Adaptable MIMO Channel Decomposition Using Channel State Information

Yi Jiang William W. Hager † Jian Li*

Abstract

Assuming the availability of the channel state information at the transmitter (CSIT) and receiver (CSIR), we consider the joint optimal transceiver design for multi-input multi-output (MIMO) communication systems. Using the recently developed generalized triangular decomposition (GTD), we propose a scheme which we refer to as adaptable channel decomposition (ACD) to decompose a MIMO channel, in a capacity lossless manner, into multiple subchannels with prescribed capacities, or equivalently, signal-to-interference-and-noise ratios (SINR). This scheme is particularly relevant to the applications where independent data streams with different qualities-of-service (QoS) share the same MIMO channel. The ACD scheme has two implementation forms. One is the combination of a linear precoder and a minimum mean-squared-error VBLAST (MMSE-VBLAST) detector, which is referred to as ACD-VBLAST, and the other includes a dirty paper (DP) precoder and a linear equalizer followed by a DP decoder, which we refer to as ACD-DP. Both forms of ACD are computationally very efficient. We also identify two seemingly distinct communications problems, the precoder design for orthogonal frequency division multiplexing (OFDM) communications and the optimal code division multiple access (CDMA) sequence design, as special cases in the unifying framework of MIMO transceiver designs. The ACD scheme can be applied to solve a considerablely wide range of problems.

Keywords

MIMO, channel capacity, joint transceiver design, adaptable channel decomposition, water filling, generalized triangular decomposition, dirty paper precoding, VBLAST, optimal CDMA sequences, quality-of-service.

This work was supported in part by the National Science Foundation Grant CCR-0104887.

Yi Jiang and Jian Li are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130, USA. Email: {yjiang, li}@dsp.ufl.edu

† William. W. Hager is with the Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Email: hager@math.ufl.edu

*Corresponding author: Dr. Jian Li, Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130, Gainesville, FL 32611-6130, USA. Email: li@dsp.ufl.edu, Phone: 352-392-2642, Fax: 352-392-0044.

I. Introduction

For a multi-input multi-output (MIMO) communication system, if the communication environment is slowly time varying, the channel state information at transmitter (CSIT) is possible via feedback or the reciprocal principle when time division duplex (TDD) is used. In the past several years, considerable research has been devoted to the joint transceiver design utilizing CSIT and channel state information at receiver (CSIR) (see, e.g., [1][2][3][4][5] and the references therein). The goal of joint transceiver design is to improve the MIMO system performance, including the channel throughput and bit-error-rate (BER) performance, through optimizing the precoder and equalizer jointly. Two classes of MIMO transceiver designs have been proposed, including linear transceiver designs [1][2][3] and nonlinear schemes [4][5]. All the MIMO linear transceiver designs start with applying the singular value decomposition (SVD) to the MIMO channel matrix. Due to the usually widely spread out singular values of the channel matrix, the SVD decomposes a MIMO channel into multiple parallel eign-subchannels with very different channel gains. To maximize the overall channel capacity, the transceiver designs allocate different power to the eign-subchannels using the "water filling" algorithm [1]. The "water filling" algorithm essentially loads more power onto the subchannels with larger gains and less otherwise. Hence the channel qualities, i.e., signal-to-noise ratios (SNR), become even more divergent. On the other hand, to make the modulation and coding procedures easier, it is desirable to apply the same constellation across all the subchannels as is adopted by the current standards such as the European standard HIPERLAN/2 and the IEEE 802.11 standards for wireless local area networks (WLANs). In this case, the overall BER performance is dominated by the BER of the worst subchannels. Thus one may want to obtain subchannels with similar (or even identical) SNRs. To achieve this goal, the linear transceivers load more power to the worse subchannels and less power to the better ones [2][3], which leads to considerable capacity loss as analyzed in [4].

In [5], we propose a uniform channel decomposition (UCD) scheme, which is based on either a VBLAST detector or a dirty paper (DP) precoder, to decompose a MIMO channel into several *identical* subchannels. Our UCD scheme is strictly capacity lossless. That is, the sum capacity of the subchannels are equal to the capacity of the MIMO channel obtained via SVD plus the "water filling" algorithm. The UCD scheme can significantly outperform its linear counterparts in terms of both BER performance and channel throughput.

All these aforementioned MIMO transceiver designs focus on improving the communication quality subject to power constraints. In this paper, we tackle a new aspect of the MIMO transceiver design problem. We regard a MIMO transceiver design as a way of decomposing a MIMO channel into multiple subchannels. As we have mentioned, the MIMO channel decomposition through SVD plus "water filling" lacks flexibility despite its optimality in terms of achieving the maximal overall channel capacity. The success of UCD motivates a much more flexible channel decomposition approach, namely the adaptable channel decomposition (ACD) scheme, which is the main result of this paper. Using the recently developed generalized triangular decomposition (GTD), we propose the ACD scheme to decompose a MIMO channel into multiple subchannels with prescribed capacities, or equivalently, signal-to-interference-and-noise ratios (SINR). The main properties of the ACD scheme are summarized as follows:

1. Given K parallel subchannels with capacities $C_1, C_2, \ldots, C_K$, which is obtained through applying SVD plus "water filling" to a rank K MIMO channel, ACD can convert the K subchannels into $L \geq K$ subchannels with capacities $R_1, R_2, \ldots, R_L$ if and only if $(C_1, \ldots, C_K, 0, \ldots, 0) \in \mathbb{R}_+^L$ majorizes $(R_1, R_2, \ldots, R_L)$ [1]. In particular, $\sum_{i=1}^K C_i = \sum_{i=1}^L R_i$, i.e., the ACD is capacity lossless.
2. The ACD scheme has two implementation forms. One is the combination of a linear precoder and a minimum mean-squared-error VBLAST (MMSE-VBLAST) detector, which is referred to

[1] The concept of majorization is introduced in Section II.

as ACD-VBLAST, and the other includes a DP precoder and a linear equalizer followed by a DP decoder, which we refer to as ACD-DP.
3. Given the SVD of the MIMO channel matrix, the computational complexity of ACD, which is to calculate the precoder and equalizer matrices, is $O(KL)$, which is computationally quite efficient.

Almost originated at the same time as the research on MIMO transceiver designs, the optimal design of symbol synchronous CDMA (S-CDMA) sequences has been under intensive study over the past decade (see, e.g., [6][7][8][9]). Although the two research topics have been studied in an apparently independent manner in the signal processing and information theory communities, we show that the CDMA sequence design problem can be viewed as a special case of the MIMO transceiver design. Hence the ACD scheme can be applied, with little modifications, to the design of optimal CDMA sequences. Moreover, the ACD-VBLAST and ACD-DP schemes can be applied to design optimal CDMA sequences in the uplink (mobile-to-base) and downlink (base-to-mobile) scenarios, respectively. Our ACD scheme, which is independently motivated by the MIMO transceiver design problem, turns out to be related to the scheme proposed in [9]. The relationship is discussed in Section IV.

The remainder of the paper is organized as follows. Section II introduces the MIMO flat fading channel model and identifies the problems of precoded orthogonal frequency division multiplexing (OFDM) communications and optimal CDMA sequence designs as two special cases in the unifying framework of MIMO transceiver designs. Several relevant results on channel capacity are also briefly reviewed there. Section III introduces the concept of majorization, the GTD theorem, and the closed-form representation of the MMSE-VBLAST detector. The ACD scheme is presented in Section IV. We discuss the application of the ACD scheme to the problem of multi-task MIMO communications with QoS constraints, which was originally studied in [10], in Section V. In Section VI, we apply the ACD scheme to the design of optimal CDMA sequences. We also present a comparative study of ACD and the scheme in [9]. Section VII gives the conclusions of this paper.

II. Channel Model, Capacity and Decomposition

A. Channel Model

We consider a MIMO communication system with M_t transmitting and M_r receiving antennas in a frequency flat fading channel. The sampled baseband signal is given by

$$\mathbf{y} = \mathbf{HFx} + \mathbf{z}, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{L\times 1}$ is the information symbols precoded by the linear precoder $\mathbf{F} \in \mathbb{C}^{M_t\times L}$ and $\mathbf{y} \in \mathbb{C}^{M_r\times 1}$ is the received signal and $\mathbf{H} \in \mathbb{C}^{M_r\times M_t}$ is the channel matrix with rank K and with its (i,j)th element denoting the fading coefficient between the jth transmitting and ith receiving antennas. We assume $E[\mathbf{xx}^*] = \sigma_x^2\mathbf{I}_L$, where $E[\cdot]$ is the expected value and $\mathbf{I}_L$ denotes the identity matrix with dimension L and $\mathbf{z} \sim N(0, \sigma_z^2\mathbf{I}_{M_r})$ is the circularly symmetric complex Gaussian noise. We define the SNR as

$$\rho = \frac{E[\mathbf{x}^*\mathbf{F}^*\mathbf{Fx}]}{\sigma_z^2} = \frac{\sigma_x^2}{\sigma_z^2}\mathrm{Tr}\{\mathbf{F}^*\mathbf{F}\} \triangleq \frac{1}{\alpha}\mathrm{Tr}\{\mathbf{F}^*\mathbf{F}\}, \tag{2}$$

where $(\cdot)^*$ is the conjugate transpose, and $\mathrm{Tr}\{\cdot\}$ stands for the trace of a matrix.

We note that the data model in (1) is generic. For an intersymbol-interference (ISI) channel with impulse response $\mathbf{h} = [h_0, h_1, \ldots, h_{M-1}]^\mathsf{T}$, if a block data with length N is transmitted using the "zero-padded" OFDM, then the received block data can also be written in the form of (1)

with

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & 0 & \dots & 0 \\ \vdots & h_0 & 0 & \dots & 0 \\ h_{M-1} & \dots & \ddots & \dots & \vdots \\ 0 & \ddots & \dots & \ddots & 0 \\ \vdots & 0 & h_{M-1} & \dots & h_0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & h_{M-1} \end{bmatrix}. \tag{3}$$

In this case, $\mathbf{H}$ is a Toeplitz matrix with its dimensionality $M_t = N$ and $M_r = N + M - 1$. If the OFDM with cyclic prefix is used, the channel matrix is a circulant Toeplitz matrix, i.e.,

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & h_{M-1} & \dots & \dots & h_1 \\ h_1 & h_0 & 0 & h_{M-1} & \ddots & h_2 \\ \vdots & \ddots & \ddots & 0 & \ddots & \vdots \\ h_{M-2} & \dots & \ddots & \ddots & \ddots & h_{M-1} \\ h_{M-1} & h_{M-2} & \dots & h_1 & h_0 & 0 \\ 0 & h_{M-1} & h_{M-2} & \dots & h_1 & h_0 \end{bmatrix}. \tag{4}$$

Here, $M_t = M_r = N$. In either case, if the block data is precoded with the linear precoder $\mathbf{F}$, then the received data is given in (1). This ISI channel problem has been studied in [11].

In an idealized synchronous CDMA (S-CDMA) system where the channel does not experience any fading or near-far effect, L mobile users modulate their information symbols via spreading sequences $\{\mathbf{s}_i\}_{i=1}^{L}$, each of which has the processing gain N. The discrete-time baseband S-CDMA signal received at the (single-antenna) base-station can be represented as [6]

$$\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{z} \tag{5}$$

where $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_L] \in \mathbb{R}^{N \times L}$ and the lth $(1 \le l \le L)$ entry of $\mathbf{x}$, x_l, stands for the information symbol from the lth user. In the downlink channel, the base station multiplexes the information dedicated to the L mobile users through the spreading sequences, which are the columns of $\mathbf{S}$. Then all the mobiles receive the same signal given in (5). We remark that (5) can also be written as (1) with $\mathbf{H} = \mathbf{I}_N$ and $\mathbf{F} = \mathbf{S}$. Here $M_r = M_t = N$ is the processing gain. Hence, optimizing the spreading sequences amounts to optimizing the precoder $\mathbf{F}$ for a MIMO system. Indeed, this problem has been under intensive research in the past several years.

In summary, both designing a precoder for OFDM transmission through an ISI channel and searching for the optimal S-CDMA sequences can be regarded as special cases in the unifying framework of MIMO transceiver designs. MIMO transceiver designs can be used in the OFMD and CDMA applications after only simple modifications. In this paper, we will concentrate on MIMO transceiver design although we will discuss the optimal design of CDMA sequences later. Throughout this paper, we assume perfect CSIR and CSIT, i.e., $\mathbf{H}$ is known exactly at both the transmitter and receiver.

B. Channel Capacity and Decomposition

Denote the SVD of a rank K channel $\mathbf{H}$ as $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$, where $\mathbf{\Lambda}$ is a $K \times K$ diagonal matrix whose diagonal elements $\{\lambda_{H,k}\}_{k=1}^{K}$ are the nonzero singular values of $\mathbf{H}$. To maximize the channel capacity with respect to $\mathbf{F}$ given the input power constraint $\mathrm{Tr}\{\mathbf{F}\mathbf{F}^*\} \le \rho\sigma_z^2/\sigma_x^2$, one needs to

solve

$$C_{IT} = \max_{\mathrm{Tr}\{\mathbf{F}\mathbf{F}^*\} \le \rho\sigma_z^2/\sigma_x^2} \log_2 |\mathbf{I} + \alpha^{-1}\mathbf{H}\mathbf{F}\mathbf{F}^*\mathbf{H}^*|. \tag{6}$$

The optimal linear precoder is [12]

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}. \tag{7}$$

Here $L = K$ and $\mathbf{\Phi}$ is diagonal whose kth ($1 \le k \le K$) diagonal element ϕ_k determines the power loaded to the kth subchannel and is found via "water filling" to be

$$\phi_k(\mu) = \left(\mu - \frac{\alpha}{\lambda_{H,k}^2}\right)^+, \tag{8}$$

with μ being chosen such that $\sigma_x^2 \sum_{k=1}^K \phi_k(\mu) = \rho\sigma_z^2$ and $(a)^+ = \max\{0, a\}$. In this case, we obtain K subchannels with capacities

$$C_k = \log_2\left(1 + \frac{\phi_k}{\alpha}\lambda_{H,k}^2\right) = \left[\log_2\left(\frac{\mu\lambda_{H,k}^2}{\alpha}\right)\right]^+ \text{ bit/s/Hz}, \quad k = 1, 2, \ldots, K. \tag{9}$$

Due to the usually large dynamic range of singular values $\{\lambda_{H,k}\}_{k=1}^K$, the SVD decomposes a MIMO channel into multiple parallel eign-subchannels with different channel capacities. Moreover, since the optimal power loading levels are fixed as given in (8), the achievable MIMO channel decomposition is rigidly given in (9) and it lacks flexibility.

Another way of decomposing a MIMO channel is to use the VBLAST detector [13]. For a MIMO system of (1) with $M_t \le M_r$ and rank $K = M_t$, the transmitter allocates independent bit streams across the M_t transmitting antennas. To decode the transmitted information symbol, VBLAST first estimates the signal with the spatial structure $\mathbf{h}_1$, where $\mathbf{h}_i$ denotes the ith column of $\mathbf{H}$, and then cancels it out of the received signal vector. Next, it estimates the signal with spatial structure $\mathbf{h}_2$ and so on. The signal estimator can be either a ZF or an MMSE estimator. The VBLAST scheme involves *sequential nulling and cancellation* and it decomposes the MIMO channel into K subchannels (or layers as coined in [13]). By changing the ordering of the signal detection, we can get $K!$ subchannel combinations, each of which is capacity lossless [14].

Theoretically, more combinations of subchannels is possible via time sharing (see [15, Ch. 14.3]). Recall that every DBLAST layer sends its data substream across the K transmitting antennas, or VBLAST layers, in a time sharing manner [16]. For example, for a system with $M_t = 2$, the transmitted data is

$$\begin{array}{lccccc} \text{Vertical Layer-I}: & x_1 & y_2 & x_3 & y_4 & \ldots \\ \text{Vertical Layer-II}: & 0 & x_2 & y_3 & x_4 & \ldots \end{array} \tag{10}$$

Let x_i and $y_i, i = 1, 2, \ldots,$ denote the symbols transmitted through the DBLAST layers I and II, respectively, at time i. The receiver first estimates x_1 and then estimates x_2 by regarding y_2 as interference. The estimates of x_1, x_2 are decoded jointly, which form the output of the diagonal layer I. After subtracting out the effect of x_1, x_2 from the received data, we can estimate and decode y_2, y_3, which form the diagonal layer II. We remark that DBLAST can be viewed as a combination of VBLAST and the time sharing technique, which decomposes the MIMO channel into multiple *identical* subchannels.

However, time sharing can be difficult to implement in practice. For instance, the major difficulty of DBLAST is the requirement of encoding the diagonal layer with short and efficient error correction codes, which limits its practical implementation despite its superb theoretical performance analyzed in [17].

If CSIT is available, more flexible and practical channel decompositions can be achieved. In [5], we propose the UCD scheme which combines the geometric mean decomposition (GMD) developed in [18] with either an MMSE-VBLAST detector or a DP precoder to decompose the MIMO channel of (1) into $L \geq K$ *identical* subchannels. Hence, the UCD scheme can achieve the theoretical performance of the DBLAST scheme without resorting to any error correcting coding.

In this paper, we generalize the results of [5] and develop a systematic channel decomposition that combines the recently proposed GTD algorithm with either an MMSE-VBLAST decoder or a DP precoder. We show that given K parallel subchannels with capacities $C_1, C_2, \ldots, C_K$, which are obtained via SVD, ACD can convert the K subchannels into $L \geq K$ subchannels[2] with capacities $R_1, R_2, \ldots, R_L$ if and only if $(R_1, R_2, \ldots, R_L)$ is majorized by $(C_1, \ldots, C_K, 0, \ldots, 0) \in \mathbb{R}^L$. This scheme is particularly relevant to the applications where independent data streams with different qualities-of-service (QoS) share the same MIMO channel [10]. For example, video services usually require higher SNRs than audio services. Decomposing a MIMO channel into multiple subchannels with prescribed capacities and transmitting independent data streams through these subchannels can provide much convenience for resource allocations.

III. Preliminaries

A. Majorization

We introduce several basic concepts and theorems of the majorization theory from [19].

Definition 1: For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, if

$$\sum_{i=1}^{j} x_{[i]} \leq \sum_{i=1}^{j} y_{[i]}, \quad 1 \leq j \leq n \tag{11}$$

with equality holds for $j = n$, where the subscript $[i]$ denotes the ith largest element of the sequence, we say that $\mathbf{x}$ is majorized by $\mathbf{y}$ and denote $\mathbf{x} \prec_+ \mathbf{y}$, or equivalently, $\mathbf{y} \succ_+ \mathbf{x}$.

Definition 2: An $n \times n$ matrix $\mathbf{P}$ is doubly stochastic if its (i,j)th entry $p_{ij} \geq 0$ for $i, j = 1, \ldots, n$, and $\sum_{i=1}^{n} p_{ij} = 1$ and $\sum_{j=1}^{n} p_{ij} = 1$.

Theorem III.1: $\mathbf{x} \prec_+ \mathbf{y}$ if and only if there exists a doubly stochastic matrix $\mathbf{P}$ such that $\mathbf{x} = \mathbf{P}\mathbf{y}$.

A square matrix $\mathbf{\Pi}$ is said to be a permutation matrix if each row and column has a single one, and all the other entries are zero. There are $n!$ permutation matrices of size $n \times n$.

Theorem III.2: The permutation matrices constitute the extreme points of the set of doubly stochastic matrices. Moreover, the set of doubly stochastic matrices is the convex hull of the permutation matrices.

It follows from Theorems III.1 and III.2 that the set $\{\mathbf{x} \,|\, \mathbf{x} \prec_+ \mathbf{y}\}$ is the convex hull spanned by the $n!$ points which are the permutations of $\mathbf{y}$.

As we have mentioned before, given K parallel subchannels with capacities $C_1, C_2, \ldots, C_K$, which are obtained via SVD, ACD can convert the K subchannels into $L \geq K$ subchannels with capacities $R_1, R_2, \ldots, R_L$ if and only if $(R_1, R_2, \ldots, R_L) \prec_+ (C_1, \ldots, C_K, 0, \ldots, 0) \in \mathbb{R}^L$. For example, for a MIMO channel $\mathbf{H}$ with rank $K = 3$, assume that the capacities of the 3 subchannels obtained via SVD are $C_1 \geq C_2 \geq C_3$. If $L = K$, then ACD can decompose the MIMO channel into 3 subchannels with a rate vector $\mathbf{r} = (R_1, R_2, R_3)$ if and only if $\mathbf{r}$ lies in the convex hull

$$Co\left\{ \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}, \begin{pmatrix} C_1 \\ C_3 \\ C_2 \end{pmatrix}, \begin{pmatrix} C_2 \\ C_1 \\ C_3 \end{pmatrix}, \begin{pmatrix} C_2 \\ C_3 \\ C_1 \end{pmatrix}, \begin{pmatrix} C_3 \\ C_2 \\ C_1 \end{pmatrix}, \begin{pmatrix} C_3 \\ C_1 \\ C_2 \end{pmatrix} \right\}. \tag{12}$$

[2]If $L < K$, some eign-subchannels are discarded, which causes capacity loss. Hence we focus on the case of $L \geq K$.

Here Co stands for the convex hull defined as

$$Co\{S\} = \{\theta_1 x_1 + \ldots + \theta_K x_K | x_i \in S, \theta_i > 0, \theta_1 + \ldots + \theta_K = 1\}. \tag{13}$$

In general, the "capacity region" is a convex hull defined by $K!$ vertices in a K-dimensional space. Since the ACD is capacity lossless, i.e., $\sum_{i=1}^{K} C_i = \sum_{i=1}^{K} R_i$, the capacity region falls into a $(K-1)$-dimensional hyperplane. The gray area in Figure 1 shows the convex hull of (12) with $C_1 = 3$, $C_2 = 2$, and $C_3 = 1$. In this case, the 6 vertices lie in the 2-D plane $\{\mathbf{x} : \sum_{i=1}^{3} x_i = 6\}$. An interesting special case is the UCD scheme [5], which achieves the rate vector corresponding to the center of the convex hull, i.e., $\mathbf{r} = (2, 2, 2)$.

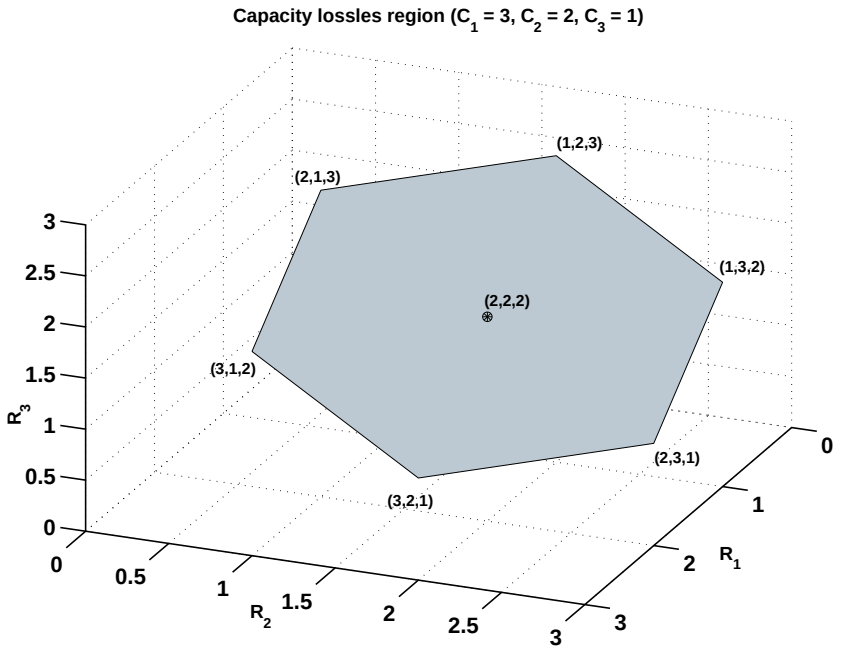

Fig. 1. Illustration of the capacity lossless region obtainable via ACD. We assume $K = 3$, $C_1 = 3$, $C_2 = 2$, and $C_3 = 1$.

Definition 3: For $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^n$, if

$$\prod_{i=1}^{j} x_{[i]} \leq \prod_{i=1}^{j} y_{[i]}, \quad 1 \leq j \leq n \tag{14}$$

with equality for $j = n$, we say that $\mathbf{x}$ is *multiplicatively* majorized by $\mathbf{y}$ and write $\mathbf{x} \prec_\times \mathbf{y}$, or equivalently, $\mathbf{y} \succ_\times \mathbf{x}$.

Obviously, if $\mathbf{x} \prec_\times \mathbf{y}$, then $\log \mathbf{x} \prec_+ \log \mathbf{y}$.

B. Generalized Triangular Decomposition

Now we are ready to introduce the GTD theorem. The following result is due to Weyl [20] (also see [21, p. 171]):

Theorem III.3: If $\mathbf{A} \in \mathbb{C}^{n \times n}$ with eigenvalues $|\gamma_1| \geq |\gamma_2| \geq \ldots \geq |\gamma_n|$ and singular values $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$, then

$$\{|\gamma_i|\}_{i=1}^{n} \prec_\times \{\lambda_i\}_{i=1}^{n}. \tag{15}$$

The following result is due to Horn [22] (also see [21, p. 220]):

Theorem III.4: If $\mathbf{r} \in \mathbb{R}^n$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^n$ satisfy $|\mathbf{r}| \prec_\times \boldsymbol{\lambda}$, where $|\mathbf{r}|$ stands for the absolute value of $\mathbf{r}$, then there exists an upper triangular matrix $\mathbf{R} \in \mathbb{C}^{K\times K}$ with singular values $\boldsymbol{\lambda}$ and with $\mathbf{r}$ on its diagonal.

We now combine Theorems III.3 and III.4 to obtain:

Theorem III.5 (GTD theorem) Let $\mathbf{H} \in \mathbb{C}^{m\times n}$ have rank K with singular values $\boldsymbol{\lambda} \in \mathbb{R}_+^K$. There exists an upper triangular matrix $\mathbf{R} \in \mathbb{C}^{K\times K}$ and matrices $\mathbf{Q}$ and $\mathbf{P}$ with orthonormal columns such that $\mathbf{H} = \mathbf{QRP}^*$ if and only if the diagonal elements of $\mathbf{R}$ satisfy $|\mathbf{r}| \prec_\times \boldsymbol{\lambda}$.

Proof: If $\mathbf{H} = \mathbf{QRP}^*$, then the eigenvalues of $\mathbf{R}$ are its diagonal elements and the singular values of $\mathbf{R}$ coincide with those of $\mathbf{H}$. By Theorem III.4, $\mathbf{r} \prec_\times \boldsymbol{\lambda}$ holds. Conversely, suppose that $\mathbf{r} \prec_\times \boldsymbol{\lambda}$ holds. Let $\mathbf{H} = \mathbf{U\Lambda V}^*$ be the singular value decomposition, where $\boldsymbol{\Lambda} \in \mathbb{R}^{K\times K}$. By Theorem III.5, there exists an upper triangular matrix $\mathbf{R} \in \mathbb{C}^{K\times K}$ with the $\mathbf{r}$ on the diagonal and with singular values $\boldsymbol{\lambda}$. Let $\mathbf{R} = \mathbf{U}_0\boldsymbol{\Lambda}\mathbf{V}_0^*$ be the singular value decomposition of $\mathbf{R}$. Substituting $\boldsymbol{\Lambda} = \mathbf{U}_0^*\mathbf{RV}_0$ in the singular value decomposition for $\mathbf{H}$, we have

$$\mathbf{H} = (\mathbf{UU}_0^*)\mathbf{R}(\mathbf{VV}_0^*)^*.$$

In other words, $\mathbf{H} = \mathbf{QRP}^*$ where $\mathbf{Q} = \mathbf{UU}_0^*$ and $\mathbf{P} = \mathbf{VV}_0^*$. ■

In [23], we propose a computationally efficient and numerically stable algorithm to achieve the GTD predicted by Theorem III.5, which is summarized in Appendix A to make this paper self-contained.

C. Closed-Form Representation of MMSE-VBLAST

The derivation of the ACD scheme for MIMO transceiver design relies on the closed-form representation of the MMSE-VBLAST detector introduced in [24].

Suppose $M_t \le M_r$ and $K = M_t$. The VBLAST scheme does not use any precoder, i.e., $\mathbf{F} = \mathbf{I}_{M_t}$ (here $L = M_t$). The data model of (1) can be rewritten as

$$\mathbf{y} = \sum_{i=1}^{M_t} \mathbf{h}_i x_i + \mathbf{z}, \tag{16}$$

where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$. To decode the transmitted information symbol, VBLAST first detects the signal with the spatial structure $\mathbf{h}_{M_t}$, which is usually referred to as the M_tth *layer*[3], and then cancels it out of the received signal vector. Next, it detects the signal with spatial structure $\mathbf{h}_{M_t-1}$ and so on. If the signals are detected based on the MMSE estimates, we refer to them as MMSE-VBLAST. This *sequential nulling and cancellation* procedure is summarized as follows [13]

$\boldsymbol{\nu}_{M_t} = \mathbf{y}$
`for` $i = M_t : -1 : 1$
  $\hat{x}_i = \mathbf{w}_i^*\boldsymbol{\nu}_i$ (nulling step)
  $\tilde{x}_i = \mathcal{C}[\hat{x}_i]$
  $\boldsymbol{\nu}_{i-1} = \boldsymbol{\nu}_i - \mathbf{h}_i\tilde{x}_i$ (cancellation step)
`end`

where $\mathcal{C}$ stands for mapping to the nearest symbol in the symbol constellation.

Consider the augmented matrix (we remind the reader that $\alpha = \dfrac{\sigma_z^2}{\sigma_x^2}$)

$$\mathbf{H}_a = \begin{bmatrix} \mathbf{H} \\ \sqrt{\alpha}\mathbf{I}_{M_t} \end{bmatrix}_{(M_r+M_t)\times M_t}. \tag{17}$$

[3]In this paper, we also refer to it as *subchannel*.

Applying the QR decomposition yields

$$\mathbf{H}_{H_a} = \mathbf{Q}_{H_a}\mathbf{R}_{H_a} \triangleq \begin{bmatrix} \mathbf{Q}^u_{H_a} \\ \mathbf{Q}^l_{H_a} \end{bmatrix} \mathbf{R}_{H_a}, \tag{18}$$

where $\mathbf{R}_a \in \mathbb{C}^{M_t \times M_t}$ is an upper triangular matrix with positive diagonal and $\mathbf{Q}^u_{H_a} \in \mathbb{C}^{M_r \times M_t}$. We can obtain the nulling vectors $\{\mathbf{w}_i\}_{i=1}^K$ using $\mathbf{Q}^u_{H_a}$ and $\mathbf{R}_{H_a}$ as shown in the following lemma [24] (also see the more detailed version [25]):

Lemma III.1: Let $\{\mathbf{q}_i\}_{i=1}^{M_t}$ denote the columns of $\mathbf{Q}^u_{H_a}$ and $\{r_{ii}\}_{i=1}^{M_t}$ the diagonal elements of $\mathbf{R}_{H_a}$ where $\mathbf{Q}^u_{H_a}$ and $\mathbf{R}_{H_a}$ are given in (18). The MMSE nulling vectors are

$$\mathbf{w}_i = r_{ii}^{-1}\mathbf{q}_i, \quad i = 1, 2, \ldots, M_t. \tag{19}$$

It is easy to verify that the output signal-to-interference-and-noise ratio (SINR) of the ith layer (i.e., the signal corresponding to $\mathbf{h}_i$) obtained via the MMSE-VBLAST detector is

$$\rho_i = \mathbf{h}_i^* \mathbf{C}_i^{-1} \mathbf{h}_i, \tag{20}$$

where $\mathbf{C}_i = \sum_{j=1}^{i-1} \mathbf{h}_j \mathbf{h}_j^* + \alpha \mathbf{I}$.

The following lemma is established in [5].

Lemma III.2: The diagonal of $\mathbf{R}_{H_a}$ given in (18) and $\{\rho_i\}_{i=1}^{M_t}$ satisfy

$$\alpha(1 + \rho_i) = r_{ii}^2, \quad i = 1, 2, \ldots, M_t. \tag{21}$$

Hence the capacity of the ith subchannel is

$$R_i = \log(1 + \rho_i) = \log(\alpha^{-1} r_{ii}^2), \quad i = 1, 2, \ldots, M_t. \tag{22}$$

IV. Adaptable Channel Decomposition

A. ACD-VBLAST

Now we are ready to establish the ACD scheme. We see from (2) that $\mathbf{F}$ can always be scaled such that $\alpha = 1$. Hence without loss of generality, we let $\alpha = 1$ in the sequel to simplify the notation. If we modify the precoder $\mathbf{F}$ given in (7) to be

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}} \tag{23}$$

where $(\cdot)$ denotes the transpose, $\mathbf{\Omega} \in \mathbb{R}^{L \times K}$ with $L \geq K$, and $\mathbf{\Omega}^{\mathsf{T}}\mathbf{\Omega} = \mathbf{I}$, then we see through inserting (23) into (6) that the $\mathbf{F}$ given in (23) is also a precoder maximizing the channel throughput. However, introducing $\mathbf{\Omega}$ brings much greater flexibility than the precoder of (7) as demonstrated in the following theorem.

Theorem IV.1 (ACD Theorem) Consider a MIMO channel of (1) with $\mathbf{F}$ given in (23). For any $L \geq K$, let $\mathbf{c} \in \mathbb{R}^L$ be a zero vector with its first K elements replaced with $\{C_k\}_{k=1}^K$, where $C_k = \log\left(1 + \lambda_{H,k}^2 \phi_k\right)$. Given any rates $\{R_k\}_{k=1}^L$, we can find a semi-unitary matrix $\mathbf{\Omega} \in \mathbb{R}^{L \times K}$ such that the combination of the linear precoder $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$ and the MMSE-VBLAST detector yields L subchannels with capacities $\{R_k\}_{k=1}^L$ if and only if $\{R_k\}_{k=1}^L \prec_+ \mathbf{c}$.

Proof: Given the precoder of (23), the virtual channel is

$$\mathbf{G} \triangleq \mathbf{H}\mathbf{F} = \mathbf{U}\mathbf{\Lambda}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}} \triangleq \mathbf{U}\mathbf{\Lambda}_G\mathbf{\Omega}^{\mathsf{T}} \tag{24}$$

where $\mathbf{\Lambda}_G = \mathbf{\Lambda}\mathbf{\Phi}^{1/2}$ is a diagonal matrix with diagonal elements

$$\lambda_{G,i} = \lambda_{H,i}\phi_i^{1/2}, \quad i = 1, \ldots, K. \tag{25}$$

Let the augmented matrix $\mathbf{G}_a$ be defined as

$$\mathbf{G}_a = \left[\begin{array}{c} \mathbf{U}\mathbf{\Lambda}_G\mathbf{\Omega}^{\mathsf{T}} \\ \mathbf{I}_L \end{array} \right]_{(M_r+L)\times L}, \tag{26}$$

which can be rewritten as

$$\mathbf{G}_a = \left[\begin{array}{cc} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_0 \end{array} \right] \left[\begin{array}{c} [\mathbf{\Lambda}_G \vdots \mathbf{0}_{K\times(L-K)}] \\ \mathbf{I}_L \end{array} \right] \mathbf{\Omega}_0^{\mathsf{T}}, \tag{27}$$

where $\mathbf{\Omega}_0 \in \mathbb{R}^{L\times L}$ is orthogonal with its first K columns forming $\mathbf{\Omega}$. Let $\mathbf{J}$ denote

$$\mathbf{J} \triangleq \left[\begin{array}{c} [\mathbf{\Lambda}_G \vdots \mathbf{0}_{K\times(L-K)}] \\ \mathbf{I}_L \end{array} \right] = \left[\begin{array}{c} [\mathbf{\Lambda}_H\mathbf{\Phi}^{1/2} \vdots \mathbf{0}_{K\times(L-K)}] \\ \mathbf{I}_L \end{array} \right]. \tag{28}$$

The singular values of $\mathbf{J}$ are

$$\lambda_{J,i} = \left\{ \begin{array}{ll} \sqrt{1+\lambda_{H,i}^2\phi_i}, & 1 \le i \le K, \\ 1, & i > K. \end{array} \right. \tag{29}$$

According to Theorem III.5, we can apply GTD to obtain

$$\mathbf{J} = \mathbf{Q}_J\mathbf{R}_J\mathbf{P}_J^{\mathsf{T}} \tag{30}$$

if and only if the diagonal elements of $\mathbf{R}_J \in \mathbb{R}^{L\times L}$, which we denote as $\{r_{J,ii}\}_{i=1}^L$, satisfy

$$\{|r_{J,ii}|\}_{i=1}^L \prec_\times \{\lambda_{J,i}\}_{i=1}^L. \tag{31}$$

Note that both $\mathbf{Q}_J$ and $\mathbf{P}_J$ are real-valued matrices because $\mathbf{J}$ is a real-valued diagonal matrix. Inserting (30) into (27) yields

$$\mathbf{G}_a = \left[\begin{array}{cc} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_0 \end{array} \right] \mathbf{Q}_J\mathbf{R}_J\mathbf{P}_J^{\mathsf{T}}\mathbf{\Omega}_0^{\mathsf{T}}. \tag{32}$$

Choose $\mathbf{\Omega}_0 = \mathbf{P}_J^{\mathsf{T}}$ and define

$$\mathbf{Q}_{G_a} = \left[\begin{array}{cc} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_0 \end{array} \right] \mathbf{Q}_J. \tag{33}$$

Then (32) can be rewritten as $\mathbf{G}_a = \mathbf{Q}_{G_a}\mathbf{R}_J$, which is the QR decomposition of $\mathbf{G}_a$. The semi-unitary matrix $\mathbf{\Omega}$ associated with $\mathbf{G}_a$ consists of the first K columns of $\mathbf{\Omega}_0 = \mathbf{P}_J^{\mathsf{T}}$. By Lemma III.2, it follows that for $\alpha = 1$, (31) is equivalent to

$$\{1+\rho_i\}_{i=1}^L = \{r_{J,ii}^2\}_{i=1}^L \prec_\times \{\lambda_{J,i}^2\}_{i=1}^L, \tag{34}$$

where ρ_i, $1 \le i \le L$, denotes the ouput SINR of the ith subchannel, and $\lambda_{J,i}$ is given in (29). If

$$\{R_i\}_{i=1}^L = \{\log(1+\rho_i)\}_{i=1}^L \prec_+ \{\log\lambda_{J,i}^2\}_{i=1}^L = \mathbf{c}, \tag{35}$$

then (31) and (34) hold, which implies the existence of $\mathbf{\Omega}$ (the first K columns of $\mathbf{P}_J^{\mathsf{T}}$).

Conversely, suppose that there exists a semi-unitary matrix $\mathbf{\Omega}$ such that the linear precoder $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$ and the MMSE-VBLAST detector yields L subchannels with capacities $\{R_k\}_{k=1}^L$. Let $\mathbf{G}_a = \mathbf{Q}_{G_a}\mathbf{R}_J$ be the QR decomposition. Since the singular values of $\mathbf{G}_a$ and $\mathbf{R}_J$ are the

same and the eigenvalues of $\mathbf{R}_J$ are its diagonal elements, it follows from Theorem III.3 of Weyl that (31) holds. Hence, by (34), we conclude that (35) holds. ■

The proof of Theorem IV.1 is insightful. Indeed, given the SVD of $\mathbf{H}$ and the "water filling" level $\mathbf{\Phi}^{1/2}$, we only need to calculate $\mathbf{\Lambda}_G$, $\mathbf{J}$, and the GTD of $\mathbf{J}$ given in (30). Then we immediately obtain the linear precoder $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$, where $\mathbf{\Omega}$ consists of the first K columns of $\mathbf{P}_J^{\mathsf{T}}$. Let $\mathbf{Q}_{G_a}^u$ denote the first M_r rows of $\mathbf{Q}_{G_a}$. According to Lemma III.1, the nulling vectors are calculated as

$$\mathbf{w}_i = r_{J,ii}^{-1}\mathbf{q}_{G_a,i}, \quad 1 \le i \le L, \tag{36}$$

where $r_{J,ii}$ is the ith diagonal element of $\mathbf{R}_J$ and $\mathbf{q}_{G_a,i}$ is the ith column of $\mathbf{Q}_{G_a}^u$.

We can exploit the structure of $\mathbf{J}$ to further reduce the computational complexity of GTD. It is easy to verify that the SVD of $\mathbf{J}$ defined in (28) is

$$\mathbf{J} = \begin{bmatrix} [\mathbf{\Lambda}_G \vdots \mathbf{0}_{K\times(L-K)}]\mathbf{\Lambda}_J^{-1} \\ \mathbf{\Lambda}_J^{-1} \end{bmatrix} \mathbf{\Lambda}_J \mathbf{I}_L, \tag{37}$$

where $\mathbf{\Lambda}_J$ is an $L \times L$ diagonal matrix with the diagonal elements given in (29). Applying the GTD matrix decomposition algorithm given in Appendix A to $\mathbf{\Lambda}_J$ yields

$$\mathbf{\Lambda}_J = (\mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_{L-1})\mathbf{R}_J(\mathbf{P}_1\mathbf{P}_2\ldots\mathbf{P}_L)^{\mathsf{T}}. \tag{38}$$

Inserting (38) into (37) yields

$$\mathbf{J} = \begin{bmatrix} [\mathbf{\Lambda}_G \vdots \mathbf{0}_{K\times(L-K)}]\mathbf{\Lambda}_J^{-1} \\ \mathbf{\Lambda}_J^{-1} \end{bmatrix} (\mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_{L-1})\mathbf{R}_J(\mathbf{P}_1\mathbf{P}_2\ldots\mathbf{P}_{L-1})^{\mathsf{T}}. \tag{39}$$

The linear precoder is

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega} = \mathbf{V}\left[\mathbf{\Phi}^{1/2} \vdots \mathbf{0}_{K\times(L-K)}\right]\mathbf{P}_1\mathbf{P}_2\ldots\mathbf{P}_{L-1}. \tag{40}$$

The nulling vectors are given by (36) with

$$\begin{aligned} \mathbf{Q}_{G_a}^u &= \mathbf{U}[\mathbf{\Lambda}_G \vdots \mathbf{0}_{K\times(L-K)}]\mathbf{\Lambda}_J^{-1}\mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_{L-1} \\ &= \mathbf{U}\left[\mathbf{\Gamma} \vdots \mathbf{0}_{K\times(L-K)}\right]\mathbf{Q}_1\mathbf{Q}_2\ldots\mathbf{Q}_{L-1}, \end{aligned} \tag{41}$$

where $\mathbf{\Gamma} \in \mathbb{R}^{K\times K}$ is diagonal with its ith diagonal element being $\gamma_i = \frac{\lambda_{G,i}}{\sqrt{\lambda_{G,i}^2+1}}$. Note that $\mathbf{Q}_l$ and $\mathbf{P}_l, l = 1, 2, \ldots, L$, are the Givens rotation matrices and hence calculating (40) and (41) needs $O(M_t(L+K))$ and $O(M_r(L+K))$ flops, respectively.

Given the SVD of $\mathbf{H}$ and the power allocation level $\mathbf{\Phi}$, then the ACD-VBLAST scheme needs to run the procedures summarized in Table I. If $M_t = M_r = K$, then the ACD-VBLAST scheme requires only $O(L^2 + K^2 + KL)$ flops, given the SVD of the channel matrix.

B. ACD-DP

As a dual form of ACD-VBLAST, the ACD scheme can be implemented by using a DP precoder, which we refer to as ACD-DP. For ACD-DP, a direct construction of the linear precoder $\mathbf{F}$ as done in Section IV-A is not obvious. Instead, we exploit the uplink-downlink duality revealed in [26] to obtain ACD-DP. This technique is also used in [5].

TABLE I
ACD ALGORITHM

step	operation	flops
1	Calculate $\mathbf{\Lambda}_G = \mathbf{\Lambda}\mathbf{\Phi}^{1/2}$	$O(K)$
2	Obtain $\mathbf{\Lambda}_J$ using (29)	$O(K)$
3	Apply GTD to $\mathbf{\Lambda}_J$ to obtain (38)	$O(L^2)$
4	Generate $\mathbf{F}$ using (40)	$O(M_t(L+K))$
5	Compute $\mathbf{Q}_{G_q}^u$ using (41)	$O(M_r(L+K))$
6	Calculate $\{\mathbf{w}_i\}_{i=1}^L$ using (36)	$O(M_r L)$

We first apply the ACD-VBLAST scheme to the reverse channel

$$\mathbf{y} = \mathbf{H}^*\mathbf{F}\mathbf{x} + \mathbf{z}, \tag{42}$$

where the roles of the transmitter and receiver are exchanged and the $\mathbf{H}$ in (1) is replaced by $\mathbf{H}^*$. Then we obtain the precoder $\mathbf{F}$ and the equalizer $\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_L]$ from $\mathbf{H}^*$ according to (40) and (36), respectively. Applying $\mathbf{F}$ and the VBLAST detector with nulling vectors $\{\mathbf{w}_i\}_{i=1}^L$, we obtain L subchannels

$$\mathbf{w}_i^*\mathbf{y} = \mathbf{w}_i^*\mathbf{H}^*\mathbf{f}_i x_i + \sum_{j=1}^{i-1} \mathbf{w}_i^*\mathbf{H}^*\mathbf{f}_j x_j + \mathbf{w}_i^*\mathbf{z}, \quad i = 1, \ldots, L, \tag{43}$$

where the ith subchannel (43) is free of interference from the jth ($j > i$) subchannels which are detected and cancelled out in advance. The SINR of the subchannel (43) is

$$\rho_i = \frac{|\mathbf{w}_i^*\mathbf{H}^*\mathbf{f}_i|^2\sigma_x^2}{\mathbf{w}_i^*\mathbf{w}_i\sigma_z^2 + \sum_{j=1}^{i-1}|\mathbf{w}_i^*\mathbf{H}^*\mathbf{f}_j|^2\sigma_x^2}. \tag{44}$$

Note that replacing $\mathbf{w}_i$ by $\bar{\mathbf{w}}_i$, which is obtained by scaling $\mathbf{w}_i$ such that $\|\bar{\mathbf{w}}_i\| = 1$, does not change ρ_i since the output SINR is invariant to the length of $\mathbf{w}_i$. Also note that $\alpha = 1$, i.e., $\sigma_z^2 = \sigma_x^2$. Hence (43) can be simplified to be

$$\rho_i = \frac{|\bar{\mathbf{w}}_i^*\mathbf{H}^*\mathbf{f}_i|^2}{1 + \sum_{j=1}^{i-1}|\bar{\mathbf{w}}_i{}^*\mathbf{H}^*\mathbf{f}_j|^2}. \tag{45}$$

Let $\bar{\mathbf{f}}_i$, $i = 1, \ldots, L$, be the scaled version of $\mathbf{f}_i$ and has unit length. Denote $p_i = \|\mathbf{f}_i\|^2$. Then

$$\rho_i = \frac{|\bar{\mathbf{w}}_i^*\mathbf{H}^*\bar{\mathbf{f}}_i|^2 p_i}{1 + \sum_{j=1}^{i-1}|\bar{\mathbf{w}}_i{}^*\mathbf{H}^*\bar{\mathbf{f}}_j|^2 p_j}, \quad i = 1, \ldots, L. \tag{46}$$

Let $a_{ij} = |\bar{\mathbf{f}}_i^*\mathbf{H}\bar{\mathbf{w}}_j|^2$. Then (46) can be represented in the matrix form

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ -\rho_2 a_{12} & a_{22} & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -\rho_L a_{1L} & -\rho_L a_{2L} & \cdots & a_{LL} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_L \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_L \end{bmatrix}. \tag{47}$$

According to the uplink-downlink duality, in the original channel, the precoder of ACD-DP should be $\tilde{\mathbf{F}} = [\sqrt{q_1}\bar{\mathbf{w}}_1, \ldots, \sqrt{q_L}\bar{\mathbf{w}}_L]$, where $\{q_i\}_{i=1}^L$ will be determined later in (51), and the receiving

vectors are $\bar{\mathbf{f}}_i,\ i=1,\ldots,L$. Then we get L subchannels whose ith scalar subchannel of the MIMO channel is

$$y_i = \bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_i \sqrt{q_i} x_i + \sum_{j=i+1}^{L} \bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_j \sqrt{q_j} x_j + \sum_{j=1}^{i-1} \bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_j \sqrt{q_j} x_j + \bar{\mathbf{f}}_i^* \mathbf{z}. \tag{48}$$

Applying the dirty paper precoder to x_i and treating $\sum_{j=1}^{i-1} \bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_j \sqrt{q_j} x_j$ as the interference known at the transmitter (note that here we precode the first layer first while for ACD-VBLAST, we detect the Lth layer first), we obtain an equivalent subchannel

$$y_i = \bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_i \sqrt{q_i} x_i + \sum_{j=i+1}^{L} \bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_j \sqrt{q_j} x_j + \bar{\mathbf{f}}_i^* \mathbf{z} \tag{49}$$

with SINR (again, recall that $\alpha = 1$ and $\sigma_x^2 = \sigma_z^2$)

$$\rho_i = \frac{q_i |\bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_i|^2}{1 + \sum_{j=i+1}^{L} q_j |\bar{\mathbf{f}}_i^* \mathbf{H} \bar{\mathbf{w}}_j|^2}, \quad \text{for } i = 1, 2, \ldots, L. \tag{50}$$

Similar to (46), (50) can also be represented as

$$\begin{bmatrix} a_{11} & -\rho_1 a_{12} & \cdots & -\rho_1 a_{1L} \\ 0 & a_{22} & \cdots & -\rho_2 a_{2L} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{LL} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_L \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_L \end{bmatrix}. \tag{51}$$

It is easy to see that $q_i > 0,\ 0 \le i \le L$. It is proven in [26] that $\sum_{i=1}^{L} q_i = \mathrm{tr}(\tilde{\mathbf{F}}\tilde{\mathbf{F}}^*) = \mathrm{tr}(\mathbf{F}\mathbf{F}^*) = \sum_{i=1}^{L} p_i$. That is, to obtain L subchannels with SINRs $\{\rho_i\}_{i=1}^{L}$, the ACD-DP needs exactly the same power as the ACD-VBLAST. To make this paper self-contained, we give below an alternative proof to this interesting and useful fact.

Let $\mathcal{U}_A$ denote a strictly upper triangular matrix whose (i,j)th entry is a_{ij} for $1 \le i < j \le L$ and zero otherwise. Let $\mathcal{D}_A$ and $\mathcal{D}_\rho$ be two $L \times L$ diagonal matrices with their ith element equal to a_{ii} and ρ_i, respectively. Then (46) can be rewritten as

$$\left(\mathcal{D}_A - \mathcal{D}_\rho \mathcal{U}_A^{\mathsf{T}}\right) \mathbf{p} = \boldsymbol{\rho} \tag{52}$$

or equivalently

$$\left(\mathcal{D}_\rho^{-1} \mathcal{D}_A - \mathcal{U}_A^{\mathsf{T}}\right) \mathbf{p} = \mathbf{1} \tag{53}$$

where $\mathbf{p} = [p_1, \ldots, p_L]^{\mathsf{T}}$, $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_L]^{\mathsf{T}}$ and $\mathbf{1}$ is a vector with unit elements. Hence

$$\mathbf{p} = \left(\mathcal{D}_\rho^{-1} \mathcal{D}_A - \mathcal{U}_A^{\mathsf{T}}\right)^{-1} \mathbf{1} \tag{54}$$

Similarly, (51) can be rewritten as

$$\left(\mathcal{D}_A - \mathcal{D}_\rho \mathcal{U}_A\right) \mathbf{q} = \boldsymbol{\rho} \tag{55}$$

or

$$\left(\mathcal{D}_\rho^{-1} \mathcal{D}_A - \mathcal{U}_A\right) \mathbf{q} = \mathbf{1}. \tag{56}$$

Hence

$$\mathbf{q} = \left(\mathcal{D}_\rho^{-1} \mathcal{D}_A - \mathcal{U}_A\right)^{-1} \mathbf{1}. \tag{57}$$

From (54) and (57),

$$\sum_{i=1}^{L} p_i = \mathbf{1}^{\mathsf{T}} \left(\mathcal{D}_\rho^{-1} \mathcal{D}_A - \mathcal{U}_A^{\mathsf{T}} \right)^{-1} \mathbf{1} = \mathbf{1}^{\mathsf{T}} \left(\mathcal{D}_\rho^{-1} \mathcal{D}_A - \mathcal{U}_A \right)^{-1} \mathbf{1} = \sum_{i=1}^{L} q_i. \tag{58}$$

We can use the Tomlinson-Harashima precoder [27][28] or the trellis precoder [29] to achieve known interference cancellation at the transmitter. For a system with high dimensionality, ACD-DP is a better choice than ACD-VBLAST since it is free of propagation errors.

V. MIMO Communications with QoS Constraints

In this section, we apply the ACD scheme to MIMO communications with QoS constraints. Suppose we want to transmit $L \geq K$ independent data streams through a MIMO channel. Instead of multiplexing all the substreams in the time division manner to share the entire MIMO channel, we apply ACD to decompose the MIMO channel into multiple subchannels whose capacities/SINRs meet the QoS requirements of the substreams, and dedicate one subchannel to each substream. In [10], the authors studied the same problem. They proposed a linear transceiver design which, similar to ACD, can also control the SINR of each subchannel via designing the precoder. However, the linear transceiver is capacity lossy and can suffer from considerable performance degradation compared with our ACD scheme as we will show at the end of this section. Given that all the subchannels meet the QoS constraints, we want to minimize the overall input power. We need to solve the following optimization problem:

$$\begin{array}{ll} \min_{\mathbf{F}} & \mathrm{tr}\,(\mathbf{F}\mathbf{F}^*) \\ \text{subject to} & \begin{pmatrix} \mathbf{H}\mathbf{F} \\ \mathbf{I}_L \end{pmatrix} = \mathbf{Q}\mathbf{R} \\ & \mathrm{diag}(\mathbf{R}) = \{\sqrt{1+\rho_i}\}_{i=1}^{L}. \end{array} \tag{59}$$

Here $\mathbf{QR}$ denotes the QR decomposition and $\mathrm{diag}(\mathbf{R})$ denotes the vector formed by the diagonal of $\mathbf{R}$. According to Lemma III.2, the diagonal of $\mathbf{R}$ determines the SINRs of the subchannels. Without loss of generality, we assume that $\rho_1 \geq \rho_2 \geq \ldots \geq \rho_L$. We now consider a problem whose constraints are more ralaxed than those of (59):

$$\begin{array}{ll} \min_{\mathbf{F}} & \mathrm{tr}\,(\mathbf{F}\mathbf{F}^*) \\ \text{subject to} & \boldsymbol{\lambda}_{G_a} \succ_\times \{\sqrt{1+\rho_i}\}_{i=1}^{L}, \quad \mathbf{G}_a = \begin{pmatrix} \mathbf{H}\mathbf{F} \\ \mathbf{I}_L \end{pmatrix}, \end{array} \tag{60}$$

where $\boldsymbol{\lambda}_{G_a}$ stands for the singular values of the augmented matrix $\mathbf{G}_a$. In general, for any matrix $\mathbf{A}$, we let $\boldsymbol{\lambda}_A$ denote the singular values of $\mathbf{A}$. By Theorem III.3 of Weyl, if $\mathbf{F}$ is feasible in (59), then $\mathbf{F}$ is feasible in (60). We now further simplify (60) and show that its solution provides a solution of (59).

Theorem V.1: If $\mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^*$ is the singular value decomposition of $\mathbf{H}$, then (60) has a solution of the form $\mathbf{F} = \mathbf{V}\boldsymbol{\Phi}^{1/2}$ where $\boldsymbol{\Phi} \in \mathbb{R}^{K \times K}$ is a diagonal matrix with diagonal elements ϕ_i, $1 \leq i \leq K$, chosen to solve the problem

$$\left.\begin{array}{ll} \min_{\boldsymbol{\Phi}} & \sum_{i=1}^{K} \phi_i \\ \text{subject to} & \prod_{i=1}^{k}(1+\lambda_{H,i}^2\phi_i) \geq \prod_{i=1}^{k}(1+\rho_i), \quad \phi_k \geq 0, \;\; 1 \leq k \leq K-1, \\ & \prod_{i=1}^{K}(1+\lambda_{H,i}^2\phi_i) = \prod_{i=1}^{L}(1+\rho_i). \end{array}\right\} \tag{61}$$

Moreover, if $\mathbf{Q}_J\mathbf{R}_J\mathbf{P}_J^{\mathsf{T}}$ is the GTD of $\mathbf{J}$ in (28) where $\mathbf{\Phi}$ is a solution of (61), then (59) has the solution $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$ where $\mathbf{\Omega}$ is the matrix formed by the first K columns of $\mathbf{P}_J^{\mathsf{T}}$.

Proof: See Appendix B. ■

We now make a change of variables to further simplify the formulation of (61). We define

$$\begin{aligned} \psi_i &= \phi_i + 1/\lambda_{H,i}^2, && 1 \le i \le K, \\ \beta_i &= \tfrac{1+\rho_i}{\lambda_{H,i}^2}, && 1 \le i < K, \\ \beta_K &= \tfrac{1}{\lambda_K^2}\textstyle\prod_{i=K}^{L}(1+\rho_i). \end{aligned} \tag{62}$$

With these definitions, (61) reduces to

$$\left.\begin{array}{ll} \min_{\boldsymbol{\psi}} & \sum_{i=1}^{K}\psi_i \\ \text{subject to} & \prod_{i=1}^{k}\psi_i \ge \prod_{i=1}^{k}\beta_i, \quad \psi_k \ge 1/\lambda_{H,k}^2, \;\; 1 \le k \le K. \end{array}\right\} \tag{63}$$

Notice that the equality constraint in (61) has been dropped since this constraint is automatically satisfied at an optimum. That is, if $\boldsymbol{\psi}$ is feasible in (63) and the inequality corresponding to $k = K$ is strictly positive, then the cost is reduced when the trailing components of $\boldsymbol{\psi}$ are lowered.

Clearly, the feasible set for (63) is nonempty and the cost function tends to infinity as any of the components of $\boldsymbol{\psi}$ tends to infinity. By continuity of the cost function and the constraints, a minimizer must exist. We now analyze the structure of the minimizer. By exploiting the structure, we obtain a fast algorithm for solving (63).

We first study a similar optimization problem with relaxed constraints.

Lemma V.1: Any solution $\boldsymbol{\psi}$ of the problem

$$\min_{\boldsymbol{\psi}} \;\sum_{i=1}^{K}\psi_i \;\text{ subject to }\; \prod_{i=1}^{k}\psi_i \ge \prod_{i=1}^{k}\beta_i, \;\; 1 \le k \le K, \tag{64}$$

has the property that $\psi_{i+1} \le \psi_i$ for each i.

Proof: We replace the inequalities in (64) by the equivalent constraints obtained by taking log's:

$$\sum_{i=1}^{k}\log(\psi_i) \ge \sum_{i=1}^{k}\log(\beta_i), \;\; 1 \le k \le T,$$

The Lagrangian $\mathcal{L}$ associated with (64), after this modification of the constraints, is

$$\mathcal{L}(\boldsymbol{\psi}, \boldsymbol{\mu}) = \sum_{k=1}^{K}\left(\psi_k - \mu_k\sum_{i=1}^{k}\left(\log(\psi_i) - \log(\beta_i)\right)\right).$$

By the first-order optimality conditions associated with $\boldsymbol{\psi}$, there exists $\boldsymbol{\mu} \ge \mathbf{0}$ with the property that the gradient of the Lagrangian with respect to $\boldsymbol{\psi}$ vanishes. Equating to zero the partial derivative of the Lagrangian with respect to ψ_j, we obtain the relations

$$\psi_j = \sum_{i=j}^{K}\mu_i, \quad j = 1, \ldots, K.$$

Hence, $\psi_j - \psi_{j+1} = \mu_j \ge 0$. ■

Using Lemma V.1, we can gain insights into the structure of a solution to (63).

Lemma V.2: There exists a solution $\boldsymbol{\psi}$ to (63) with the property that for some integer $j \in [1, K]$,

$$\psi_{i+1} \leq \psi_i \text{ for all } i < j, \quad \psi_{i+1} \geq \psi_i \text{ for all } i \geq j, \quad \psi_i = \frac{1}{\lambda_{H,i}^2} \text{ for all } i > j. \tag{65}$$

In particular, $\psi_j \leq \psi_i$ for all i.

Proof: If $\boldsymbol{\psi}$ is a solution of (63) with the property that $\psi_i > \frac{1}{\lambda_{H,i}^2}$ for all $1 \leq j \leq K$, then by the convexity of the constraints, it follows that $\boldsymbol{\psi}$ is a solution of (64). By Lemma V.1, we conclude that Lemma V.2 holds with $j = K + 1$. Now, suppose that $\boldsymbol{\psi}$ is a solution of (63) with $\psi_i = 1/\lambda_{H,i}^2$ for some i. We wish to show that $\psi_k = 1/\lambda_{H,k}^2$ for all $k > i$. Suppose, to the contrary, that there exists an index $k \geq i$ with the property that $\psi_k = 1/\lambda_{H,k}^2$ and $\psi_{k+1} > 1/\lambda_{H,k+1}^2$. We show that components k and $k+1$ of $\boldsymbol{\psi}$ can be modified so as to satisfy the constraints and make the cost strictly smaller. In particular, let $\boldsymbol{\psi}(\epsilon)$ be identical with $\boldsymbol{\psi}$ except for components k and $k+1$:

$$\psi_k(\epsilon) = (1+\epsilon)\psi_k \quad \text{and} \quad \psi_{k+1}(\epsilon) = \frac{\psi_{k+1}}{1+\epsilon}. \tag{66}$$

For $\epsilon > 0$ small, $\boldsymbol{\psi}(\epsilon)$ satisfies the constraints of (63). The change $\Delta(\epsilon)$ in the cost function of (63) is

$$\Delta(\epsilon) = (1+\epsilon)\psi_k + \frac{\psi_{k+1}}{1+\epsilon} - \psi_k - \psi_{k+1}.$$

The derivative of $\Delta(\epsilon)$ evaluated at zero is

$$\Delta'(0) = \psi_k - \psi_{k+1}.$$

Since $1/\lambda_{H,k}^2$ is an increasing function of k and since $\psi_k = 1/\lambda_{H,k}^2$, we conclude that $\psi_{k+1} > \psi_k$ and $\Delta'(0) < 0$. Hence, for $\epsilon > 0$ near zero, $\boldsymbol{\psi}(\epsilon)$ has a smaller cost than $\boldsymbol{\psi}(0)$, which yields a contradiction. Hence, there exists an index j with the property that $\psi_i = 1/\lambda_{H,i}^2$ for all $i > j$ and $\psi_i > 1/\lambda_{H,i}^2$ for all $i \leq j$.

According to Lemma V.1, $\psi_i \geq \psi_{i+1}$ for any $i < j$. To complete the proof, we need to show that $\psi_j \leq \psi_{j+1}$. As noted previously, any solution of (63) satisfies

$$\prod_{i=1}^{K} \psi_i = \prod_{i=1}^{K} \beta_i,$$

which implies (cf. (62))

$$\prod_{i=1}^{j} \psi_i = \prod_{i=1}^{j} \beta_i \left(\prod_{i=j+1}^{K} \beta_i \lambda_{H,i}^2 \right) > \prod_{i=1}^{j} \beta_i.$$

That is, the constraint $\prod_{i=1}^{j} \psi_i \geq \prod_{i=1}^{j} \beta_i$ in (63) is inactive. If $\psi_j > \psi_{j+1}$, we will decrease the j-th component and increase the $j+1$ component, while leaving the other components unchanged. Letting $\boldsymbol{\psi}(\delta)$ be the modified vector, we set

$$\psi_{j+1}(\delta) = (1+\delta)\psi_{j+1} \quad \text{and} \quad \psi_j(\delta) = \frac{\psi_j}{1+\delta}.$$

Since the j-th constraint in (63) is inactive, $\boldsymbol{\psi}(\delta)$ is feasible for δ near zero. And if $\psi_j > \psi_{j+1}$, then the cost decreases as δ increases. It follows that $\psi_j \leq \psi_{j+1}$. ■

We refer to the index j in Lemma V.2 as the "break point." At the break point, the lower bound constraint $\psi_i \geq 1/\lambda_{H,i}^2$ changes from inactive to active. We now use Lemma V.2 to obtain an algorithm for (63).

```
function ψ = acdPow (β,λ)
L = 1 ; R = length (β) ; ψ = zeros (1, R) ;
γ = cumsum (log (β)) ;
while R ≥ L
     [t, l] = max (γ(L:R)./[1:R-L+1]) ;
     γl = exp (t) ; L1 = L + l - 1 ;
     if γl ≥ 1/λ(L1)^2
          ψ(L:L1) = γl ;
          L = L + l ;
          γ(L:R) = γ(L:R) - γ(L-1) ;
     else
          ψ(L1:R) = 1./(λ(L1:R).^2) ;
          γ(L1-1) = γ(R) - sum (log (ψ(L1:R))) ;
          R = L1 - 1 ;
     end
end
```

Fig. 2. A Matlab function to solve (63).

Lemma V.3: Let γ_k denote the k-th geometric mean of the β_i:

$$\gamma_k = \left(\prod_{i=1}^{k} \beta_i \right)^{1/k},$$

and let l denote an index for which γ_k is the largest:

$$l = \arg\max\{\gamma_k : 1 \le k \le K\}. \tag{67}$$

If $\gamma_l \ge 1/\lambda_l^2$, then putting $\psi_i = \gamma_l$ for all $i \le l$ is optimal in (63). If $\gamma_l < 1/\lambda_l^2$, then $\psi_i = 1/\lambda_{H,i}^2$ for all $i \ge l$ at an optimal solution of (63).

Proof: See Appendix C. ■

Based on Lemma V.3, we can use the following strategy to solve (63). We form the geometric mean described in Lemma V.3 and we evaluate l. If $\gamma_l \ge 1/\lambda_{H,l}^2$, then we set $\phi_i = \gamma_l$ for $i \le l$, and we simplify (63) by removing ψ_i, $1 \le i \le l$, from the problem. If $\gamma_l < 1/\lambda_{H,l}^2$, then we set $\psi_i = 1/\lambda_{H,i}^2$ for $i \ge l$, and we simplify (63) by removing ψ_i, $l \le i \le K$, from the problem. A Matlab code implementing this algorithm appears in Figure 2.

After obtaining the power loading level $\phi_i = \psi_i - 1/\lambda_{H,i}^2$, $1 \le i \le K$, we calculate the precoder $\mathbf{F}$ and the nulling vectors $\{\mathbf{w}_i\}_{i=1}^{L}$ according to Table I in Section IV. Note that one of the possible paths through the `acdPow` routine makes the leading elements of $\boldsymbol{\psi}$ all equal while setting the trailing elements of $\psi_i = 1/\lambda_{H,i}^2$. This path coincides with the standard water filling algorithm. Note that the role of $\boldsymbol{\psi}$ is analogous to that of the Lagrange multipler μ in (8). In this case, the ACD scheme is optimal in terms of maximizing the overall throughput given the input power. On the other hand, if some substream has a very high prescribed SINR such that the l given in (67) is less than the "break point" j, then $\boldsymbol{\psi}$ leads to be a multi-level water filling power allocation, which suffers from overall capacity loss. This happens when the target rate vector $[R_1, \ldots R_L]$ falls out of the convex hull spanned by the $L!$ permutations of $[C_1, \ldots, C_K, 0, \ldots, 0]$ (cf. Figure 1), where $C_k, k = 1, \ldots, K$, are the capacities of the eigen subchannels with water filling power allocation. As a remedy to this issue, one can "break" (if it is practically allowable) the *oversized* substream

into more than one substreams with smaller rates, or equivalently, lower SINR requirements. Note that ACD can decompose a MIMO channel into an arbitrarily large number of subchannels.

An interesting special case is that $\rho_1 = \rho_2 = \ldots = \rho_L$, i.e., the substream shares the same SINR requirements. In this case, $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_K$ since the singular values $\{\lambda_{H,i}\}_{i=1}^{K}$ are in nonincreasing order, and `acdPow` yields a standard water filling solution. In this case, ACD becomes UCD.

We present one numerical example to conclude this section. We assume Rayleigh independent flat fading channels with $M_t = 5$ and $M_r = 6$. We consider equal QoS requirements for $L = 5$ independent substreams. Figure 3 compares the input power needed by our ACD scheme and the linear transceiver scheme of [10]. Our scheme can save about 2.5 dB for any prescribed output SINR.

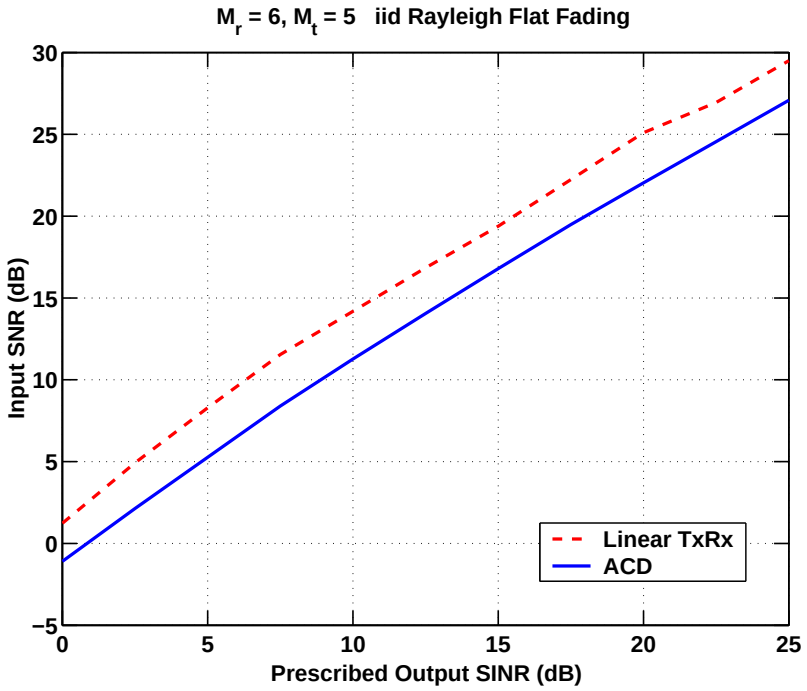

Fig. 3. Input SNR vs. Output SNR. Result is based on the average of 500 Monte Carlo trials of a i.i.d. Rayleigh flat fading channel with $M_t = 5$ and $M_r = 6$.

VI. CDMA Sequence Design

As we have mentioned before, the problem of optimal CDMA sequence design is a special case in the framework of designing the MIMO transceivers. Hence the application of the ACD scheme to MIMO transceiver design, which we have developed in Section V, can be used for the CDMA sequence design with small modifications. Indeed, due to the simple channel matrix ($\mathbf{H} = \mathbf{I}$), some procedures of the ACD scheme can be simplified. We shall show that the ACD scheme turns out to be an improved solution to the sequence design proposed in [9]. At the end of this section, we will compare our ACD scheme and the scheme proposed in [9].

A. CDMA Sequences Maximizing Sum Capacity

Recall that the precoder maximizing the overall MIMO channel capacity is $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$. For an S-CDMA channel, $\mathbf{H} = \mathbf{I}$, then $\mathbf{V} = \mathbf{I}$ and the optimal power loading level is the uniform power allocation. Hence the CDMA sequence maximizing the sum capacity is $\mathbf{S} = \sqrt{\rho}\mathbf{\Omega}^{\mathsf{T}}$. Since

$\mathbf{\Omega}$ has orthonormal columns, we obtain $\mathbf{S}\mathbf{S}^{\mathsf{T}} = \rho\mathbf{I}$. This observation coincides with the findings in [6], in which the authors show that the CDMA sequences maximizing the sum capacity are the Welch-Bound-Equality sequences.

B. Uplink Case

For the uplink scenario, i.e., the mobiles to base station case, the base station calculates the optimal CDMA sequences for each mobile user and the associated successive nulling vectors needed by itself. Then the base station informs the mobile users their designated CDMA sequences.

First, we need to calculate the power loading levels $\mathbf{\Phi} \in \mathbb{R}^{N\times N}$ such that the following GTD matrix decomposition is possible:

$$\mathbf{\Phi}_a \triangleq \begin{pmatrix} [\mathbf{\Phi}^{1/2} \vdots \mathbf{0}_{N\times(L-N)}] \\ \mathbf{I}_L \end{pmatrix} = \mathbf{Q}\mathbf{R}\mathbf{P}^{\mathsf{T}}, \tag{68}$$

where the diagonal elements of $\mathbf{R}$, $r_{ii}, i = 1, 2, \ldots, L$, satisfy the QoS constraints. Note that the singular values of $\mathbf{\Phi}_a$ form a sequence whose first N elements are $\sqrt{1+\phi_i}, i = 1, 2, \ldots, N$, followed by $L - N$ ones. From Theorem III.5, (68) exists if and only if

$$(\{1+\phi_i\}_{i=1}^{N}, 1, \ldots, 1) \succ_{\times} \{1+\rho_i\}_{i=1}^{L}. \tag{69}$$

Similar to (61), we need to solve the problem

$$\begin{array}{ll} \min_{\mathbf{\Phi}} & \sum_{i=1}^{N} \phi_i \\ \text{subject to} & (\{1+\phi_i\}_{i=1}^{N}, 1, \ldots, 1) \succ_{\times} \{1+\rho_i\}_{i=1}^{L}. \\ & \phi_i \geq 0, \quad \forall i \end{array} \tag{70}$$

Similar to (63), (70) can be further simplified using the variables

$$\psi_i = \phi_i + 1, \quad \beta_i = 1 + \rho_i \text{ for } i < K, \quad \text{and} \quad \beta_K = \prod_{i=K}^{L}(1+\rho_i).$$

The simplified problem is

$$\left.\begin{array}{ll} \min_{\boldsymbol{\psi}} & \sum_{i=1}^{K} \psi_i \\ \text{subject to} & \prod_{i=1}^{k} \psi_i \geq \prod_{i=1}^{k} \beta_i, \quad \psi_k \geq 1, \;\; 1 \leq k \leq K. \end{array}\right\} \tag{71}$$

The algorithm `acdPow` simplifies immensely when we apply it to (71). Since $\beta_i \geq 1 = \frac{1}{\lambda_{H,i}}$ for all i, the constraints $\psi_i \geq 1$ are inactive. Since $\beta_i \leq \beta_{i-1}$ for all $i < K$, the geometric means satisfy $\gamma_i \leq \gamma_{i-1}$ for all $i < K$. Hence, in Lemma V.3, the value of l is either 1 or K. If $l = 1$, then we set $\psi_1 = \beta_1$ and we remove ψ_1 from the problem. If $l = K$, then $\psi_i = \gamma_K$ for all i. It follows that there exists an index j with the property that

$$\psi_i = \beta_i \text{ for all } i \leq j \quad \text{and} \quad \psi_i = \left(\prod_{i=j+1}^{K} \beta_i\right)^{1/(K-j)} \quad \text{for all } i > j.$$

This observation coincides with the solution obtained in [9].

Let $\mathbf{\Psi}$ denote an $L \times L$ identity matrix with its first K diagonal elements replaced by $\psi_i, 1 \leq i \leq K$. According to the ACD scheme presented in Section IV-A, we then apply the GTD algorithm to $\mathbf{\Psi}^{1/2}$ to obtain

$$\mathbf{\Psi}^{1/2} = \mathbf{Q}_{\Psi}\mathbf{R}_{\Psi}\mathbf{P}_{\Psi}^{\mathsf{T}}. \tag{72}$$

According to (40),

$$\mathbf{S} = \mathbf{F} = \left[\mathbf{\Phi}^{1/2} \vdots \mathbf{0}_{N\times(L-N)}\right] \mathbf{P}_{\Psi}. \tag{73}$$

Let

$$[\mathbf{v}_1, \ldots, \mathbf{v}_L] = [\mathbf{\Phi}^{1/2} \vdots \mathbf{0}_{N\times(L-N)}]\mathbf{\Psi}^{-1/2}\mathbf{Q}_{\Psi}. \tag{74}$$

By (41) and (36), the nulling vectors used at the base station is

$$\mathbf{w}_i = r_{\Psi,ii}^{-1}\mathbf{v}_i, \tag{75}$$

where $r_{\Psi,ii}$ is the ith diagonal element of $\mathbf{R}_{\Psi}$. In summary, the base station needs to run the following three steps:

1. Solve the optimization problem (71).
2. Apply the GTD algorithm to $\mathbf{\Psi}^{1/2}$ in (72).
3. Obtain the spreading sequences for all mobile users, $[\mathbf{s}_1, \ldots, \mathbf{s}_L] = \mathbf{S}$, and the nulling vectors $\{\mathbf{w}_i\}_{i=1}^{L}$ (cf. (73) and (75) for the base station).

C. Downlink Case

In the downlink case, since the mobiles cannot cooperate with each other for decision feedback. Hence the VBLAST detection is impractical at receivers. However, we can apply ACD-DP as introduced in Section IV-B to cancel out known interferences at the transmitter, i.e., the base station. We can convert the downlink problem as an uplink one and exploit the downlink-uplink duality as we have done in Section IV-B. Note that $\mathbf{H} = \mathbf{H}^* = \mathbf{I}$, i.e., the downlink and uplink channels are the same! Consider the case where the uplink and downlink communications are symmetric, i.e., for each mobile user, the QoS of the communications from the user to the base station and the base station to the user are the same. After obtaining the spreading sequences $[\mathbf{s}_1, \ldots, \mathbf{s}_L]$ for the mobile users, and the nulling vectors $[\mathbf{w}_1, \ldots, \mathbf{w}_L]$ used at the base station for the uplink case, we immediately know that in the downlink case the spreading sequences transmitted from the base station are exactly $[\mathbf{w}_1, \ldots, \mathbf{w}_L]$ and the nulling vectors used at the mobiles are the spreading sequences, $[\mathbf{s}_1, \ldots, \mathbf{s}_L]$, used in the uplink case. The only parameters we need to calculate are $q_1, \ldots, q_N$ (cf. (51)). Hence in this symmetric case, the base station only needs to inform the mobiles their designated spreading sequences once in the two-way communications. Each mobile uses the same sequence for both data transmission in the uplink channel and interference nulling in the downlink channel.

D. Further Remarks

The ACD scheme, which was originally motivated by MIMO transceiver designs, turns out to be similar to the scheme of [9] in several aspects. Both schemes are based on the nonlinear decision feedback operations. Hence both are optimal in terms of maximizing the channel throughput and minimizing the overall input power. Both the GTD algorithm, on which the ACD scheme is based, and the construction of the Hermitian matrix with prescribed eigenvalues and Cholesky values as done in [9] rely on the Weyl-Horn theorem. However, our ACD scheme enjoys several remarkable advantages over the scheme of [9]. First, note that if we obtain the GTD $\mathbf{H} = \mathbf{QRP}^*$, where $\mathbf{R}$ has the prescribed diagonal elements, then it follows immediately that $\mathbf{A} \triangleq \mathbf{P}^*\mathbf{H}^*\mathbf{HP} = \mathbf{RR}^*$ is the desired Cholesky decomposition. However, the information associated with $\mathbf{Q}$ is lost in the Cholesky decomposition. Hence the nulling vectors used at the receivers of [9] cannot be calculated explicitly as our ACD does (cf. (36)). Furthermore, the correlation matrix $\mathbf{A}$ is only an intermediate result. To get the CDMA sequences, one has to decompose $\mathbf{A} = \mathbf{RR}^*$ explicitly. The ACD scheme, however, can be used to obtain both the precoder (CDMA sequences), which

are the columns of $\mathbf{P}$, and the equalizer from $\mathbf{Q}$ simultaneously. Second, our ACD scheme is a solution to the more general MIMO transceiver design problem. The Cholesky decomposition algorithm provided in Appendix C of [9] is only applicable to the scenario where the singular values are only of two values. Hence it is not applicable to the general design of MIMO transceivers. The more general Cholesky factorization algorithm suggested in the proofs is computationally far less efficient. Third, the ACD scheme has two implementation forms, i.e., ACD-VBLAST and ACD-DP, which makes it applicable to both the downlink and uplink scenarios. Finally, the ACD scheme provides insights that unify the CDMA sequence design problem as well as the precoded OFDM communication problem as special cases of the MIMO transceiver design.

VII. Conclusions

Based on the recently developed GTD matrix decomposition algorithm, we have proposed the ACD scheme utilizing the CSIT and CSIR. ACD can be used to decompose a MIMO channel into multiple subchannels with prescribed capacities. The ACD scheme has two implementation forms. One is the combination of a linear precoder and a minimum mean-squared-error VBLAST (MMSE-VBLAST) detector, which is referred to as ACD-VBLAST, and the other includes a dirty paper (DP) precoder and a linear equalizer followed by a DP decoder, which we refer to as ACD-DP. Both forms of ACD are computationally very efficient. We have also determined the subchannel capacity region such that a capacity lossless decomposition is possible. The applications of the ACD scheme for MIMO communications with QoS constraints have been investigated. We have also identified the problems of designing precoders for OFDM communications and designing CDMA sequences as special cases in the unifying framework of MIMO transceiver designs. In particular, we have shown that the CDMA sequence design problem in the uplink and downlink scenarios can be solved using ACD-VBLAST and ACD-DP, respectively.

Appendix A
Generalized Triangular Decomposition

We summarize the steps of the GTD algorithm as follows. To make it easier to distinguish between the elements of the matrix $\mathbf{R}$ and the elements of the given diagonal vector $\mathbf{r}$, we use R_{ij} to denote the (i,j) element of $\mathbf{R}$ and r_i to denote the i-th element of $\mathbf{r}$.

1. Let $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$ be the singular value decomposition of $\mathbf{H}$, and suppose we are given $\mathbf{r} \in \mathbb{C}^{\mathbb{K}}$ with $\mathbf{r} \preceq \boldsymbol{\sigma}$. Initialize $\mathbf{Q} = \mathbf{U}$, $\mathbf{P} = \mathbf{V}$, $\mathbf{R} = \mathbf{\Lambda}$, and $k = 1$.

2. Let p and q be defined as follows:

$$\begin{aligned} p &= \arg \min_i \{|R_{ii}| : k \le i \le K,\ |R_{ii}| \ge |r_k|\}, \\ q &= \arg \max_i \{|R_{ii}| : k \le i \le K,\ |R_{ii}| \le |r_k|,\ i \ne p\}. \end{aligned}$$

In $\mathbf{R}$, $\mathbf{P}$, and $\mathbf{Q}$, perform the following exchanges:

$$\begin{aligned} (R_{kk}, R_{k+1,k+1}) &\leftrightarrow (R_{pp}, R_{qq}) \\ (\mathbf{R}_{1:k-1,k}, \mathbf{R}_{1:k-1,k+1}) &\leftrightarrow (\mathbf{R}_{1:k-1,p}, \mathbf{R}_{1:k-1,q}) \\ (\mathbf{P}_{:,k}, \mathbf{P}_{:,k+1}) &\leftrightarrow (\mathbf{P}_{:,p}, \mathbf{P}_{:,q}) \\ (\mathbf{Q}_{:,k}, \mathbf{Q}_{:,k+1}) &\leftrightarrow (\mathbf{Q}_{:,p}, \mathbf{Q}_{:,q}) \end{aligned}$$

3. Construct the matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ shown below:

$$\underbrace{\frac{r_k}{|r_k|^2}\begin{bmatrix} c\delta_1^* & s\delta_2^* \\ -s\delta_2 & c\delta_1 \end{bmatrix}}_{(\mathbf{G}_2^*)} \underbrace{\begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}}_{(\mathbf{\Pi}^*\mathbf{R}^{(k)}\mathbf{\Pi})} \underbrace{\begin{bmatrix} c & -s \\ s & c \end{bmatrix}}_{(\mathbf{G}_1)} = \underbrace{\begin{bmatrix} r_k & x \\ 0 & y \end{bmatrix}}_{(\mathbf{R}^{(k+1)})} \tag{76}$$

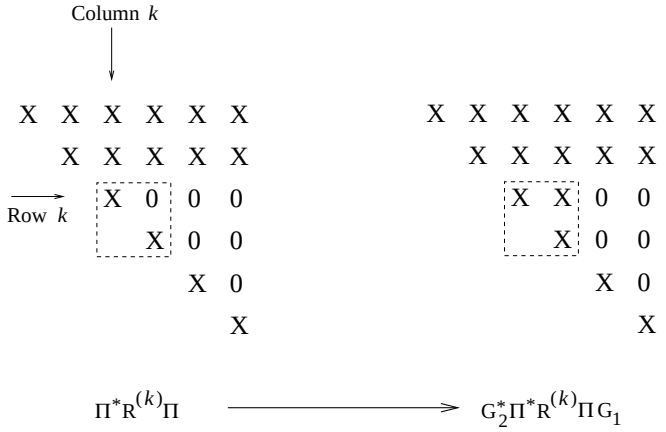

Fig. 4. The operation displayed in (76).

If $|\delta_1| = |\delta_2| = |r_k|$, we take $c = 1$ and $s = 0$; if $|\delta_1| \neq |\delta_2|$, we take

$$c = \sqrt{\frac{|r_k|^2 - |\delta_2|^2}{|\delta_1|^2 - |\delta_2|^2}} \quad \text{and} \quad s = \sqrt{1 - c^2}. \tag{77}$$

Replace $\mathbf{R}$ by $\mathbf{G}_2^* \mathbf{R} \mathbf{G}_1$, replace $\mathbf{Q}$ by $\mathbf{Q}\mathbf{G}_2$, and replace $\mathbf{P}$ by $\mathbf{P}\mathbf{G}_1$.

4. If $k = K - 1$, then go to Step 5. Otherwise, replace k by $k + 1$ and go to Step 2.

5. Multiply column K of $\mathbf{Q}$ by R_{KK}/r_K; replace R_{KK} by r_K. The product $\mathbf{QRP}^*$ is the GTD of $\mathbf{H}$ based on $\mathbf{r}$.

To write $\mathbf{Q}, \mathbf{P}$ explicitly, we define $\mathbf{Q}_k = \mathbf{\Pi} \mathbf{G}_2$ and $\mathbf{P}_k = \mathbf{\Pi} \mathbf{G}_1$ (cf. (76)). Then

$$\mathbf{H} = \mathbf{U} \mathbf{Q}_1 \mathbf{Q}_2 \dots \mathbf{Q}_{k-1} \mathbf{C} \mathbf{R} \mathbf{P}_{k-1}^* \dots \mathbf{P}_2^* \mathbf{P}_1^* \mathbf{V}^*.$$

with

$$\mathbf{Q} = \mathbf{U} \left(\prod_{i=1}^{K-1} \mathbf{Q}_i \right) \mathbf{C} \quad \text{and} \quad \mathbf{P} = \mathbf{V} \left(\prod_{i=1}^{K-1} \mathbf{P}_i \right).$$

Here $\mathbf{C}$ is a K by K identity matrix with its Kth diagonal element replaced by R_{KK}/r_K. Finally, note that if $\mathbf{r}$ is real-valued, then $\mathbf{G}_1$ and $\mathbf{G}_2$ are real, which implies $\mathbf{R}$ is real-valued and $\mathbf{C} = \mathbf{I}$.

Figure 4 depicts the transformation from $\mathbf{\Pi}^* \mathbf{R}^{(k)} \mathbf{\Pi}$ to $\mathbf{G}_2^* \mathbf{\Pi}^* \mathbf{R}^{(k)} \mathbf{\Pi} \mathbf{G}_1$. The dashed box is the 2 by 2 submatrix displayed in (76). The numerical stability of this algorithm is analyzed in [23]. In particular, we have shown that the division by the possibly small denominator in (77) is safe, and the algorithm is stable. A Matlab implementation of our GTD algorithm appears in [23].

Appendix B
Proof of Theorem V.1

Observe that for $\mathbf{F} = \mathbf{V} \mathbf{\Phi}^{1/2}$, we have

$$\text{tr } (\mathbf{F}\mathbf{F}^*) = \sum_{i=1}^{K} \phi_i \quad \text{and} \quad \mathbf{HF} = \mathbf{U} \mathbf{\Lambda} \mathbf{\Phi}^{1/2}. \tag{78}$$

Hence, $\lambda_{HF,i}^2 = \lambda_{H,i}^2 \phi_i$ for $1 \leq i \leq K$, and

$$\lambda_{G_a,i} = \begin{cases} \sqrt{1 + \lambda_{H,i}^2 \phi_i}, & 1 \leq i \leq K, \\ 1, & i > K. \end{cases}$$

Since $1+\rho_i \geq 1$, the last $L-K$ inequalities in the multiplicative majorization condition in (60) are implied by the single equality constraint in (61). Hence, the problem (60) reduces to (61) where $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}$, which gives an upper bound for the minimum in (60).

Let $\mathbf{F} = \mathbf{U}_F \mathbf{\Phi}^{1/2} \mathbf{\Omega}^{\mathsf{T}}$ denote the singular value decomposition for any given $\mathbf{F} \in \mathbb{C}^{M_t \times L}$. Once again, tr $(\mathbf{F}\mathbf{F}^*)$ is given by the sum in (78). By [21, Theorem 3.3.14], the singular values of the product $\mathbf{H}\mathbf{F}$ of two matrices are multiplicatively majorized by the product of the singular values of $\mathbf{H}$ and $\mathbf{F}$:

$$\prod_{i=1}^{k} \lambda_{H,i}^2 \phi_i \geq \prod_{i=1}^{k} \lambda_{HF,i}^2, \quad 1 \leq k \leq K. \tag{79}$$

Taking log's, we have

$$\sum_{i=1}^{k} \log(\lambda_{H,i}^2 \phi_i) \geq \sum_{i=1}^{k} \log(\lambda_{HF,i}^2), \quad 1 \leq k \leq K. \tag{80}$$

By [21, Lemma 3.3.8]) and (80),

$$\sum_{i=1}^{k} f(\log(\lambda_{H,i}^2 \phi_i)) \geq \sum_{i=1}^{k} f(\log(\lambda_{HF,i}^2)), \quad 1 \leq k \leq K, \tag{81}$$

whenever f is a real-valued, increasing convex function. The function $f(t) = \log(e^t + 1)$ is convex since its second derivative is positive. Making this choice for f in (81) and exponentiating both sides, we obtain:

$$\prod_{i=1}^{k} (\lambda_{H,i}^2 \phi_i + 1) \geq \prod_{i=1}^{k} (\lambda_{HF,i}^2 + 1), \quad 1 \leq k \leq K. \tag{82}$$

Since $\mathbf{F}$ is feasible in (60),

$$\begin{aligned} \prod_{i=1}^{k} (\lambda_{HF,i}^2 + 1) &\geq \prod_{i=1}^{k} (\rho_i + 1), \quad 1 \leq k < K, \\ \prod_{i=1}^{K} (\lambda_{HF,i}^2 + 1) &= \prod_{i=1}^{L} (\rho_i + 1). \end{aligned}$$

Combining this with (82), we get

$$\begin{aligned} \prod_{i=1}^{k} (\lambda_{H,i}^2 \phi_i + 1) &\geq \prod_{i=1}^{k} (\rho_i + 1), \quad 1 \leq k < K, \\ \prod_{i=1}^{K} (\lambda_{H,i}^2 \phi_i + 1) &\geq \prod_{i=1}^{L} (\rho_i + 1). \end{aligned} \tag{83}$$

Since $\lambda_{H,i}^2 \phi_i + 1$ is the square of the i-th singular value of the augmented matrix $\mathbf{G}_a$ corresponding to the choice $\mathbf{U}_F$, we conclude that $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}$ satisfies all the inequality constraints in (60). If the inequality (83) is strict, then ϕ_K should be decreased in order to satisfy the equality constraint in (61). Since decreasing ϕ_K only lowers $\mathrm{tr}(\mathbf{F}\mathbf{F}^*)$, we deduce that the minimum in (60) is achieved by a matrix of the form $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}$. If $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}$ is optimal in (60), then so is $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$ whenever $\mathbf{\Omega}$ has K orthonormal columns (since the constraints are satisfied and the value of the cost does not change). We now make the choice for $\mathbf{\Omega}$ given in Theorem IV.1. That is, if $\mathbf{Q}_J \mathbf{R}_J \mathbf{P}_J^{\mathsf{T}}$

is the GTD of $\mathbf{J}$ in (28) where $\mathbf{\Phi}$ is a solution of (61), then $\mathbf{\Omega}$ is the matrix formed by the first K columns of $\mathbf{P}_J^{\mathsf{T}}$. For this choice of $\mathbf{\Omega}$, the constraints of (59) are satisfied. As noted earlier, the minimum in (59) can be no smaller than the minimum in (60). Since this choice for $\mathbf{F}$ yields the same cost in both (59) and (60), we conclude that $\mathbf{F} = \mathbf{V}\mathbf{\Phi}^{1/2}\mathbf{\Omega}^{\mathsf{T}}$ is optimal in (59).

Appendix C
Proof of Lemma V.3

First suppose that $\gamma_l \geq 1/\lambda_l^2$. By the arithmetic/geometric mean inequality, the problem

$$\min \sum_{i=1}^{l} \psi_i \quad \text{subject to} \quad \prod_{i=1}^{l} \psi_i \geq \prod_{i=1}^{l} \beta_i, \;\; \boldsymbol{\psi} \geq \mathbf{0}, \tag{84}$$

has the solution $\psi_i = \gamma_l$, $1 \leq i \leq l$. Since $\lambda_{H,i}$ is a decreasing function of i and $\gamma_l \geq 1/\lambda_{H,i}^2$, we conclude that $\psi_i = \gamma_l$ satisfies the constraints $\psi_i \geq 1/\lambda_{H,i}^2$ for $1 \leq i \leq l$. Since l attains the maximum in (67),

$$\gamma_l^k \geq \prod_{i=1}^{k} \beta_k$$

for all $k \leq l$. Hence, by taking $\psi_i = \gamma_l$ for $1 \leq i \leq l$, the first l inequalities in (63) are satisfied, with equality for $k = l$, and the first l lower bound constraints $\psi_i \geq 1/\lambda_{H,i}^2$ are satisfied.

Let $\boldsymbol{\psi}^*$ denote any optimal solution of (63). If

$$\prod_{i=1}^{l} \psi_i^* = \prod_{i=1}^{l} \beta_i, \tag{85}$$

then by the unique optimality of $\psi_i = \gamma_l$, $1 \leq i \leq l$, in (84), and by the fact that the inequality constraints in (63) are satisfied for $k \in [1, l]$, we conclude that $\psi_i^* = \gamma_l$ for all $i \in [1, l]$. On the other hand, suppose that

$$\prod_{i=1}^{l} \psi_i^* > \prod_{i=1}^{l} \beta_l = \gamma_l^l. \tag{86}$$

We show below that this leads to a contradiction; consequently, (85) holds and $\psi_i^* = \gamma_l$ for $i \in [1, l]$.

Define the quantity

$$\gamma_* = \left(\prod_{i=1}^{l} \psi_i^* \right)^{1/l}.$$

By (86) $\gamma_* > \gamma_l$. Again, by the arithmetic/geometric mean inequality, the solution of the problem

$$\min \sum_{i=1}^{l} \psi_i \quad \text{subject to} \quad \prod_{i=1}^{l} \psi_i \geq \gamma_*^l, \;\; \boldsymbol{\psi} \geq \mathbf{0}, \tag{87}$$

is $\psi_i = \gamma_*$ for $i \in [1, l]$. By (86), $\gamma_* > \gamma_l$ and $\boldsymbol{\psi}$ satisfies the inequality constraints in (63) for $k \in [1, l]$.

Let M be the first index with the property that

$$\prod_{i=1}^{M} \psi_i^* = \prod_{i=1}^{M} \beta_i. \tag{88}$$

Such an index exists since $\boldsymbol{\psi}^*$ is optimal, which implies that

$$\prod_{i=1}^{K} \psi_i^* = \prod_{i=1}^{K} \beta_i.$$

First, suppose that $M \leq j$, where j is the break point given in Lemma V.2. By complementary slackness, $\mu_i = 0$ and $\psi_i^* - \psi_{i+1}^* = \mu_i$ for $1 \leq i < M$. We conclude that $\psi_i = \gamma_*$ for $1 \leq i \leq M$. By (88) we have

$$\gamma_*^M = \prod_{i=1}^{M} \beta_i.$$

It follows that

$$\left(\prod_{i=1}^{M} \beta_i \right)^{1/M} = \gamma_* > \gamma_l,$$

which contradicts the fact that l achieves the maximum in (67).

In the case $M > j$, we have $\psi_i = \gamma_*$ for $1 \leq i \leq j$. Again, this follows by complementary slackness. However, we need to stop when $i = j$ since the lower bound constraints become active for $i > j$. In Lemma V.2, we show that $\psi_i^* \geq \psi_j^* = \gamma_*$ for $i \geq j$. Consequently, we have

$$\prod_{i=1}^{M} \beta_i = \prod_{i=1}^{M} \psi_i^* \geq \gamma_*^M > \gamma_l^M.$$

Again, this contradicts the fact that l achieves the maximum in (67). This completes the analysis of the case where $\gamma_l \geq 1/\lambda_{H,l}^2$.

Now consider the case $\gamma_l < 1/\lambda_{H,l}^2$. By the definition of γ_l, we have

$$\gamma_l \geq \left(\prod_{i=1}^{K} \beta_i \right)^{1/K} \quad \text{or} \quad \gamma_l^K \geq \prod_{i=1}^{K} \beta_i. \tag{89}$$

If j is the break point described in Lemma V.2, then $\psi_i^* \geq \psi_j^*$ for all i; it follows that

$$\prod_{i=1}^{K} \psi_i^* \geq \left(\psi_j^* \right)^K. \tag{90}$$

Since the product of the components of $\boldsymbol{\psi}^*$ is equal to the product of the components of $\boldsymbol{\beta}$, from (89) and (90) we get

$$\gamma_l^K \geq \prod_{i=1}^{K} \beta_i = \prod_{i=1}^{K} \psi_i^* \geq \left(\psi_j^* \right)^K.$$

Hence, $\gamma_l \geq \psi_j^* \geq 1/\lambda_{H,j}^2 \geq 1/\lambda_{H,i}^2$ for all $i \leq j$. In particular, if $l \leq j$, then $\gamma_l \geq 1/\lambda_{H,l}^2$, or, $l > j$ when $\gamma_l < 1/\lambda_{H,l}^2$. As a consequence, $\psi_l^* = \frac{1}{\lambda_{H,l}^2}$.

References

[1] G. G. Raleigh and J. M. Cioffi, "Spatial-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, pp. 357–366, March 1998.

[2] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Filterbank transceiver optimizating information rate in block transmissions over dispersive channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 1019–1032, April 1999.

[3] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2381–2401, September 2003.

[4] Y. Jiang, J. Li, and W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Transactions on Signal Processing*, To appear. Available on line: http://www.sal.ufl.edu/yjiang/papers/gmdCommR2.pdf.

[5] Y. Jiang, J. Li, and W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Transactions on Signal Processing, Available on line: http://www.sal.ufl.edu/yjiang/papers/ucdR1.pdf*, Submitted. 2004.

[6] M. Rupf and J. L. Massey, "Optimal sequence multisets for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 40, pp. 1261–1266, July 1994.

[7] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous cdma systems," *IEEE Transactions on Information Theory*, vol. 45, pp. 1984–1991, September 1999.

[8] P. Viswanath, V. Anantharam, and D. Tse, "Optimal sequences, power control and user capacity of synchronous CDMA systems with linear mmse multiuser receivers," *IEEE Transactions on Information Theory*, vol. 45, pp. 1968–1983, September 1999.

[9] T. Guess, "Optimal sequence for CDMA with decision-feedback receivers," *IEEE Transactions on Information Theory*, vol. 49, pp. 886–900, April 2003.

[10] D. Palomar, M. Lagunas, and J. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1179–1197, May 2004.

[11] J.-K. Zhang, A. Kavčić, X. Ma, and K. M. Wong, "Design of unitary precoders for ISI channels," in *Proceedings IEEE International Conference on Acoustics Speech and Signal Processing*, vol. III, (Orlando, Florida), pp. 2265–2268, 2002.

[12] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[13] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multiple-element arrays," *Wireless Personal Communications*, vol. 6, pp. 311–335, March 1999.

[14] M. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1405 – 1409, Nov 2-5 1997.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc, 1991.

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

[17] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[18] Y. Jiang, W. Hager, and J. Li, "The geometric mean decomposition," *Linear Algebra and Its Applications*, To appear. Available on line: http://www.math.ufl.edu/∼hager/papers/gmd.pdf.

[19] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization*. New York: Academic, 1979.

[20] H. Weyl, "Inequalities between two kinds of eigenvalues of a linear transformation," *Proc. Nat. Acad. Sci. U. S. A.*, vol. 35, pp. 408–411, 1949.

[21] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge: Cambridge University Press, 1991.

[22] A. Horn, "On the eigenvalues of a matrix with prescribed singular values," *Proc. Amer. Math. Soc.*, vol. 5, pp. 4–7, 1954.

[23] Y. Jiang, W. W. Hager, and J. Li, "The generalized triangular decomposition," *SIAM Journal on Matrix Analysis and Applications*, Available on line: http://www.math.ufl.edu /hager/papers/gtd.ps. Submitted.

[24] B. Hassibi, "A fast square-root implementation for BLAST," *Thirty-Fourth Asilomar Conf. Signals, Systems and Computers*, pp. 1255–1259, November 2000.

[25] B. Hassibi, "An efficient square-root implementation for BLAST," *http://cm.bell-labs.com/who/hochwald/papers/squareroot/*, 2000.

[26] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, pp. 1912–1921, August 2003.

[27] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, March 1971.

[28] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Communications*, pp. 774–780, August 1972.

[29] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," *Global Telecommunications Conference*, vol. 2, pp. 1344–1348, November 2001.