Deriving Optimal Successive Decoders for the Asynchronous CDMA Channel Using Information Theory

Tommy Guess
Dept. of Electrical Engineering
University of Virginia
Charlottesville, VA 22904-4743

Mahesh K. Varanasi
Dept. of Electrical and Computer Engineering
Boulder, CO 80309-0425

Abstract — **The symbol-asynchronous, Gaussian Code-Division Multiple Access (CDMA) channel is considered where the users transmit channel coded information independently of each other using quadrature-amplitude modulation. The receiver is of a successive decoding type and therefore allows the users to employ independent single-user coding and decoding. We derive successive decoding structures that are optimal in that they maximize average mutual information subject to certain causality constraints. Depending on these constraints, the resulting successive decoder may or may not preserve the sum-rate (i. e., sum of the users' rates) capacity of the channel. Moreover, it is found that each set of constraints corresponds to a structured factorization of a multivariate spectrum that is both unique and optimal over the constraints.**

I. Introduction

The asynchronous Gaussian CDMA channel is considered in which each of K simultaneously active users transmits a sequence of information symbols using passband pulse-amplitude modulation (PAM). The sequence of complex-valued information symbols sent by the k-th user is denoted as $\{(X_k)_n\}_{n=-\infty}^{\infty}$, or more succinctly as $(X_k)_{-\infty}^{\infty}$, where $(X_k)_n$ is the n-th symbol transmitted by the k-th user.[1] It is assumed that these sequences are mutually independent, wide-sense stationary (w.s.s.) processes. The receiver observes the waveform $\sum_{n=-\infty}^{\infty}(X_k)_n u_k(t-nT)$ due to k^{th} user's transmission (in the absence of noise) where the complex-valued pulse $u_k(t)$ is the k-th user's signature waveform convolved with the channel impulse response. When all K users are active, the receiver observes an asynchronous superposition of such waveforms corrupted by noise so that

$$y(t) = \sum_{n=-\infty}^{\infty}\sum_{k=1}^{K}(X_k)_n u_k(t-nT-\tau_k) + n(t). \tag{1}$$

where τ_k is the delay associated with the arrival of the k-th user's transmission and $n(t)$ is a zero-mean, complex-valued, additive white Gaussian noise (AWGN) process with a two-sided power-spectral density of N_0. All complex-valued random variables and processes are assumed to be *proper* [1].

The received power of the k-th user in units of energy per symbol is denoted as P_k, and we incorporate this into the model by first normalizing the energy of each received pulse to be unity (i. e., $\int_{-\infty}^{\infty}|u_k(t)|^2 dt = 1$), so that $\mathbb{E}[|(X_k)_n|^2] = P_k$, where $\mathbb{E}[\cdot]$ denotes expectation.

The receiver front-end is any bank of M matched filters that projects the received signal into the space spanned by all time-delayed versions of all the users' received pulses. An equivalent discrete-time model is a vector channel of the form

$$(\underline{Y})_n = \sum_{l=-\infty}^{\infty}(\mathbf{A})_l(\underline{X})_{n-l} + (\underline{N})_n, \tag{2}$$

where the under-bar notation denotes a column vector. Specifically, $(\underline{X})_n$ is a $K\times 1$ vector formed by stacking the users' transmitted symbols $(X_1)_n$ through $(X_K)_n$, and the noise sequence, $(\underline{N})_{-\infty}^{\infty}$, is an $M\times 1$ vector process that is zero-mean, w.s.s., and Gaussian; its auto-correlation sequence is $(\mathbf{R}_{\underline{N}})_{-\infty}^{\infty}$, where each $(\mathbf{R}_{\underline{N}})_n$ is an $M\times M$ matrix. The channel is given by the sequence of $M\times K$ matrices, $(\mathbf{A})_{-\infty}^{\infty}$, where it is assumed that $\sum_{l=-\infty}^{\infty}\mathrm{trace}((\mathbf{A})_l(\mathbf{A}^\dagger)_l) < \infty$ (the superscript † denotes the Hermitian transpose) so that the $M\times 1$ output sequence, $(\underline{Y})_{-\infty}^{\infty}$, is also w.s.s. .

To facilitate the presentation, we make formal use of the z-transform to represent the convolution operator. Thus for any sequence $\{h_n\}$, stochastic or deterministic, we write that

$$h(z) = \sum_{n=-\infty}^{\infty}h_n z^{-n}. \tag{3}$$

This allows the channel in (2) to be expressed as

$$\underline{Y}(z) = \mathbf{A}(z)\underline{X}(z) + \underline{N}(z). \tag{4}$$

Since $(\underline{X})_{-\infty}^{\infty}$, $(\underline{Y})_{-\infty}^{\infty}$, and $(\underline{N})_{-\infty}^{\infty}$ are w.s.s. processes, their auto- and cross-correlation sequences possess z-transforms in the sense that z is a complex variable. For example, if we let $(\mathbf{R}_{\underline{X}\,\underline{Y}})_n = \mathbb{E}[(\underline{X})_{l+n}(\underline{Y}^\dagger)_l]$, then the cross spectrum of $\underline{X}(z)$ and $\underline{Y}(z)$ is given by

$$\mathbf{S}_{\underline{X}\,\underline{Y}}(z) = \sum_{n=-\infty}^{\infty}(\mathbf{R}_{\underline{X}\,\underline{Y}})_n z^{-n}. \tag{5}$$

In fact, it is easy to see that

$$\mathbf{S}_{\underline{Y}}(z) = \mathbf{A}(z)\mathbf{S}_{\underline{X}}(z)\mathbf{A}^\dagger(1/z^*) + \mathbf{S}_{\underline{N}}(z) \tag{6}$$

$$\mathbf{S}_{\underline{Y}\,\underline{X}}(z) = \mathbf{A}(z)\mathbf{S}_{\underline{X}}(z), \tag{7}$$

since $\underline{X}(z)$ and $\underline{N}(z)$ are independent. We have used the notation $\mathbf{A}^\dagger(1/z^*)$ to represent $\sum_{n=-\infty}^{\infty}(\mathbf{A})_n^\dagger z^n$. We assume that $\mathbf{S}_{\underline{N}}(z)$ is a full-rank, regular (i. e., non-predictable) process. This is equivalent to stating that the Szegö condition, $\int_{-\pi}^{\pi}\log|\mathbf{S}_{\underline{N}}(e^{j\theta})|d\theta > -\infty$, is met, where $|\mathbf{S}_{\underline{N}}(e^{j\theta})|$ denotes the determinant of $\mathbf{S}_{\underline{N}}(e^{j\theta})$ (see for example [2]).

The capacity region of the symbol-asynchronous CDMA channel was derived in [3]. The user inputs that maximize average mutual information are Gaussian processes, but no single set of input spectra allows every point of the capacity region to be achieved. This results in a capacity region that is a K-dimensional pentagon with "rounded" vertices. For the purposes of this paper, it is assumed for each k that $\{(X_k)_n\}_{n=-\infty}^{\infty}$ is a Gaussian process, the collection of which provides an average mutual information equal to the sum-rate capacity of the channel. This average mutual information,

$$\frac{1}{N}\lim_{N\to\infty}\mathcal{I}((\underline{X})_1^N;(\underline{Y})_1^N), \tag{8}$$

[0] This work was supported in part by NSF grants NCR-99725778 and NCR-9706591.

[1] In general we shall let notation of the form $(Z)_\alpha^\beta$, where $\alpha\le\beta$ are integers, denote $\{(Z)_\alpha,(Z)_{\alpha+1},\cdots,(Z)_\beta\}$.

we represent by $\mathcal{I}(\underline{X}(z);\underline{Y}(z))$. It is given analytically by [4] [5]

$$\mathcal{I}(\underline{X}(z);\underline{Y}(z)) = \int_{-\pi}^{\pi} \log\left(\frac{|\mathbf{S}_{\underline{Y}}(e^{j\theta})|}{|\mathbf{S}_{\underline{N}}(e^{j\theta})|}\right)\frac{d\theta}{2\pi} \tag{9}$$

in units of nats per K-user channel use.

The remainder of this paper is organized as follows. In Section II, the general successive-decoding structure is described. In Section III, a successive decoder that preserves mutual information is derived. Section IV considers several cases where causality constraints are imposed to yield information-lossy successive decoders. Section V makes some concluding remarks.

II. Successive Decoding for the Asynchronous CDMA Channel

A successive-decoding structure for the symbol-asynchronous CDMA channel is shown in Figure 1 (cf. [6]). It is parameterized by

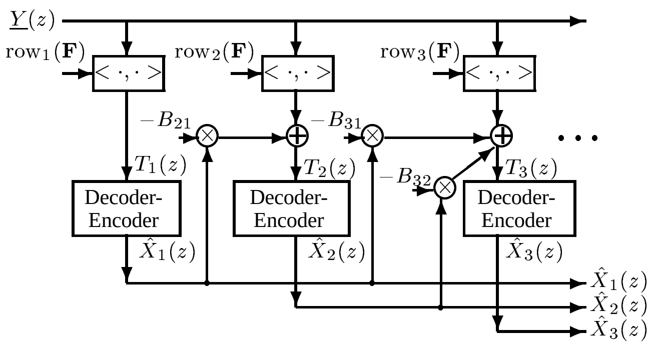

Figure 1: A successive decoder for the asynchronous CDMA channel.

a sequence of $K \times M$ feedforward matrices $(\mathbf{F})_{-\infty}^{\infty}$ and a sequence $K \times K$ lower-triangular feedback matrices $(\mathbf{B})_{-\infty}^{\infty}$. The first row of $\mathbf{F}(z)$ is multiplied by $\underline{Y}(z)$ to yield the Gaussian scalar process $T_1(z)$ so that the first user effectively sees the single-user intersymbol interference (ISI) channel with $X_1(z)$ as the input and $T_1(z)$ as the output. It is assumed that an appropriate single-user coder/decoder is implemented for this channel, and that the resulting symbol decisions are then re-encoded to yield $\hat{X}_1(z)$. This processing is taken into account before decoding the second user. Specifically, the second row of $\mathbf{F}(z)$ is multiplied by $\underline{Y}(z)$ and from this is subtracted $\hat{X}_1(z)$ after being scaled by the (2,1) element of $\mathbf{B}(z)$. This yields $T_2(z) = \mathrm{row}_2(\mathbf{F}(z))\underline{Y}(z) - B_{21}(z)\hat{X}_1(z)$. As with the first user, the second user chooses an appropriate single-user coder/decoder for this channel. The procedure is successively applied until all K users have been decoded.

Figure 2 redraws the successive decoder without explicitly identifying the successive-decoding nature of the receiver. This figure also assumes that the feedback of decoded symbols is always error free.

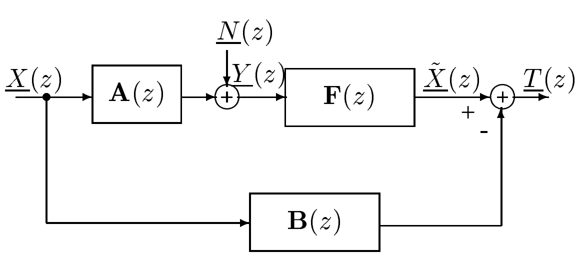

Figure 2: A matrix representation of the successive decoder.

III. A Successive Decoder that Preserves Mutual Information

We formally apply the chain rule to the average mutual information of the channel to get

$$\mathcal{I}(\underline{X}(z);\underline{Y}(z)) = \sum_{j=1}^{K}\mathcal{I}(X_k(z);\underline{Y}(z)|X_1^{k-1}(z)), \tag{10}$$

where we have used the notation $X_1^{k-1}(z)$ to denote the sequences of users 1 through $k-1$, i. e., $\{X_1(z), X_2(z), \cdots, X_{k-1}(z)\}$. Equation (10) must still be validated. Consider the two-user case, for which we have[2]

$$\mathcal{I}(X_2(z);\underline{Y}(z)|X_1(z)) = \mathcal{I}((X_2)_{-\infty}^{\infty};(\underline{Y})_n|(X_1)_{-\infty}^{\infty},(\underline{Y})_{-\infty}^{n-1}). \tag{11}$$

Now $\mathcal{I}(X_1^2(z);\underline{Y}(z))$ is known to equal $\mathcal{I}((X_1^2)_{-\infty}^{\infty};(\underline{Y})_n|(\underline{Y})_{-\infty}^{n-1})$ [5] [7], which can be decomposed via the chain rule of mutual information as

$$\mathcal{I}((X_1^2)_{-\infty}^{\infty};(\underline{Y})_n|(\underline{Y})_{-\infty}^{n-1}) = \mathcal{I}((X_1)_{-\infty}^{\infty};(\underline{Y})_n|(\underline{Y})_{-\infty}^{n-1}) + \mathcal{I}((X_2)_{-\infty}^{\infty};(\underline{Y})_n|(X_1)_{-\infty}^{\infty},(\underline{Y})_{-\infty}^{n-1}). \tag{12}$$

The first of these terms is equal to $\mathcal{I}(X_1(z);\underline{Y}(z))$, and the latter is equal to $\mathcal{I}(X_2(z);\underline{Y}(z)|X_1(z))$ by (11). Thus the notation used in (10) is indeed well-defined.

We proceed as in [8] to express each of the conditional mutual-information terms of the right-hand side of (10) as $\mathcal{I}(X_k(z);\underline{Y}(z)|X_1^{k-1}(z)) =$

$$\mathcal{I}(X_k(z);\underline{Y}(z)) + \mathcal{I}(X_k(z);X_1^{k-1}(z)|\underline{Y}(z)), \tag{13}$$

where we have used the fact that $\mathcal{I}(X_k(z);X_1^{k-1}(z)) = 0$ since the users signal independently of each other. Because the processes of interest are jointly Gaussian and have zero mean, the mutual information $\mathcal{I}(X_k(z);\underline{Y}(z))$ can be rewritten as $\mathcal{I}(X_k(z);\tilde{X}_k(z))$ where the sequence $(\tilde{X}_k)_{-\infty}^{\infty}$ is defined by $(\tilde{X}_k)_n = \mathbb{E}[(X_k)_n|(\underline{Y})_{-\infty}^{\infty}]$ [5] (see [8] for a more accessible discussion of this result). Essentially we have orthogonally projected $X_2(z)$ onto $\underline{Y}(z)$ to yield a scalar sequence that is sufficient for the vector sequence $\underline{Y}(z)$. Combining $\tilde{X}_1(z)$ through $\tilde{X}_K(z)$ into the vector $\underline{\tilde{X}}(z)$, we have

$$\underline{\tilde{X}}(z) = \mathbf{S}_{\underline{X}\,\underline{Y}}(z)\mathbf{S}_{\underline{Y}}^{-1}(z)\underline{Y}(z). \tag{14}$$

Similarly, we remove the conditioning in the second term of the right-hand side of (13) by expressing the mutual information in terms of those parts of $X_k(z)$ and $X_1^{k-1}(z)$ that are orthogonal to $\underline{Y}(z)$ (see [5] [8]). That is,

$$\mathcal{I}(X_k(z);X_1^{k-1}(z)|\underline{Y}(z)) = \mathcal{I}(E_k(z);E_1^{k-1}(z)), \tag{15}$$

where for each $j \in \{1,\ldots,K\}$, $E_j(z)$ is the sequence whose n-th term is given by $(E_j)_n = (X_j)_n - (\tilde{X}_j)_n$. We now incorporate the hat notation to represent the projection of $(E_k)_n$ onto $(E_1^{k-1})_{-\infty}^{\infty}$. That is, $(\hat{E}_k)_n = \mathbb{E}[(E_k)_n|(E_1^{k-1})_{-\infty}^{\infty}]$. Then we can write $\mathcal{I}(E_k(z);E_1^{k-1}(z)) = \mathcal{I}(E_k(z);\hat{E}_k(z))$ so that

$$\begin{aligned}
&\mathcal{I}(\underline{X}(z);\underline{Y}(z))\\
&= \sum\nolimits_{k=1}^{K}\left(\mathcal{I}(X_k(z);\tilde{X}_k(z)) + \mathcal{I}(E_k(z);\hat{E}_k(z))\right)\\
&= \sum\nolimits_{k=1}^{K}\mathcal{I}(X_k(z);\tilde{X}_k(z)+\hat{E}_k(z))\\
&= \sum\nolimits_{k=1}^{K}\mathcal{I}(X_k(z);X_k(z)-I_k(z)),
\end{aligned} \tag{16}$$

[2] Although there are several interpretations of the quantity $\mathcal{I}(X_k(z);\underline{Y}(z)|X_1^{k-1}(z))$ [5, Sec. 6.1], they are all equivalent by an extension of Theorem 10.2.1 from the same reference since we are working with a full-rank regular process.

where $I_k(z) = E_k(z) - \hat{E}_k(z)$. Since $I_j(z)$ and $I_l(z)$ are independent sequence whenever $j \neq l$, we can think of $\{I_1(z), I_2(z), \ldots, I_K(z)\}$ as a K-length "innovations" sequence. This has effectively decomposed the multiple-access channel into K single-user channels that can be encoded and decoded independently of each other.

At this juncture we consider in detail how the preceding discussion relates to spectral factorization. Let $\underline{E}(z)$ be the $K \times 1$ vector formed by stacking $E_1(z)$ through $E_K(z)$ and consider its multivariate spectrum $\mathbf{S}_{\underline{E}}(z) = \mathbf{S}_{\underline{X}}(z) - \mathbf{S}_{\underline{X}\,\underline{Y}}(z)\mathbf{S}_{\underline{Y}}^{-1}(z)\mathbf{S}_{\underline{Y}\,\underline{X}}(z)$. From Woodbury's identity[3] and equations (6) and (7), this is equal to

$$\mathbf{S}_{\underline{E}}(z) = \left(\mathbf{S}_{\underline{X}}^{-1}(z) + \mathbf{A}^{\dagger}(1/z^*)\mathbf{S}_{\underline{N}}^{-1}(z)\mathbf{A}(z)\right)^{-1}. \qquad (17)$$

Factor this spectrum uniquely as follows,

$$\mathbf{S}_{\underline{E}}(z) = \mathbf{L}(z)\mathbf{D}(z)\mathbf{L}^{\dagger}(1/z^*), \qquad (18)$$

where $\mathbf{D}(z)$ is diagonal and $\mathbf{L}(z)$ is a lower-triangular matrix whose diagonal elements are each unity. With this Cholesky decomposition (in the z-domain), we have that $\underline{I}(z) = \mathbf{L}^{-1}(z)\underline{E}(z)$ and $\underline{\hat{E}}(z) = (\mathbf{I} - \mathbf{L}^{-1}(z))\underline{E}(z)$, where $\underline{I}(z)$ is the column vector formed by stacking $I_1(z)$ through $I_K(z)$ (similarly for $\underline{\hat{E}}(z)$) and $\mathbf{I}$ is the identity matrix.

Moreover, (18) represents a structured spectral decomposition that is optimal in the following sense.

Lemma 1. *Given any lower-triangular $\mathbf{C}(z)$ with unity-valued diagonal elements (i. e., $C_{kk}(z) = 1$), the k-th diagonal element of $\mathbf{S}'(z) \triangleq \mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^{\dagger}(1/z^*)$ is more positive definite than the k-th diagonal element of $\mathbf{D}(z)$ (i. e., $S'_{kk}(z) - D_{kk}(z)$ is non-negative definite). Equality occurs for all z if and only if $\mathbf{C}(z) = \mathbf{L}^{-1}(z)$.* ⋄

As we shall see, $\mathcal{I}(\underline{X}(z); \underline{Y}(z))$ depends inversely on the geometric means of the diagonal elements of $\mathbf{D}(z)$, so this lemma can be proved rather simply from the fact that the receiver preserves mutual information. But this lemma is of algebraic interest in and of itself, so we consider the following proof as well.

Proof. Let $\mathbf{C}(z)$ be any matrix satisfying the hypotheses of the lemma. With $\mathbf{S}'(z) \triangleq \mathbf{C}(z)\mathbf{S}(z)\mathbf{C}^{\dagger}(1/z^*)$, and the factorization of $\mathbf{S}(z)$ as given in (18), we write

$$\mathbf{S}'(z) = \mathbf{C}(z)\mathbf{L}(z)\mathbf{D}(z)\mathbf{L}^{\dagger}(1/z^*)\mathbf{C}^{\dagger}(1/z^*). \qquad (19)$$

Note that $\mathbf{C}(z)$ can always be expressed as $\mathbf{C}(z) = \mathbf{A}(z)\mathbf{L}^{-1}(z)$ for some $\mathbf{A}(z)$ that satisfies the same hypotheses as $\mathbf{C}(z)$. With this $\mathbf{A}(z)$, we find that

$$\begin{aligned} S'_{kk}(z) &= \left(\mathbf{A}(z)\mathbf{D}(z)\mathbf{A}^{\dagger}(1/z^*)\right)_{kk} && (20)\\ &= D_{kk}(z) + \sum_{i=1}^{k-1} A_{ki}(z)D_{ii}(z)A_{ki}^*(1/z^*) && (21)\\ &\succeq D_{kk}(z), && (22)\end{aligned}$$

where equality in the last step requires $A_{ki}(z) = 0$ for all $i \neq k$. □

Analogously, this lemma is similar to the optimality property of multivariate spectral factorization [2]. In that context, $\mathbf{C}(z)$ is required to be causal with $(\mathbf{C})_0 = \mathbf{I}$, in which case the arithmetic mean[4] of $\mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^{\dagger}(1/z^*)$ is more positive definite than $\mathbf{G}_{\underline{E}}$ (i. e., $(\mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^{\dagger}(1/z^*))_0 - \mathbf{G}_{\underline{E}}$ is non-negative definite), where $\mathbf{S}_{\underline{E}}(z) = \mathbf{\Phi}_{\underline{E}}(z)\mathbf{G}_{\underline{E}}\mathbf{\Phi}_{\underline{E}}^{\dagger}(1/z^*)$ is the unique multivariate spectral factorization such that $\mathbf{\Phi}_{\underline{E}}(z)$ and $\mathbf{\Phi}_{\underline{E}}^{-1}(z)$ are both causal/stable, and $(\mathbf{\Phi}_{\underline{E}})_0 = \mathbf{I}$. Equality occurs if and only if $\mathbf{C}(z) = \mathbf{\Phi}_{\underline{E}}^{-1}(z)$.

The information-lossless decomposition of equation (16) can now be viewed in terms of Figures 1 and 2. By appropriately choosing $\mathbf{F}(z)$ and $\mathbf{B}(z)$, the k-th user is provided with the effective single-input single-output (SISO) channel with $X_k(z)$ as the input and $T_k(z) = X_k(z) - I_k(z)$ as the output. These feedforward and feedback filters are given by

$$\begin{aligned} \mathbf{F}(z) &= \mathbf{L}^{-1}(z)\mathbf{S}_{\underline{X}\,\underline{Y}}(z)\mathbf{S}_{\underline{Y}}^{-1}(z) && (23)\\ \mathbf{B}(z) &= \mathbf{L}^{-1}(z) - \mathbf{I}. && (24)\end{aligned}$$

It is of interest to note that just as we have decomposed the K-user ISI channel into K single-user ISI channels, we can decompose, without loss in mutual information, each of these single-user ISI channels into single-user memoryless channels by means of a perfect-feedback minimum mean-squared error (MMSE) decision-feedback equalizer (DFE). We follow the approach found in [8] (see also [10]) to find the appropriate feedforward and feedback equalizers by first calculating $\mathbf{S}_{X_k\,T_k}(z)\mathbf{S}_{T_k}^{-1}(z)$. This filter is used to project X_k onto $(T_k)_{-\infty}^{\infty}$. Some algebra leads to

$$\begin{aligned} \mathbf{S}_{\underline{T}}(z) &= \mathbf{S}_{\underline{X}}(z) - \mathbf{D}(z)\mathbf{L}^{\dagger}(1/z^*) - \mathbf{L}(z)\mathbf{D}(z) + \mathbf{D}(z)\\ \mathbf{S}_{\underline{X}\,\underline{T}}(z) &= \mathbf{S}_{\underline{X}}(z) - \mathbf{D}(z)\mathbf{L}^{\dagger}(1/z^*), && (25)\end{aligned}$$

so that $\mathbf{S}_{X_k\,T_k}(z)\mathbf{S}_{T_k}^{-1}(z) = 1$. Thus the error $X_k(z) - \mathbf{S}_{X_k\,T_k}(z)\mathbf{S}_{T_k}^{-1}(z)T_k(z)$ is found to equal $I_k(z) = X_k(z) - T_k(z)$. Recall from (18) and its subsequent discussion, that the spectrum of $I_k(z)$ is equal to the k-th diagonal element of $\mathbf{D}(z)$, i. e., $D_{kk}(z)$. The minimum-phase spectral factorization of this spectrum provides the feedforward and feedback filters for the k-th user's perfect-feedback MMSE DFE. This is given by

$$D_{kk}(z) = \phi_k(z)g_k\phi_k^*(1/z^*), \qquad (26)$$

where $\phi_k(z)$ and its inverse are both causal/stable and monic (i. e., $(\phi_k)_0 = 1$) (see, for example, [9]). Then, the feedforward filter is $\phi_k^{-1}(z)$ and the feedback filter is equal to $\phi_k^{-1}(z) - 1$. Thus, we conclude that

$$\mathcal{I}(\underline{X}(z); \underline{Y}(z)) = \sum_{k=1}^{K} \log\left(\frac{P_k}{g_k}\right). \qquad (27)$$

We consider now the corresponding factorization. Let $\mathbf{H}(z)$ be the diagonal matrix whose k-th diagonal element is $\phi_k(z)$, and let $\mathbf{G}$ be the constant diagonal matrix whose k-th element is g_k. Then we have the following structured factorization,

$$\mathbf{S}_{\underline{E}}(z) = \tilde{\mathbf{L}}(z)\mathbf{G}\tilde{\mathbf{L}}^{\dagger}(1/z^*), \qquad (28)$$

where $\tilde{\mathbf{L}}(z) = \mathbf{L}(z)\mathbf{H}(z)$.

Lemma 2. *Given any lower-triangular $\mathbf{C}(z)$ with diagonal elements that are causal and monic, the arithmetic mean of the k-th diagonal element of $\mathbf{S}'(z) \triangleq \mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^{\dagger}(1/z^*)$ is no less than the k-th diagonal element of $\mathbf{G}$ which is g_k. Equality occurs if and only if $\mathbf{C}(z) = \tilde{\mathbf{L}}^{-1}(z)$.* ⋄

Again the information-theoretic proof has essentially been done. An algebraic proof follows.

[3] $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$.

[4] The arithmetic mean of some $\mathbf{S}(z)$ is the zeroth term of its Fourier expansion, which satisfies $(\mathbf{S}(z))_0 = \int_{-\pi}^{\pi} \mathbf{S}(e^{j\theta})\frac{d\theta}{2\pi}$.

Proof. Let $\mathbf{C}(z)$ be any matrix satisfying the hypotheses of the lemma. With $\mathbf{S}'(z) \stackrel{\triangle}{=} \mathbf{C}(z)\mathbf{S}(z)\mathbf{C}^\dagger(1/z^*)$, and the factorization of $\mathbf{S}(z)$ as given in (28), we write

$$\mathbf{S}'(z) = \mathbf{C}(z)\tilde{\mathbf{L}}(z)\mathbf{G}\tilde{\mathbf{L}}^\dagger(1/z^*)\mathbf{C}^\dagger(1/z^*). \quad (29)$$

Note that $\mathbf{C}(z)$ can always be expressed as $\mathbf{C}(z) = \mathbf{A}(z)\tilde{\mathbf{L}}^{-1}(z)$ for some $\mathbf{A}(z)$ that satisfies the same hypotheses as $\mathbf{C}(z)$. With this $\mathbf{A}(z)$, we find that

$$\begin{aligned} S'_{kk}(z) &= \Big(\mathbf{A}(z)\mathbf{G}\mathbf{A}^\dagger(1/z^*)\Big)_{kk} \\ &= A_{kk}(z)g_k A^*_{kk}(1/z^*) + \sum_{i=1}^{k-1} A_{ki}(z)g_i A^*_{ki}(1/z^*) \\ &\geq A_{kk}(z)g_k A^*_{kk}(1/z^*), \end{aligned} \quad (30)$$

where equality in the last step requires $A_{ki}(z) = 0$ for all $i \neq k$. By Parseval's relationship and the causality of $A_{kk}(z)$, the arithmetic mean of $A_{kk}(z)g_k A^*_{kk}(1/z^*)$ is equal to $g_k \sum_{n=0}^{\infty} |(A_{kk})_n|^2$, which is greater than or equal to unity since $A_{kk}(z)$ is also monic. In fact, equality occurs when and only when $A_{kk}(z) = 1$. □

In summary, then, the high-level structure of this optimum (i. e., information-lossless) successive decoder converts the multi-input multi-output (MIMO) ISI channel into a set of K independent SISO ISI channels, and the low-level structure converts each of these single-user ISI channels into memoryless SISO channels.

IV. Lossy Successive Decoders for the Asynchronous CDMA Channel

The previous section placed no constraints on causality. For example, $(\underline{\tilde{X}})_n$ is given by the projection of $(\underline{X})_n$ onto the entire sequence $(\underline{Y})_{-\infty}^{\infty}$. This is, of course, unrealistic from a practical viewpoint. However, we can also require that the successive decoder be subject to such causality constraints. In general, a capacity penalty is incurred, but there exists an optimum successive decoder over the constrained class of receivers. Just as the unconstrained optimum successive decoder of the previous section corresponds to a unique factorization of a multivariate spectrum (e. g., (18) and Lemma 1), each constrained optimum successive decoder also corresponds to a unique structured factorization of a multivariate spectrum.

An Infinite-Dimensional Case:

Consider first the case where the feedback matrix $\mathbf{B}(z)$ is constrained to be lower triangular and causal, and to have diagonal elements that are each equal to unity. This differs from the previous example in Section III in that there is the added condition of causality. This means that the receiver feeds back only the already-decoded symbols of past users. Since $(X_k)_n$ is independent of $(X_1^{k-1})_{-\infty}^{\infty}$, we have the succeeding representation of a necessary loss in mutual information,[5]

$$\begin{aligned} &\mathcal{I}(\underline{X}(z);\underline{Y}(z)) \\ &= \sum_{k=1}^{K} \mathcal{I}((X_k)_n;(\underline{Y})_{-\infty}^{\infty}|(X_1^{k-1})_{-\infty}^{\infty},(X_k)_{-\infty}^{n-1}) \\ &\geq \sum_{k=1}^{K} \mathcal{I}((X_k)_n;(\underline{Y})_{-\infty}^{\infty}|(X_1^{k-1})_{-\infty}^{n},(X_k)_{-\infty}^{n-1}). \end{aligned} \quad (31)$$

[5] To process the received vector sequence causally as well, we can replace $(\underline{Y})_{-\infty}^{\infty}$ by $(\underline{Y})_{-\infty}^{n}$. This reduces the mutual information even further. The presentation that follows remains essentially unchanged except that $(\underline{\tilde{X}})_n$ becomes $\mathbb{E}[(\underline{X})_n|(\underline{Y})_{-\infty}^{n}]$, which in turn modifies the error spectrum $\mathbf{S}_{\underline{E}}(z)$.

That is, for past users we make use of only their current and past symbols, $(X_1^{k-1})_{-\infty}^{n}$, as opposed to all symbols of past users, $(X_1^{k-1})_{-\infty}^{\infty}$. Now make the following two definitions: $(\tilde{X}_k)_n = \mathbb{E}[(X_k)_n|(\underline{Y})_{-\infty}^{\infty}]$ and $(E_k)_n = (X_k)_n - (\tilde{X}_k)_n$. Then, without further loss of mutual information, (31) can be expressed as

$$\begin{aligned} &\mathcal{I}(\underline{X}(z);\underline{Y}(z)) \\ &\geq \sum_{k=1}^{K} \mathcal{I}((X_k)_n;(\tilde{X}_k)_n) + \mathcal{I}((E_k)_n;(E_1^{k-1})_{-\infty}^{n},(E_k)_{-\infty}^{n-1}) \end{aligned} \quad (32)$$

where the the projection techniques of [8] have been employed. Furthermore, if we let $(\hat{E}_k)_n = \mathbb{E}[(E_k)_n|(E_1^{k-1})_{-\infty}^{n},(E_k)_{-\infty}^{n-1}]$ and $(I_k)_n = (E_k)_n - (\hat{E}_k)_n$, then the k-th term in this sum can be expressed as $I((X_k)_n;(X_k)_n - (I_k)_n)$.

The sequence $(\hat{E}_k)_n$ can be produced by a filter that is created with the help of a multivariate factorization. Specifically, let $\mathbf{S}_{(k)}(z)$ be the principal submatrix of the spectrum of $\mathbf{S}_{\underline{E}}(z)$ formed using the first k indices. It possesses the unique factorization [2]

$$\mathbf{S}_{(k)}(z) = \mathbf{\Phi}_{(k)}(z)\mathbf{G}_{(k)}(z)\mathbf{\Phi}^\dagger_{(k)}(1/z^*), \quad (33)$$

where $\mathbf{\Phi}_{(k)}(z)$ and its inverse are causal/stable, $(\mathbf{\Phi}_{(k)})_0$ is lower triangular with diagonal elements each equal to unity, and $\mathbf{G}_{(k)}$ is a constant diagonal matrix. The k-th row of $\mathbf{\Phi}^{-1}_{(k)}(z)$ is used to form a filter that projects $(E_k)_n$ onto $\{(E_1^{k-1})_{-\infty}^{n},(E_k)_{-\infty}^{n-1}\}$. It is given by $\hat{E}_k(z) = ([\mathbf{0}_{1\times k-1}\ 1\ \mathbf{0}_{1\times K-k}] - \mathrm{row}_k(\mathbf{\Phi}^{-1}_{(k)}(z))\underline{E}(z)$, where $\mathbf{0}_{1\times k-1}$ is a zero vector of dimensions $1 \times k-1$. Note that the (k,k)-th element of $\mathbf{G}_{(k)}$ is the power of $(I_k)_n$.

To derive a factorization of $\mathbf{S}_{\underline{E}}(z)$, let

$$\mathbf{\Psi}(z) = \begin{bmatrix} \mathrm{row}_1(\mathbf{\Phi}^{-1}_{(1)}(z)) & 0 & \cdots & 0 \\ \mathrm{row}_2(\mathbf{\Phi}^{-1}_{(2)}(z)) & & 0 \cdots & 0 \\ \vdots & & & \\ \mathrm{row}_K(\mathbf{\Phi}^{-1}_{(K)}(z)) & & & \end{bmatrix}^{-1}. \quad (34)$$

The desired structured factorization is

$$\mathbf{S}_{\underline{E}}(z) = \mathbf{\Psi}(z)\mathbf{D}(z)\mathbf{\Psi}^\dagger(1/z^*), \quad (35)$$

where $\mathbf{D}(z) = \mathbf{\Psi}^{-1}(z)\mathbf{S}_{\underline{E}}(z)(\mathbf{\Psi}^\dagger(1/z^*))^{-1}$. The k-th diagonal element of $\mathbf{D}(z)$ is obviously equal to the k-th diagonal element of $\mathbf{G}_{(k)}$. It should be noted that $\mathbf{\Psi}(z)$ is lower triangular, causal, causally invertible, and has monic diagonal entries. This factorization is essentially the so-called *partial spectral factorization* due to [6], which was derived by maximizing the effective signal-to-noise ratios of the users.

Lemma 3. *Given any causal lower-triangular* $\mathbf{C}(z)$ *with* $C_{kk}(z)$ *monic for all* k*, then the arithmetic mean of the* k*-th diagonal element of* $\mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^\dagger(1/z^*)$ *is greater than or equal to the arithmetic mean of the* k*-th diagonal element of* $\mathbf{D}(z)$*. Equality occurs if and only if* $\mathbf{C}(z) = \mathbf{\Psi}^{-1}(z)$*.* ⋄

Proof. Let $\mathbf{S}'(z) = \mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^\dagger(1/z^*)$. If we let $\mathbf{c}_k(z)$ denote the row matrix formed by retaining only the first k elements of the k-th row of $\mathbf{C}(z)$, we have that $S'_{kk}(z) = \mathbf{c}_k(z)\mathbf{S}_{(k)}(z)\mathbf{c}_k^\dagger(1/z^*)$, where $\mathbf{S}_{(k)}(z)$ is the principal submatrix of $\mathbf{S}(z)$ formed by the first k indices. If can be shown that the factorization in (33) implies there is a row vector $\mathbf{a}(z)$ such that $\mathbf{c}_k(z) = \mathbf{a}(z)\mathbf{\Phi}^{-1}_{(k)}(z)$ with $\mathbf{a}(z)$ being causal and its k-th element additionally monic. Therefore, $S'_{kk}(z) = \mathbf{a}(z)\mathbf{G}_{(k)}\mathbf{a}^\dagger(1/z^*)$, where $\mathbf{G}_{(k)}$ is the diagonal matrix from (33).

Clearly, then,

$$S'_{kk}(z) = \sum_{i=1}^{k} a_i(z)(\mathbf{G}_{(k)})_{ii} a_i^*(1/z^*) \quad (36)$$
$$\geq a_k(z)(\mathbf{G}_{(k)})_{kk} a_k^*(1/z^*), \quad (37)$$

with equality for all z if and only if every element of $\mathbf{a}(z)$ is zero save the k-th element. Since $a_k(z)$ is causal and monic, the arithmetic mean of $S'_{kk}(z)$ is no less than $(\mathbf{G}_{(k)})_{kk}$, the k-th diagonal element of $\mathbf{G}_{(k)}$ and of $\mathbf{D}(z)$. This equality requires $a_k(z) = 1$. □

The optimal feedforward and feedback equalizers are thus

$$\mathbf{F}(z) = \mathbf{\Psi}^{-1}(z)\mathbf{S}_{\underline{X}\,\underline{Y}}(z)\mathbf{S}_{\underline{Y}}^{-1}(z) \quad (38)$$
$$\mathbf{B}(z) = \mathbf{\Psi}^{-1}(z) - \mathbf{I}. \quad (39)$$

A Finite-Dimensional Case:

For our final example we consider the situation where the feedforward and feedback filters have finite impulse responses (FIR). That is, $\mathbf{F}(z)$ and $\mathbf{B}(z)$ are matrix polynomials in z. We begin with

$$\mathcal{I}(\underline{X}(z); \underline{Y}(z)) \geq \sum_{k=1}^{K} \mathcal{I}((X_k)_n; (\underline{Y})_{n-N_y}^{n} | (X_1^k)_{n-N_x}^{n-1})$$

where N_x and N_y are non-negative integers. We now make the following definitions,

$$(\tilde{X}_l)_{n-j} = \mathbb{E}[(X_l)_{n-j} | (\underline{Y})_{n-N_y}^{n}] \quad (40)$$
$$(E_l)_{n-j} = (X_l)_{n-j} - (\tilde{X}_l)_{n-j}, \quad (41)$$

for $l = 1, 2, \ldots, k$ and $j = 0, 1, \ldots, N_x$;

$$(\hat{E}_k)_n = \mathbb{E}[(E_k)_n | (E_1^k)_{n-N_x}^{n-1}] \quad (42)$$
$$(I_k)_n = (E_k)_n - (\hat{E}_k)_n. \quad (43)$$

With these definitions, we find that without further reduction in mutual information we have

$$\mathcal{I}(\underline{X}(z); \underline{Y}(z)) \geq \sum_{k=1}^{K} \mathcal{I}((X_k)_n; (X_k)_n - (I_k)_n). \quad (44)$$

To calculate these quantities it is easiest to form the column vectors

$$\mathcal{X} = \left[\begin{array}{cccc} (\underline{X})_{n-N_x}^{\mathsf{T}} & (\underline{X})_{n-N_x+1}^{\mathsf{T}} & \cdots & (\underline{X})_{n}^{\mathsf{T}} \end{array}\right]^{\mathsf{T}} \quad (45)$$
$$\mathcal{Y} = \left[\begin{array}{cccc} (\underline{Y})_{n-N_y}^{\mathsf{T}} & (\underline{Y})_{n-N_y+1}^{\mathsf{T}} & \cdots & (\underline{Y})_{n}^{\mathsf{T}} \end{array}\right]^{\mathsf{T}} \quad (46)$$
$$\mathcal{E} = \left[\begin{array}{cccc} (\underline{E})_{n-N_x}^{\mathsf{T}} & (\underline{E})_{n-N_x+1}^{\mathsf{T}} & \cdots & (\underline{E})_{n}^{\mathsf{T}} \end{array}\right]^{\mathsf{T}}. \quad (47)$$

We also need the following block-Toeplitz matrix as derived from the channel $\mathbf{A}(z)$,

$$\mathcal{A} = \left[\begin{array}{cccc} (\mathbf{A})_0 & (\mathbf{A})_{-1} & \cdots & (\mathbf{A})_{-N_x} \\ (\mathbf{A})_1 & & & \\ \vdots & & \ddots & \\ (\mathbf{A})_{N_y} & & & \end{array}\right]. \quad (48)$$

Then the covariance of $\mathcal{E}$ is given by

$$\mathbf{R}_{\mathcal{E}} = \mathcal{P} - \mathcal{P}\mathcal{A}^{\dagger}\mathbf{R}_{\mathcal{Y}}^{-1}\mathcal{A}\mathcal{P}, \quad (49)$$

where $\mathbf{R}_{\mathcal{Y}}$ is the covariance of $\mathcal{Y}$ and $\mathcal{P}$ is a block-diagonal matrix with each diagonal block equal to $\mathbf{P}$, the matrix of the users' received powers. This covariance matrix can be viewed as a block matrix (the size of each block is $K \times K$) of dimensions $N_x + 1 \times N_x + 1$.

To evaluate the projection of $(E_k)_n$ onto $(E_1^k)_{n-Nx}^{n-1}$, we first reduce $\mathbf{R}_{\mathcal{E}}$ by retaining only the principal sub-matrices of each block formed by the first k indices. Call this reduced covariance matrix $\mathbf{R}_{\mathcal{E},k}$, and perform a Cholesky factorization so that $\mathbf{R}_{\mathcal{E},k} = \mathbf{L}_{\mathcal{E},k}\mathbf{G}_{\mathcal{E},k}\mathbf{L}_{\mathcal{E},k}^{\dagger}$, where $\mathbf{L}_{\mathcal{E},k}$ is lower triangular with diagonal elements that are unity, and $\mathbf{G}_{\mathcal{E},k}$ is a diagonal matrix. We now parse the last row of $\mathbf{L}_{\mathcal{E},k}^{-1}$ into row vectors which we label $\tilde{\mathbf{b}}_{k0}$ through $\tilde{\mathbf{b}}_{kNx}$; each is of length k. In z notation we have $\tilde{\mathbf{b}}_k(z) = \sum_{j=0}^{N_x} \tilde{\mathbf{b}}_{kj} z^{-N_x+j}$. After repeating this procedure for $k = 1, 2, \ldots, K$, we form the feedback matrix

$$\tilde{\mathbf{B}}(z) = \left[\begin{array}{cccc} \tilde{\mathbf{b}}_1(z) & 0 & \cdots & 0 \\ \tilde{\mathbf{b}}_2(z) & & 0 \cdots & 0 \\ \vdots & & & \\ \tilde{\mathbf{b}}_K(z) & & & \end{array}\right]. \quad (50)$$

The feedback matrix is then given by

$$\mathbf{B}(z) = \tilde{\mathbf{B}}(z) - \mathbf{I}; \quad (51)$$

it is strictly causal, lower triangular, and polynomial. To evaluate the feedforward equalizer we first find

$$\mathbf{P}\left[\begin{array}{cccc} (\mathbf{A})_{-N_x} & (\mathbf{A})_{-N_x+1} & \cdots & (\mathbf{A})_{-N_x+N_y} \end{array}\right]\mathbf{R}_{\mathcal{Y}}^{-1} = \left[\begin{array}{cccc} (\mathbf{H})_{N_y} & (\mathbf{H})_{N_y-1} & \cdots & (\mathbf{H})_0 \end{array}\right], \quad (52)$$

where $\mathbf{P}$ is a diagonal matrix containing the users' powers and the $K \times M$ matrices $(\mathbf{H})_j$ are implicitly defined. We represent these as the $K \times M$ polynomial matrix $\mathbf{H}(z) = \sum_{j=0}^{N_y} (H)_j z^{-j}$. Then we have that

$$\mathbf{F}(z) = \tilde{\mathbf{B}}(z)\mathbf{H}(z). \quad (53)$$

The corresponding structured spectral factorization is embodied in this lemma.

Lemma 4. *Consider the structured factorization*

$$\mathbf{S}_{\underline{E}}(z) = \tilde{\mathbf{B}}^{-1}(z)\mathbf{D}(z)(\tilde{\mathbf{B}}^{\dagger}(1/z^*))^{-1},$$

where $\tilde{\mathbf{B}}(z)$ is as given in (50) and $\mathbf{D}(z) = \tilde{\mathbf{B}}(z)\mathbf{S}_{\underline{E}}(z)\tilde{\mathbf{B}}^{\dagger}(1/z^*)$. Given any causal, polynomial of order N_x, lower-triangular $\mathbf{C}(z)$ with monic diagonal elements, then the arithmetic mean of the k-th diagonal element of $\mathbf{S}'(z) = \mathbf{C}(z)\mathbf{S}_{\underline{E}}(z)\mathbf{C}^{\dagger}(1/z^*)$ is greater than or equal to the arithmetic mean of $D_{kk}(z)$. Equality occurs if and only if $\mathbf{C}(z) = \tilde{\mathbf{B}}(z)$. ⋄

V. Concluding Remarks

In this paper we considered a coded asynchronous CDMA channel and gave information theoretic derivations of several optimal successive decoders. The corresponding receivers for the uncoded CDMA channel were derived algebraically in [6] by maximizing the effective SNR of each user. In contrast, our approach applies information-preserving transformations to the average mutual information of the channel. Since for each case the analytical representation of the resulting successive decoder requires a structured, multivariate, spectral factorization, we are also able to state the optimality algebraically in terms of this decomposition.

Though not developed here, it is of interest to note that each of the factorization lemmas we presented has a finite-dimensional analog. For example, consider the Cholesky factorization $\mathbf{S} = \mathbf{LDL}^{\dagger}$ of a positive-definite matrix $\mathbf{S}$. That is, $\mathbf{L}$ is lower triangular with unity-valued diagonal elements and $\mathbf{D}$ is diagonal. This can be viewed in the following manner. For any lower-triangular $\mathbf{C}$ with $C_{kk} = 1$ for each k, we have $(\mathbf{CSC}^{\dagger})_{kk} \geq D_{kk}$ for each k, with equality if and only if $\mathbf{C} = \mathbf{L}^{-1}$. Besides being of interest algebraically, such finite-dimensional representations possibly provide a method for numerically evaluating the infinite-dimensional structured factorizations discussed in this paper. Specifically, this would be done by generalizing results in [11] where Bauer's method is used to numerically evaluate multivariate spectral factorizations.

REFERENCES

[1] F. D. Neeser and J. L. Massey, " Proper Complex Random Processes with Applications to Information Theory ," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1293–1302, Jul. 1993.

[2] N. Wiener and P. Masani, "The Prediction Theory of Multivariate Processes, Part I: The Regularity Condition," *Acta Mathematica*, vol. 98, pp. 111–150, 1957.

[3] S. Verdú, " The Capacity Region of the Symbol-Asynchronous Gaussian Multiple-Access Channel ," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 733–751, Jul. 1989.

[4] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian Channel with Memory: The Multivariate Case," *Bell Systems Technical Journal*, vol. 53, pp. 745–778, May-Jun. 1974.

[5] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, Inc., 1964.

[6] A. Duel-Hallen, "A Family of Multiuser Decision-Feedback Detectors for Asynchronous Code-Division Multiple Access Channels," *IEEE Trans. Commun.*, pp. 421–434, Feb./Mar./Apr. 1995.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[8] T. Guess and M. K. Varanasi, " An Information-Theoretic Derivation of the MMSE Decision-Feedback Equalizer," in *Proc. Thirty-Sixth Allerton Conf. Commun., Contr. and Comput.*, Allerton, IL, Sept. 1998.

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, Second Edition*. New York, NY: McGraw-Hill Book Co., 1984.

[10] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and J. Forney, G. D., " MMSE Decision-Feedback Equalizers and Coding- Part I: Equalization Results and Part II: Coding Results," *IEEE Trans. Commun.*, vol. COM-43, pp. 2582–2604, Oct. 1995.

[11] D. C. Youla and N. N. Kazanjian, "Bauer-Type Factorization of Positive Matrices and the Theory of Matrix Polynomials Orthogonal to the Unit Circle," *IEEE Trans. Circ. and Syst.*, vol. 25, pp. 57–69, Feb. 1978.