An Information-Theoretic Framework for Deriving Canonical Decision-Feedback Receivers in Gaussian Channels

Tommy Guess, *Member, IEEE,* and Mahesh K. Varanasi, *Senior Member, IEEE*

***Abstract*—A framework is presented that allows a number of known results relating feedback equalization, linear prediction, and mutual information to be easily understood. A lossless, additive decomposition of mutual information in a general class of Gaussian channels is introduced and shown to produce an information-preserving canonical decision-feedback receiver. The approach is applied to intersymbol interference (ISI) channels to derive the well-known minimum mean-square error (MMSE) decision-feedback equalizer (DFE). When applied to the synchronous code-division multiple-access (CDMA) channel, the result is the MMSE (or signal-to-interference ratio (SIR) maximizing) decision-feedback detector, which is shown to achieve the channel sum-capacity at the vertices of the capacity region. Finally, in the case of the asynchronous CDMA channel we are able to give new connections between information theory, decision-feedback receivers, and structured factorizations of multivariate spectra.**

***Index Terms*—Decision-feedback equalizer (DFE), Gaussian channel, intersymbol interference (ISI), minimum mean-squared error (MMSE), multiple access, prediction, projection, spectral factorization, Wiener filter.**

I. INTRODUCTION

THOUGH originally postulated for data transmission without any error-control coding, it was later recognized that the decision-feedback equalizer (DFE) possesses some rather remarkable information-theoretic properties. With regard to channel capacity, there is no loss in assuming that the receiver for an intersymbol interference (ISI) channel with additive Gaussian noise is the perfect-feedback minimum mean-square error (MMSE) DFE [1]–[5, Sec. 10.5.5]. Similarly, information-lossless DFEs are associated with the multivariate ISI channel [6], the ISI channel with periodic zero padding to create an equivalent memoryless multivariate channel [7] and, as shown by the authors, the Gaussian code-division multiple-access (CDMA) channel [8], [9].

Manuscript received September 9, 2002; revised July 7, 2004. This work was supported in part by the National Science Foundation under Grants NCR-9706591, CCR-0112977, CCR-0093114, and by grants from the Colorado Center for Information Storage at the University of Colorado, Boulder. The material in this paper was presented in part at the 36th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 1998 and the 2000 Conference on Information Sciences and Systems, Princeton, NJ, March 2000.

T. Guess is with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904-4743 USA (e-mail: trg6f@virginia.edu).

M. K. Varanasi is with the Department of Electrical and Computer Engineering, University of Colorado at Boulder, Boulder, CO 80309-0425 USA (e-mail: varanasi@colorado.edu).

Communicated by G. Caire, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2004.839506

In this paper, we show that these are all special cases of a fundamental decision-feedback receiver structure that applies to any linear Gaussian channel. Our approach is to begin with mutual information and decompose it losslessly in a particular manner using the chain rule and orthogonal projections. From this decomposition, a decision-feedback receiver is naturally exposed and easily seen to consist of Wiener filtering and Wiener prediction. Since this decomposition is lossless with regard to mutual information, the capacity-achieving property of the decision-feedback receiver becomes self-evident and so, in this sense, the receiver structure is canonical. The generality of the setup implies its applicability to the cases mentioned above, and we use our result to explicitly derive the decision-feedback receivers for these specific instances. We also forge new ground by considering the asynchronous CDMA channel. Not only do we derive a variety of decision-feedback receivers starting from mutual information, but we also discover new connections between mutual information and various factorizations of multivariate spectra.

We also remark here that our result in [8] and [9] that the MMSE decision-feedback receiver achieves the sum capacity of the synchronous CDMA additive white Gaussian noise (AWGN) channel has found application in the burgeoning field of space–time or multiple-antenna wireless communication. For instance, the result in [8] and [9] can be used to show that with coding the so-called vertical Bell Labs layered space–time (vertical-BLAST) architecture of [10] (also known as horizontal-BLAST [11]) with MMSE front-end filtering and successive cancellation achieves the capacity of the multiple-antenna (or multiple-input multiple-output (MIMO)) channel for ergodic fading processes, thereby implying that coding across space is not necessary to achieve optimum performance in this case. Moreover, the result of [8] and [9] can also be used to show that the diagonal-BLAST architecture of [12] with MMSE instead of zero-forcing filtering and decision feedback achieves the outage capacity of the multiple-antenna quasi-static fading channel in the limit of large frame lengths where the loss due to frame initialization and frame termination becomes negligible. For detailed discussions of information lossless (and lossy) MIMO space–time architectures, the reader is referred to [11], [13]–[16].

The remainder of this paper is organized as follows. Section II gives some background concerning geometric interpretations of mutual information. Section III derives a generally applicable canonical decision-feedback receiver for Gaussian channels via a particular additive decomposition of mutual information. Section IV applies the theory of Section III to the symbol-synchronous CDMA and multivariate ISI channels. Section V considers the symbol-asynchronous CDMA channel and uses the theory to derive relationships between various decision-feedback receivers, information theory, and multivariate spectral factorization. Finally, Section VI provides some closing comments and the Appendix shows a technique for evaluating structured multivariate spectral factorizations.

II. Geometry of Mutual Information for Gaussian Variables

We begin with a brief review of some useful concepts from the theory of Hilbert spaces. When mutual information between Gaussian variables is viewed in this context, one obtains some powerful but simply stated properties that were explored in depth by Pinsker in [17].[1]

A. The Hilbert Space of Second-Order Random Variables

Let $\mathcal{H}$ denote the set of all finite-variance, zero-mean, complex (scalar) random variables. It is well known that $\mathcal{H}$ becomes a Hilbert space under the inner product mapping

$$\langle x, y \rangle = E[xy^*] : \mathcal{H} \times \mathcal{H} \mapsto \mathbb{C}$$

where $E[\cdot]$ denotes expectation, e.g., [18]. (This corresponds to the L^2 space from measure theory, e.g., [19, Ch. 9].) This Hilbert space possesses a property known as separability, the primary importance for our purposes being its implication that $\mathcal{H}$ has a countable basis. It also provides a means for generating a linear subspace from any subset of the Hilbert space. For example, suppose that $\mathcal{X} \subset \mathcal{H}$, then $\mathcal{X}$ has a countable basis since it is contained in $\mathcal{H}$. Thus, if $\{x_i\}$ is a basis of $\mathcal{X}$, then

$$\mathcal{H}_{\mathcal{X}} \triangleq \left\{ \sum_i a_i x_i : a_i \in \mathbb{C} \, \forall i \right\}$$

is a linear subspace of $\mathcal{H}$. Clearly, if $\mathcal{X}$ has a finite number of elements, then $\mathcal{H}_{\mathcal{X}}$ is a finite-dimensional subspace.

We are interested in bounded linear operators that map elements of $\mathcal{H}$ back into $\mathcal{H}$. Let $\mathbf{\Gamma} : \mathcal{H} \mapsto \mathcal{H}$ be such an operator. Its linearity means that $\mathbf{\Gamma}(\alpha x + \beta y) = \alpha \mathbf{\Gamma} x + \beta \mathbf{\Gamma} y$ for all $x, y \in \mathcal{H}$ and all $\alpha, \beta \in \mathbb{C}$. For each $\mathbf{\Gamma}$, there exists a unique adjoint operator $\mathbf{\Gamma}^* : \mathcal{H} \mapsto \mathcal{H}$ that is also linear and satisfies $\langle \mathbf{\Gamma} x, y \rangle = \langle x, \mathbf{\Gamma}^* y \rangle$ for all $x, y \in \mathcal{H}$. The type of linear operator that we will have occasion to employ is known as an orthogonal projection. The foundation of such operators is the geometrical nature of $\mathcal{H}$ that allows us to work with the notion of orthogonality between elements of $\mathcal{H}$. Specifically, we say that random variables $x, y \in \mathcal{H}$ are orthogonal if $\langle x, y \rangle = 0$. An orthogonal projection operator $\mathbf{\Pi}$ satisfies two properties—it is a projection operator and a self-adjoint operator. A projection operator is described as follows. After operating on x with $\mathbf{\Pi}$ to yield $y = \mathbf{\Pi} x$, the result is unchanged with a second application of the operator. That is, $\mathbf{\Pi} y = y$. An operator is said to be self-adjoint if $\mathbf{\Pi}$ and its adjoint $\mathbf{\Pi}^*$ are in fact the same operator so that $\mathbf{\Pi} x = \mathbf{\Pi}^* x$ for all $x \in \mathcal{H}$. This property boosts a mere projection operator into the class of orthogonal projection operators. This terminology is used since $\mathbf{\Pi}$ satisfies $\langle \mathbf{\Pi} x, x - \mathbf{\Pi} x \rangle = 0$ for all $x \in \mathcal{H}$. In practical terms, we may think of this as the estimate of x, namely, $\mathbf{\Pi} x$ being orthogonal to $x - \mathbf{\Pi} x$, the error associated with the estimate.

For any $\mathcal{X} \subset \mathcal{H}$, there exists an operator that orthogonally projects onto the subspace $\mathcal{H}_{\mathcal{X}}$ that it generates. We shall denote this operator by $\mathbf{\Pi}_{\mathcal{X}} : \mathcal{H} \mapsto \mathcal{H}_{\mathcal{X}}$. If y is orthogonally projected onto $\mathcal{H}_{\mathcal{X}}$ to yield $z = \mathbf{\Pi}_{\mathcal{X}} y$, then z may be expressed as a linear combination of elements of $\mathcal{H}_{\mathcal{X}}$. Observe that if two sets $\mathcal{X}, \mathcal{Y} \subset \mathcal{H}$ are orthogonal to each other, in the sense that $\langle x, y \rangle = 0$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, then the operators $\mathbf{\Pi}_{\mathcal{X}}$ and $\mathbf{\Pi}_{\mathcal{Y}}$ project onto orthogonal subspaces. In other words, $\mathbf{\Pi}_{\mathcal{X}} \mathbf{\Pi}_{\mathcal{Y}} z = 0$ for any $z \in \mathcal{H}$ so that the concatenated operators $\mathbf{\Pi}_{\mathcal{X}} \mathbf{\Pi}_{\mathcal{Y}}$ and $\mathbf{\Pi}_{\mathcal{Y}} \mathbf{\Pi}_{\mathcal{X}}$ are in fact the zero operator since they take every element of $\mathcal{H}$ to zero. Given a subspace $\mathcal{H}_{\mathcal{X}}$, there is an orthogonal subspace $\mathcal{H}_{\mathcal{X}}^{\perp}$ such that $\mathcal{H} = \mathcal{H}_{\mathcal{X}} \oplus \mathcal{H}_{\mathcal{X}}^{\perp}$. That is to say, if we let $\boldsymbol{I}$ be the identity operator (i. e., $\boldsymbol{I} z = z$ for all $z \in \mathcal{H}$), then every $z \in \mathcal{H}$ has the unique additive decomposition $z = \mathbf{\Pi}_{\mathcal{X}} z + (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{X}}) z$ where $\hat{z} = \mathbf{\Pi}_{\mathcal{X}} z$ is an element of $\mathcal{H}_{\mathcal{X}}$ and $e = (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{X}}) z$ is an element of $\mathcal{H}_{\mathcal{X}}^{\perp}$. The estimate $\hat{z}$ is orthogonal to the error e. Moreover, since $\mathcal{H}$ is populated with second-order, zero-mean random variables, $\hat{z}$ is the linear MMSE estimate of z conditional on $\mathcal{X}$ [18].

It is convenient to have a notation to cover cases in which we need to operate on a collection of random variables. Suppose that $\mathcal{X} = \{x_\alpha\}$ is a subset of $\mathcal{H}$ whose elements are indexed by α from some set. We define $\mathcal{Y} = \mathbf{\Pi} \mathcal{X}$ as the set $\{y_\alpha\}$, where $y_\alpha = \mathbf{\Pi} x_\alpha$; this is simply the collection of the elements of $\mathcal{X}$ after each has been operated on by $\mathbf{\Pi}$, with the indexing on $\mathcal{Y}$ induced by the indexing on $\mathcal{X}$.

B. Mutual Information Between Sets of Random Variables

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are sets of random variables in $\mathcal{H}$ with respective denumerable bases $\{x_i\}$ and $\{y_i\}$. Here, and in the sequel, we shall use notation of the type x_i^j to represent a consecutive string of elements $(x_i, x_{i+1}, \ldots, x_{j-1}, x_j)$. The expression $I(\mathcal{X}; \mathcal{Y})$ will be used to denote either the mutual information or the information rate between $\mathcal{X}$ and $\mathcal{Y}$ depending on whether the basis of the "input" $\mathcal{X}$ is finite or infinite dimensional. If the basis of $\mathcal{X}$ is finite dimensional, say $\{x_1, \ldots, x_k\}$, then we define

$$I(\mathcal{X}; \mathcal{Y}) \triangleq \begin{cases} I(x_1^k; y_1^l), & \text{if the dimension of } \{y_i\} \text{ is } l \\ \lim_{n \to \infty} I(x_1^k; y_1^n), & \text{if the dimension of } \{y_i\} \text{ is infinity} \end{cases} \tag{1}$$

where the right-hand $I(\cdot;\cdot)$ terms are simply the mutual information between two finite sets of random variables, and the second definition holds whenever the limit exists. If instead the

[1]Cioffi and Forney also work with the geometry of mutual information in [7].

basis of $\mathcal{X}$ is infinite dimensional, then we define the information rates[2]

$$I(\mathcal{X};\mathcal{Y}) \triangleq \begin{cases} \lim_{n\to\infty} \frac{1}{n} I(x_1^n; y_1^l), & \text{if the dimension of } \{y_i\} \text{ is } l \\ \lim_{n\to\infty} \frac{1}{n} I(x_1^n; y_1^n), & \text{if the dimension of } \{y_i\} \text{ is infinity} \end{cases} \tag{2}$$

when the limits exist. Conditional information between sets is defined similarly. For example, when the basis of $\mathcal{X}$ is finite dimensional, say $\{x_1,\ldots,x_k\}$, we get (3) at the bottom of the page. And when the basis of $\mathcal{X}$ is infinite dimensional, we get (4) at the bottom of the page.

C. Properties of Mutual Information for Gaussian Variables

We now focus our attention toward jointly Gaussian random variables in $\mathcal{H}$. A set of random variables is jointly Gaussian if every finite subset possesses a multivariate Gaussian distribution. The following lemma may be found in [17, Sec.9.3]. We include a proof to introduce the use of Hilbert space concepts.

Lemma 1: Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be sets of random variables for which $\{\mathcal{X},\mathcal{Y},\mathcal{Z}\}$ is a set of jointly Gaussian random variables. Then the mutual information between $\mathcal{X}$ and $\mathcal{Y}$ conditional on $\mathcal{Z}$ can be expressed as

$$I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) = I((\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{Z}})\mathcal{X};(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{Z}})\mathcal{Y}). \tag{5}$$

Proof: Take $\{x_i\}$, $\{y_i\}$, and $\{z_i\}$ to be countable bases of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. We shall assume these bases are all infinite dimensional, but the proof easily lends itself to the finite-dimensional cases as well. From our definition in (4) we have

$$I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) = \lim_{n\to\infty} \frac{1}{n} I(x_1^n; y_1^n | z_1^n). \tag{6}$$

Working with the mutual information term under the limit, we can say that

$$I(x_1^n;y_1^n|z_1^n) = I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n|z_1^n) \tag{7}$$

since $\boldsymbol{\Pi}_{z_1^n}x_i$ and $\boldsymbol{\Pi}_{z_1^n}y_i$ are linear combinations of the elements of z_1^n, the random variables on which we are conditioning. With the chain rule of mutual information, this may be expressed as

$$\begin{aligned} I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n|z_1^n) &= I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n,z_1^n) \\ &\quad - I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;z_1^n). \end{aligned} \tag{8}$$

The last term on the right-hand side is obviously equal to zero since $\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n \subset \mathcal{H}_{z_1^n}^{\perp}$ and $z_1^n \subset \mathcal{H}_{z_1^n}$; for jointly Gaussian random variables this orthogonality implies independence, which further implies a mutual information of zero. The first term on the right-hand side of (8), again using the chain rule of mutual information, becomes

$$\begin{aligned} I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n,z_1^n) &= I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n) \\ &\quad + I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;z_1^n|\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n). \end{aligned} \tag{9}$$

The rightmost mutual information term in this relationship is also zero. This follows since both $\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n$ and $\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n$ are orthogonal to z_1^n. So, finally, we conclude that

$$\begin{aligned} I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) &= \lim_{n\to\infty}\frac{1}{n} I(\{x_i-\boldsymbol{\Pi}_{z_1^n}x_i\}_{i=1}^n;\{y_i-\boldsymbol{\Pi}_{z_1^n}y_i\}_{i=1}^n) \\ &= I((\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{Z}})\mathcal{X};(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{Z}})\mathcal{Y}). \end{aligned} \tag{10}$$

□

The next lemma follows from Lemma 1, and is used to reduce the number of terms involved in evaluating a mutual information quantity.

Lemma 2: Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of random variables for which $\{\mathcal{X},\mathcal{Y}\}$ is a set of jointly Gaussian random variables. Then the mutual information between $\mathcal{X}$ and $\mathcal{Y}$ may be expressed as

$$I(\mathcal{X};\mathcal{Y}) = I(\mathcal{X};\boldsymbol{\Pi}_{\mathcal{Y}}\mathcal{X}). \tag{11}$$

Proof: Let $\{x_i\}$ and $\{y_i\}$ denote bases of $\mathcal{X}$ and $\mathcal{Y}$, respectively. We have

$$\begin{aligned} I(\mathcal{X};\mathcal{Y}) &= \lim_{n\to\infty}\frac{1}{n} I(x_1^n;y_1^n) \\ &= \lim_{n\to\infty}\frac{1}{n} I(x_1^n;y_1^n,\boldsymbol{\Pi}_{y_1^n}x_1^n) \end{aligned} \tag{12}$$

[2]We use information rate here because, at least for the cases considered later in this paper, the limiting mutual-information quantities would be infinity without the $\frac{1}{n}$ factor. However, for cases where $\lim_{n\to\infty} I(x_1^n;y_1^n)$ is finite, one would of course want to define this limit as $I(\mathcal{X};\mathcal{Y})$.

$$I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) \triangleq \begin{cases} I(x_1^k;y_1^l|z_1^m), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are } l \text{ and } m \\ \lim_{n\to\infty} I(x_1^k;y_1^n|z_1^m), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are infinity and } m \\ \lim_{n\to\infty} I(x_1^k;y_1^l|z_1^n), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are } l \text{ and infinity} \\ \lim_{n\to\infty} I(x_1^k;y_1^n|z_1^n), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are both infinity.} \end{cases} \tag{3}$$

$$I(\mathcal{X};\mathcal{Y}|\mathcal{Z}) \triangleq \begin{cases} \lim_{n\to\infty} \frac{1}{n} I(x_1^n;y_1^l|z_1^m), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are } l \text{ and } m \\ \lim_{n\to\infty} \frac{1}{n} I(x_1^n;y_1^n|z_1^m), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are infinity and } m \\ \lim_{n\to\infty} \frac{1}{n} I(x_1^n;y_1^l|z_1^n), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are } l \text{ and infinity} \\ \lim_{n\to\infty} \frac{1}{n} I(x_1^n;y_1^n|z_1^n), & \text{if the dimensions of } \{y_i\} \text{ and } \{z_i\} \text{ are both infinity.} \end{cases} \tag{4}$$

where the second equality follows as a byproduct of $\mathbf{\Pi}_{y_1^n} x_1^n$ being a function of the elements of y_1^n. With the chain rule of mutual information, the term under the limit becomes

$$\frac{1}{n} I(x_1^n; y_1^n, \mathbf{\Pi}_{y_1^n} x_1^n) = \frac{1}{n} \left(I(x_1^n; \mathbf{\Pi}_{y_1^n} x_1^n) + I(x_1^n; y_1^n | \mathbf{\Pi}_{y_1^n} x_1^n) \right) \tag{13}$$

leaving us to conclude in the limit as n goes to infinity that

$$I(\mathcal{X}; \mathcal{Y}) = I(\mathcal{X}; \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X}) + I(\mathcal{X}; \mathcal{Y} | \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X}). \tag{14}$$

Let us now define $\mathcal{Z} = \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X}$ and apply Lemma 1 to determine that

$$I(\mathcal{X}; \mathcal{Y} | \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X}) = I((\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Z}})\mathcal{X}; (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Z}})\mathcal{Y}). \tag{15}$$

But since $\mathcal{Z}$ represents the orthogonal projection of $\mathcal{X}$ onto $\mathcal{Y}$, it is clear that $(\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Z}})\mathcal{X} = (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}$, which lies in $\mathcal{H}_{\mathcal{Y}}^{\perp}$, the subspace that is orthogonal to $\mathcal{H}_{\mathcal{Y}}$. Meanwhile, the argument $(\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Z}})\mathcal{Y}$ is an element of $\mathcal{H}_{\mathcal{Y}}$. So the expression in (15) is zero as it represents the mutual information between independent quantities. Therefore, we must conclude that

$$I(\mathcal{X}; \mathcal{Y}) = I(\mathcal{X}; \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X}). \qquad \square$$

Recall that an orthogonal projection acts as a linear MMSE estimator. Thus, when $\{\mathcal{X}, \mathcal{Y}\}$ is a jointly Gaussian set, we have the following conditional-mean representations since in this situation the MMSE estimator turns out to be a linear MMSE estimator:

$$\mathbf{\Pi}_{\mathcal{Y}} \mathcal{X} = E[\mathcal{X} | \mathcal{Y}] \tag{16}$$

$$(\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\mathcal{X} = \mathcal{X} - E[\mathcal{X} | \mathcal{Y}] \tag{17}$$

where $E[\mathcal{X}|\mathcal{Y}] = \{E[x_\alpha | \mathcal{Y}]\}_\alpha$ and

$$\mathcal{X} - E[\mathcal{X}|\mathcal{Y}] = \{x_\alpha - E[x_\alpha|\mathcal{Y}]\}_\alpha$$

when the elements of $\mathcal{X}$ are indexed by α from some set.

To close this subsection, we point out that all lemmas and corollaries are equally applicable when the elements of the sets are indexed in terms of vectors instead of scalars. As an example, suppose that $\mathcal{X} = \{x_1, \dots, x_{2n}\}$ is arranged as $\mathcal{X} = \{\underline{x}_1, \underline{x}_2\}$ where $\underline{x}_1 = [x_1 \cdots x_n]^\mathsf{T}$ and $\underline{x}_2 = [x_{n+1} \cdots x_{2n}]^\mathsf{T}$ are column vectors. Then $\mathcal{Z} = \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X}$ may be given by $\mathcal{Z} = \{\underline{z}_1, \underline{z}_2\}$ where $\underline{z}_i = \mathbf{\Pi}_{\mathcal{Y}} \underline{x}_i$, and Lemma 2 says that $I(\mathcal{X}; \mathcal{Y}) = I(\underline{x}_1, \underline{x}_2; \underline{z}_1, \underline{z}_2)$.

D. Evaluation of the Mutual Information Between Gaussian Variables

For a set $\mathcal{X}$ of random variables with basis $\{x_i\}$, we define

$$\langle \mathcal{X} \rangle \triangleq \begin{cases} \left|\mathrm{Cov}[x_1^k]\right| & \text{if the dimension of } \{x_i\} \text{ is } k \\ \lim\limits_{n\to\infty} \left|\mathrm{Cov}[x_1^n]\right|^{1/n} & \text{if the dimension of } \{x_i\} \text{ is infinity} \end{cases} \tag{18}$$

where $\mathrm{Cov}[x_1^n]$ is the covariance matrix of the random column vector $\underline{x} = [x_1 \cdots x_n]^\mathsf{T}$ (i.e., $\mathrm{Cov}[x_1^n] = E[\underline{x}\, \underline{x}^\dagger]$ when $\underline{x}$ is zero mean and $\underline{x}^\dagger$ is the Hermitian transpose of $\underline{x}$) and $|\cdot|$ denotes the determinant operation. Now if $\mathcal{X}$ and $\mathcal{Y}$ are sets for which $\{\mathcal{X}, \mathcal{Y}\}$ is a set of jointly Gaussian random variables, then we have that the mutual information between them is given by

$$I(\mathcal{X}; \mathcal{Y}) = \log \frac{\langle \mathcal{X} \rangle}{\langle (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\mathcal{X} \rangle}. \tag{19}$$

To see why this is the case, we note first that an application of Lemma 2 allows us to rewrite $I(\mathcal{X}; \mathcal{Y})$ as $I(\mathcal{X}; \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X})$. By properties of orthogonal projections, we know that the additive decomposition $\mathcal{X} = \mathbf{\Pi}_{\mathcal{Y}} \mathcal{X} + (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}$ is such that the two addends are orthogonal (and hence independent) sets of random variables since they come from orthogonal subspaces. We view the first addend as "signal" and the second addend as "noise." If $\{x_i\} = x_1$ is a one-dimensional basis, then (19) is an expression for Shannon's well-known result that mutual information in Gaussian channels is the logarithm of the ratio of signal-plus-noise power and noise power

$$I(x_1; \mathcal{Y}) = \log \frac{E[|x_1|^2]}{E[|(\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})x_1|^2]}. \tag{20}$$

If $\{x_i\}$ is k dimensional, then the generalization of this formula is [20, Sec.10.5]

$$I(x_1^k; \mathcal{Y}) = \log \frac{|\mathrm{Cov}[x_1^k]|}{|\mathrm{Cov}[(\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})x_1^k]|}. \tag{21}$$

Another special case is when $\mathcal{X} = \{\underline{x}_i\}_{-\infty}^{\infty}$ and $\mathcal{Y} = \{\underline{y}_i\}_{-\infty}^{\infty}$ are jointly wide-sense stationary (w.s.s.) and jointly Gaussian multivariate processes. Because of stationarity, the information rate is equivalently given by (e.g., [21], [17])

$$I(\mathcal{X}; \mathcal{Y}) = I(\underline{x}_i; \mathcal{Y} | \underline{x}_{-\infty}^{i-1}). \tag{22}$$

Following an approach introduced by Pinsker [17], we now apply the chain rule of mutual information and Lemma 1 to get

$$\begin{aligned} I(\underline{x}_i; \mathcal{Y} | \underline{x}_{-\infty}^{i-1}) &= I(\underline{x}_i; \mathcal{Y}) + I(\underline{x}_i; \underline{x}_{-\infty}^{i-1} | \mathcal{Y}) - I(\underline{x}_i; \underline{x}_{-\infty}^{i-1}) \\ &= I(\underline{x}_i; \mathcal{Y}) + I((\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\underline{x}_i; (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\underline{x}_{-\infty}^{i-1}) - I(\underline{x}_i; \underline{x}_{-\infty}^{i-1}). \end{aligned} \tag{23}$$

If we define $\underline{e}_i = (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\underline{x}_i$ and apply the analytical expression for mutual information given in (21), we obtain

$$\begin{aligned} I(\underline{x}_i; \mathcal{Y} | \underline{x}_{-\infty}^{i-1}) &= \log \frac{|\mathrm{Cov}[\underline{x}_i]|}{|\mathrm{Cov}[\underline{e}_i]|} + \log \frac{|\mathrm{Cov}[\underline{e}_i]|}{\left|\mathrm{Cov}[(\boldsymbol{I} - \mathbf{\Pi}_{\underline{e}_{-\infty}^{i-1}})\underline{e}_i]\right|} \\ &\quad - \log \frac{|\mathrm{Cov}[\underline{x}_i]|}{\left|\mathrm{Cov}[(\boldsymbol{I} - \mathbf{\Pi}_{\underline{x}_{-\infty}^{i-1}})\underline{x}_i]\right|} \\ &= \log \frac{\left|\mathrm{Cov}[(\boldsymbol{I} - \mathbf{\Pi}_{\underline{x}_{-\infty}^{i-1}})\underline{x}_i]\right|}{\left|\mathrm{Cov}[(\boldsymbol{I} - \mathbf{\Pi}_{\underline{e}_{-\infty}^{i-1}})\underline{e}_i]\right|} \qquad (24) \\ &= \log \frac{\left|\mathrm{Cov}[\underline{x}_i | \underline{x}_{-\infty}^{i-1}]\right|}{\left|\mathrm{Cov}[\underline{e}_i | \underline{e}_{-\infty}^{i-1}]\right|}. \qquad (25) \end{aligned}$$

It is of interest to point out that, for a zero-mean w.s.s. process, the determinant of $\mathrm{Cov}[\underline{x}_i | \underline{x}_{-\infty}^{i-1}]$ is equal to the geometric mean of the determinant of the multivariate power-spectral density of

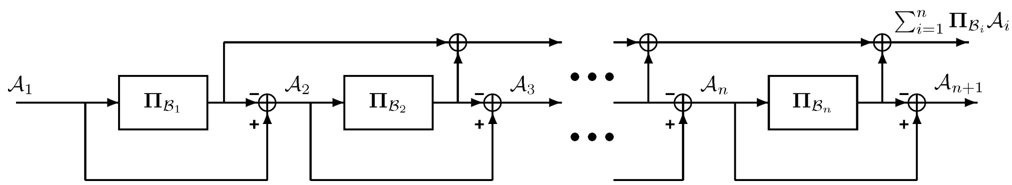

Fig. 1. Illustration of Proposition 1. The summation of the mutual information quantities across each $\Pi_{\mathcal{B}_i}$ is equal to the mutual information between $\mathcal{A}_1$ and $\sum_{i=1}^{n} \Pi_{\mathcal{B}_i}\mathcal{A}_i$.

the process [22]. That is, if we define the autocorrelation sequence $(R_{\underline{x}})_n = E[\underline{x}_{j+n}\underline{x}_j^{\dagger}]$ and the power spectral density

$$S_{\underline{x}}(e^{j\theta}) = \sum_{n=-\infty}^{\infty} (R_{\underline{x}})_n e^{-jn\theta}$$

then

$$\left|\mathrm{Cov}[\underline{x}_i|\underline{x}_{-\infty}^{i-1}]\right| = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|S_{\underline{x}}(e^{j\theta})\right| d\theta\right).$$

III. Derivation of the Canonical Decision Feedback Receiver

In this section, we start with a mutual information term and manipulate it using the chain rule of mutual information and orthogonal projections to produce a useful additive decomposition. This allows us to derive a generalized canonical decision-feedback structure that applies to Gaussian channels.

A. An Additive Decomposition of Mutual Information

We commence with a proposition that will enable us to derive the canonical decision-feedback receiver structure. It employs the general expression for the mutual information between Gaussian quantities found in (19).

Proposition 1: Let $\mathcal{A}_1, \mathcal{B}_1, \ldots, \mathcal{B}_n$ be sets of random variables in $\mathcal{H}$ for which $\{\mathcal{A}_1, \mathcal{B}_1, \ldots, \mathcal{B}_n\}$ is a set of jointly Gaussian random variables and the sets $\mathcal{B}_1, \ldots, \mathcal{B}_n$ are mutually orthogonal to each other. If we define $\mathcal{A}_{i+1} = (\boldsymbol{I} - \Pi_{\mathcal{B}_i})\mathcal{A}_i$ for $i = 1, \ldots, n$, then

$$I\left(\mathcal{A}_1; \sum_{i=1}^{n} \Pi_{\mathcal{B}_i}\mathcal{A}_i\right) = \sum_{i=1}^{n} I(\mathcal{A}_i; \Pi_{\mathcal{B}_i}\mathcal{A}_i). \tag{26}$$

Proof: Let us first define $\mathcal{Z}_i = \Pi_{\mathcal{B}_i}\mathcal{A}_i$ for each i. Since $\mathcal{A}_i$ and $\mathcal{Z}_i$ are jointly Gaussian, we know from (19) that

$$I(\mathcal{A}_i; \mathcal{Z}_i) = \log\frac{\langle\mathcal{A}_i\rangle}{\langle(\boldsymbol{I} - \Pi_{\mathcal{Z}_i})\mathcal{A}_i\rangle}. \tag{27}$$

Now $(\boldsymbol{I} - \Pi_{\mathcal{Z}_i})\mathcal{A}_i$ is equal to $(\boldsymbol{I} - \Pi_{\mathcal{B}_i})\mathcal{A}_i$ (cf. discussion following (15) in the proof of Lemma 2), which is itself called $\mathcal{A}_{i+1}$. Thus, the right-hand side of (26) becomes

$$\begin{aligned}\sum_{i=1}^{n} I(\mathcal{A}_i; \Pi_{\mathcal{B}_i}\mathcal{A}_i) &= \sum_{i=1}^{n} \log\frac{\langle\mathcal{A}_i\rangle}{\langle\mathcal{A}_{i+1}\rangle}\\ &= \log\frac{\langle\mathcal{A}_1\rangle}{\langle\mathcal{A}_{n+1}\rangle}.\end{aligned} \tag{28}$$

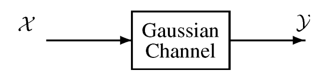

Fig. 2. The Gaussian channel. If the set $\mathcal{X}$ contains jointly Gaussian variables, then the set $\{\mathcal{X}, \mathcal{Y}\}$ is jointly Gaussian.

Now notice that $\mathcal{A}_i = (\boldsymbol{I} - \Pi_{\mathcal{B}_{i-1}})(\boldsymbol{I} - \Pi_{\mathcal{B}_{i-2}})\cdots(\boldsymbol{I} - \Pi_{\mathcal{B}_1})\mathcal{A}_1$ simplifies to $(\boldsymbol{I} - \sum_{j=1}^{i-1}\Pi_{\mathcal{B}_j})\mathcal{A}_1$ because of the mutual orthogonality of $\mathcal{B}_1, \ldots, \mathcal{B}_n$. This provides

$$\begin{aligned}\Pi_{\mathcal{B}_i}\mathcal{A}_i &= \Pi_{\mathcal{B}_i}\left(\boldsymbol{I} - \sum_{j=1}^{i-1}\Pi_{\mathcal{B}_j}\right)\mathcal{A}_1\\ &= \Pi_{\mathcal{B}_i}\mathcal{A}_1\end{aligned} \tag{29}$$

where we have used the fact that $\mathcal{B}_i$ is orthogonal to $\{\mathcal{B}_j\}_{j=1}^{i-1}$. This leads us to conclude that the left-hand side of (26) satisfies

$$I\left(\mathcal{A}_1; \sum_{i=1}^{n} \Pi_{\mathcal{B}_i}\mathcal{A}_i\right) = I\left(\mathcal{A}_1; \tilde{\Pi}\mathcal{A}_1\right) \tag{30}$$

where $\tilde{\Pi} \triangleq \sum_{i=1}^{n}\Pi_{\mathcal{B}_i}$; it is easily verified that the operator $\tilde{\Pi}$ is an orthogonal projection. Hence, we have that

$$\begin{aligned}I\left(\mathcal{A}_1; \sum_{i=1}^{n} \Pi_{\mathcal{B}_i}\mathcal{A}_i\right) &= \log\frac{\langle\mathcal{A}_1\rangle}{\langle(\boldsymbol{I} - \tilde{\Pi})\mathcal{A}_1\rangle}\\ &= \log\frac{\langle\mathcal{A}_1\rangle}{\langle\mathcal{A}_{n+1}\rangle}.\end{aligned} \tag{31}$$

Since (28) and (31) are equivalent, we have the desired result. □

A block diagram of the quantities involved in this proposition is given in Fig. 1.

B. A Decomposition of the Gaussian Channel

Consider the Gaussian channel shown in Fig. 2. The input into the channel is $\mathcal{X}$ and the output is $\mathcal{Y}$, where $\{\mathcal{X}, \mathcal{Y}\}$ is a jointly Gaussian subset of $\mathcal{H}$ whenever $\mathcal{X}$ consists of jointly Gaussian variables. The mutual information $I(\mathcal{X}; \mathcal{Y})$ gives the maximum rate at which data can be reliably transmitted across the channel under the given input distribution. If the input is partitioned as $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_n$, then the chain rule of mutual information provides

$$I(\mathcal{X}; \mathcal{Y}) = \sum_{i=1}^{n} I(\mathcal{X}_i; \mathcal{Y}|\mathcal{X}_1, \ldots, \mathcal{X}_{i-1}). \tag{32}$$

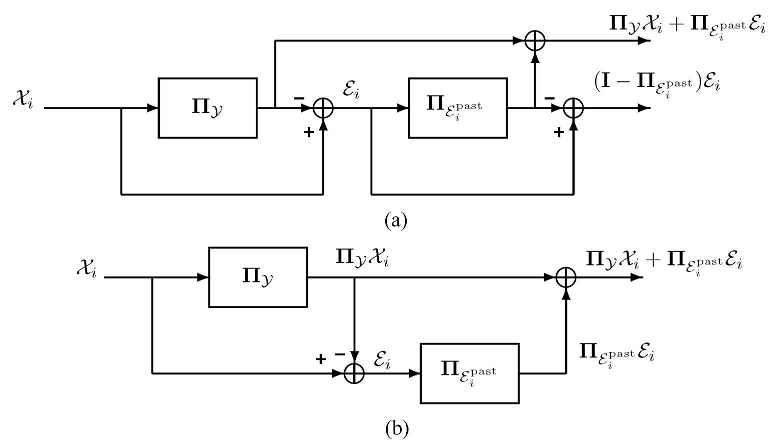

Fig. 3. Decomposition of mutual information. (b) is equivalent to (a) except that it does not include the output $(\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{E}_i^{\text{past}}})\mathcal{E}_i$.

Let us refer to the collection $(\mathcal{X}_1,\ldots,\mathcal{X}_{i-1})$ as $\mathcal{X}_i^{\text{past}}$. Taking any one of the terms in the summation of (32), we have

$$\begin{aligned} I(\mathcal{X}_i;\mathcal{Y}|\mathcal{X}_i^{\text{past}}) &= I(\mathcal{X}_i;\mathcal{Y}) + I(\mathcal{X};\mathcal{X}_i^{\text{past}}|\mathcal{Y}) - I(\mathcal{X}_i;\mathcal{X}_i^{\text{past}}) \\ &= I(\mathcal{X}_i;\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i) + I((\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}_i;(\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}_i^{\text{past}}) \\ &\quad - I(\mathcal{X}_i;\mathbf{\Pi}_{\mathcal{X}_i^{\text{past}}}\mathcal{X}_i) \end{aligned} \tag{33}$$

where the first line comes from the chain rule of mutual information and the second line uses Lemmas 1 and 2. We now make a further assumption that $\mathcal{X}_i$ and $\mathcal{X}_i^{\text{past}}$ are orthogonal sets of random variables so that the term $I(\mathcal{X}_i;\mathbf{\Pi}_{\mathcal{X}_i^{\text{past}}}\mathcal{X}_i)$ becomes zero.[3] Now define $\mathcal{E}_i = (\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}_i$ and $\mathcal{E}_i^{\text{past}} = (\boldsymbol{I}-\mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}_i^{\text{past}}$. This notation allows us to write

$$\begin{aligned} I(\mathcal{X}_i;\mathcal{Y}|\mathcal{X}_i^{\text{past}}) &= I(\mathcal{X}_i;\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i) + I(\mathcal{E}_i;\mathcal{E}_i^{\text{past}}) \\ &= I(\mathcal{X}_i;\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i) + I(\mathcal{E}_i;\mathbf{\Pi}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i). \end{aligned} \tag{34}$$

Note that $\mathbf{\Pi}_{\mathcal{Y}}$ and $\mathbf{\Pi}_{\mathcal{E}_i^{\text{past}}}$ project onto orthogonal subspaces of $\mathcal{H}$. This follows since $\mathcal{X}_i = \mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i + \mathcal{E}_i^{\text{past}}$ where $\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i \subset \mathcal{H}_{\mathcal{Y}}$ and $\mathcal{E}_i \subset \mathcal{H}_{\mathcal{Y}}^{\perp}$. It is now clear that we have expressed $I(\mathcal{X}_i;\mathcal{Y}|\mathcal{X}_i^{\text{past}})$ as a summation of terms that satisfy the hypotheses of Proposition 1 by identifying $\mathcal{A}_1 = \mathcal{X}_i$, $\mathcal{A}_2 = \mathcal{E}_i$, $\mathcal{B}_1 = \mathcal{Y}$, and $\mathcal{B}_2 = \mathcal{E}_i^{\text{past}}$. So we are able to state that

$$I(\mathcal{X}_i;\mathcal{Y}|\mathcal{X}_i^{\text{past}}) = I(\mathcal{X}_i;\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i + \mathbf{\Pi}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i). \tag{35}$$

This decomposition is pictured in Fig. 3 in two equivalent forms.

C. The Canonical Decision-Feedback Receiver

At this juncture, we interpret the decomposition of the Gaussian channel pictured in Fig. 3 from the viewpoint of how the linear operators are realized in an actual channel. Recall that since $\mathbf{\Pi}_{\mathcal{Y}}$ is a linear operator, each element of $\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i$ can be expressed as a linear combination of the elements of $\mathcal{Y}$ (e. g., matrix multiplication, convolution). In other words, we may view $\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i$ as a filtered version of the input $\mathcal{Y}$. We will represent this linear filter by $\boldsymbol{M}_1$. Since we need $\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i$ for each i, we will take the filter output $\boldsymbol{M}_1\mathcal{Y}$ to be $\{\mathbf{\Pi}_{\mathcal{Y}}\mathcal{X}_i\}_{i=1}^n$. That is, the filter $\boldsymbol{M}_1$ takes $\mathcal{Y}$ as its input, and its output is the orthogonal projection of $\mathcal{X}_i$ onto $\mathcal{H}_{\mathcal{Y}}$ for each i.

Similarly, $\mathbf{\Pi}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i$ is given by a linear combination of elements of $\mathcal{E}_i^{\text{past}}$. This we capture with the strictly "causal" linear filter $\boldsymbol{M}_2$. This filter takes $\mathcal{E} = (\mathcal{E}_1,\ldots,\mathcal{E}_n)$ as its input and produces $\{\mathbf{\Pi}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i\}_{i=1}^n$. Note that $\boldsymbol{M}_2$ can have $\mathcal{E}$ as its input because, for each i, $\mathcal{E}_i$ is being projected onto $\mathcal{H}_{\mathcal{E}_1^{i-1}}$ which is a subset of $\mathcal{H}_{\mathcal{E}}$. The notion of causality for the filter $\boldsymbol{M}_2$ derives from the fact that it projects an element of $\mathcal{E}_i$ onto is "past" $\{\mathcal{E}_j\}_{j=1}^{i-1}$.

By employing the linear filters $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$, Fig. 4 illustrates the data flow for the decomposition of the Gaussian channel that is implicit in Fig. 3. Note that Fig. 4 (b) possesses the structure of a decision-feedback receiver as typically defined, with its feedforward filter $(\boldsymbol{I}-\boldsymbol{M}_2)\boldsymbol{M}_1$ and its feedback filter $-\boldsymbol{M}_2$, except that the feedback is coming directly from the input since the derivations have explicitly assumed the feedback is perfect.

The generality of the derived canonical decision-feedback structure indicates that it applies to all Gaussian channels of the type shown in Fig. 2. Some important special cases are when the input $\mathcal{X}$ is a scalar, a vector, a w.s.s. process, or a multivariate w.s.s. process. Application of the result to these and other instances requires only that one determine the appropriate linear filters $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ for the particular case of interest. This is the subject of the next two sections.

IV. Realization of the Decision-Feedback Receiver for Synchronous CDMA and ISI Channels

We now illustrate how the development of the previous section applies to the particular cases of symbol-synchronous CDMA and the ISI channel. The net result is an information-theoretic derivation of their canonical decision-feedback receivers. This may be contrasted to works in which canonicity is derived by starting with an MMSE decision-feedback receiver and proving its information-theoretic optimality (e. g., [6], [2], [9]).

[3]It is easy to address the case where the inputs $(\mathcal{X}_1,\ldots,\mathcal{X}_n)$ are statistically dependent. The inputs are whitened at the transmitter to produce statistically independent inputs $(\tilde{\mathcal{X}}_1,\ldots,\tilde{\mathcal{X}}_n)$, and the particular coloring of the $(\mathcal{X}_1,\ldots,\mathcal{X}_n)$ is absorbed into the Gaussian channel. So this assumption is without loss of generality.

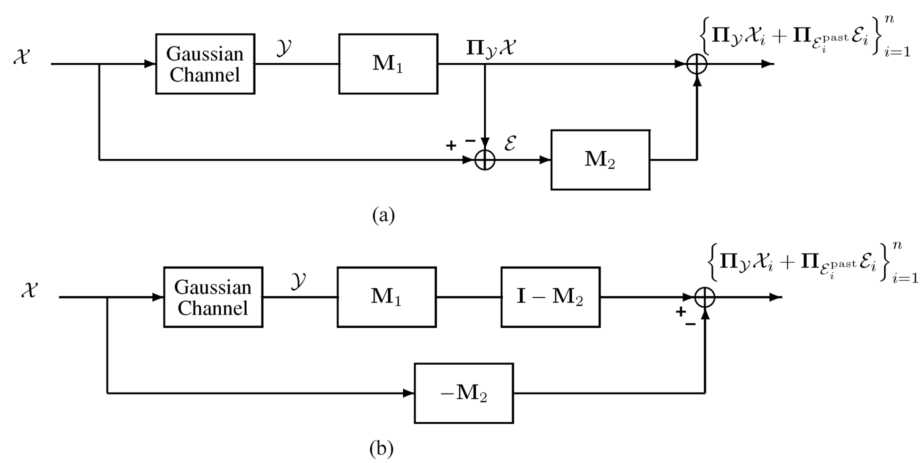

Fig. 4. A representation of the decomposition of mutual information with linear filtering; (a) and (b) are equivalent, with the latter having the well-known decision-feedback structure when the feedback is perfect. The linear filter represented by $\boldsymbol{I}$ is one for which the output is equal to the input.

A. Synchronous CDMA

In a symbol-synchronous CDMA channel, each user transmits a digital symbol every T seconds using multilevel quadrature-amplitude modulation (QAM). We assume that the superposition of their transmitted waveforms is corrupted by AWGN, and that the receiver consists of a parallel bank of M filters whose outputs are sampled at the symbol rate. The corresponding memoryless discrete-time channel is given by [23]

$$\underline{y} = \boldsymbol{A}\underline{x} + \underline{\eta} \tag{36}$$

where $\underline{x}$ is a length-K column vector containing the symbols of the K users, $\boldsymbol{A}$ is the $M \times K$ channel matrix, and the $M \times 1$ Gaussian noise vector $\underline{\eta}$ is zero-mean and proper.[4] To model the received power of the kth user, we let $p_k = E[|x_k|^2]$, with the diagonal matrix containing the users' powers given by $\boldsymbol{P} = \operatorname{diag}(p_1, p_2, \ldots, p_K)$.

The capacity region of this Gaussian CDMA channel was derived in [25] and the (decorrelating) decision-feedback detector was introduced in [26]. The authors showed that the MMSE decision-feedback detector has the property of achieving the sum-capacity of this channel at vertices on the dominant face of the capacity region [8], [9]. We now apply the results of Section III-C to succinctly yield the optimal decision-feedback structure.

In the notation of Section III-C, we have that $\mathcal{X} = \{x_1, \ldots, x_K\}$ and $\mathcal{Y} = \{y_1, \ldots, y_M\}$. The linear filter $\boldsymbol{M}_1$ takes $\underline{y}$ as its input and calculates $\mathbf{\Pi}_{\underline{y}}\underline{x}$. In the subsequent discussion we make use of the following notation. For a zero-mean random vector $\underline{a}$, denote its covariance by $\boldsymbol{R}_{\underline{a}} = E[\underline{a}\,\underline{a}^\dagger]$, and for zero-mean random vectors $\underline{a}$ and $\underline{b}$, denote their cross covariance by $\boldsymbol{R}_{\underline{a}\,\underline{b}} = E[\underline{a}\,\underline{b}^\dagger]$. Since $\boldsymbol{M}_1\underline{y} = E[\underline{x}|\underline{y}]$, this allows us to state that in this case the linear filter $\boldsymbol{M}_1$ corresponds to (see, e.g., [27], [18])

$$\boldsymbol{M}_1 = \boldsymbol{R}_{\underline{x}\,\underline{y}}\boldsymbol{R}_{\underline{y}}^{-1}. \tag{37}$$

[4] A random vector is proper if its *pseudo* covariance $E[(\underline{x} - E[\underline{x}])(\underline{x} - E[\underline{x}])^{\mathsf{T}}] = \mathbf{0}$ (as opposed to $E[(\underline{x} - E[\underline{x}])(\underline{x} - E[\underline{x}])^\dagger)] = \mathbf{0}$), where the superscripts T and † denote matrix and Hermitian transposition, respectively. See [5, Sec. 8.1.1] and [24].

That is, $\boldsymbol{M}_1$ takes a vector input and multiplies it by the matrix $\boldsymbol{R}_{\underline{x}\,\underline{y}}\boldsymbol{R}_{\underline{y}}^{-1}$. The error $\mathcal{E} = (\boldsymbol{I} - \mathbf{\Pi}_{\mathcal{Y}})\mathcal{X}$ in this context is the vector $\underline{e} = (\boldsymbol{I} - \mathbf{\Pi}_{\underline{y}})\underline{x}$. In terms of filtering this becomes $\underline{e} = \underline{x} - \boldsymbol{M}_1\underline{y}$, so that its covariance is evidently

$$\boldsymbol{R}_{\underline{e}} = \boldsymbol{R}_{\underline{x}} - \boldsymbol{R}_{\underline{x}\,\underline{y}}\boldsymbol{R}_{\underline{y}}^{-1}\boldsymbol{R}_{\underline{y}\,\underline{x}}. \tag{38}$$

To determine the filter $\boldsymbol{M}_2$, we begin with the unique Cholesky decomposition $\boldsymbol{R}_{\underline{e}} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^\dagger$, where $\boldsymbol{L}$ is a lower-triangular matrix with each diagonal entry equal to unity and $\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_n)$ is a diagonal matrix. We now argue that the ith element of the vector $\boldsymbol{L}^{-1}\underline{e}$ calculates $(\boldsymbol{I} - \mathbf{\Pi}_{e_1^{i-1}})e_i$. To see this, note first that $(\boldsymbol{I} - \mathbf{\Pi}_{e_1^{i-1}})e_i$ is equal to e_i plus a linear combination of e_1^{i-1}, and second that the set $\{(\boldsymbol{I} - \mathbf{\Pi}_{e_1^{i-1}})e_i\}_{i=1}^n$ contains mutually orthogonal random variables since $(\boldsymbol{I} - \mathbf{\Pi}_{e_1^{i-1}})e_i$ is that part of e_i that is orthogonal to its past. Analogously, the ith element of the vector $\boldsymbol{L}^{-1}\underline{e}$ is equal to e_i plus a linear combination of e_1^{i-1}, and the covariance of $\boldsymbol{L}^{-1}\underline{e}$ is equal to the diagonal matrix $\boldsymbol{D}$, meaning that the elements of $\boldsymbol{L}^{-1}\underline{e}$ are mutually orthogonal random variables. It is clear, then, that ith element of $(\boldsymbol{I} - \boldsymbol{L}^{-1})\underline{e}$ is equal to $\mathbf{\Pi}_{e_1^{i-1}}e_i$. Thus, we have that the linear filter $\boldsymbol{M}_2$ corresponds to matrix multiplication by

$$\boldsymbol{M}_2 = \boldsymbol{I} - \boldsymbol{L}^{-1}. \tag{39}$$

From (36), the required covariances and cross-covariances for evaluating the matrix filters $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are easily determined to be

$$\begin{aligned} \boldsymbol{R}_{\underline{x}} &= \boldsymbol{P} \\ \boldsymbol{R}_{\underline{x}} &= \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^\dagger + \boldsymbol{R}_{\underline{\eta}} \\ \boldsymbol{R}_{\underline{x}\,\underline{y}} &= \boldsymbol{R}_{\underline{y}\,\underline{x}}^\dagger = \boldsymbol{P}\boldsymbol{A}^\dagger \\ \boldsymbol{R}_{\underline{e}} &= \boldsymbol{P} - \boldsymbol{P}\boldsymbol{A}^\dagger(\boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^\dagger + \boldsymbol{R}_{\underline{\eta}})^{-1}\boldsymbol{A}\boldsymbol{P} \\ &= (\boldsymbol{P}^{-1} + \boldsymbol{A}^\dagger\boldsymbol{R}_{\underline{\eta}}^{-1}\boldsymbol{A})^{-1} \end{aligned} \tag{40}$$

where the last line is an application of the matrix-inversion lemma [28, Sec. 0.7.4].

The following theorem shows an important connection between the canonical decision-feedback receiver and the dominant face the CDMA capacity region.

Theorem 1: The canonical decision-feedback receiver for the Gaussian CDMA channel, under an assumption of error-free feedback, achieves the sum-capacity of the channel at a vertex of the dominant face of the capacity region.

Proof: From (20), we see that the achievable rate, r_i, for the ith user is

$$I(x_i;\underline{y}|x_1^{i-1}) = \log \frac{E[|x_i|^2]}{E[|(\boldsymbol{I}-\boldsymbol{\Pi}_{e_1^{i-1}})e_i|^2]} = \log \frac{p_i}{d_i} \tag{41}$$

where d_i is the ith diagonal element of $\boldsymbol{D}$ in the decomposition $\boldsymbol{R}_{\underline{e}} = \boldsymbol{LDL}^\dagger$. Clearly, the achieved rate tuple

$$(r_1,\ldots,r_K) = (\log(p_1/d_1),\ldots,\log(p_K/d_K))$$

must lie on the dominant face of the capacity region since the corresponding sum-rate $\sum_{i=1}^K r_i$ is equal to the sum-capacity of the channel (i. e., $I(\underline{x};\underline{y})$) because of the canonical nature of the decision-feedback receiver derived in Section III-C. In fact, since $r_i = I(x_i;\underline{y}|x_1^{i-1})$ for each i, we have that the rate tuple is a vertex or corner point of the dominant face of the capacity region [25]. There are generally $K!$ distinct vertices, one for each permutation of the the K indices. Each index permutation yields a decision-feedback receiver that processes the users in that order. Thus, by considering all possible orderings, every vertex of the dominant face is achievable. □

Before closing this section, we point the reader also to the generalized DFE (GDFE) developed by Cioffi and Forney in [7]. By periodically transmitting zeros in a scalar ISI channel, it is converted into a memoryless vector channel. This vector channel is parallelized to allow for the use of single-input single-output codecs in parallel. If the parallel scalar channels are viewed as those associated with users in a synchronous CDMA channel, the GDFE is equivalent to the per-user decision-feedback receiver discussed in this section. Conversely, if the channel matrix and noise covariance in (36) happen to be Toeplitz, then the CDMA decision-feedback receiver is an instance of the GDFE.

B. Multivariate ISI Channel

Consider the multivariate Gaussian ISI channel

$$\underline{y}_i = \sum_{j=-\infty}^{\infty} \boldsymbol{A}_j \underline{x}_{i-j} + \underline{\eta}_i \tag{42}$$

where the input sequence $\{\underline{x}_i\}$ consists of $K\times 1$ vectors, and the noise $\{\underline{\eta}_i\}$ is a sequence of $M\times 1$ vectors that are statistically independent of the input. The scalar Gaussian ISI channel is a special case occurring when $K = M = 1$ and arises in multilevel QAM signaling over a time-dispersive channel with matched filtering and symbol-rate sampling at the receiver (e. g., [29], [5, Sec. 6.2.1], [30, Sec. 6.2]).

We take the input and noise to be jointly Gaussian w.s.s. processes whose means are both zero. Furthermore, as per Footnote 3, the input is assumed to be a white process and the noise is assumed to be a full-rank regular process.[5] The sequence of $K\times M$ matrices $\{\boldsymbol{A}_i\}$ is assumed to satisfy $\sum_{i=-\infty}^{\infty} \mathrm{trace}(\boldsymbol{A}_i\boldsymbol{A}_i^\dagger) < \infty$, so that the Gaussian output process $\{\underline{y}_i\}$ is also w.s.s. Information-theoretic aspects under the Gaussian assumption are adequately covered in several places (e.g., see [21] for the scalar case and [31] for the multivariate case). Connections between the well-known DFE and capacity are found in [2], [3], and [5] for the scalar channel, and in [6] for the multivariate channel.

Before we apply canonical decision-feedback to this channel, we introduce some more notation. For a zero-mean, multivariate w.s.s. process $\{\underline{a}_i\}$, denote its autocorrelation sequence by $\{(\boldsymbol{R}_{\underline{a}})_i\}$, where $(\boldsymbol{R}_{\underline{a}})_i = E[\underline{a}_{j+i}\underline{a}_j^\dagger]_i$, and the z-transform of this sequence by

$$\boldsymbol{S}_{\underline{a}}(z) = \sum_{i=-\infty}^{\infty} (\boldsymbol{R}_{\underline{a}})_i z^{-i}.$$

When $\boldsymbol{S}_{\underline{a}}(z)$ is evaluated at $z = e^{j\theta}$ we obtain the multivariate power spectral density of the process. Similarly, for zero-mean, multivariate processes $\{\underline{a}_i\}$ and $\{\underline{b}_i\}$ that are jointly w.s.s., their cross-correlation sequence and cross spectrum are given by

$$\{(\boldsymbol{R}_{\underline{a}\,\underline{b}})_i\} = \{E[\underline{a}_{j+i}\underline{b}_j^\dagger]\}_i$$

and

$$\boldsymbol{S}_{\underline{a}\,\underline{b}}(z) = \sum_{i=-\infty}^{\infty} (\boldsymbol{R}_{\underline{a}\,\underline{b}})_i z^{-i}$$

respectively.

We now derive the canonical decision-feedback receiver with our information-lossless decomposition of the information rate. In the notation of Section III-C, we have $\mathcal{X} = \{\underline{x}_i\}$ and $\mathcal{Y} = \{\underline{y}_i\}$. The information rate is denoted by $I(\mathcal{X};\mathcal{Y})$, which because of stationarity can be expressed as $I(\underline{x}_i;\mathcal{Y}|\underline{x}_{-\infty}^{i-1})$. Thus, we make the partition $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ where $\mathcal{X}_1 = \underline{x}_i$ and $\mathcal{X}_2 = \underline{x}_{-\infty}^{i-1}$. The filter $\boldsymbol{M}_1$ evaluates $\boldsymbol{\Pi}_{\mathcal{Y}}\mathcal{X}$ by converting the input $\{\underline{y}_i\}$ to $\{E[\underline{x}_i|\mathcal{Y}]\}$. This is effected with multivariate Wiener filtering according to

$$\boldsymbol{M}_1 = \boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z). \tag{43}$$

The error sequence $\{\underline{e}_i\}$ is produced by passing the process $\{\underline{x}_i\}$ through the filter $\boldsymbol{I}-\boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z)$, and from (42), we find the spectrum of this process to be

$$\boldsymbol{S}_{\underline{e}}(z) = \boldsymbol{S}_{\underline{x}}(z) - \boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z)\boldsymbol{S}_{\underline{y}\,\underline{x}}(z). \tag{44}$$

We must now project $\underline{e}_i$ onto its past $\underline{e}_{-\infty}^{i-1}$, and this is done with a multivariate one-step prediction filter [22]. Finding this filter

[5]That is, the noise process satisfies the Szëgo (or Paley–Wiener) condition, $\int_{-\pi}^{\pi} \log|\boldsymbol{S}_{\underline{\eta}}(e^{j\theta})|d\theta > -\infty$, where $\boldsymbol{S}_{\underline{\eta}}(e^{j\theta})$ is the multivariate power spectral density of $\{\underline{\eta}_i\}$. This technical condition guarantees that the noise is not perfectly predictable from its past [22]; otherwise, the information rate would be infinite.

requires the unique "minimum-phase" multivariate spectral factorization[6]

$$\boldsymbol{S}_{\underline{e}}(z) = \boldsymbol{\Phi}(z)\boldsymbol{G}\boldsymbol{\Phi}^{\dagger}(1/z^*) \tag{45}$$

where $\boldsymbol{\Phi}_{\underline{e}}(z)$ and its inverse are both monic, causal, and stable. In other words, $\boldsymbol{\Phi}_{\underline{e}}(z)$ can be represented in the form $\boldsymbol{I} + \sum_{i=1}^{\infty} \boldsymbol{C}_i z^{-i}$ where the matrix Fourier coefficients satisfy

$$\sum_{i=1}^{\infty} \mathrm{trace}(\boldsymbol{C}_i \boldsymbol{C}_i^{\dagger}) < \infty$$

and similarly

$$\boldsymbol{\Phi}_{\underline{e}}^{-1}(z) = \boldsymbol{I} + \sum_{i=1}^{\infty} \boldsymbol{D}_i z^{-i}$$

with $\sum_{i=1}^{\infty} \mathrm{trace}(\boldsymbol{D}_i \boldsymbol{D}_i^{\dagger}) < \infty$. The one-step prediction filter is thus

$$\boldsymbol{M}_2 = \boldsymbol{I} - \boldsymbol{\Phi}_{\underline{e}}^{-1}(z). \tag{46}$$

Given the covariance $\boldsymbol{P} = E[\underline{x}_i \underline{x}_i^{\dagger}]$ of the white process $\{\underline{x}_i\}$, the terms necessary to fully describe $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ in this channel are

$$\begin{aligned}
\boldsymbol{S}_{\underline{x}}(z) &= \boldsymbol{P} \\
\boldsymbol{S}_{\underline{y}}(z) &= \boldsymbol{A}(z)\boldsymbol{P}\boldsymbol{A}^{\dagger}(1/z^*) + \boldsymbol{S}_{\underline{\eta}}(z) \\
\boldsymbol{S}_{\underline{x}\,\underline{y}}(z) &= \boldsymbol{S}_{\underline{y}\,\underline{x}}^{\dagger}(1/z^*) = \boldsymbol{P}\boldsymbol{A}^{\dagger}(1/z^*) \\
\boldsymbol{S}_{\underline{e}}(z) &= \boldsymbol{P} - \boldsymbol{P}\boldsymbol{A}^{\dagger}(1/z^*)\Big(\boldsymbol{A}(z)\boldsymbol{P}\boldsymbol{A}^{\dagger}(1/z^*) + \boldsymbol{S}_{\underline{\eta}}(z)\Big)^{-1}\boldsymbol{A}(z)\boldsymbol{P} \\
&= \Big(\boldsymbol{P}^{-1} + \boldsymbol{A}^{\dagger}(1/z^*)\boldsymbol{S}_{\underline{\eta}}^{-1}(z)\boldsymbol{A}(z)\Big)^{-1}.
\end{aligned} \tag{47}$$

From (25), the information rate of the channel is given by

$$\begin{aligned}
I(x_i; \underline{y}_{-\infty}^{\infty} | x_{-\infty}^{i-1}) &= \log \frac{|\mathrm{Cov}[\underline{x}_i]|}{\left|\mathrm{Cov}[(\boldsymbol{I} - \boldsymbol{\Pi}_{\underline{e}_{-\infty}^{i-1}})\underline{e}_i]\right|} \\
&= \log \frac{|\boldsymbol{P}|}{|\boldsymbol{G}|}.
\end{aligned} \tag{48}$$

Recall from Section II-D that $|\boldsymbol{G}|$ is the geometric mean of the determinant of the spectrum of $\{\underline{e}_i\}$ [22]; that is,

$$|\boldsymbol{G}| = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|S_{\underline{e}}(e^{j\theta})\right| d\theta\right).$$

We are also able to give some additional insight into the following proposition from the literature.

Proposition 2: Suppose we have a scalar Gaussian ISI channel with capacity C_{ISI}. Let C_{DFE} denote the resulting channel capacity when the receiver is a perfect-feedback MMSE DFE. The two capacities are related by $C_{\mathrm{DFE}} = C_{\mathrm{ISI}}$. $\diamondsuit$

In [2] it was shown only that $C_{\mathrm{DFE}} \geq C_{\mathrm{ISI}}$, and it was conjectured that the inequality cannot be replaced by equality because of the paradoxical result from [3] that perfect cancellation of post-cursor ISI is generally an information-increasing operation. But from our derivation of the MMSE DFE, since it begins with mutual information, we can see that the inequality may indeed be replaced by equality. Note that the canonical decision-feedback receiver converts the information rate $I(\mathcal{X};\mathcal{Y})$ into the mutual information between two scalar random variables in accordance with $I(x_i; \boldsymbol{\Pi}_{\mathcal{Y}} x_i + \boldsymbol{\Pi}_{e_{-\infty}^{i-1}} e_i)$. In order to view the perfect-feedback canonical decision-feedback receiver as information increasing requires that one instead deal with the information rate between the two w.s.s. random processes $\{x_i\}_{i=-\infty}^{\infty}$ and $\{\boldsymbol{\Pi}_{\mathcal{Y}} x_i + \boldsymbol{\Pi}_{e_{-\infty}^{i-1}} e_i\}_{i=-\infty}^{\infty}$, but such an understanding is in violation of what the decision-feedback receiver is effecting.[7]

V. The Symbol-Asynchronous CDMA Channel

The final channel that we consider is the symbol-asynchronous CDMA channel. Information-theoretic aspects and decision-feedback receivers have been explored separately for this channel in the literature [32], [33], and the authors considered them jointly in [34]. Deriving canonical decision-feedback receivers from decompositions of the information rate in this context borrows ideas from the synchronous CDMA channel discussed in Section IV-A and the multivariate ISI channel discussed in Section IV-B. In addition to developing lossless decision-feedback receivers in this section, we also discuss lossy receivers that meet certain causality constraints. Finally, we show some connections between multivariate spectral factorization, decision-feedback receivers, and information theory.

Since we must deal with both users and time, notation of the following type will be used in this section. Let $(x_k)_i$ be the symbol transmitted by the kth user at time i, and $(\underline{x})_i$ the $K \times 1$ vector of symbols transmitted by the users at time i. The sequence of symbols transmitted by the kth user is thus $(x_k)_{-\infty}^{\infty}$, and the vector sequence of symbols transmitted by all users is $(\underline{x})_{-\infty}^{\infty}$. We shall often find it convenient to denote these by $x_k(z)$ and $\underline{x}(z)$, respectively. In contrast to the synchronous CDMA channel discussed in Section IV-A, now the users' transmit pulses arrive asynchronously at the receiver, though it is assumed that the receiver knows the timing offsets. The received signal with AWGN $\eta(t)$ is

$$r(t) = \sum_{i=-\infty}^{\infty} \sum_{k=1}^{K} p_k^{1/2}(x_k)_i s(t - \tau_k - iT) + \eta(t) \tag{49}$$

where p_k, $s_k(t)$, and τ_k are the power, complex signature waveform, and relative timing offset of the kth user, respectively; and T is the symbol interval. To obtain a discrete-time model, we take the receiver frontend to consist of a parallel bank of M filters $\{f_m(t)\}_{m=1}^{M}$ whose outputs are sampled at the symbol rate. The sample of the mth filter at the ith time is

$$\int_{-\infty}^{\infty} r(t) f_m(iT - t) dt.$$

Stacking the M sampled outputs of the filters at the ith time interval we get $\underline{y}_i$. It takes the form

$$\underline{y}_i = \sum_{j=-\infty}^{\infty} \boldsymbol{A}_j (\underline{x})_{i-j} + \underline{\eta}_i \tag{50}$$

[6]If $\boldsymbol{A}(z) = \sum_{i=-\infty}^{\infty} \boldsymbol{A}_i z^{-i}$, we use the notation $\boldsymbol{A}^{\dagger}(1/z^*)$ to mean $\sum_{i=-\infty}^{\infty} \boldsymbol{A}_i^{\dagger} z^i$.

[7]In [3], the difference between the information rate and $I(x_i; \boldsymbol{\Pi}_{\mathcal{Y}} x_i + \boldsymbol{\Pi}_{e_{-\infty}^{i-1}} e_i)$ is interpreted as a mandatory precoding loss when the feedback is housed at in the transmitter.

where the (m,k) element of the matrix $\boldsymbol{A}_i$ is given by

$$\int_{-\infty}^{\infty} s_k(t-\tau_k+iT)f_m(-t)dt$$

and the nth element of $\underline{\eta}_i$ is the Gaussian variable

$$\int_{-\infty}^{\infty} \eta(t+iT)f_m(-t)dt.$$

A specific instance would be when there are $M=K$ receive filters that are matched to the users' delay-shifted signature waveforms, in which case $f_m(-t)=s_m^*(t-\tau_m)$ [23]. Note that (50) is of the same form as the multivariate ISI channel considered in Section IV-B, except that the covariance of $(\underline{x})_i$ is a diagonal matrix $\boldsymbol{P}=\mathrm{diag}(p_1,p_2,\ldots,p_K)$ since the users signal independently of each other, where p_i is the power of the ith user.

The capacity region of the symbol-asynchronous CDMA channel was derived in [32]. The user inputs that maximize the sum information rate 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 our purposes, then, we assume that the sequence of symbols transmitted by each user is a white Gaussian process since any spectral shaping can be absorbed into the channel model (cf. Footnote 3). The input into the Gaussian channel is $\mathcal{X}=(\underline{x})_{-\infty}^{\infty}$ and the channel output is $\mathcal{Y}=\underline{y}_{-\infty}^{\infty}$. The sum information rate is $I(\mathcal{X};\mathcal{Y})$.

A. The Information-Lossless Decision-Feedback Receiver

Let us first partition the input as $\mathcal{X}=\mathcal{X}_1\cup\cdots\cup\mathcal{X}_K$, where $\mathcal{X}_k=(x_k)_{-\infty}^{\infty}$ is the sequence of symbols transmitted by the kth user. The first filter is

$$\boldsymbol{M}_1=\boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z) \tag{51}$$

the same expression as in the multivariate ISI channel (cf. Section IV-B) since in both cases $\mathcal{X}$ and $\mathcal{Y}$ are multivariate w.s.s. processes. Similarly, the error $(\underline{e})_{-\infty}^{\infty}$ (or, equivalently, $\underline{e}(z)$) is a K-variate w.s.s. process whose kth sequence $e_k(z)$ is $(e_k)_{-\infty}^{\infty}=\{(x_k)_i-\boldsymbol{\Pi}_{\mathcal{Y}}(x_k)_i\}_{i=-\infty}^{\infty}$. The multivariate spectrum of $\underline{e}(z)$ is $\boldsymbol{S}_{\underline{e}}(z)=\boldsymbol{S}_{\underline{x}}(z)-\boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z)\boldsymbol{S}_{\underline{y}\,\underline{x}}(z)$. To find the filter $\boldsymbol{M}_2$, we follow a path similar to that taken in deriving it for the synchronous CDMA channel in Section IV-A. Toward this end, start with the decomposition

$$\boldsymbol{S}_{\underline{e}}(z)=\boldsymbol{L}(z)\boldsymbol{D}(z)\boldsymbol{L}^{\dagger}(1/z^*) \tag{52}$$

in which $\boldsymbol{D}(z)=\mathrm{diag}(d_1(z),\ldots,d_K(z))$ is a diagonal matrix and $\boldsymbol{L}(z)$ is lower-triangular and has all of its diagonal elements equal to unity. Suppose that the error sequence $\underline{e}(z)$ is filtered by $\boldsymbol{L}^{-1}(z)$, then the kth element of the vector $\boldsymbol{L}^{-1}(z)\underline{e}(z)$ gives the sequence $e_k(z)-\boldsymbol{\Pi}_{e_1^{k-1}(z)}\underline{e}(z)$, where $e_1^{k-1}(z)=(e_1(z),\ldots,e_{k-1}(z))$. Thus, the kth element of $(\boldsymbol{I}-\boldsymbol{L}^{-1}(z))\underline{e}(z)$ is the sequence $\boldsymbol{\Pi}_{e_1^{k-1}(z)}e_k(z)$ so that the second filter is

$$\boldsymbol{M}_2=\boldsymbol{I}-\boldsymbol{L}^{-1}(z). \tag{53}$$

The spectra required for finding $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are the same as those used for the multivariate ISI channel in (47) when $\boldsymbol{P}$ is a constant, diagonal power matrix.

It may be observed that our particular partition of the input $\mathcal{X}$ has effectively decomposed the K-user ISI channel into K single-user ISI channels. The kth single-user ISI channel from this decomposition has $x_k(z)$ as the input and $\tilde{y}_k(z)=\boldsymbol{\Pi}_{\mathcal{Y}}x_k(z)+\boldsymbol{\Pi}_{e_1^{k-1}(z)}e_k(z)$ as its output. Clearly, since these are ISI channels, their information rates can each be decomposed as was done in Section IV-B for an arbitrary ISI channel. Thus, from (43) and (46), the two required filters $\boldsymbol{M}_{k1}$ and $\boldsymbol{M}_{k2}$ for the kth effective single-user channel are

$$\boldsymbol{M}_{k1}=\boldsymbol{S}_{x_k\tilde{y}_k}(z)\boldsymbol{S}_{\tilde{y}_k}^{-1}(z) \tag{54}$$

$$\boldsymbol{M}_{k2}=1-\phi_k^{-1}(z) \tag{55}$$

where $\phi_k(z)$ comes from the minimum-phase spectral factorization of the kth diagonal element of $\boldsymbol{D}(z)$. That is, $d_k(z)=\phi_k(z)g_k\phi_k^*(1/z^*)$ where $\phi_k(z)$ and its inverse are both monic (i.e., the zeroth Fourier coefficient is unity), causal, and stable. It is easy to show that the filter $\boldsymbol{M}_{k1}$ simplifies to unity (i.e., the output equals the input); to see this, observe that $\boldsymbol{M}_{k1}$ evaluates $\boldsymbol{\Pi}_{\tilde{y}_k(z)}x_k(z)$, which is the orthogonal projection of $x_k(z)$ onto $\mathcal{H}_{\tilde{y}_k(z)}$, and then note that $\mathcal{H}_{\tilde{y}_k(z)}\subset\mathcal{H}_{\mathcal{Y}}$ and $\tilde{y}_k(z)=\boldsymbol{\Pi}_{\mathcal{Y}}x_k(z)$. The corresponding partition of the input for the kth user is in this case $\mathcal{X}_k=\mathcal{X}_{k1}\cup\mathcal{X}_{k2}$ where $\mathcal{X}_{k1}=(x_k)_i$ is the ith symbol of user k and $\mathcal{X}_{k2}=\left((x_k)_{-\infty}^{i-1},(x_1^{k-1})_{-\infty}^{\infty}\right)$ consists of the past symbols of user k and all symbols of users 1 through $k-1$. The resulting decomposition of the multiuser channel information rate yields

$$I(\mathcal{X};\mathcal{Y})=\sum_{k=1}^{K}\log\left(\frac{p_k}{g_k}\right). \tag{56}$$

In summary, then, the high-level structure of this canonical decision-feedback receiver converts the MIMO asynchronous channel into a set of K independent single-input single-output ISI channels, while its low-level structure converts each of these scalar ISI channels into memoryless channels. Combining these levels together, we may express the two filters as

$$\boldsymbol{M}_1=\boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z) \tag{57}$$

$$\boldsymbol{M}_2=\boldsymbol{I}-\tilde{\boldsymbol{L}}^{-1}(z) \tag{58}$$

where $\tilde{\boldsymbol{L}}(z)=\boldsymbol{L}(z)\mathrm{diag}(\phi_1(z),\ldots,\phi_K(z))$ comes from the structured spectral factorization

$$\boldsymbol{S}_{\underline{e}}(z)=\tilde{\boldsymbol{L}}(z)\boldsymbol{G}\tilde{\boldsymbol{L}}^{\dagger}(1/z^*) \tag{59}$$

in which $\boldsymbol{G}=\mathrm{diag}(g_1,\ldots,g_K)$. Note that $\tilde{\boldsymbol{L}}(z)$ and its inverse are lower-triangular with diagonal elements that are monic, causal, and stable.

B. Some Lossy Decision-Feedback Receivers

While the previous section placed causality constraints on the feedback only in the sense of feeding back "past" users (i. e., $\mathcal{X}_1,\ldots,\mathcal{X}_{k-1}$ are fed back for user k), one may also subject the feedback to causal restrictions in time. In general, a capacity penalty is incurred, but still there exists a corresponding canonical decision-feedback receiver.

1) A Case in Which Feedback is Causal for Both Users and Time: Consider the following:

$$\begin{aligned} I(\mathcal{X};\mathcal{Y}) &= \sum_{k=1}^{K} I((x_k)_i; \underline{y}_{-\infty}^{\infty} | (x_1^{k-1})_{-\infty}^{\infty}, (x_k)_{-\infty}^{i-1}) \\ &\geq \sum_{k=1}^{K} I((x_k)_i; \underline{y}_{-\infty}^{\infty} | (x_1^{k-1})_{-\infty}^{i}, (x_k)_{-\infty}^{i-1}). \quad (60) \end{aligned}$$

If we now apply the techniques of Section III-C, the result is a decomposition of the lossy information rate that has causal feedback in both users and time. That is, when processing the kth user at time i, we use only the past and current symbols of users 1 through $k-1$ (i.e., $(x_1^{k-1})_{-\infty}^{i}$ as opposed to all of their symbols, $(x_1^{k-1})_{-\infty}^{\infty}$) and the past symbols of user k (i.e., $(x_k)_{-\infty}^{i-1}$).

The filter $\boldsymbol{M}_1$ that projects onto $\mathcal{Y}$ remains unchanged from the lossless cases just discussed

$$\boldsymbol{M}_1 = \boldsymbol{S}_{\underline{x}\,\underline{y}}(z)\boldsymbol{S}_{\underline{y}}^{-1}(z). \quad (61)$$

The multivariate error sequence is still given by $(\underline{e})_{-\infty}^{\infty} = \underline{e}(z)$ with $\underline{e}_i = \underline{x}_i - \mathbf{\Pi}_{\mathcal{Y}}\underline{x}_i$. The filter $\boldsymbol{M}_2$ must allow us to evaluate $\mathbf{\Pi}_{\{(e_1^{k-1})_{-\infty}^{i}, (e_k)_{-\infty}^{i-1}\}}(e_k)_i$. To accomplish this, let $\boldsymbol{S}_{(k)}(z)$ be the principal submatrix of the spectrum of $\boldsymbol{S}_{\underline{e}}(z)$ formed using the first k indices. Its unique minimum-phase spectral factorization is [22]

$$\boldsymbol{S}_{(k)}(z) = \mathbf{\Phi}_{(k)}(z)\boldsymbol{G}_{(k)}\mathbf{\Phi}_{(k)}^{\dagger}(1/z^*) \quad (62)$$

where $\mathbf{\Phi}_{(k)}(z)$ and its inverse are both causal, stable, and have zeroth Fourier coefficients that are lower-triangular with diagonal elements that are unity, and $\boldsymbol{G}_{(k)} = \mathrm{diag}(g_{k1}, \ldots g_{kk})$ is a constant diagonal matrix.[8] Note that filtering $[e_1(z) \cdots e_k(z)]^{\mathsf{T}}$ by the kth row of $\mathbf{\Phi}_{(k)}^{-1}(z)$ produces the sequence

$$\left\{ \left(\boldsymbol{I} - \mathbf{\Pi}_{\{(e_1^{k-1})_{-\infty}^{i}, (e_k)_{-\infty}^{i-1}\}} \right) (e_k)_i \right\}_{i=-\infty}^{\infty}. \quad (63)$$

Clearly, then, the following multivariate filter:

$$\boldsymbol{M}_2 = \boldsymbol{I} - \begin{bmatrix} [\mathrm{row}_1(\mathbf{\Phi}_{(1)}^{-1}(z)) & \mathbf{0}_{1\times K-1}] \\ [\mathrm{row}_2(\mathbf{\Phi}_{(2)}^{-1}(z)) & \mathbf{0}_{1\times K-2}] \\ \vdots & \\ \mathrm{row}_K(\mathbf{\Phi}_{(K)}^{-1}(z)) & \end{bmatrix} \quad (64)$$

is lower-triangular, causal (strictly so on the diagonal elements), stable, and converts $\underline{e}(z)$ to a vector whose kth element represents the sequence $\{\mathbf{\Pi}_{(e_1^{k-1})_{-\infty}^{i}, (e_k)_{-\infty}^{i-1}}(e_k)_i\}_{i=-\infty}^{\infty}$. The resulting lossy information rate in this case is the right-hand side of

$$I(\mathcal{X};\mathcal{Y}) \geq \sum_{k=1}^{K} \log \frac{p_k}{g_{kk}}. \quad (65)$$

[8]The only difference between this minimum-phase multivariate spectral factorization and the one discussed in (45) is that we have performed a Cholesky decomposition $\boldsymbol{G} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\dagger}$ on the $\boldsymbol{G}$ of (45), absorbing $\boldsymbol{L}$ and $\boldsymbol{L}^{\dagger}$ into $\mathbf{\Phi}(z)$ and $\mathbf{\Phi}^{\dagger}(1/z^*)$, respectively, to obtain (62).

If we express $\boldsymbol{M}_2$ in (64) as $\boldsymbol{I} - \mathbf{\Psi}^{-1}(z)$ with $\mathbf{\Psi}(z)$ defined implicitly, then we have a corresponding *structured* factorization of the error spectrum

$$\boldsymbol{S}_{\underline{e}}(z) = \mathbf{\Psi}(z)\boldsymbol{F}(z)\mathbf{\Psi}^{\dagger}(1/z^*), \quad (66)$$

where $\boldsymbol{F}(z) = \mathbf{\Psi}^{-1}(z)\boldsymbol{S}_{\underline{e}}(z)(\mathbf{\Psi}^{\dagger}(1/z^*))^{-1}$. The kth diagonal element of $\boldsymbol{F}(z)$ is obviously equal to g_{kk}, the kth diagonal element of $\boldsymbol{G}_{(k)}$. It should be noted that $\mathbf{\Psi}(z)$ and its inverse are lower-triangular, causal, stable, and have monic diagonal entries. This factorization is essentially the so-called partial spectral factorization due to Duel-Hallen [33], which was derived therein by maximizing the effective signal-to-noise ratios of the users.

2) A Case in Which Both Filters Have Finite Impulse Responses: For our final example we consider a lossy case for which the derived decision-feedback receiver can be implemented with finite impulse response (FIR) filters. We begin with the lossy information rate

$$I(\mathcal{X};\mathcal{Y}) \geq \sum_{k=1}^{K} I((x_k)_i; \underline{y}_{n-N_y}^{i} | (x_1^{k-1})_{i-N_x}^{i-1}, (x_k)_{i-N_x}^{i-1})$$

where N_x and N_y are nonnegative integers that will indicate the number of taps in the filters $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$. The filter $\boldsymbol{M}_1$ projects $(\underline{x})_i$ onto $\underline{y}_{i-N_y}^{i}$. To determine this filter let

$$\underline{Y} = \left[\underline{y}_{i-N_y}^{\mathsf{T}} \quad \underline{y}_{i-N_y+1}^{\mathsf{T}} \quad \cdots \quad \underline{y}_i^{\mathsf{T}} \right]^{\mathsf{T}}. \quad (67)$$

Note that by stationarity the covariance of this vector is independent of i. Thus, we must evaluate $\boldsymbol{R}_{(\underline{x})_i\underline{Y}}\boldsymbol{R}_{\underline{Y}}^{-1}$ where

$$\boldsymbol{R}_{(\underline{x})_i\underline{Y}} = \boldsymbol{P}\left[\boldsymbol{A}_{-N_y}^{\dagger} \quad \boldsymbol{A}_{-N_y+1}^{\dagger} \quad \cdots \quad \boldsymbol{A}_0^{\dagger} \right].$$

To represent $\boldsymbol{M}_1$ as an FIR filter, we identify

$$\boldsymbol{M}_1 = \boldsymbol{J}(z) \quad (68)$$

where $\boldsymbol{J}(z) = \sum_{j=0}^{N_y} \boldsymbol{J}_j z^{-j}$ and the $K \times M$ matrix coefficients $\{\boldsymbol{J}_j\}$ are implicitly defined by

$$\boldsymbol{R}_{(\underline{x})_i\underline{Y}}\boldsymbol{R}_{\underline{Y}}^{-1} \triangleq [\, \boldsymbol{J}_{N_y} \quad \boldsymbol{J}_{N_y-1} \quad \cdots \quad \boldsymbol{J}_0 \,]. \quad (69)$$

To find filter $\boldsymbol{M}_2$, let

$$\underline{X} = [\, \underline{x}_{i-N_x}^{\mathsf{T}} \quad (\underline{x})_{i-N_x+1}^{\mathsf{T}} \quad \cdots \quad (\underline{x})_i^{\mathsf{T}} \,]^{\mathsf{T}}. \quad (70)$$

The ith error vector is $(\underline{e})_i = (\boldsymbol{I} - \mathbf{\Pi}_{\underline{y}_{i-N_y}^{i}})(\underline{x})_i$, and these may be stacked to form the vector

$$\underline{E} = [\, (\underline{e})_{i-N_x}^{\mathsf{T}} \quad (\underline{e})_{i-N_x+1}^{\mathsf{T}} \quad \cdots \quad (\underline{e})_i^{\mathsf{T}} \,]^{\mathsf{T}}. \quad (71)$$

Its covariance is

$$\boldsymbol{R}_{\underline{E}} = \mathbb{P} - \mathbb{P}\mathbb{A}^{\dagger}\boldsymbol{R}_{\underline{Y}}^{-1}\mathbb{A}\mathbb{P} \quad (72)$$

where $\mathbb{P}$ is the $(N_x+1) \times (N_x+1)$ block-diagonal matrix

$$\mathbb{P} = \text{block-diag}\,(\boldsymbol{P}, \ldots, \boldsymbol{P}) \quad (73)$$

and

$$\mathbb{A} = \begin{bmatrix} \boldsymbol{A}_{N_x-N_y} & \cdots & & \boldsymbol{A}_{-N_y} \\ \vdots & & & \vdots \\ & & \ddots & \boldsymbol{A}_{-1} \\ \boldsymbol{A}_{N_x} & \cdots & \boldsymbol{A}_1 & \boldsymbol{A}_0 \end{bmatrix} \quad (74)$$

is an $(N_y+1)\times(N_x+1)$ block-Toeplitz matrix formed from the matrix coefficients of the channel response in (50). $\boldsymbol{R}_{\underline{E}}$ can be viewed as an $(N_x+1)\times(N_x+1)$ block matrix, the size of each block being $K\times K$.

For filter $\boldsymbol{M}_2$ we must evaluate the projection of $(e_k)_i$ onto $(e_1^k)_{i-N_x}^{i-1}$. To accomplish this, we first reduce $\boldsymbol{R}_{\underline{E}}$ by retaining only the principal submatrices formed by using the first k indices of each block. Call this reduced covariance matrix $\boldsymbol{R}_{(k)}$, and perform a Cholesky factorization so that $\boldsymbol{R}_{(k)}=\boldsymbol{L}_{(k)}\boldsymbol{G}_{(k)}\boldsymbol{L}_{(k)}^{\dagger}$, where $\boldsymbol{L}_{(k)}$ is lower-triangular with diagonal elements that are unity and $\boldsymbol{G}_{(k)}=\mathrm{diag}(g_{k1},\ldots,g_{kk})$ is a diagonal matrix. We now parse the last row of $\boldsymbol{L}_{(k)}^{-1}$ into row vectors, each of length k, which we label from left to right as $\underline{l}_{kN_x}^{\mathsf{T}}$ down to $\underline{l}_{k0}^{\mathsf{T}}$. In z notation, we have $\underline{l}_k(z)=\sum_{j=0}^{N_x}\underline{l}_{kj}z^{-j}$. After repeating this procedure for $k=1,2,\ldots,K$, we form the lower-triangular matrix

$$\boldsymbol{\Sigma}(z)=\begin{bmatrix}[\underline{l}_1^{\mathsf{T}}(z) \quad \mathbf{0}_{1\times K-1}]\\ [\underline{l}_2^{\mathsf{T}}(z) \quad \mathbf{0}_{1\times K-2}]\\ \vdots \\ \underline{l}_K^{\mathsf{T}}(z)\end{bmatrix}^{-1}. \tag{75}$$

This gives us

$$\boldsymbol{M}_2=\boldsymbol{I}-\boldsymbol{\Sigma}^{-1}(z) \tag{76}$$

for the second filter, which is strictly causal, lower-triangular, and polynomial (i.e., an FIR filter). The corresponding spectral decomposition is

$$\boldsymbol{S}_{\underline{e}}(z)=\boldsymbol{\Sigma}(z)\boldsymbol{H}(z)\boldsymbol{\Sigma}^{\dagger}(1/z^*) \tag{77}$$

where $\boldsymbol{H}(z)=\boldsymbol{\Sigma}^{-1}(z)\boldsymbol{S}_{\underline{e}}(z)\left(\boldsymbol{\Sigma}^{\dagger}(1/z^*)\right)^{-1}$.

C. Connections to Multivariate Spectral Factorization

Let us revisit the decomposition of mutual information for the general Gaussian channel that allowed us to derive (34). Of primary interest here is the relationship

$$I(\mathcal{E}_i;\mathcal{E}_i^{\text{past}})=I(\mathcal{E}_i;\boldsymbol{\Pi}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i).$$

Suppose that instead of an orthogonal projection we were to use some linear operator $\boldsymbol{\Gamma}_{\mathcal{E}_i^{\text{past}}}$ whose output is expressible as a linear combination of the elements of $\mathcal{E}_i^{\text{past}}$. In general, the data processing theorem of mutual information tells us that

$$I(\mathcal{E}_i;\mathcal{E}_i^{\text{past}})\geq I(\mathcal{E}_i;\boldsymbol{\Gamma}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i) \tag{78}$$

with equality if and only if $\mathcal{E}_i$ and $\mathcal{E}_i^{\text{past}}$ are independent conditional on $\boldsymbol{\Gamma}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i$. If we let $\mathcal{A}=\boldsymbol{\Gamma}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i$, then from (19), the inequality in (78) has the following equivalent representations:

$$\log\frac{\langle\mathcal{E}_i\rangle}{\langle(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{E}_i^{\text{past}}})\mathcal{E}_i\rangle}\geq\log\frac{\langle\mathcal{E}_i\rangle}{\langle(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{A}})\mathcal{E}_i\rangle}$$
$$\langle(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{E}_i^{\text{past}}})\mathcal{E}_i\rangle\leq\langle(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{A}})\mathcal{E}_i\rangle. \tag{79}$$

Finally, since $\boldsymbol{\Pi}_{\mathcal{A}}\mathcal{E}_i$ is a linear combination of elements of $\mathcal{E}_i^{\text{past}}$, then for some linear operator $\tilde{\boldsymbol{\Gamma}}_{\mathcal{E}_i^{\text{past}}}$ that takes inputs onto $\mathcal{E}_i^{\text{past}}$ we have that $\boldsymbol{\Pi}_{\mathcal{A}}\mathcal{E}_i=\tilde{\boldsymbol{\Gamma}}_{\mathcal{E}_i^{\text{past}}}\mathcal{E}_i$. This allows us to conclude that

$$\langle(\boldsymbol{I}-\boldsymbol{\Pi}_{\mathcal{E}_i^{\text{past}}})\mathcal{E}_i\rangle\leq\langle(\boldsymbol{I}-\tilde{\boldsymbol{\Gamma}}_{\mathcal{E}_i^{\text{past}}})\mathcal{E}_i\rangle \tag{80}$$

for *any* linear operator $\tilde{\boldsymbol{\Gamma}}_{\mathcal{E}_i^{\text{past}}}$ that takes inputs onto $\mathcal{E}_i^{\text{past}}$. For multivariate w.s.s. processes, this may be expressed as follows: the orthogonal projection onto $\mathcal{E}_i^{\text{past}}$ minimizes the geometric mean relative to all other linear operators that take their input onto $\mathcal{E}_i^{\text{past}}$. (An example of a suboptimal linear processing is the zero-forcing decision-feedback receiver; see [4] for the ISI channel and [35] for synchronous CDMA.) We now show that this maximization of mutual information corresponds to a spectral factorization.

Recall the ISI channel developed in Section IV-B. In the context of (80), the terms of interest in this case are the determinants of the covariances of $(\boldsymbol{I}-\boldsymbol{\Pi}_{\underline{e}_{-\infty}^{i-1}})\underline{e}_i$ and $(\boldsymbol{I}-\tilde{\boldsymbol{\Gamma}}_{\underline{e}_{-\infty}^{i-1}})\underline{e}_i$. A corresponding statement concerning the optimality of the minimum-phase spectral factorization is the following proposition which follows from [22, Theorems 7.10 and 7.12].

Proposition 3: Let $\boldsymbol{S}_{\underline{e}}(z)=\boldsymbol{\Phi}(z)\boldsymbol{G}\boldsymbol{\Phi}^{\dagger}(1/z^*)$ be the unique multivariate spectral factorization from (45) that was used in determining the canonical decision-feedback receiver for the ISI channel, where $\boldsymbol{\Phi}(z)$ and its inverse are monic (i.e., the zeroth Fourier coefficient is the identity), causal, and stable. For any $\boldsymbol{C}(z)$ that is also monic and causal, we have that $\boldsymbol{S}(z)=\boldsymbol{C}(z)\boldsymbol{S}_{\underline{e}}(z)\boldsymbol{C}^{\dagger}(1/z^*)$ is at least as positive definite as $\boldsymbol{G}$ for all z lying on the unit circle, with equality for all such z occurring if and only if $\boldsymbol{C}(z)=\boldsymbol{\Phi}^{-1}(z)$. ◊

Observe that the arbitrary multivariate filter $\boldsymbol{C}(z)$ in this proposition plays the role of the arbitrary linear operator $\tilde{\boldsymbol{\Gamma}}_{\underline{e}_{-\infty}^{i-1}}$ in the preceding discussion.

Proof: Any $\boldsymbol{C}(z)$ that satisfies the hypotheses given in the statement can, by virtue of properties associated with $\boldsymbol{\Phi}(z)$, be expressed as $\boldsymbol{C}(z)=\boldsymbol{A}(z)\boldsymbol{\Phi}^{-1}(z)$ for some $\boldsymbol{A}(z)$ that satisfies the same hypotheses. From this and the spectral factorization of $\boldsymbol{S}_{\underline{e}}(z)$, it follows immediately that $\boldsymbol{S}(z)=\boldsymbol{A}(z)\boldsymbol{G}\boldsymbol{A}^{\dagger}(1/z^*)$. Evaluating this on the unit circle we find

$$\boldsymbol{S}(e^{j\theta})=\boldsymbol{G}+\sum_{i=1}^{\infty}\boldsymbol{A}(e^{j\theta})\boldsymbol{G}\left(\boldsymbol{A}(e^{j\theta})\right)^{\dagger} \tag{81}$$
$$\preceq\boldsymbol{G} \tag{82}$$

where $\boldsymbol{S}(e^{j\theta})\preceq\boldsymbol{G}$ means that $\boldsymbol{G}-\boldsymbol{S}(e^{j\theta})$ is positive semidefinite. To obtain equality for all θ requires that $\boldsymbol{A}(z)=\boldsymbol{I}$. □

We now state in similar terms the optimality of the structured multivariate spectral factorizations used in Section V-A for the information lossless decision-feedback receivers of the asynchronous CDMA channel. The first of these occurred in our high-level decomposition of mutual information into K single-input single-output ISI channels. The sequences of interest in this case are $(\boldsymbol{I}-\boldsymbol{\Pi}_{\underline{e}_1^{i-1}}(z))\underline{e}_i(z)$ and $(\boldsymbol{I}-\tilde{\boldsymbol{\Gamma}}_{\underline{e}_1^{i-1}}(z))\underline{e}_i(z)$. We have the following result.

Proposition 4: Let $\boldsymbol{S}_{\underline{e}}(z)=\boldsymbol{L}(z)\boldsymbol{D}(z)\boldsymbol{L}^{\dagger}(1/z^*)$ be the structured multivariate spectral factorization from (52), where $\boldsymbol{L}(z)$ and its inverse are lower-triangular with unity-valued diagonal elements, and $\boldsymbol{D}(z)$ is diagonal. For any $\boldsymbol{C}(z)$ that is also lower-triangular with diagonal elements equal to unity, we have that the kth diagonal element of $\boldsymbol{S}(z)=\boldsymbol{C}(z)\boldsymbol{S}_{\underline{e}}(z)\boldsymbol{C}^{\dagger}(1/z^*)$ is at least as large as the kth diagonal element of $\boldsymbol{D}(z)$ for all

z lying on the unit circle. Equality occurs for all such z if and only if $\boldsymbol{C}(z) = \boldsymbol{L}^{-1}(z)$.

Proof: Here we verify the result with an algebraic proof similar to that used in our proof of Proposition 3. Using $\boldsymbol{S}_{\underline{e}}(z) = \boldsymbol{L}(z)\boldsymbol{D}(z)\boldsymbol{L}^{\dagger}(1/z^*)$ as given in (52), we find that $\boldsymbol{S}(z)$ in the proposition statement becomes

$$\boldsymbol{S}(z) = \boldsymbol{C}(z)\boldsymbol{L}(z)\boldsymbol{D}(z)\boldsymbol{L}^{\dagger}(1/z^*)\boldsymbol{C}^{\dagger}(1/z^*). \tag{83}$$

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

$$\begin{aligned} S_{kk}(e^{j\theta}) &= \left(\boldsymbol{A}(e^{j\theta})\boldsymbol{D}(e^{j\theta})\left(\boldsymbol{A}(e^{j\theta})\right)^{\dagger}\right)_{kk} && (84)\\ &= D_{kk}(e^{j\theta}) + \sum_{i=1}^{k-1} |A_{ki}(e^{j\theta})|^2 D_{i,i}(e^{j\theta}) && (85)\\ &\geq D_{kk}(e^{j\theta}), && (86)\end{aligned}$$

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

The second structured spectral factorization in Section V-A has the following property.

Proposition 5: Let $\boldsymbol{S}_{\underline{e}}(z) = \tilde{\boldsymbol{L}}(z)\boldsymbol{G}\tilde{\boldsymbol{L}}^{\dagger}(1/z^*)$ be the structured multivariate spectral factorization in (59), where $\tilde{\boldsymbol{L}}(z)$ and its inverse are lower-triangular with diagonal entries that are monic, causal, and stable, and $\boldsymbol{G}$ is diagonal. For any $\boldsymbol{C}(z)$ that is lower-triangular with diagonal elements that are monic and causal, the kth diagonal element of

$$\boldsymbol{S}(z) = \boldsymbol{C}(z)\boldsymbol{S}_{\underline{e}}(z)\boldsymbol{C}^{\dagger}(1/z^*)$$

for all z on the unit circle is no less than g_k, the kth diagonal element of $\boldsymbol{G}$. Equality occurs for all such z if and only if $\boldsymbol{C}(z) = \tilde{\boldsymbol{L}}^{-1}(z)$.

Proof: To show this algebraically, consider

$$\boldsymbol{S}(z) = \boldsymbol{C}(z)\tilde{\boldsymbol{L}}(z)\boldsymbol{G}\tilde{\boldsymbol{L}}^{\dagger}(1/z^*)\boldsymbol{C}^{\dagger}(1/z^*). \tag{87}$$

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

$$\begin{aligned} S_{kk}(e^{j\theta}) &= \left(\boldsymbol{A}(e^{j\theta})\boldsymbol{G}\boldsymbol{A}^{\dagger}(e^{j\theta})\right)_{kk}\\ &= |A_{kk}(e^{j\theta})|^2 g_k + \sum_{i=1}^{k-1} |A_{ki}(e^{j\theta})|^2 g_i\\ &\geq |A_{kk}(e^{j\theta})|^2 g_k\\ &\geq g_k. && (88)\end{aligned}$$

Equality in the penultimate step requires $A_{ki}(z) = 0$ for all $i \neq k$, while equality in the final step (which holds because $A_{kk}(z)$ is monic and causal) requires that $A_{kk}(z) = 1$. □

Similarly, the structured spectral factorizations that we used to derive the lossy canonical decision-feedback receivers in Section V-B also satisfy optimality properties. These are now summarized.

Proposition 6: Let $\boldsymbol{S}_{\underline{e}}(z) = \boldsymbol{\Psi}(z)\boldsymbol{F}(z)\boldsymbol{\Psi}^{\dagger}(1/z^*)$ be the structured multivariate spectral factorization given in (66), where $\boldsymbol{\Psi}(z)$ and its inverse are causal, stable, and lower-triangular with diagonal entries that are monic. For any $\boldsymbol{C}(z)$ that is causal, stable, and lower-triangular with diagonal elements that are monic, the kth diagonal element of

$$\boldsymbol{S}(z) = \boldsymbol{C}(z)\boldsymbol{S}_{\underline{e}}(z)\boldsymbol{C}^{\dagger}(1/z^*)$$

is no less than the kth diagonal element of $\boldsymbol{F}(z)$ for all z on the unit circle. Equality occurs for all such z if and only if $\boldsymbol{C}(z) = \boldsymbol{\Psi}^{-1}(z)$. ◇

Proof: Denote by $\underline{c}^{\mathsf{T}}(z)$ the first k elements of the kth row of $\boldsymbol{C}(z)$ (recall that elements $k+1$ through K of this row are zero). All elements of $\underline{c}^{\mathsf{T}}(z)$ are causal and stable with the kth element also monic. From (62), we know that there always exists a vector $\underline{a}^{\mathsf{T}}(z)$ that possesses these same properties and is related to $\underline{c}^{\mathsf{T}}(z)$ according to $\underline{c}^{\mathsf{T}}(z) = \underline{a}^{\mathsf{T}}(z)\boldsymbol{\Phi}_{(k)}^{-1}(z)$. With this substitution and (62), we find that the kth diagonal element of $\boldsymbol{S}(z)$ can be expressed as

$$S_{kk}(z) = \underline{a}^{\mathsf{T}}(z)\boldsymbol{G}_{(k)}\left(\underline{a}(1/z^*)\right)^* \tag{89}$$

where $\boldsymbol{G}_{(k)} = \operatorname{diag}(g_{k1}, \ldots, g_{kk})$ is a diagonal matrix. On the unit circle, then, we have

$$\begin{aligned} S_{kk}(e^{j\theta}) &= \sum_{i=1}^{k-1} |a_i(e^{j\theta})|^2 g_{ki} + |a_k(e^{j\theta})|^2 g_{kk} && (90)\\ &\geq |a_k(e^{j\theta})|^2 g_{kk} && (91)\\ &\geq g_{kk}. && (92)\end{aligned}$$

The fact that the kth element of $\underline{a}^{\mathsf{T}}(z)$ is monic was used to obtain the final inequality. Clearly, equality occurs only when $\underline{a}^{\mathsf{T}}(z)$ is zero for every entry but the kth entry which is unity. From the discussion following (66) we recall that g_{kk} is equal to the kth diagonal element of $\boldsymbol{F}(z)$. Since the above reasoning holds for every row of $\boldsymbol{C}(z)$, we have confirmed the proposition. □

Proposition 7: Let $\boldsymbol{S}_{\underline{e}}(z) = \boldsymbol{\Sigma}(z)\boldsymbol{H}(z)\boldsymbol{\Sigma}^{\dagger}(1/z^*)$ be the structured multivariate spectral factorization given in (77), where $\boldsymbol{\Sigma}^{-1}(z)$ is lower-triangular, causal and FIR of order N_x. For any $\boldsymbol{C}(z)$ that is lower-triangular, causal and FIR of order N_x, the kth diagonal element of

$$\boldsymbol{S}(z) = \boldsymbol{C}(z)\boldsymbol{S}_{\underline{e}}(z)\boldsymbol{C}^{\dagger}(1/z^*)$$

is no less than the kth diagonal element of $\boldsymbol{H}(z)$ for all z on the unit circle. Equality occurs for all such z if and only if $\boldsymbol{C}(z) = \boldsymbol{\Sigma}^{-1}(z)$. ◇

Proof: This result is shown algebraically in a manner very similar to the proof of Proposition 6 except that one works with the block-vector and block-matrix representations of Section V-B2 rather than working directly in the z-domain. □

D. Evaluating Structured Spectral Factorizations

There are efficient methods for numerically evaluating the unstructured multivariate spectral factorization used in (45) when the spectrum is rational. The quadratically converging Newton map given in [36] is implemented with a fast algorithm in [37]. For the structured factorizations that we have encountered, evaluation by similar techniques becomes too cumbersome to be useful in practice. There is, however, another approach known as Bauer's method for evaluating spectral factorizations that was developed in [38] for the case of multivariate spectral factorizations. This technique calculates the spectral factorization in (45) to arbitrary accuracy by performing the Cholesky factorization of a large enough finite-dimensional matrix. An application of this idea to numerically determine structured multivariate spectral factorizations is given in the Appendix.

VI. Conclusion

We have derived an information-lossless decision-feedback receiver structure that applies to a general class of Gaussian channels by starting from mutual information. The underlying building block for the resulting canonical decision-feedback receiver is a particular additive decomposition of mutual information. The receiver effects this decomposition by performing information-lossless orthogonal projections. These projections correspond to Wiener filtering and prediction, so the net result is a receiver that takes advantage of an important bridge between mutual information, optimal filtering, and prediction.

From the generality of the result, the information-preserving property of known canonical decision-feedback receivers for a variety Gaussian channels may be inferred. It also enabled us to derive a number of information-lossless decision-feedback receivers for use in asynchronous CDMA channels. A byproduct of this endeavor was the discovery of information-theoretic derivations of a variety of structured decompositions of multivariate spectra. Given that the canonical decision-feedback receiver employs Wiener prediction, from which originated the concept of spectral factorization, the intimate connection between mutual information and spectral factorization is not surprising.

Appendix
Numerical Evaluation of Structured Spectral Factorizations

We first review the Bauer technique for evaluating an unstructured multivariate spectral factorization as developed in [38]. Let $\boldsymbol{S}(z) = \sum_{i=-\infty}^{\infty} \boldsymbol{R}_i z^{-i}$ be a multivariate spectrum of dimensions $K \times K$ for which we would like to find the decomposition $\boldsymbol{S}(z) = \boldsymbol{\Phi}(z)\boldsymbol{G}\boldsymbol{\Phi}^{\dagger}(1/z^*)$ where $\boldsymbol{\Phi}(z)$ and its inverse are monic, causal, and stable, and the matrix $\boldsymbol{G}$ is a constant. This particular factorization was the subject of Proposition 3 in Section V-C. Start by creating the $KP \times KP$ block-Toeplitz matrix

$$\mathbb{R} = \begin{bmatrix} \boldsymbol{R}_0 & \boldsymbol{R}_{-1} & \cdots & \boldsymbol{R}_{-P+1} \\ \boldsymbol{R}_1 & \ddots & & \\ \vdots & & & \\ \boldsymbol{R}_{P-1} & & & \end{bmatrix}. \tag{93}$$

On this matrix, perform the Cholesky decomposition $\mathbb{R} = \mathbb{L}\mathbb{G}\mathbb{L}^{\dagger}$ where block-matrix $\mathbb{L}$ (the blocks are still $K \times K$) is lower-triangular with diagonal blocks equal to the identity and $\mathbb{G}$ is a block-diagonal matrix. Now use the last block-row of $\mathbb{L}$ to create the filter

$$\boldsymbol{\Phi}_P(z) = \sum_{i=0}^{P-1} \mathbb{L}_{P,P-i} z^{-i} \tag{94}$$

where $\mathbb{L}_{i,j}$ is the (i,j) block of $\mathbb{L}$. Also, let $\boldsymbol{G}_P = \mathbb{G}_{P,P}$, the last diagonal block of $\mathbb{G}$. As the the size of $\mathbb{R}$ grows by increasing the number of blocks P, then $\boldsymbol{\Phi}_P(z) \to \boldsymbol{\Phi}(z)$ and $\boldsymbol{G}_P \to \boldsymbol{G}$.

We now show how this idea can be used to determine the structured multivariate spectral factorization of Proposition 6. Recall that $\boldsymbol{S}(z) = \boldsymbol{\Psi}(z)\boldsymbol{F}(z)\boldsymbol{\Psi}^{\dagger}(1/z^*)$, where $\boldsymbol{\Psi}(z)$ and its inverse are causal, stable, and lower-triangular with diagonal entries that are monic. For $k = 1, \ldots, K$, we let $\mathbb{R}_{(k)}$ denote the matrix formed by replacing each block of $\mathbb{R}$ in (93) with its $k \times k$ principal submatrix formed by the first k indices. This means that $\mathbb{R}_{(k)}$ is a $P \times P$ block-matrix with blocks of dimensions $k \times k$. A Cholesky decomposition yields $\mathbb{R}_{(k)} = \boldsymbol{L}_{(k)}\boldsymbol{D}_{(k)}\boldsymbol{L}_{(k)}^{\dagger}$ where $\boldsymbol{L}_{(k)}$ is lower-triangular with unity-valued diagonal elements. We now take every kth row of $\boldsymbol{L}_{(k)}$ (i.e., rows $k, 2k, \ldots, Pk$) and use them to create rows $k, k+K, \ldots, k+(P-1)K$ of a $KP \times KP$ matrix $\tilde{\mathbb{L}}$. This is done by taking every kth row of $\boldsymbol{L}_{(k)}$, parsing it into words of k entries, and then inserting $K - k$ zeros between parsed words. The first of these expanded rows is the kth row of $\tilde{\mathbb{L}}$, the second is the $(k+K)$th row of $\tilde{\mathbb{L}}$, and so on, until the last of these is row $k+(P-1)K$ of $\tilde{\mathbb{L}}$. The construction of $\tilde{\mathbb{L}}$ is complete after we have done this for all k. Note that if we now view $\tilde{\mathbb{L}}$ as a block matrix with $K \times K$ blocks and $\tilde{\mathbb{L}}_{i,j}$ denoting its (i,j) block, then it has the following properties:

- $\tilde{\mathbb{L}}_{i,j}$ is lower-triangular for all i, j to satisfy that $\boldsymbol{\Psi}(z)$ is lower-triangular;
- $\tilde{\mathbb{L}}_{i,j}$ is the zero matrix whenever $i < j$ to satisfy that $\boldsymbol{\Psi}(z)$ is causal;
- $\tilde{\mathbb{L}}_{i,i}$ has diagonal elements that are each unity for all i to satisfy that the diagonal elements of $\boldsymbol{\Psi}(z)$ are monic.

We then have the factorization $\mathbb{R} = \tilde{\mathbb{L}}\mathbb{F}\tilde{\mathbb{L}}^{\dagger}$ where $\mathbb{F} = \tilde{\mathbb{L}}^{-1}\mathbb{R}(\tilde{\mathbb{L}}^{\dagger})^{-1}$. The last block-row of $\tilde{\mathbb{L}}$ ($K \times K$ blocks) is used to create the filter

$$\boldsymbol{\Psi}_P(z) = \sum_{i=0}^{P-1} \tilde{\mathbb{L}}_{P,P-i} z^{-i}. \tag{95}$$

Also, let

$$\boldsymbol{F}_P(z) = \mathbb{F}_{P,P} + \sum_{i=1}^{P-1} \left(\mathbb{F}_{P,P-i} z^{-i} + \mathbb{F}_{P,P-i}^{\dagger} z^{i} \right). \tag{96}$$

As the the size of $\mathbb{R}$ grows by increasing the number of blocks P, then $\boldsymbol{\Psi}_P(z) \to \boldsymbol{\Psi}(z)$ and $\boldsymbol{F}_P(z) \to \boldsymbol{F}(z)$.

Finally, we point out that the factorizations of Propositions 4 and 5 can be handled similarly by enforcing the appropriate properties when determining $\tilde{\mathbb{L}}$. For Proposition 4 we need

- $\tilde{\mathbb{L}}_{i,j}$ is lower-triangular for all i, j to satisfy that $\boldsymbol{L}(z)$ is lower-triangular;
- $\tilde{\mathbb{L}}_{i,i}$ is the identity matrix for all i and the diagonal elements of $\tilde{\mathbb{L}}_{i,j}$ are all zero whenever $i \neq j$ to satisfy that the diagonal elements of $\boldsymbol{L}(z)$ are all unity.

For Proposition 5 we need

- $\tilde{\mathbb{L}}_{i,j}$ is lower-triangular for all i, j to satisfy that $\tilde{\boldsymbol{L}}(z)$ is lower-triangular;
- $\tilde{\mathbb{L}}_{i,i}$ has diagonal elements that are each unity for all i to satisfy that the diagonal elements of $\tilde{\boldsymbol{L}}(z)$ are all monic;
- $\tilde{\mathbb{L}}_{i,j}$ has all diagonal elements equal to zero whenever $i < j$ to satisfy that the diagonal elements of $\tilde{\boldsymbol{L}}(z)$ are all causal.

ACKNOWLEDGMENT

The authors would like to thank the referees for their helpful comments which have improved the presentation of this paper.

REFERENCES

[1] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., “MMSE decision-feedback equalizers and coding- Part I: Equalization results,” *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[2] ——, “MMSE decision-feedback equalizers and coding- Part II: Coding results,” *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.

[3] S. Shamai (Shitz) and R. Laroia, “The intersymbol interference channel: Lower bounds on capacity and channel precoding loss,” *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1388–1404, Sep. 1996.

[4] J. R. Barry, E. Lee, and D. G. Messerschmitt, “Capacity penalty due to ideal zero-forcing decision-feedback equalization,” *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1062–1071, Jul. 1996.

[5] E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Norwell, MA: Kluwer Academic, 1994.

[6] J. Yang and S. Roy, “Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback,” *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1334–1347, Sep. 1994.

[7] J. M. Cioffi and G. D. Forney, Jr., “Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise,” in *Communications, Computation, Control and Signal Processing, a Tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury, and C. D. Schaper, Eds. Boston, London, Dordrecht: Kluwer Academic, 1997.

[8] M. K. Varanasi and T. Guess, “Achieving vertices of the capacity region of the synchronous Gaussian correlated-waveform multiple-access channel with decision-feedback receivers,” in *Proc. IEEE Int. Symp. Information Theory*, Ulm, Germany, Jul. 1997, p. 270.

[9] ——, “Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel,” in *Proc. 31st Asilomar Conf. Signals, Systems, and Computers*, Monticello, IL, Nov. 1997, pp. 1405–1409.

[10] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, “Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture,” *Electron. Lett.*, vol. 35, pp. 14–16, Jan. 1999.

[11] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela, “Analysis and performance of some basic space-time architectures,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 303–320, Apr. 2003.

[12] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Tech, J.*, vol. 1, pp. 41–59, 1996.

[13] L. Zheng and D. N. C. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[14] T. Guess, H. Zhang, and T. V. Kotchiev, “The outage capacity of blast in mimo channels,” in *Proc. 2003 (CDROM) IEEE Int. Conf. Communications (ICC 2003): Communication Theory Symp.*, Anchorage, AK, May 11–15, 2003, paper CT07–2.

[15] T. Guess and H. Zhang, “Asymptotical analysis of the outage capacity of rate-tailored blast,” in *Proc. 2003 (CDROM) IEEE Global Communications Conf. (GLOBECOM 2003): Communication Theory Symp.*, San Francisco, CA, Dec. 1–5, 2003, paper CT7–4.

[16] N. Prasad and M. K. Varanasi, “Outage capacities of space-time architectures,” in *Proc. IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004.

[17] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day, 1964.

[18] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[19] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. New York: Wiley, 1984.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

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

[22] N. Wiener and P. Masani, “The prediction theory of multivariate processes, Part I: The regularity condition,” *Acta Math.*, vol. 98, pp. 111–150, 1957.

[23] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[24] F. D. Neeser and J. L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[25] S. Verdú, “Capacity region of Gaussian CDMA channels: The symbol-synchronous case,” in *Proc. 24th Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 1986, pp. 1025–1034.

[26] A. Duel-Hallen, “Decorrelating decision-feedback multiuser detector for synchronous code-division multiple access channel,” *IEEE Trans. Commun.*, vol. 41, no. 2, pp. 285–290, Feb. 1993.

[27] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Second ed. New York: McGraw-Hill, 1984.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Melbourne, Australia: Cambridge Univ. Press, 1993.

[29] G. D. 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.

[30] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill Book Co., 2000.

[31] L. H. Brandenburg and A. D. Wyner, “Capacity of the Gaussian channel with memory: The multivariate case,” *Bell Syst. Tech. J.*, vol. 53, pp. 745–778, May-Jun. 1974.

[32] S. Verdú, “The capacity region of the symbol-asynchronous Gaussian multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 733–751, Jul. 1989.

[33] A. Duel-Hallen, “A family of multiuser decision-feedback detectors for asynchronous code-division multiple access channels,” *IEEE Trans. Commun.*, no. 2–4, pp. 421–434, Feb.–Apr. 1995.

[34] T. Guess and M. K. Varanasi, “Deriving optimal successive decoders for the asynchronous CDMA channel using information theory,” in *Proc. 2000 Conf. Information Sciences and Systems (CISS’2000)*, Princeton, NJ, Mar. 15–17, 2000, pp. WP6–11.

[35] ——, “Multiuser decision-feedback receivers for the general Gaussian multiple-access channel,” in *Proc. 34th Allerton Conf. Communication, Control and Computing*, Allerton, IL, Oct. 1996, pp. 190–199.

[36] G. T. Wilson, “The factorization of matricial spectral densities,” *SIAM J. Appl. Math.*, vol. 23, pp. 420–426, Dec. 1972.

[37] J. Jezek and V. Kucera, “Efficient algorithm for matrix spectral factorization,” *Automatica*, vol. 21, no. 6, pp. 663–669, 1985.

[38] 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. Circuits Syst.*, vol. CAS–25, no. 2, pp. 57–69, Feb. 1978.