The Diversity-Multiplexing Tradeoff of RF Chain Limited MIMO System with Antenna Selection: Part I Theoretical Framework

Yi Jiang   Mahesh K. Varanasi

Abstract

The large gain promised by the multi-input multi-output (MIMO) technology comes with a cost. In particular multiple analog radio frequency (RF) chains, which are expensive and power consuming, are required at both the transmitter and receiver sides. On the other hand, the antennas connecting to the RF chains are less expensive. Hence one engineering compromise is to implement more antennas than RF chains and to use only a subset of them based on some antenna selection (AS) algorithm. An interesting question therefore arises: given a RF chain limited MIMO system, what is the fundamental performance gain by adding more antennas. In this two-part paper, we answer this question by using the diversity-multiplexing (D-M) gain tradeoff metric. Consider a Rayleigh fading channel with M_t antennas and L_t ($L_t \leq M_t$) RF chains at the transmitter while M_r antennas and L_r ($L_r \leq M_r$) RF chains at the receiver. We obtain the fundamental D-M tradeoff as a function of M_t, M_r, and $\min(L_r, L_t)$. Referring to the special case where $L_t = M_t$ and $L_r = M_r$ as the RF unlimited system (or full system) and RF limited system (or pruned system) otherwise, we prove that the pruned system with optimal channel-dependent AS has the same D-M tradeoff as the full system if the multiplexing gain is less than some integer threshold P, while it suffers from some diversity gain loss for multiplexing gains larger than P. In particular, if $\min(L_r, L_t) = K \triangleq \min(M_r, M_t)$, then $P = K$, i.e. the D-M tradeoffs of the pruned system and the full system are the same. Moreover, this result can be extended to more general fading channels such as Nakagami channel. A fast and D-M tradeoff-optimal AS algorithm is proposed as a byproduct of our analysis.

Index Terms

MIMO, antenna selection, diversity gain, spatial multiplexing gain, tradeoff, outage probability.

This work is supported in part by NSF Grants CCF-0423842 and CCF-0434410.
The authors are with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: yjiang@dsp.colorado.edu; varanasi@colorado.edu.) The material of this paper was presented in Asilomar Conference on Signals, Systems, and Computers 2006, IEEE International Conference on Communications 2007, and IEEE International Symposium on Information Theory 2007.

I. Introduction

Multiple-input multiple-output (MIMO) wireless communication systems have well-known advantages over their single-input single-output (SISO) counterpart, i.e., much higher spectral efficiency and greatly improved reliability [1][2][3][4]. By quantifying spectral efficiency as *multiplexing gain* and reliability as *diversity gain*, it is shown in [5] that a MIMO system can achieve both diversity gain and spatial multiplexing gain simultaneously but there is a fundamental tradeoff between them. Such a tradeoff is referred to as the diversity-multiplexing (D-M) gain tradeoff. The D-M tradeoff metric has since then been a popular performance measure for the existing MIMO communication schemes [5] and has motivated some interesting new designs [6][7][8].

The dramatic performance gain of the MIMO system comes with increased hardware complexity, i.e. more expensive and power consuming analog radio frequency (RF) chains at both sides of the channel [9]. On the other hand, the antennas connecting to the RF chains are less expensive. Hence one engineering compromise is to implement more antennas than the RF chains and to use only a subset of them based on some antenna selection (AS) algorithm [9] [10]. An interesting question therefore arises: given a RF chain limited MIMO system, what is the fundamental performance gain by adding more antennas. In this two-part paper, we answer this question using the D-M gain tradeoff metric. Assuming the L_t transmitting antennas and L_r receiving antennas are chosen according to the maximization of the instantaneous channel mutual information, we derive the D-M tradeoff corresponding to the optimal AS approach. To facilitate our discussion, we refer to a channel with M_t antennas and L_t ($L_t \leq M_t$) RF chains at the transmitter while M_r antennas and L_r ($L_r \leq M_r$) RF chains at the receiver as an $(M_t \times M_r, L_t \times L_r)$ channel. Note that the conventional channel where the number of RF chains equal to the number of antennas can be denoted as $(M_t \times M_r, M_t \times M_r)$.

One of the major results of this paper is illustrated in Figurer 1. We highlight the major points as follows.

- P1 The D-M tradeoff curve of the $(M_t \times M_r, L_t \times L_r)$ channel $(-\circ-)$ is strictly higher than $(L_t \times L_r, L_t \times L_r)$ $(-\triangle-)$ except for the end point $(N, 0)$, where $N \triangleq L_t \wedge L_r$. [1]
- P2 The D-M tradeoff curves of the $(M_t \times M_r, L_t \times L_r)$ and $(M_t \times M_r, M_t \times M_r)$ channels overlap for the multiplexing gain $r \in (0, P)$, where P is an integer between 0 and N.
- P3 The D-M tradeoff curve of $(M_t \times M_r, L_t \times L_r)$ is linear for $r \in (P, N)$. And it is tangent to the D-M tradeoff curve of $(M_t \times M_r, M_t \times M_r)$.
- P4 The D-M tradeoff curve of $(M_t \times M_r, L_t \times L_r)$ is a function of M_t, M_r, and N, but does not depend on the individual values of L_r and L_t.
- P5 If $N = K \triangleq M_t \wedge M_r$, then the D-M tradeoffs of $(M_t \times M_r, L_t \times L_r)$ and $(M_t \times M_r, M_t \times M_r)$ fully overlap.

P1-P3 agree with the intuition that the D-M gain tradeoff of a $(M_t \times M_r, L_t \times L_r)$ channel should lie somewhere

[1]Throughout this paper, we denote $a \wedge b$ as $\min\{a, b\}$ and $a \vee b$ as $\max\{a, b\}$

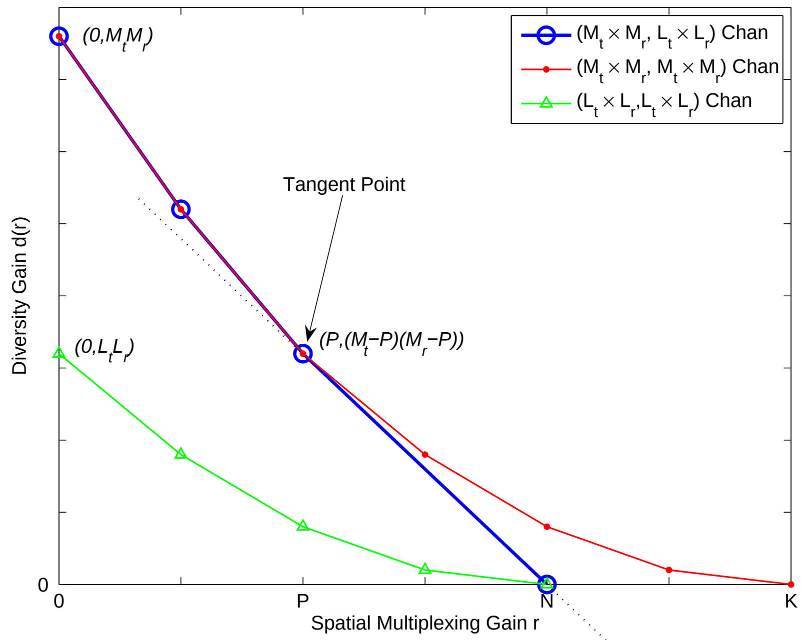

Fig. 1. The optimal D-M tradeoff the standard $(M_t \times M_r, M_t \times M_r)$ channel, $(L_t \times L_r, L_t \times L_r)$ channel, and RF chain limited $(M_t \times M_r, L_t \times L_r)$ channel. Here $N = L_t \wedge L_r$ and $K = M_t \wedge M_r$.

between those of $(M_t \times M_r, M_t \times M_r)$ and $(L_t \times L_r, L_t \times L_r)$. However, P1-P3 clearly represent a rather optimistic result. It can be seen that introducing extra antennas can greatly boost the D-M tradeoff although the maximal spatial multiplexing gain is limited by the number of RF chains. Figure 2 shows the simple case of a system with two RF chains at both transmitter and receiver side. An immediate corollary of P2 is the known result that pruning a $M_t \times M_r$ channel to a smaller $L_t \times L_r$ one can still maintain the maximal diversity gain $M_t M_r$ (corresponding to the spatial multiplexing gain $r = 0$) [12][13]. But the D-M tradeoff analysis is clearly a more complete characterization. It follows from P4 that one can always design the system with $L_t = L_r$ for reduced hardware complexity but without incurring D-M tradeoff loss.

It is not difficult to see that the major points summarized above also constitutes an answer to the following question: what is the optimal D-M tradeoff of an $L_r \times L_t$ channel pruned from an $M_r \times M_t$ original channel. Although considerable research work has been done on quantifying the influence of AS upon system performance, including channel capacity, outage probability, and diversity gain (see [9] [10] [11] [14] and the references therein), the D-M tradeoff analysis of the pruned systems has remained elusive, which is mainly due to the fact that the channel-dependent antenna selection complicates the distribution of the pruned channel. This gap is now closed.

The rest of the paper is organized as follows. In Section II we introduce the MIMO fading channel and the concept of D-M gain tradeoff. A fast AS algorithm inspired by Greedy QR decomposition is introduced in Section III. Interestingly, this algorithm turns out to be optimal in terms of the D-M gain tradeoff. In Section IV, we obtain the fundamental D-M gain tradeoff of a MIMO system with optimal AS. The D-M tradeoff result is further

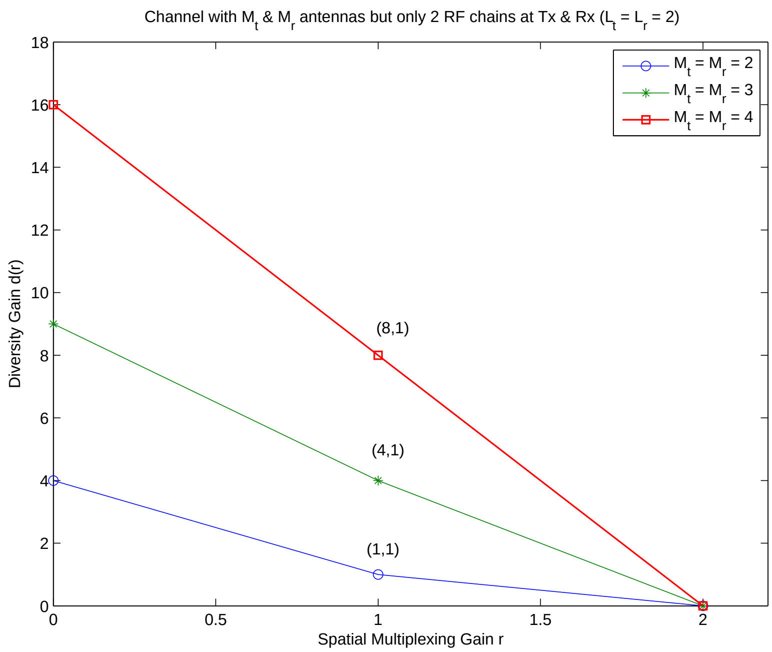

Fig. 2. The improvement of the D-M gain tradeoff by introducing more transmit and receive antennas.

extended to the more general fading channel than Rayleigh fading channel. The summary of the theoretical results and their applications are relegated to the second part of this paper [23].

II. MIMO Channel Model and D-M Tradeoff

A. Channel Model

Consider a communication system with M_t transmit and M_r receive antennas in an independent, identically distributed (iid) Rayleigh flat fading channel. The sampled baseband signal is given by

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

where $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the fading channel matrix, $\mathbf{x} \in \mathbb{C}^{M_t \times 1}$ is the information symbol vector, and $\mathbf{y} \in \mathbb{C}^{M_r \times 1}$ is the received signal. Without loss of generality, we assume that $\mathbf{z} \sim N(0, \mathbf{I}_{M_r})$ is the circularly symmetric complex Gaussian noise where $\mathbf{I}_{M_r}$ denotes the identity matrix with dimension M_r. Given $\mathbb{E}[\mathbf{x}^*\mathbf{x}] = \rho$, where $\mathbb{E}[\cdot]$ is the expectation and $(\cdot)^*$ is the conjugate transpose, the input SNR is ρ

B. Diversity-Multiplexing Gain Tradeoff and Approximate Universality

In [5], the authors established the framework of D-M tradeoff analysis in the asymptotically high SNR regime. Denote $R(\rho)$ as the data rate of any communication scheme with input SNR ρ. The diversity gain and multiplexing gain are defined as follows.

Definition 2.1: A scheme is said to have multiplexing gain r and diversity gain d if the data rate $R(\rho)$ satisfies

$$\lim_{\rho \to \infty} \frac{R(\rho)}{\log \rho} = r, \tag{2}$$

and the average error probability $P_e(\rho)$ satisfies

$$\lim_{\rho\to\infty}\frac{\log P_e(\rho)}{\log\rho}=-d. \tag{3}$$

Following [5], we adopt the symbol $\doteq$ to denote *exponential equality*, i.e., we write $f(\rho)\doteq\rho^b$ if

$$\lim_{\rho\to\infty}\frac{\log f(\rho)}{\log\rho}=b.$$

Thus (3) can be rewritten as $P_e(\rho)\doteq\rho^{-d}$.

Define the outage probability as [2]

$$P_{\text{outage}}(\rho)\triangleq\inf_{\mathbf{Q}\succeq 0,\text{tr}(\mathbf{Q})<\rho}\mathbb{P}\left(\log|\mathbf{I}+\mathbf{HQH}^*|<R(\rho)\right), \tag{4}$$

where $\mathbb{P}(\mathcal{E})$ stands for the probability of the event $\mathcal{E}$. It is rigorously proven in [24] that in the limit of increasing frame length

$$P_e=P_{\text{outage}}. \tag{5}$$

That is, not only is $P_{\text{outage}}(\rho)$ achievable in the sense that there exist codes with average frame error probability that is arbitrarily close to the outage probability but that it is also a fundamental limit in that a lower frame error probability cannot be achieved for arbitrary $\epsilon>0$, there exists a code of sufficient length for which $P_e(\rho)<P_{\text{outage}}(\rho)+\epsilon$. Conversely, for sufficiently long codes, $P_e(\rho)>P_{\text{outage}}(\rho)-\epsilon$. Moreover, this relationship is independent of the channel fading statistics [24].

Hence in the limit of long frame length, we have from (3) and (5) that the optimal diversity gain as a function of the multiplexing gain r is

$$d_{\text{opt}}(r)=-\lim_{\rho\to\infty}\frac{\log P_{\text{outage}}(\rho)}{\log\rho}. \tag{6}$$

It is easy be shown that water filling power allocation in the spatial domain yields no D-M tradeoff improvement over the isotropic transmission with the input covariance matrix $\frac{\rho}{M_t}\mathbf{I}$ [5]. Therefore

$$d_{\text{opt}}(r)=-\lim_{\rho\to\infty}\frac{\log\mathbb{P}\left(\log\left|\mathbf{I}+\mathbf{HH}^*\frac{\rho}{M_t}\right|<r\log\rho\right)}{\log\rho}. \tag{7}$$

For an iid Rayleigh channel given in (1), the optimal D-M gain tradeoff is shown to be a piece-wise linear curve obtained by connecting the following $K+1$ points [5]

$$\{(r,(M_r-r)(M_t-r))\}_{r=0}^{K}, \tag{8}$$

where $K\triangleq M_r\wedge M_t$. More recently, the D-M tradeoff analysis is extended to more general fading channels such as Rician and Nakagami fadings [19]. It is shown that some channel fadings such as Nakagami-m ($m\geq 1$) has better D-M tradeoff than Rayleigh fading channels in the regime of $r\in[0,1)$.

[2]We write $\mathbf{A}\succeq 0$ if $\mathbf{A}$ is a positive semi-definite (p.s.d.) matrix, and $\mathbf{A}\succeq\mathbf{B}$ or $\mathbf{B}\preceq\mathbf{A}$ if $\mathbf{A}-\mathbf{B}\succeq 0$.

The relationship (5) is true only for the limiting case that the frame length (code length) goes to infinity. But in terms of the coarser scale of D-M tradeoff, the constraint of frame length can be relaxed. For instance, it is proven in [5] that there exist tradeoff-optimal codes with length $T \geq M_t + M_r - 1$ in iid Rayleigh fading channels. Furthermore, [25] shows that (6) holds independent of frame length and channel fading distribution if the codes are *approximately universal*, i.e. the pairwise error probability for every pair of codewords decays exponentially with SNR given that the channel realization is not in outage. Indeed, based on [25, Theorem 3.1] it can be shown that the recently proposed cyclic division algebra (CDA) based $M_t \times M_t$ space-time block codes (STBC) with non-vanishing determinant (NDV) property [26][8][7] is approximately universal [25]. Therefore, the NDV STBCs achieve the optimal D-M gain tradeoff for any fading channels with delay $T = M_t$.

III. Fast Antenna Selection Algorithm

Suppose the MIMO system is RF chain limited, i.e. the number of RF chain is less than the number of antennas. We consider using only $L_t \leq M_t$ transmit antennas and $L_r \leq M_r$ receive antennas for data transmission. Denote $\mathcal{S}_t \subset \{1, 2, \cdots, M_t\}$ and $\mathcal{S}_r \subset \{1, 2, \cdots, M_r\}$ the sets of the indices of antennas selected at the transmitter and receiver sides, respectively. The cardinality of the sets $|\mathcal{S}_t| = L_t$ and $|\mathcal{S}_r| = L_r$. We focus on isotropic transmission since it does not incur the loss of D-M gain tradeoff. To select the antenna subsets incurring the smallest capacity loss, one needs to solve the following optimization problem

$$\begin{aligned} \mathcal{S}_t^{\text{opt}}, \mathcal{S}_r^{\text{opt}} \quad &= \quad \arg\max\nolimits_{\mathcal{S}_t, \mathcal{S}_r} \log \left| \mathbf{I} + \mathbf{H}_{\mathcal{S}_r, \mathcal{S}_t} \mathbf{H}^*_{\mathcal{S}_r, \mathcal{S}_t} \frac{\rho}{L_t} \right| \\ \text{subject to} \quad & \quad \mathcal{S}_t \subset \{1, 2, \cdots, M_t\}, \mathcal{S}_r \subset \{1, 2, \cdots, M_r\} \\ & \quad |\mathcal{S}_t| = L_t, |\mathcal{S}_r| = L_r. \end{aligned} \tag{9}$$

Here $\mathbf{H}_{\mathcal{S}_r, \mathcal{S}_t} \in \mathbb{C}^{L_r \times L_t}$ is the submatrix of $\mathbf{H}$ obtained by keeping only the rows and columns whose indices are in $\mathcal{S}_r$ and $\mathcal{S}_t$, respectively. No solution to (9) is known other than exhaustive search over the $\binom{M_r}{L_r} \cdot \binom{M_t}{L_t}$ combinations, which makes the optimal solution computationally difficult especially in the systems with many antennas [27]. Several fast yet suboptimal AS algorithms have been proposed in the literature, e.g., [16] [18]. We introduce next another new computationally efficient AS algorithm which is later shown to be D-M tradeoff-optimal.

A. Algorithm Description

The fast AS algorithm applies the same AS routine to the transmit antennas and receive antennas separately. We remark that the AS routine is closely related to the Greedy QR decomposition which plays an important role in the Greedy ordering Rate Tailored V-BLAST (GRT-VB) scheme [28]. We first show how to select $L_t < M_t$ transmit antennas. The transmit AS routine consists of L_t steps. We elaborate the first step. The subsequent steps are easily inferred.

In the first step, we go through the following procedure.

(i) Calculate the Euclidean norms $\{\|\mathbf{h}_i\|\}_{i=1}^{M_t}$ where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$.

(ii) Permute $\mathbf{h}_1$ and $\mathbf{h}_j$ where $j = \arg\max_{1\le i\le M_t}\{\|\mathbf{h}_i\|\}$. This operation can be represented by $\mathbf{H}_1 = \mathbf{H}\mathbf{\Pi}_1$ with $\mathbf{\Pi}_1$ being a permutation matrix. ($\mathbf{\Pi}_1$ degrades to be $\mathbf{I}_{M_t}$ if $j = 1$.)

(iii) Apply a Householder matrix $\mathbf{Q}_1$ to transform the first column of $\mathbf{H}_1$ to a scaled $\mathbf{e}_1$, where $\mathbf{e}_1$ is the first column of $\mathbf{I}_{M_r}$.

The procedure (i–iii) can be illustrated as follows

$$\begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix} \xrightarrow{\mathbf{Q}_1^*\mathbf{H}\mathbf{\Pi}_1} \begin{pmatrix} r_{11} & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{pmatrix}. \tag{10}$$

Note that $r_{11} = \max\{\|\mathbf{h}_i\|, 1 \le i \le M_t\}$. In the next step, the same procedure is applied to the trailing $(M_r - 1) \times (M_t - 1)$ submatrix on the right hand side of (10), which yields a permutation matrix $\mathbf{\Pi}_2$ and a Householder matrix $\mathbf{Q}_2$. After L_t recursive steps, we obtain

$$\mathbf{H}\mathbf{\Pi} = \mathbf{Q}\mathbf{R} \tag{11}$$

where $\mathbf{\Pi} = \mathbf{\Pi}_1\mathbf{\Pi}_2\cdots\mathbf{\Pi}_{L_t}$ is a permutation matrix, $\mathbf{Q} = \mathbf{Q}_1\mathbf{Q}_2\cdots\mathbf{Q}_{L_t}$ is a unitary matrix, and $\mathbf{R}$ is a matrix whose first L_t columns form an upper triangular matrix with positive diagonal elements $\{r_{ii}\}_{i=1}^{L_t}$. If the procedure is repeated for $M_t \wedge M_r$ steps, we obtain a *Greedy QR* decomposition which is used in GRT-VB [28]. Denoting $\tilde{\mathbf{\Pi}}$ and $\tilde{\mathbf{R}}$ the submatrices consisting of the first L_t columns of $\mathbf{\Pi}$ and $\mathbf{R}$, respectively, we have

$$\mathbf{H}\tilde{\mathbf{\Pi}} = \mathbf{Q}\tilde{\mathbf{R}}. \tag{12}$$

We select the transmit antennas whose indices correspond to the nonzero rows of $\tilde{\mathbf{\Pi}}$, and denote the set of their indices as $\tilde{\mathcal{S}}_t$. We denote the channel matrix after transmit AS as $\mathbf{H}_{:,\tilde{\mathcal{S}}_t} \triangleq \mathbf{H}\tilde{\mathbf{\Pi}} \in \mathbb{C}^{M_r\times L_t}$. This algorithm is computationally quite efficient as it involves only $O(L_t M_r M_t)$ complex multiplications, which is detailed in Appendix.

To select the $L_r < M_r$ receive antennas, we apply the same procedure to $\mathbf{H}^T_{:,\tilde{\mathcal{S}}_t} \in \mathbb{C}^{L_t\times M_r}$. In this case L_r recursive steps are involved. We denote by $\tilde{\mathcal{S}}_r$ the set of indexes of selected receive antennas. Hence, the pruned channel matrix can be denoted by $\mathbf{H}_{\tilde{\mathcal{S}}_r,\tilde{\mathcal{S}}_t} \in \mathbb{C}^{L_r\times L_t}$. In general, $\tilde{\mathcal{S}}_t \ne \mathcal{S}_t^{\text{opt}}$ and $\tilde{\mathcal{S}}_r \ne \mathcal{S}_r^{\text{opt}}$, i.e., the fast AS algorithm is suboptimal. But we shall see later that this algorithm is optimal in terms of D-M gain tradeoff.

B. Bounds on Singular Values

Besides the close-to-optimal performance, a major significance of this fast AS algorithm is that it enables us to reveal the important relationship between the singular values of $\mathbf{H}$ and $\mathbf{H}_{\tilde{\mathcal{S}}_r,\tilde{\mathcal{S}}_t}$.

Theorem 3.1: Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$ be the singular values of $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$. Let $\breve{\lambda}_1 \geq \cdots \geq \breve{\lambda}_N$ ($N \triangleq L_t \wedge L_r$) be the singular values of the pruned channel matrix $\mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t}$ obtained using the proposed fast AS algorithm. Then

$$\lambda_n^2 \prod_{i=1}^{n} \frac{1}{(M_r - i + 1)(M_t - i + 1)} \leq \breve{\lambda}_n^2 \leq \lambda_n^2, \quad n = 1, \cdots, N, \tag{13}$$

and

$$\prod_{n=1}^{N} \frac{\lambda_n^2}{(M_r - n + 1)(M_t - n + 1)} \leq \prod_{n=1}^{N} \breve{\lambda}_n^2 \leq \prod_{n=1}^{N} \lambda_n^2. \tag{14}$$

An important corollary of Theorem 3.1 is the following; compared to the transmission over the N strongest eigen-subchannels of the full system, the mutual information loss of the pruned channel obtained via the fast AS algorithm is upper bounded by a finite constant without regard to SNR, which can be seen as follows. After transmit AS, the mutual information of input and output of the pruned channel with input SNR ρ is

$$\begin{aligned} I(\mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t}, \rho) &= \log \left| \mathbf{I} + \mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t} \mathbf{H}^*_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t} \tfrac{\rho}{L_t} \right| \\ &= \textstyle\sum_{n=1}^{N} \log \left(1 + \frac{\breve{\lambda}_n^2 \rho}{L_t} \right). \end{aligned} \tag{15}$$

In contrast, if the input power is uniformly loaded on the strongest N eigen-subchannels of the full system, the channel mutual information is

$$I(\mathbf{\Lambda}_N, \rho) = \sum_{n=1}^{N} \log \left(1 + \frac{\lambda_n^2 \rho}{N} \right), \tag{16}$$

where $\mathbf{\Lambda}_N$ is a diagonal matrix consisting of the N largest eigenvalues of $\mathbf{H}\mathbf{H}^*$. Therefore,

$$\begin{aligned} & I(\mathbf{\Lambda}_N, \rho) - I(\mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t}, \rho) \\ &= \sum_{n=1}^{N} \log \left(\frac{1 + \frac{\rho \lambda_n^2}{N}}{1 + \frac{\rho \breve{\lambda}_n^2}{L_t}} \right) \\ &\leq \sum_{n=1}^{N} \log \left(\frac{L_t \lambda_n^2}{N \breve{\lambda}_n^2} \right) \\ \text{(by (14))} \quad &\leq \sum_{n=1}^{N} \log \left(\frac{L_t (M_t - n + 1)(M_r - n + 1)}{N} \right) \triangleq C. \end{aligned} \tag{17}$$

On the other hand, it follows from the upper bound $\lambda_n \geq \breve{\lambda}_n$ that $I(\mathbf{\Lambda}_N, \rho) \geq I(\mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t}, \rho)$. Hence we have obtained the following bounds:

$$0 \leq I(\mathbf{\Lambda}_N, \rho) - \log \left| \mathbf{I} + \mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t} \mathbf{H}^*_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t} \frac{\rho}{N} \right| \leq C, \tag{18}$$

where C is a finite constant. We deduce from these bounds that the D-M tradeoff of the pruned channel is the same as that of the transmission constrained over the N strongest eigen-subchannels of the full system, since a finite mutual information gap amounts to a finite scaling of input SNR, which does not influence the SNR exponent of the outage probability. This observation is indeed the cornerstone of our analysis in the next section. It is also worth emphasizing that Theorem 3.1 is purely a linear algebra result and is irrelevant to the distribution of $\mathbf{H}$.

To prove Theorem 3.1, we need to establish the following two lemmas.

Lemma 3.2: Consider a matrix $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ with singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$ $(K = M_t \wedge M_r)$. Denote a_i the Euclidean norm of the ith column of $\mathbf{H}$, and b_i the Euclidean norm of the ith row of $\mathbf{H}$. We have

$$\sum_{i=1}^{k} \lambda_i^2 \geq \sum_{i=1}^{k} a_{[i]}^2, \quad k = 1, 2, \cdots, K, \tag{19}$$

and

$$\sum_{i=1}^{k} \lambda_i^2 \geq \sum_{i=1}^{k} b_{[i]}^2, \quad k = 1, 2, \cdots, K, \tag{20}$$

where $a_{[i]}$ and $b_{[i]}$ are the ith largest elements of the sequences $\{a\}_{i=1}^{M_t}$ and $\{b\}_{i=1}^{M_r}$, respectively.

Proof: Note that $\{\lambda_i^2\}_{i=1}^{K}$ are the largest K singular values of the Hermitian matrices $\mathbf{H}\mathbf{H}^*$ and $\mathbf{H}^*\mathbf{H}$. Also note that $\mathbf{H}^*\mathbf{H}$ has diagonal elements $\{a_i^2\}_{i=1}^{M_t}$ and $\mathbf{H}\mathbf{H}^*$ has diagonal elements $\{b_i^2\}_{i=1}^{M_r}$. The lemma follows immediately from Schur-Horn's Theorem [29, Theorems 4.3.26], which says that the diagonal elements of a positive semi-definite matrix is additively majorized by the singular (eigen) values [30]. ■

Lemma 3.3: Applying the Greedy QR decomposition (i.e. the procedure (10) is applied K times) to $\mathbf{H}$ yields

$$\mathbf{H}\mathbf{\Pi} = \mathbf{Q}\mathbf{R}. \tag{21}$$

Denote r_{kk}^2 the squared kth diagonal elements of $\mathbf{R}$. Then

$$r_{kk}^2 \geq \frac{\sum_{i=k}^{K} \lambda_i^2}{M_t - k + 1}, \quad k = 1, 2, \cdots, K. \tag{22}$$

Proof: Recall that the Greedy QR decomposition is achieved by successively applying the procedure illustrated in (10) K times. According to the procedure (i)-(iii) given in Section III-A,

$$r_{11}^2 = \max_{1 \leq i \leq M_t} \{\|\mathbf{h}_i\|^2\} \geq \frac{1}{M_t} \|\mathbf{H}\|_F^2 = \frac{\sum_{i=1}^{K} \lambda_i^2}{M_t} \tag{23}$$

where $\|\cdot\|_F$ stands for the Frobenius norm. Hence (22) is true for $k = 1$. At the kth step $(2 \leq k \leq K)$, we have $\mathbf{H}\mathbf{\Pi}_1 \cdots \mathbf{\Pi}_{k-1} = \mathbf{Q}_1 \cdots \mathbf{Q}_{k-1}\mathbf{R}^{(k)}$ with

$$\mathbf{R}^{(k)} = \begin{pmatrix} r_{11} & \times \cdots & \cdots & \cdots & \cdots & \times \\ 0 & \ddots & \ddots & \cdots & \cdots & \times \\ 0 & 0 & r_{k-1,k-1} & \cdots & \cdots & \times \\ 0 & \cdots & 0 & * & \cdots & * \\ 0 & \cdots & 0 & * & \cdots & * \\ 0 & \cdots & 0 & * & \cdots & * \end{pmatrix}. \tag{24}$$

Because left and right multiplying a matrix by any unitary matrix does not change its singular values, $\mathbf{R}^{(k)}$ has the same singular values as $\mathbf{H}$. According to Lemma 3.2, the first $k-1$ rows of $\mathbf{R}^{(k)}$ have Frobenius norm less than $\sum_{i=1}^{k-1} \lambda_i^2$. Therefore the trailing $(M_r - k + 1) \times (M_t - k + 1)$ submatrix (denoted by $*$'s) has Frobenius norm larger than $\sum_{i=k}^{K} \lambda_i^2$. It follows that

$$r_{kk}^2 \geq \frac{1}{M_t - k + 1} \sum_{i=k}^{K} \lambda_i^2, \quad 2 \leq k \leq K. \tag{25}$$

Combining (23) and (25), we have proven the lemma. ■

Now we are ready to prove Theorem 3.1.

Proof: (of Theorem 3.1) We first prove the upper bound in (13). Consider the pruned matrix $\mathbf{H}_{:,\mathcal{S}_t}$ which consists of the columns of $\mathbf{H}$ whose indices belong to $\mathcal{S}_t$. Denote $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_N$ the N largest singular values of $\mathbf{H}_{:,\mathcal{S}_t}$. Because $\mathbf{H}_{:,\mathcal{S}_t}\mathbf{H}_{:,\mathcal{S}_t}^* \preceq \mathbf{H}\mathbf{H}^*$, $\tilde{\lambda}_k^2 \leq \lambda_k^2$ for $1 \leq k \leq N$ [31]. Because $\mathbf{H}_{\mathcal{S}_r,\mathcal{S}_t}^*\mathbf{H}_{\mathcal{S}_r,\mathcal{S}_t} \preceq \mathbf{H}_{:,\mathcal{S}_t}^*\mathbf{H}_{:,\mathcal{S}_t}$, $\breve{\lambda}_k^2 \leq \tilde{\lambda}_k^2$ for $1 \leq k \leq N$. Hence $\breve{\lambda}_k^2 \leq \lambda_k^2$ and the upper bound in (13) is proven.

Let $\mathbf{H}_{:,\mathcal{S}_t} = \mathbf{Q}\tilde{\mathbf{R}}$ be the QR decomposition where $\mathbf{Q}$ is given in (21), and $\tilde{\mathbf{R}}$ is the submatrix consisting of the first L_t columns of $\mathbf{R}$ there. Clearly $\tilde{\mathbf{R}}$ has the same diagonal as $\mathbf{R}$. It follows from Lema 3.3 that $r_{ii}^2 \geq \frac{\lambda_i^2}{M_t - i + 1}$. Hence

$$\prod_{i=1}^{n} r_{ii}^2 \geq \prod_{i=1}^{n} \frac{\lambda_{ii}^2}{M_t - i + 1}. \tag{26}$$

Note that the singular values of $\tilde{\mathbf{R}}$ and $\mathbf{H}_{:,\mathcal{S}_t}$ are the same. Recall the fact that for an upper triangular matrix, the squared diagonal elements are multiplicatively majorized by its squared singular values [32][30], i.e.,

$$\prod_{i=1}^{n} \tilde{\lambda}_{ii}^2 \geq \prod_{i=1}^{n} r_{ii}^2 \geq \prod_{i=1}^{n} \frac{\lambda_i^2}{M_t - i + 1}, \quad \text{for} \quad 1 \leq i \leq N. \tag{27}$$

Combining (27) and the proven upper bound that $\tilde{\lambda}_i \leq \lambda_i$, we have

$$\tilde{\lambda}_n^2 \geq \lambda_n^2 \prod_{i=1}^{n} \frac{1}{M_t - i + 1}, \quad n = 1, 2, \cdots, N. \tag{28}$$

Applying the same AS procedure to $\mathbf{H}_{:,\mathcal{S}_t}^T$, we obtain $\mathbf{H}_{\mathcal{S}_r,\mathcal{S}_t}$ with singular values $\breve{\lambda}_1 \geq \ldots \geq \breve{\lambda}_N$. Following the same argument leading to (28), we can prove that

$$\breve{\lambda}_n^2 \geq \tilde{\lambda}_n^2 \prod_{i=1}^{n} \frac{1}{M_r - i + 1}, \quad n = 1, 2, \cdots, N. \tag{29}$$

Combining (28) and (29), we have proven the lower bound in (13).

The upper bound in (14) is trivial given the proven upper bound $\breve{\lambda}_n \leq \lambda_n$ for $\forall n$. We see from (27) that

$$\prod_{n=1}^{N} \tilde{\lambda}_n^2 \geq \prod_{n=1}^{N} \frac{\lambda_n^2}{M_t - n + 1}. \tag{30}$$

In a similar vein,

$$\prod_{n=1}^{N} \breve{\lambda}_n^2 \geq \prod_{n=1}^{N} \frac{\tilde{\lambda}_n^2}{M_r - n + 1}. \tag{31}$$

It follows from (30) and (31) that

$$\prod_{n=1}^{N} \breve{\lambda}_n^2 \geq \prod_{n=1}^{N} \frac{\lambda_n^2}{(M_r - n + 1)(M_t - n + 1)}. \tag{32}$$

The theorem is proven. ■

IV. Diversity-Multiplexing Tradeoff Analysis

In this section, we derive the fundamental D-M tradeoff of an optimally pruned channel. We first establish the result in the case of iid Rayleigh fading channel. The extension to the general fading channels is presented in Section IV-B.

A. Rayleigh Fading Channel

For the pruned iid Rayleigh fading channel, the optimal D-M tradeoff is summarized in the following theorem.

Theorem 4.1: Consider pruning the $M_r \times M_t$ Rayleigh fading MIMO channel given in (1) to a smaller one with L_t transmit and L_r receive antennas using the proposed fast AS algorithm. The optimal D-M gain tradeoff of the pruned system is a piecewise linear curve obtained by connecting the following $P + 2$ points

$$\{n, (M_r - n)(M_t - n)\}_{n=0}^{P}, \ (N, 0), \tag{33}$$

where $N \triangleq L_r \wedge L_t$, and

$$\begin{aligned} P \ = \ & \arg\min_p \ \tfrac{(M_r - p)(M_t - p)}{N - p} \\ & \text{subject to} \ \ 0 \leq p \leq N - 1, \ p \in \mathbb{Z}. \end{aligned} \tag{34}$$

Moreover, this D-M gain tradeoff is an upper bound for any AS strategy.

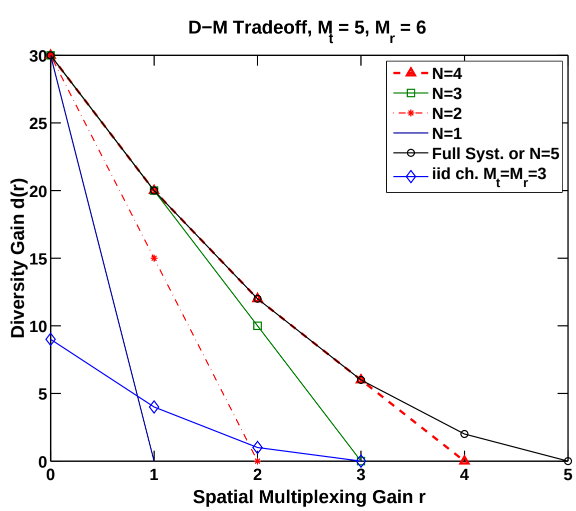

Fig. 3. Optimal D-M tradeoffs of full and pruned MIMO systems

As illustrated in Figure 3, Theorem 4.1 says that a good AS algorithm incurs no diversity gain loss if the multiplexing gain is less than P. In particular, if $N = 1$ ($N = L_t \wedge L_r$), the tradeoff is the line connecting $(0, M_t M_r)$ and $(1, 0)$. On the other hand, if the antenna selection is made such that $N = K$ ($K = M_t \wedge M_r$), then $P = K - 1$ since the slope of the curve corresponding to $d \in (K-1, K)$ is the smallest among all the K pieces. Therefore the pruned system has the same D-M tradeoff as the full system. For instance, as in Figure 3, if $L_t = L_r = 5$, the pruned system still has the D-M tradeoff as shown in $-\circ-$.

On the other hand, Theorem 4.1 also quantifies the *improvement* of the D-M tradeoff by introducing additional antennas at transmitter and receiver without increasing the number of RF chains and the size of codes. In Figure 3 the D-M tradeoff of a 3×3 iid Rayleigh channel is presented ($-\Diamond-$). Comparing the lines $-\Diamond-$ and $-\Box-$, we see the dramatic improvement by introducing additional three receive antennas and two transmit antennas, meanwhile the increase of hardware complexity is minimal by using AS.

We now prove Theorem 4.1.

Proof: The proof of the theorem contains two parts. The first part is the derivations leading to (41) which lean heavily on the techniques used in [5]. Hence we only give a sketch for this part. The second part is the solution of the optimization problem (41).

Denote

$$P_{\text{outage,p}}(r, \rho) = \mathbb{P}\left(\log\left|\mathbf{I} + \mathbf{H}_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t}\mathbf{H}^*_{\tilde{\mathcal{S}}_r, \tilde{\mathcal{S}}_t}\frac{\rho}{L_t}\right| < r\log\rho\right)$$

the outage probability of the pruned channel with input SNR ρ and multiplexing gain r. Here the subscript "p" stands for "pruned". The diversity gain of the pruned system with multiplexing gain r is

$$d_{\text{p}}(r) = -\lim_{\rho\to\infty}\frac{\log P_{\text{outage,p}}(r, \rho)}{\log\rho}.$$

Since a finite mutual information gap amounts to a finite scaling of input SNR, which does not influence the SNR exponent of the outage probability, it follows from (18) that

$$P_{\text{outage,p}}(r, \rho) \doteq \mathbb{P}\left(\sum_{n=1}^{N}\log\left(1 + \rho\lambda_n^2\right) < r\log\rho\right). \tag{35}$$

For an i.i.d Rayleigh fading channel, the joint distribution of the ordered squared singular values of $\mathbf{H}$, $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_K^2 > 0$, is [33]

$$f(\lambda_1^2, \cdots, \lambda_K^2) = C_{M_t, M_r} \cdot \prod_{k=1}^{K}\lambda_k^{2(M-K)}\prod_{k<j}(\lambda_k^2 - \lambda_j^2)^2 e^{-\sum_i \lambda_i^2},$$

where C_{M_t, M_r} is a normalizing constant and $M \triangleq M_r \vee M_t$. Define $\lambda_k^2 = \rho^{-\alpha_k}, 1 \leq k \leq K$ where $\alpha_1 \leq \alpha_1 \leq \cdots \leq \alpha_K$ as $0 < \lambda_1 \leq \cdots \leq \lambda_K$. The distribution of $\boldsymbol{\alpha}$ is

$$f(\alpha_1, \ldots, \alpha_K) = C_{M_t, M_r}(\log\rho)^K\prod_{k=1}^{K}\rho^{-[(M-K)+1]\alpha_k}\prod_{k<j}(\rho^{-\alpha_k} - \rho^{-\alpha_j})^2 e^{-\sum_i \rho^{-\alpha_i}}. \tag{36}$$

At high SNR, $(1+\rho\lambda_n^2) \doteq \rho^{(1-\alpha_n)^+}$ with $(x)^+ = x \vee 0$. We see from (35) that

$$\begin{aligned} P_{\text{outage,p}}(r,\rho) &\doteq \mathbb{P}\left(\textstyle\sum_{n=1}^{N}(1-\alpha_n)^+ < r\right) \\ &\doteq \textstyle\int_{\mathcal{A}} \prod_{k=1}^{K} \rho^{-[(M-K)+1]\alpha_k} \prod_{k<j}(\rho^{-\alpha_k} - \rho^{-\alpha_j})^2 e^{-\sum_i \rho^{-\alpha_i}} d\boldsymbol{\alpha}. \end{aligned} \tag{37}$$

where

$$\mathcal{A} = \left\{ \boldsymbol{\alpha} : \alpha_1 \le \alpha_2 \le \cdots \le \alpha_K, \sum_{n=1}^{N}(1-\alpha_n)^+ < r \right\}. \tag{38}$$

Also note that $\exp(-\rho^{-\alpha_n})$ decreases with ρ exponentially for any $\alpha_n < 0$ and $\exp(-\rho^{-\alpha_n}) \to 1$ as $\rho \to \infty$ for $\alpha_n > 0$. Therefore

$$P_{\text{outage,p}}(r,\rho) \doteq \int_{\mathcal{A}_+} \prod_{k=1}^{K} \rho^{-[(M-K)+1]\alpha_k} \prod_{k<j}(\rho^{-\alpha_k} - \rho^{-\alpha_j})^2 d\boldsymbol{\alpha}, \tag{39}$$

where

$$\mathcal{A}_+ = \left\{ \boldsymbol{\alpha} : 0 < \alpha_1 \le \cdots \le \alpha_K, \sum_{n=1}^{N}(1-\alpha_n)^+ < r \right\}.$$

Since $\alpha_i \le \alpha_j$ for $j > i$, we can replace the term $\prod_{k<j}(\rho^{-\alpha_k} - \rho^{-\alpha_j})^2$ by $\rho^{-2(K-k)\alpha_k}$ at high SNR. Hence

$$d_{\text{p}}(r) = -\lim_{\rho\to\infty} \frac{\log \int_{\mathcal{A}_+} \prod_{k=1}^{K} \rho^{-(M+K-2k+1)\alpha_k} d\alpha_1 \cdots d\alpha_K}{\log \rho}. \tag{40}$$

According to Laplace's principle, the integral is dominated by the term corresponding to the largest SNR exponent as $\rho \to \infty$. Hence

$$\begin{aligned} d_{\text{p}}(r) &= \inf_{\alpha_k} \textstyle\sum_{k=1}^{K}(M+K-2k+1)\alpha_k \\ \text{subject to} &\quad \{\alpha_k\}_{k=1}^{K} \in \mathcal{A}_+. \end{aligned} \tag{41}$$

We observe that at an optimal solution to (41) it must be true that (i) $\alpha_k \le 1$ for $\forall k$ and (ii) $\alpha_k = \alpha_N$ for $N \le k \le K$. The argument for (ii) is trivial. As for (i), if there are some elements of $\boldsymbol{\alpha}$ which are greater than one, without violating the constraint, we can set them to be one and reduce the objective function in (41). Also note that $M+K = M_t + M_r$ since $M = M_t \vee M_r$ and $K = M_t \wedge M_r$. Using these observations and denoting

$$c_n = \begin{cases} M_t + M_r - 2n + 1 & n = 1, \cdots, N-1 \\ (M_t - N + 1)(M_r - N + 1) & n = N, \end{cases} \tag{42}$$

we can rewrite (41) as

$$\begin{aligned} d_{\text{p}}(r) &= \inf_{\alpha_n} \textstyle\sum_{n=1}^{N} c_n \alpha_n \\ \text{subject to} &\quad 0 < \alpha_1 \le \alpha_2 \le \cdots \le \alpha_N < 1 \\ &\quad \textstyle\sum_{n=1}^{N} \alpha_n > N - r. \end{aligned} \tag{43}$$

To make the problem physically meaningful, we focus on the case $0 \le r \le N$. It is straightforward to show that (43) is a convex problem whose Lagrangian is

$$\mathcal{L}(\boldsymbol{\alpha}, \mu, \boldsymbol{\gamma}) = \sum_{n=1}^{N} c_n \alpha_n - \mu \left[\sum_{n=1}^{N} \alpha_n - (N-r) \right] - \gamma_1 \alpha_1 - \sum_{n=2}^{N} \gamma_n (\alpha_n - \alpha_{n-1}) - \gamma_{N+1}(1 - \alpha_N), \tag{44}$$

where the multipliers $\mu, \boldsymbol{\gamma} \geq 0$. It can been seen that the Slater's condition is satisfied for $r > 0$, i.e., the constraint set has nonempty interior [34]. Hence according to the convex optimization theory [34]

$$d_{\mathrm{p}}(r) = \inf_{\boldsymbol{\alpha}} \sup_{\mu, \boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\alpha}, \mu, \boldsymbol{\gamma}). \tag{45}$$

According to the complementary slackness condition,

$$\gamma_1 \alpha_1 = \sum_{n=2}^{N} \gamma_n (\alpha_n - \alpha_{n-1}) = \gamma_{N+1}(1 - \alpha_N) = 0. \tag{46}$$

Equating to zero the partial derivative of the Lagrangian with respect to α_n's, we obtain

$$c_n = \mu + \gamma_n - \gamma_{n+1}, \quad 1 \leq n \leq N. \tag{47}$$

Inserting (46) and (47) into (45) yields

$$\begin{aligned} d_{\mathrm{p}}(r) &= \inf_{\boldsymbol{\alpha}} \sup_{\mu, \boldsymbol{\gamma}} \textstyle\sum_{n=1}^{N} (\mu + \gamma_n - \gamma_{n+1}) \alpha_n \\ &= \inf_{\boldsymbol{\alpha}} \sup_{\mu, \boldsymbol{\gamma}} \mu \textstyle\sum_{n=1}^{N} \alpha_n + \textstyle\sum_{n=1}^{N} (\gamma_n - \gamma_{n+1}) \alpha_n. \end{aligned} \tag{48}$$

As we can rewrite $\sum_{n=1}^{N} (\gamma_n - \gamma_{n+1}) \alpha_n = \gamma_1 \alpha_1 + \sum_{n=2}^{N} \gamma_n (\alpha_n - \alpha_{n-1}) - \gamma_{N+1} \alpha_N$, it follows from (46) that

$$\sum_{n=1}^{N} (\gamma_n - \gamma_{n+1}) \alpha_n = -\gamma_{N+1}. \tag{49}$$

Combining (48) and (49), we obtain

$$\begin{aligned} d_{\mathrm{p}}(r) &= \inf_{\boldsymbol{\alpha}} \sup_{\mu, \boldsymbol{\gamma}} \mu \textstyle\sum_{n=1}^{N} \alpha_n - \gamma_{N+1} \\ &= \sup_{\mu, \boldsymbol{\gamma}} \mu (N - r) - \gamma_{N+1}. \end{aligned} \tag{50}$$

It follows from the N equations in (47) that

$$\gamma_{n+1} = \gamma_1 + n\mu - \sum_{i=1}^{n} c_i, \quad 1 \leq n \leq N. \tag{51}$$

In particular $\gamma_{N+1} = \gamma_1 + N\mu - \sum_{i=1}^{N} c_i$. Therefore

$$\begin{aligned} d_{\mathrm{p}}(r) &= \sup_{\mu, \boldsymbol{\gamma}} \mu (N - r) - (\gamma_1 + N\mu - \textstyle\sum_{i=1}^{N} c_i) \\ &= \textstyle\sum_{i=1}^{N} c_i - \inf_{\mu, \boldsymbol{\gamma}} (\gamma_1 + \mu r) \\ &= M_t M_r - \inf_{\mu, \gamma_1} (\gamma_1 + \mu r). \end{aligned} \tag{52}$$

The constraints of μ, γ_1 are implied in (51), i.e.,

$$\gamma_1 + n\mu - \sum_{i=1}^{n} c_i \geq 0, \quad 1 \leq n \leq N. \tag{53}$$

Let us consider the optimization problem

$$\begin{aligned} & \inf_{\mu, \gamma_1} (\gamma_1 + \mu r) \\ \text{subject to} \quad & \gamma_1 + n\mu - \textstyle\sum_{i=1}^{n} c_i \geq 0, \quad 1 \leq n \leq N, \end{aligned} \tag{54}$$

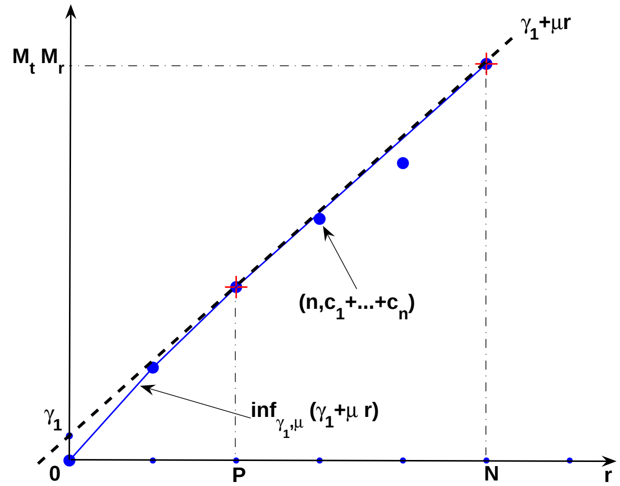

Fig. 4. Visualization of $\gamma_1 + n\mu$ and $\sum_{i=1}^{n} c_i$.

which is depicted in Figure 4. The cost function $\gamma_1 + \mu r$ can be represented by the dashed line passing the point $(0, \gamma_1)$ with slope μ. The $N+1$ dots in Figure 4 have coordinates $(n, \sum_{i=1}^{n} c_n)$, $0 \le n \le N$, which are indexed from left to right as the zero-th to the Nth point. The constraints ($\gamma_1 + n\mu \ge \sum_{i=1}^{n} c_i$) mean that the dashed line should always be above the $N+1$ points. Hence we see that the function $\inf_{\gamma_1,\mu} \gamma_1 + \mu x$ subject to the constraints given in (54) is the upper edge of the convex hull spanned by the $N+1$ points. The slope of the straight line passing the two points with indices p ($p < N$) and N is

$$\mu = \frac{\sum_{n=p}^{N} c_n}{N-p} = \frac{(M_r - p)(M_t - p)}{N-p}$$

We refer to the points on the edge of the convex hull as "active points". The Nth point is an "active point". Then the adjacent "active point" to the Nth point must have the index defined in (34). Otherwise a line passing the Nth and the pth ($p \ne P$) points would be below the Pth point, which violates the constraint of (53). Because c_n decreases as n increases for $n \le N-1$, all the points with index less than P are "active". According to (42),

$$\sum_{i=1}^{n} c_i = \begin{cases} (M_t + M_r - n)n & 1 \le n \le N-1 \\ M_t M_r & n = N. \end{cases}$$

Therefore the edge of the convex hull, i.e., $\inf_{\gamma_1,\mu} \mu + \gamma_1 r$, is obtained by connecting the $P+2$ points with coordinates

$$\{(n, (M_t + M_r - n)n)\}_{n=0}^{P}, \quad (N, M_t M_r). \tag{55}$$

Since $d_{\mathrm{p}}(r) = M_t M_r - (\inf_{\gamma_1,\mu} \mu + \gamma_1 r)$ and $M_t M_r - (M_t + M_r - n)n = (M_r - n)(M_t - n)$, we conclude that $d_{\mathrm{p}}(r)$ is also a piecewise linear curve obtained by connecting

$$\{(n, (M_t - n)(M_r - n))\}_{n=0}^{P}, \quad (N, 0). \tag{56}$$

To see that this D-M tradeoff is an upper bound for any AS strategy, it is sufficient to note that

$$I(\mathbf{\Lambda}_N, \rho) \geq \log \left| \mathbf{I} + \mathbf{H}_{\mathcal{S}_r, \mathcal{S}_t} \mathbf{H}^*_{\mathcal{S}_r, \mathcal{S}_t} \frac{\rho}{N} \right|$$

holds for any $\mathcal{S}_r, \mathcal{S}_t$ with $|\mathcal{S}_r| = L_r$ and $|\mathcal{S}_t| = L_t$, where $I(\mathbf{\Lambda}_N, \rho)$ is defined in (16). We have proved the theorem.

■

B. Extension to Non-Rayleigh Fading Channel

The D-M tradeoff analysis in [5] is based on the assumption of iid Rayleigh fading channel. Recently, [35] extends the result of [5] to the more general cases where the entries of $\mathbf{H}$ have distribution:

$$f_{|h_{ij}|}(x) = a x^t e^{-b|x-c|^\beta}, \quad x \geq 0, \tag{57}$$

which includes the Rician and Nakagami fading channels as special cases. It is shown that for an $M_r \times M_t$ MIMO channel with distribution given in (57), the optimal D-M tradeoff is a piece-wise linear curve obtained by connecting the $K+1$ points:

$$\left(0, \left(1 + \frac{t}{2}\right) M_t M_r\right), \ \{(k, (M_r - k)(M_t - k))\}_{k=1}^{K}. \tag{58}$$

That is, the channel with distribution (57) may have better D-M tradeoff than the Rayleigh channel for $r \in [0, 1)$, while it has the same tradeoff as the Rayleigh channel for $r \in [1, K]$.

Note that the bound in (18) is obtained without assuming the distribution of $\mathbf{H}$. Therefore according to (35),

$$P_{\text{outage,p}}(r, \rho) \doteq \mathbb{P}\left(\sum_{n=1}^{N} \log\left(1 + \rho\lambda_n^2\right) < r \log \rho\right). \tag{59}$$

Combining the results in [35] and the derivations leading to (39), we can show that

$$P_{\text{outage,p}}(r, \rho) \doteq \int_{\mathcal{A}_+} \prod_{k=1}^{K} \rho^{-[(M-K)+1]\alpha_k} \prod_{k<j} (\rho^{-\alpha_k} - \rho^{-\alpha_j})^2 \rho^{-\frac{t}{2} M_r M_t \alpha_1} d\boldsymbol{\alpha}, \tag{60}$$

Similar to the derivations from (39) to (43), we can obtain the optimal D-M tradeoff of the pruned channel as

$$\begin{array}{rcl} d_{\text{p}}(r) & = & \inf_{\alpha_n} \sum_{n=1}^{N} c_n \alpha_n \\ \text{subject to} & & 0 < \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_N < 1 \\ & & \sum_{n=1}^{N} \alpha_n > N - r. \end{array} \tag{61}$$

The only difference between (61) and (43) is that here

$$c_n = \begin{cases} \frac{t}{2} M_r M_t + M_t + M_r - 1 & n = 1 \\ M_t + M_r - 2n + 1 & n = 2, \cdots, N-1 \\ (M_t - N + 1)(M_r - N + 1) & n = N. \end{cases}$$

Following the arguments similar to those in the proof of Theorem 4.1, we can show that for the general fading channel, the optimal D-M tradeoff of the pruned channel is also a piecewise linear curve obtained by connecting the following $P+2$ points

$$\left\{n, (M_r-n)(M_t-n)+\frac{tM_tM_r}{2}\delta(n)\right\}_{n=0}^{P}, \ (N,0), \tag{62}$$

where

$$\delta(n)=\begin{cases}1 & \text{for } n=0\\ 0 & \text{for } n\neq 0,\end{cases}$$

and

$$\begin{aligned} P \quad = \quad & \arg\min_p \quad \frac{(M_r-p)(M_t-p)+\frac{tM_tM_r}{2}\delta(p)}{N-p} \\ & \text{subject to} \quad 0\le p\le N-1,\ p\in\mathbb{Z}. \end{aligned} \tag{63}$$

Consider an iid channel with the same dimension as the one in Figure 3 but with distribution given in (57) where $t=2$. Figure 5 shows that the optimal D-M gain tradeoff of the pruned MIMO channels. Comparing Figures 3 and 5, we see that for $N=2$, the pruned non-Rayleigh channel has better D-M tradeoff than Rayleigh channel even for $r\in(1,2)$.

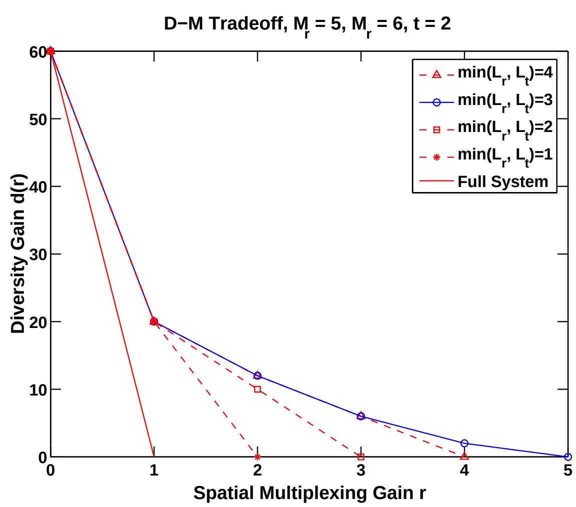

Fig. 5. Optimal D-M tradeoffs of full and pruned non-Rayleigh MIMO channels whose entries have distribution (57) with $t=2$.

We conclude this section by emphasizing that the CDA based STBCs [8] with minimum delay $T=L_t$ achieve the optimal D-M gain tradeoff of the pruned channel since they are approximately universal and hence is D-M tradeoff-optimal in any fading channel [25].

V. Conclusion

The conclusion is given at the end of [23].

APPENDIX

We analyze the numbers of complex multiplication involved in the proposed AS algorithm. At the ith step, the AS algorithm compares the column norms of an $(M_r - i + 1) \times (M_t - i + 1)$ submatrix, which requires $(M_r - i+1)(M_t - i+1)$ multiplications. Calculating the Householder $\mathbf{Q}_i$ and left multiplying the channel matrix by $\mathbf{Q}_i$ requires about $3(M_r - i + 1) + 4(M_r - i+1)(M_t - i+1)$ multiplications (see, e.g., [**?**, Sections 5.1.3 and 5.1.4]). Hence the total number of complex multiplications in the transmit AS is

$$\begin{aligned} & \sum_{i=1}^{L_t} [5(M_r - i+1)(M_t - i+1) + 3(M_r - i+1)] \\ = \ & \frac{1}{6} L_t \left(30 M_r M_t + 10 L_t^2 - 15 L_t M_t - 15 L_t M_r + 33 M_r + 15 M_t - 24 L_r + 14\right). \end{aligned} \tag{64}$$

We see that the computational complexity of our proposed AS algorithm is of order $O(L_t M_r M_t)$, which is similar to the algorithm proposed in [17] which has complexity $O((M_r + M_t) M_t L_t)$. Clearly, the computational complexity of receiving antennas selection after transmit antenna selection is of order $O(L_r M_r L_t)$.

REFERENCES

[1] I. E. Telatar, "Capacity of multiple antenna Gaussian channels," *AT&T Technical Memorandum*, June 1995.

[2] 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.

[3] 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.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[5] 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.

[6] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 968–985, June 2004.

[7] F. Oggier, G. Rekaya, J.-C. Belfiore, and E. Viterbo, "Perfect space-time block codes," *IEEE Transactions on Information Theory*, vol. 52, pp. 3903– 3912, September 2006.

[8] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, "Explicit construction of space-time block codes achieving the diversity-multiplexing gain tradeoff," *IEEE Transactions on Information Theory*, vol. 52, pp. 3869– 3884, September 2006.

[9] A. Molisch and M. Win, "MIMO systems with antenna selection," *IEEE Microwave Magazine*, vol. 5, pp. 46 – 56, March 2004.

[10] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Communications Magazine*, pp. 68 – 73, October 2004.

[11] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO flat-fading channels: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 49, pp. 2687 – 2696, October 2003.

[12] I. Bahceci, T. M. Duman, and Y. Altunbasak, "Antenna selection for multiple-antenna transmission systems: Performance analysis and code construction," *IEEE Transactions on Information Theory*, vol. 49, pp. 2669–2681, October 2003.

[13] H. Shen and A. Ghrayeb, "Analysis of the outage probability for mimo systems with receive antenna selection," *IEEE Transactions on Vehicular Technology*, vol. 55, pp. 1435 – 1441, July 2006.

[14] H. Zhang, H. Dai, Q. Zhou, and B. L. Hughes, "On the diversity order of spatial multiplexing systems with transmit antenna selection: A geometrical approach," *IEEE Transactions on Information Theory*, vol. 52, pp. 5297–5311, December 2006.

[15] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic space time architectures," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 303–320, April 2003.

[16] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: Theory and algorithms," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2796 – 2807, November 2003.

[17] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Transactions on Signal Processing*, vol. 52, pp. 339 – 347, February 2005.

[18] A. Molisch, M. Win, and J. H. Choi, Y. S.and Winters, "Capacity of MIMO systems with antenna selection," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 1759 – 1772, July 2005.

[19] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *IEEE Transactions on Information Theory*, vol. 53, pp. 1549 – 1557, April 2007.

[20] Y. Jiang, W. Hager, and J. Li, "The geometric mean decomposition," *Linear Algebra and Its Applications*, vol. 396, pp. 373–384, February 2005.

[21] Y. Jiang, J. Li, and W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3791 – 3803, October 2005.

[22] J.-K. Zhang, A. Kavčić, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Transactions on Information Theory*, vol. 51, pp. 154–172, January 2005.

[23] Y. Jiang and M. Varanasi, "Diversity-multiplexing tradeoff of MIMO systems with antenna selection part ii: Case study," *IEEE Transactions on Wireless Communication*, submitted.

[24] N. Prasad and M. K. Varanasi, "Outage theorems for mimo block fading channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 5284–5296, Dec. 2006.

[25] S. Tavildar and P. Viswanath, "Approximately universal codes over slow fading channels," *IEEE Transactions on Information Theory*, vol. 52, pp. 3233–3258, July 2006.

[26] T. Kiran and B. S. Rajan, "Stbc-schemes with nonvanishing determinant for certain number of transmit antennas," *IEEE Transactions on Information Theory*, vol. 51, p. 2984C2992, August 2005.

[27] T. L. Marzetta, "The case for many (¿16) antennas at the base station," *Information Theory and Application Colloquium*, Feb. 08 2007.

[28] Y. Jiang and M. Varanasi, "A class of spatial multiplexing architecture combining rate-tailored transmission and ordered BLAST detection Part I: On ordered detection," *IEEE Transactions on Wireless Communications*, submitted.

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

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

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

[32] Y. Jiang, W. Hager, and J. Li, "Tunable channel decomposition for MIMO communications using channel state information," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4405 – 4418, November 2006.

[33] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley Series in Probability and Statistics, 1982.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2002.

[35] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *Conference on Information Sciences and Systems*, March 22 - 24, 2006.