, solving (24) and (25) incurs a considerable computational penalty. Fortunately, our equalizer structure is quite suitable for use with stochastic gradient–descent algorithms, such as LMS, to adaptively determine the MMSE equalizer coefficients when training data is available. The LMS algorithm is a stochastic gradient–descent algorithm which uses the instantaneous gradient [13] of the MSE as an

[1]An $n \times m$ matrix $\boldsymbol{A}^-$ is a 1-inverse of an $m \times n$ matrix $\boldsymbol{A}$ for which $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$.

estimate for the true gradient (i.e., by ignoring the expectation operator).

Taking the instantaneous derivative of (8) results in the matrices

$$\frac{1}{2}\frac{\partial \hat{J}_{\mathrm{mse2}}}{\partial \boldsymbol{\mathcal{F}}} = \bar{\boldsymbol{y}}[n]\left[\tilde{\boldsymbol{x}}[n] - \bar{\boldsymbol{U}}_M \boldsymbol{x}[n-\Delta]\right]^\top$$
$$\frac{1}{2}\frac{\partial \hat{J}_{\mathrm{mse2}}}{\partial \boldsymbol{\mathcal{G}}} = \bar{\hat{\boldsymbol{x}}}[n-1]\left[\tilde{\boldsymbol{x}}[n] - \boldsymbol{U}_M \boldsymbol{x}[n-\Delta]\right]^\top$$

which results in the LMS update equations as

$$\begin{aligned} \boldsymbol{\epsilon}_{\mathrm{lms}}[n] &= \tilde{\boldsymbol{x}}[n] - \boldsymbol{U}_M \boldsymbol{x}[n-\Delta] \\ \boldsymbol{\mathcal{F}}_{\mathrm{lms}}[n+1] &= \boldsymbol{\mathcal{F}}_{\mathrm{lms}}[n] - \mu_1 \bar{\boldsymbol{y}}[n]\boldsymbol{\epsilon}_{\mathrm{lms}}^\top[n] \\ \boldsymbol{\mathcal{G}}_{\mathrm{lms}}[n+1] &= \boldsymbol{\mathcal{G}}_{\mathrm{lms}}[n] - \mu_2 \bar{\tilde{\boldsymbol{x}}}[n-1]\boldsymbol{\epsilon}_{\mathrm{lms}}^\top[n] \end{aligned} \tag{28}$$

where μ_1 and μ_2 are small positive step sizes which serve to average out the noise in the gradient estimate, and the presence of $\boldsymbol{x}[n-\Delta]$ in the error term implies the availability of training data. When training data is unavailable, we can feed back the output of the decision device $\hat{\boldsymbol{x}}[n]$ instead, arriving at the decision-directed DD-LMS update equation for the block equalizer; this is accomplished by simply replacing $\boldsymbol{x}[n-\Delta]$ with $\hat{\boldsymbol{x}}[n]$ in the equation for the error term (28).

We recall that $\boldsymbol{R}_{\hat{x}\hat{x}}$ is noninvertible, and consequently, the optimal feedback filter $\boldsymbol{\mathcal{G}}$ can assume an infinite number of possibilities. While this is really only a concern for direct calculation of the MMSE equalizer taps via (24) and (25), it does have implications in the adaptive implementation. In Section II-D, we hinted at a possible choice of $\boldsymbol{\mathcal{G}}$ that involved constraining some of the rows to zero [11]. During adaptation with LMS, this constraint can be imposed quite simply, as well, to benefit from the computational savings it provides. Without a constraint, the system is underdetermined and $\boldsymbol{\mathcal{G}}_{\mathrm{lms}}[n]$ will wander through the subspace of allowable MMSE solutions.

From an error propagation perspective, however, it could be preferable not to impose a constraint during LMS adaptation. While the choice of constraint has no affect on the performance of the DFE in the absence of decision errors, some constraints may be better than others in the presence of decision errors. It was reported in [14] that the MMSE adaptation mechanism in the presence of error propagation can find a better answer than the solution computed in the absence of decision errors. Thus, it may be preferable not to impose any constraint at all, and instead rely on the LMS algorithm to find the best solution in the presence of decision errors. Extensive simulation, however, has not shown any observable difference among the different contraints (or lack thereof) in terms of error rate or convergence speed.

Finally, we point out that DD algorithms are notoriously sensitive to initialization. Thus, in the absence of training data, it is by no means a guarantee that the DD-LMS algorithm will converge to the MMSE solution. However, the same structure of the PPM signal which we have exploited here in the equalization context permits subspace-based blind channel estimation for PPM without oversampling, so long as the same channel length and disparity conditions given in Section II-E are satisfied [15]. Such blind subspace-based estimation schemes require perfect knowledge of the channel length, however, and incur a severe performance penalty in the presence of channel overmodeling. While it may be possible to develop alternative blind adaptive gradient–descent algorithms with better convergence properties than DD-LMS, and without the overmodeling problems of subspace-based estimators, preliminary attempts [11] suggest that traditional blind algorithms like the constant modulus algorithm (CMA) or vector-CMA are not suitable for use with PPM signals.

TABLE I
COMPUTATIONAL COMPLEXITY OF EQUALIZERS

equalizer	number of multiplies
traditional chip-rate scalar	$M(N_f+N_g)$
BDFE-compare scheme of [3]	$M(M+N_h-1)$
proposed MMSE scheme	$(M-1)(N_f+N_g+2)-1$
proposed MMSE scheme with constraint on feedback taps	$(M-1)(N_f+N_g+2)-N_g/M-2$

III. SIMULATIONS AND NUMERICAL EXAMPLES

A. Computational Complexity of Equalizer

As computational complexity is typically dominated by multiplication operations, we consider the number of multiplies required to equalize one PPM symbol with three equalizer structures: a traditional scalar equalizer, the previously proposed ZF scheme [3], and the newly proposed MMSE-DFE. For comparison, we first consider a traditional scalar equalizer, where N_f+N_g multiplies are required per chip, thus $M(N_f+N_g)$ multiplies are required per symbol. In the block ZF scheme of [3], the block equalizer also requires $M(N_f+N_g)$ multiplications for one symbol; however, the equalizer lengths are not design parameters, but are dependent on the alphabet size and channel length, i.e., $N_f = M$ and $N_g = N_h - 1$, thus requiring $M(M+N_h-1)$ multiplications.

The equalizers in the BDFE proposed in this paper are of reduced size, i.e., $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{N_f \times M-1}$ and $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{N_g \times M-1}$. Thus, equalization of a symbol with the proposed equalizer requires $(M-1)(N_f+N_g)$ multiplications before accounting for subsequent multiplication by $\boldsymbol{U}_M^\top$, as shown in Fig. 2. We note from Appendix IV that $\boldsymbol{U}_M^\top$ is lower triangular, only has two unique values per column, and that the last column only has one unique value; thus, it requires $2(M-2)+1$ multiplications. Further computational savings can be obtained by constraining rows of $\boldsymbol{\mathcal{G}}$ to be zero, as discussed in Section II-D. The computational complexity of each scheme is presented in Table I. We observe that for very short equalizer and channel lengths, the various structures all have roughly the same complexity. As N_f, N_g, and N_h grow large, however, the proposed structure has slightly less complexity by roughly the factor $(M-1)/M$.

B. Equalizer Performance Comparison With [3]

Here, we consider the same simulation setup as was performed in [3]. That is, $M=2$, $\boldsymbol{h} = [1\ -1\ 1]^\top$, and the noise is AWGN. We are forced to consider channels that are monic

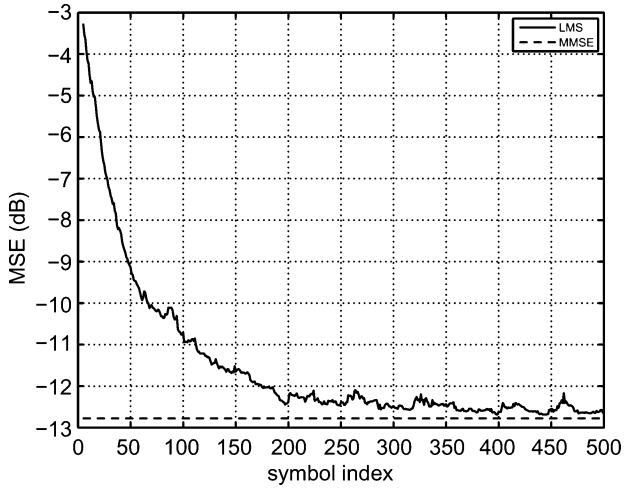

Fig. 3. Convergence of LMS equalizer to MMSE solution at 8-dB SNR.

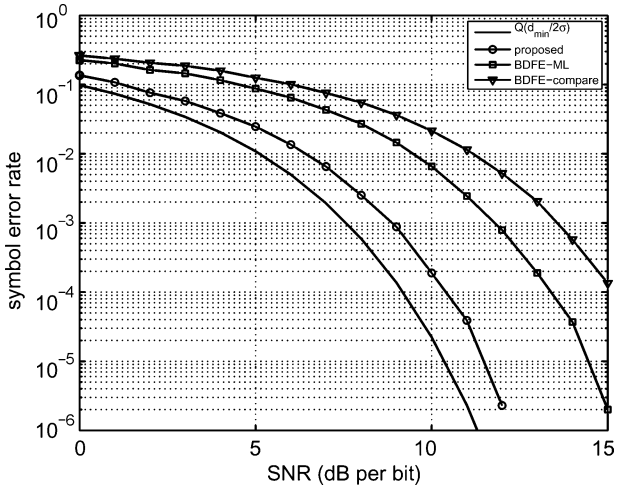

Fig. 4. Comparison of ZF and MMSE bit error rate with $M = 2$.

and minimum phase to permit comparison of our scheme with that of [3], and this simple channel from [3] falls into this class. In addition, while this channel seems a bit unrealistic, it has roots on the unit circle, which is a situation known to pose problems in equalization. For the proposed MMSE equalizer, we chose $N_f = 6$, $N_g = 2$, and $\Delta = 2$. The proposed equalizer was trained with LMS with $\mu_1 = \mu_2 = 0.01$, and given 1000 symbols of random training data before calculating the symbol error rate. While 1000 training symbols may seem excessively large, we note that the LMS algorithm typically converges in much shorter time, and such a large number of samples was chosen only as a matter of practicality to ensure algorithm convergence before the symbol error rate calculation. An example of LMS convergence is shown in Fig. 3 for an signal-to-noise ratio (SNR) of 8 dB. The ZF equalizer in [3], on the other hand, was given full "genie-aided" channel knowledge. The results are shown in Fig. 4 and the proposed MMSE equalizer demonstrates approximately a 4.5-dB performance gain over the ZF BDFE-compare scheme [3], even at high SNR. While several alternate equalizer structures were proposed in [3], the so-called BDFE-compare is the closest to our structure. The BDFE-ML scheme, which our proposed structure also outperforms, was the best performing equalizer among those presented in [3], as it substitutes a (considerably more complex) ML decision device for the simple Euclidean distance detector but uses the same equalizer coefficients as the BDFE-compare. As mentioned in the introduction, the optimal detector is the pure MLSE; the symbol error probability of the MLSE is well-approximated at high SNR by [16]

$$P_e = Q\left(\frac{d_{\min}}{2\sigma_w^2}\right)$$

where $Q(\cdot)$ is the Gaussian Q-function, and the minimum Euclidean distance is $d_{\min} = 10$ for this channel [3]. This bound is also plotted in Fig. 4 for comparison.

Typically, ZF equalizers have similar performance to MMSE equalizers in an ISI-dominated regime where the SNR is high. However, this is not the case here because our proposed structure is fundamentally different in the way we perform the feedforward equalization. In the scheme of [3], the feedfoward equalizer $\boldsymbol{\mathcal{F}}$ will always be a square lower triangular Tœplitz matrix, so the Mth chip of each symbol effectively has no feedforward equalization—only a gain. Since our proposed scheme does not impose this structure, it has a better ability to suppress ISI in addition to having less noise enhancement. We observed similar performance gains at $M = 8$, with the proposed MMSE-DFE again outperforming the ZF BDFE-compare by approximately 4.5 dB.

C. Performance on Random Channels With $M = 8$

Next, we consider the performance of the equalizers on random channels for the case where $M = 8$. We chose the channels to have length $N_h = 12$ and generated a set of 100 random channels where each of the 12 channel taps are chosen to be i.i.d. standard Gaussian random variables. The parameters for the proposed MMSE-DFE were chosen to be $N_f = 16$, $N_g = 8$, and $\Delta = 1$. For these simulations, no LMS adaptation was used and the MMSE equalizer coefficients were calculated with full channel knowledge. The results are shown in Fig. 5, where we see from the error rate that the ZF BDFE-compare of [3] did not succeed in equalizing the channels. The authors in [3] claim that the ZF BDFE is not intended for nonminimum phase channels, and indeed this is evident from the performance in Fig. 5. Since the ZF BDFE feedforward section effectively inverts the leading channel coefficient, relying on the feedback section to subtract all the interference, the noise becomes amplified immensely in situations where the first channel tap is small.

A performance comparison on nonminimum phase channels is not completely fair since it employs the ZF BDFE in a situation for which it was not designed. Consequently, we repeated the exact same experiment, but with one modification. Taking the same set of 100 channels, we first reflected all of their roots to be inside the unit circle, resulting in a set of minimum phase channels with unchanged frequency response. The performance on this modified set of channels is shown in Fig. 6, where we see

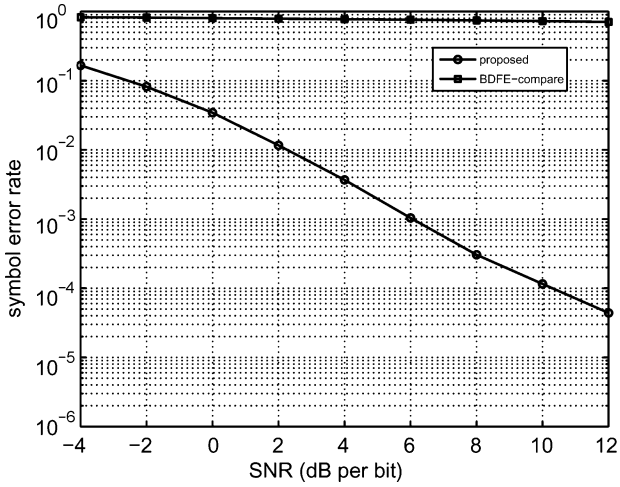

Fig. 5. Comparison of proposed MMSE-DFE with ZF BDFE on random channels with $M = 8$.

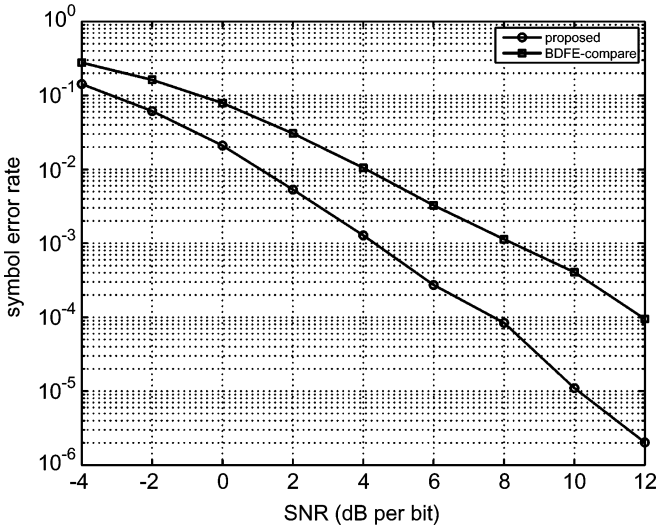

Fig. 6. Comparison of proposed MMSE-DFE with ZF BDFE on minimum phase random channels with $M = 8$.

that the proposed MMSE equalizer again demonstrates approximately a 4.5-dB performance gain over the ZF BDFE-compare scheme.

IV. Conclusion

Now that more than a decade has passed since the first look at equalization for PPM [3], we have taken a second look at the topic. We have proposed an MMSE-DFE that significantly outperforms the previous works, that does not require strict assumptions about the channel, that typically involves less computational complexity, and that lends itself to adaptive calculation of the equalizer coefficients via LMS. While this work has treated PPM, we note that the extension to the general class of orthogonal modulations is quite straightforward [11]. Future work could consider the development of stochastic gradient–descent algorithms for blind equalization of PPM, as such algorithms are often preferred in practice since they do not suffer from the overmodeling problems of blind subspace estimators [15]. As far as we know, the existence of such an algorithm for PPM that exhibits global convergence is an outstanding open problem.

Appendix I
Proof of Lemma 1

Proof: The decision device amounts to simply choosing the index of the largest element of the input vector. Consequently, the decision device is unaffected by the addition of a constant to all elements of the input vector. Thus, for any scalar b

$$\mathcal{Q}(\tilde{\boldsymbol{x}}[n]) = \mathcal{Q}(\tilde{\boldsymbol{x}}[n] + \mathbf{1}_{M\times 1} b)\,.$$

Note that this is true when b is any scalar—even one that is a function of $\tilde{\boldsymbol{x}}[n]$. Consider the particular choice $b = -1/M \mathbf{1}_{M\times 1}^{\top} \tilde{\boldsymbol{x}}[n]$, so that b equals the negated average of the elements of $\tilde{\boldsymbol{x}}[n]$. For this choice

$$\begin{aligned}\mathcal{Q}(\tilde{\boldsymbol{x}}[n]) &= \mathcal{Q}\left(\tilde{\boldsymbol{x}}[n] - \frac{1}{M}\mathbf{1}_{M\times 1}\mathbf{1}_{M\times 1}^{\top}\tilde{\boldsymbol{x}}[n]\right)\\ &= \mathcal{Q}(\boldsymbol{J}_M \tilde{\boldsymbol{x}}[n])\,.\end{aligned}$$

Thus, premultiplication of the decision device input by $\boldsymbol{J}_M$ has no effect on the decision device output. ■

Appendix II
Proof of Lemma 2

Proof: To see that $\boldsymbol{J}_M$ is idempotent

$$\begin{aligned}\boldsymbol{J}_M^2 &= \boldsymbol{I}_M - \frac{2}{M}\mathbf{1}_{M\times M} + \frac{1}{M^2}\mathbf{1}_{M\times M}^2\\ &= \boldsymbol{J}_M.\end{aligned}$$

As for symmetry, it is easy to see that $\boldsymbol{J}_M = \boldsymbol{J}_M^{\top}$. It is well known that the eigenvalues of an idempotent matrix can only assume values zero or one [17], thus the matrix is positive semidefinite. The trace of a matrix equals the sum of its eigenvalues and the trace of $\boldsymbol{J}_M$ is $M-1$. Thus, the matrix $\boldsymbol{J}_M$ has $M-1$ eigenvalues that are 1, one eigenvalue that is 0, and hence, it has rank $M-1$. ■

Appendix III
Proof of Lemma 3

Proof: Note that this proof is valid for any square symmetric positive–semidefinite matrix, which includes our particular choice $\boldsymbol{J}_M$ in (3). The Cholesky decomposition is typically only defined for strictly positive–definite matrices [18], and here, we extend the definition to positive–semidefinite matrices. Since $\boldsymbol{J}_M$ is symmetric and positive semidefinite, the singular value decomposition can be written as

$$\boldsymbol{J}_M = \boldsymbol{V}^{\top}\boldsymbol{\Sigma}\boldsymbol{V}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix containing the nonnegative singular values, assumed to be in decreasing order. Then, $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^{\top}\boldsymbol{\Gamma}$ where $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times M}$ is the extraction of the first r rows of $\boldsymbol{\Sigma}^{(1/2)}$. Note that the remaining $M-r$ rows of $\boldsymbol{\Sigma}^{(1/2)}$ are all zeros and are thrown out. Next, let $\boldsymbol{Q}$ and $\boldsymbol{U}_M$ be the QR-decomposition of $\boldsymbol{\Gamma V}$ defined as

$$\boldsymbol{Q}\boldsymbol{U}_M = \boldsymbol{\Gamma V} \tag{29}$$

where $\boldsymbol{Q} \in \mathbb{R}^{r\times r}$ is orthogonal and $\boldsymbol{U}_M \in \mathbb{R}^{r\times M}$ is upper triangular. Summarizing this development, then

$$\begin{aligned}\boldsymbol{J}_M &= \boldsymbol{V}^\top \boldsymbol{\Sigma} \boldsymbol{V} \\ &= \boldsymbol{V}^\top \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma} \boldsymbol{V} \\ &= \boldsymbol{U}_M^\top \boldsymbol{Q}^\top \boldsymbol{Q} \boldsymbol{U} \\ &= \boldsymbol{U}_M^\top \boldsymbol{U}_M\end{aligned}$$

where $\boldsymbol{U}_M$ is upper triangular and of appropriate dimension. ■

APPENDIX IV
EXPLICIT EXPRESSION FOR $\boldsymbol{U}_M$

First, we note that, for $M = 2$

$$\boldsymbol{U}_2 = \frac{1}{\sqrt{2}}[1 \quad -1]$$

which satisfies $\boldsymbol{U}_2^\top \boldsymbol{U}_2 = \boldsymbol{J}_2$ and is trivially upper triangular. For $M > 2$, $\boldsymbol{U}_M$ has the recursive definition

$$\boldsymbol{U}_M = \begin{bmatrix} \sqrt{\frac{M-1}{M}} & -\sqrt{\frac{1}{M(M-1)}}\mathbf{1}_{1\times M-1} \\ \mathbf{0}_{M-2\times 1} & \boldsymbol{U}_{M-1} \end{bmatrix}$$

which, through induction, can be shown to be upper triangular and to satisfy $\boldsymbol{U}_M^\top \boldsymbol{U}_M = \boldsymbol{J}_M \triangleq \boldsymbol{I}_M - (1/M)\mathbf{1}_{M\times M}$.

REFERENCES

[1] J. Kahn and J. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

[2] M. Z. Win and R. A. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multipleaccess communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–689, Apr. 2000.

[3] J. Barry, "Sequence detection and equalization for pulse-position modulation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 1994, pp. 1561–1565.

[4] M. Audeh, J. Kahn, and J. Barry, "Decision-feedback equalization of pulse-position modulation on measured nondirected indoor infrared channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 500–503, Apr. 1999.

[5] M. Audeh, J. Kahn, and J. Barry, "Performance of pulse-position modulation on measured non-directed indoor infrared channels," *IEEE Trans. Commun.*, vol. 44, no. 6, pp. 654–659, Jun. 1996.

[6] L. Tong, G. Xu, and T. Kailaith, "Fast blind equalization via antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 1993, pp. 272–275.

[7] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.

[8] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[9] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[10] R. Kennedy, B. Anderson, and R. Bitmead, "Tight bounds on the error probabilities of decision feedback equalizers," *IEEE Trans. Commun.*, vol. COMM-35, no. 10, pp. 1022–1028, Oct. 1987.

[11] A. G. Klein, "Equalization for energy efficient modulation," Ph.D. dissertation, School Electr. Comp. Eng., Cornell Univ., Ithaca, NY, Jan. 2006.

[12] Z. Ding, "Characteristics of band-limited channels unidentifiable from second-order cyclostationary statistics," *IEEE Signal Process. Lett.*, vol. 3, no. 5, pp. 150–152, May 1996.

[13] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, N.J.: Prentice-Hall, 2001.

[14] M. Ghosh, "Analysis of the MMSE DFE with error propagation," in *Proc. IEEE GLOBECOM Commun. Theory Mini-Conf.*, Nov. 1997, vol. 4, pp. 85–89.

[15] A. Klein and P. Duhamel, "Subspace-based blind channel identification for orthogonal modulation," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2006.

[16] G. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[18] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

Andrew G. Klein (S'95–M'98) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1998, the M.S. degree in electrical engineering from the University of California, Berkeley, in 2000, and the Ph.D. degree in electrical and computer engineering from Cornell University, in 2005.

He worked at several Bay Area wireless startup companies. In 2006, he was a Chateaubriand Fellow conducting postdoctoral research at Laboratoire des Signaux et Systémes (LSS), SUPELEC, Gif-sur-Yvette, France. Starting in Fall 2007, he will be an Assistant Professor at Worcester Polytechnic Institute, Worcester, MA. His areas of interest are statistical signal processing and adaptive parameter estimation for wireless communication systems.

Pierre Duhamel (M'87–SM'87–F'98) was born in France in 1953. He received the Eng. degree in electrical engineering from the National Institute for Applied Sciences (INSA) Rennes, France, in 1975 and the Dr.Eng. and the Doctorat ès Sciences degrees from Orsay University, Orsay, France, in 1978 and 1986, respectively.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive charge-coupled device (CCD) filters. Later, he worked on fast algorithms for computing Fourier transforms and convolutions and applied similar techniques to adaptive filtering, spectral analysis, and wavelet transforms. From 1993 to September 2000, he was the Professor at the Ecole Nationale Supérieure des Télécommunications (ENST, National School of Engineering in Telecommunications), Paris, France, with research activities focused on signal processing for communications, where he headed the Signal and Image Processing Department from 1997 to 2000. He is now with Laboratoire des Signaux et Systémes (LSS), SUPELEC, Gif-sur-Yvette, France, where he develops studies in signal processing for communications (including equalization, iterative decoding, and multicarrier systems) and signal/image processing for multimedia applications, including source coding, joint source/channel coding, watermarking, and audio processing.

Dr. Duhamel was Chairman of the Distinguished Speakers Program (DSP) committee from 1996 to 1998 and a Member of the Signal Processing for Communications committee until 2001. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, an Associate Editor for IEEE SIGNAL PROCESSING LETTERS, and a Guest Editor for the special issue on wavelets of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was Distnguished Lecturer, IEEE, for 1999, and was Co-General Chair of the 2001 International Workshop on Multimedia Signal Processing, Cannes, France. The paper on subspace-based methods for blind equalization, which he coauthored, received the Best Paper Award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1998. He was awarded the Grand Prix France Telecom by the French Science Academy in 2000.