On the Diversity Order of Spatial Multiplexing Systems With Transmit Antenna Selection: A Geometrical Approach

Hongyuan Zhang, *Member, IEEE*, Huaiyu Dai, *Member, IEEE*, Quan Zhou, *Member, IEEE*, and Brian L. Hughes, *Member, IEEE*

***Abstract*—In recent years, the remarkable ability of multiple-input–multiple-output (MIMO) wireless communication systems to provide spatial diversity or multiplexing gains has been clearly demonstrated. For MIMO diversity schemes, it is well known that antenna selection methods that optimize the postprocessing signal-to-noise ratio (SNR) can preserve the diversity order of the original full-size MIMO system. On the other hand, the diversity order achieved by antenna selection in spatial multiplexing systems, especially those exploiting practical coding and decoding schemes, has not thus far been rigorously analyzed. In this paper, a geometrical framework is proposed to theoretically analyze the diversity order achieved by transmit antenna selection for separately encoded spatial multiplexing systems with linear and decision-feedback receivers. When two antennas are selected from the transmitter, the exact achievable diversity order is rigorously derived, which previously only appears as conjectures based on numerical results in the literature. If more than two antennas are selected, we give lower and upper bounds on the achievable diversity order. Furthermore, the same geometrical approach is used to evaluate the diversity-multiplexing tradeoff in spatial multiplexing systems with transmit antenna selection.**

***Index Terms*—Antenna selection, diversity order, diversity and multiplexing tradeoff, multiple-input–multiple-output (MIMO), spatial multiplexing.**

I. Introduction

MULTIPLE-INPUT–MULTIPLE-OUTPUT (MIMO) techniques are expected to be widely employed in future wireless communications to address the ever-increasing demand for capacity. A major potential problem with MIMO is increased hardware cost due to multiple analog/RF front-ends, which has recently motivated the investigation of antenna selection techniques for MIMO systems [2]. In many scenarios, judicious antenna selection may incur little or no loss in system performance, while significantly reducing system cost.

Manuscript received June 8, 2005; revised March 15, 2006. This work was supported in part by the National Science Foundation by Grant CCR 0312696 and CCF-0515164. The work of H. Zhang was done during his Ph.D. study at North Carolina State University. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia,September 2005.

H. Zhang is with Marvell Semiconductor, Inc., Santa Clara, CA 95054 USA (e-mail: hongyuan@marvell.com).

H. Dai, Q. Zhou, and B. L. Hughes are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA (e-mail: hdai@ncsu.edu; qzhou@ncsu.edu; blhughes@ncsu.edu).

Communicated by R. R. Müller, Associate Editor for Communications.

Color versions of Figs. 3–5 available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2006.885531

MIMO systems can be exploited for spatial diversity (SD) or spatial multiplexing (SM) gains [7]. The majority of work on MIMO antenna selection focuses on the former, including selection combining, hybrid selection-maximum ratio combining (HS-MRC) [2], [3], and antenna selection with space-time coding [4], [5]. Essentially in these works, with independent and identically distributed (i.i.d.) Rayleigh fading, the system error performance or outage probability can be readily analyzed through order statistics [19]. It has been shown that the diversity order of the original full-size system can be maintained through the signal-to-noise ratio (SNR) maximization selection criterion.

In contrast, antenna selection for MIMO systems with multiple data streams, i.e., spatial multiplexing systems, has received less attention. The few existing analytical results generally assume capacity-achieving joint space–time coding and optimal decoding. Capacity-maximizing receive antenna selection is analyzed in [6] and shown to achieve the same diversity order as the full-size system. In [8], it is shown that the fundamental tradeoff between diversity and spatial multiplexing of the full-size system, obtained in [7], holds as well for MIMO systems with antenna selection.

In practice, the multiple streams in a SM system may be uncoded or separately encoded and suboptimally decoded due to complexity and feedback overhead concerns. In [1], several transmit antenna selection algorithms for SM with linear receivers are proposed, and some conjectures on the achieved diversity orders are made based on numerical results. To the best of our knowledge, the exact diversity order achieved by antenna selection for practical SM systems has not been rigorously obtained. In contrast to MIMO diversity schemes, the key challenge that hinders accurate performance analysis is that selection is performed among a list of interdependent random quantities, which are correlated in a complex manner.

In this paper, we propose a new framework to analyze the diversity order achieved by transmit antenna selection for SM systems with separately (and independently) encoded data streams and linear or decision-feedback (DF) receivers (i.e., the V-BLAST structure [9]). In particular, we rigorously show that the optimal diversity order is $(N_T-1)(N_R-1)$ for an $N_R \times N_T$ separately encoded SM system employing linear or DF receivers, when $L=2$ antennas are selected from the transmit side. This should be compared with the diversity order of a two-stream SM system without antenna selection: N_R-1. Such a diversity gain can be tremendous for downlink high-data-rate communications, where there may be a large number of transmit antennas at the base stations but few receive antennas at the mobiles (e.g., $N_R = 2$). For DF receivers, it is also clarified that, the antenna selection rule that maximizes the performance of the first decoded data stream is *not* optimal, as the freedom to choose transmit antennas for subsequent data streams is restricted. Furthermore, following the same geometrical approach, we give an upper bound, $(N_T - L+1)\cdot(N_R-1)$ and a lower bound, $(N_T - L+1)(N_R - L+1)$, on the diversity order for general L, which coincide when $L = 2$. The corresponding diversity and multiplexing tradeoff curves are also derived. Generally speaking, our results confirm and generalize some of the conjectures in [1], thus verifying that transmit antenna selection can achieve high data rates and robust error performance in practical SM systems without complex coding. Furthermore, the proposed geometrical approach may be used to solve other open problems related to MIMO communication systems.

Among feasible suboptimal receivers for separately encoded SM systems, we first investigate the linear zero-forcing (ZF) receiver [10]. The analysis for the zero-forcing decision-feedback (ZF-DF) receiver relies heavily on the former. Furthermore, their diversity order analysis, a study at high SNR regimes, hold for linear minimum mean-square-error (MMSE) and MMSE decision-feedback (MMSE-DF) receivers as well.

The rest of the paper is organized as follows. The system model and problem formulation are given in Section II. The main ideas of our approach are illustrated in Section III for $L = 2$, where new notations and concepts can be more easily appreciated. Then extension to the general L scenario is discussed in Section IV, the analysis of which nonetheless becomes more involved. Finally, Section V concludes the paper with future research directions.

II. System Model and Problem Formulation

We consider a frequency nonselective block Rayleigh fading channel model, in which the $N_R \times N_T$ channel matrix is represented by $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_T}]$; where $\{\mathbf{h}_k\}$ are i.i.d. complex Gaussian column vectors, whose entries are i.i.d. with zero mean and unit variance. With transmit antenna selection, there are a total of $N_U = \binom{N_T}{L}$ possible antenna subsets, defined as $U_1 \sim U_{N_U}$[1]

$$\begin{aligned} U_1 &= \{\mathbf{h}_1, \mathbf{h}_2, \ldots \mathbf{h}_L\} \\ U_2 &= \{\mathbf{h}_1, \mathbf{h}_2, \ldots \mathbf{h}_{L-1}, \mathbf{h}_{L+1}\} \\ &\vdots \\ U_{N_U} &= \{\mathbf{h}_{N_T-L+1}, \ldots, \mathbf{h}_{N_T}\}. \end{aligned} \tag{1}$$

If the subset U_j is selected, of which the L column vectors compose the channel matrix $\mathbf{H}_j$, the SM system can then be expressed as

$$\mathbf{y} = \sqrt{\frac{\rho_0}{L}}\mathbf{H}_j\mathbf{s} + \mathbf{n} \tag{2}$$

where the $N_R \times T$ matrix $\mathbf{y}$ is the received signal block with T the block length; $\mathbf{n}$ is the circularly symmetric complex Gaussian background noise with zero mean and identity covariance matrix; and the $L \times T$ matrix $\mathbf{s}$ represents the transmitted signal block, with independent rows each assuming unit average energy per symbol. Therefore, ρ_0 is the average SNR per receive antenna. As in [7], we assume the fading channel keeps constant for the entire data frame. Each stream independently adopts a code of fixed data rate R_0.[2] As will be shown (see proof of Lemma I), channel coding in each data stream can provide coding gain but not diversity gain. Throughout the paper we assume $N_T \geq L$ and $N_R \geq L$. Therefore given the assumption of i.i.d. Rayleigh fading, the selected channel matrix $\mathbf{H}_j$ has full column rank with probability one [18].

For a ZF receiver, a spatial equalizer

$$\mathbf{G}_{\mathrm{ZF}} = \mathbf{H}_j^{+} = \left[\mathbf{H}_j^H\mathbf{H}_j\right]^{-1}\mathbf{H}_j^H$$

is applied to the received signal $\mathbf{y}$ to obtain an estimate of the transmitted signal

$$\hat{\mathbf{s}} = \mathbf{G}_{\mathrm{ZF}}\mathbf{y} = \sqrt{\frac{\rho_0}{L}}\mathbf{s} + \mathbf{G}_{\mathrm{ZF}}\mathbf{n} \tag{3}$$

where $\dagger$ denotes the pseudo-inverse of a matrix, and ^{H} stands for conjugate transpose. Therefore L equivalent decoupled single-input–single-output (SISO) data links are formed as

$$\hat{s}_k = \sqrt{\frac{\rho_0}{L}}s_k + [\mathbf{G}_{\mathrm{ZF}}]_{k^*}\mathbf{n}, \quad 1 \leq k \leq L \tag{4}$$

where $[\ ]_{k^*}$ denotes the kth row in matrix. The instantaneous channel capacity and error performance of these subchannels are determined by the corresponding postprocessing SNRs, which can be directly derived from (4) as

$$\begin{aligned} \rho_{k,\mathrm{ZF}}^{(j)} &= \left(\frac{\rho_0}{L}\right)\Big/\|[\mathbf{G}_{\mathrm{ZF}}]_{k*}\|^2 \\ &= \left(\frac{\rho_0}{L}\right)\Big/\left[\mathbf{H}_j^H\mathbf{H}_j\right]_{kk}^{-1}, \quad 1 \leq k \leq L \end{aligned} \tag{5}$$

where $[\ ]_{kk}$ represents the kth diagonal element of a matrix. Let $\mathbf{h}_k^{(j)}$ be the kth column of $\mathbf{H}_j$, and $\mathbf{H}_k^{(j)}$ the $N_R \times (L-1)$ matrix resulting from leaving out $\mathbf{h}_k^{(j)}$ from $\mathbf{H}_j$. It is known that [10], $1/[\mathbf{H}_j^H\mathbf{H}_j]_{kk}^{-1}$ is equal to $R_k^{(j)}$, the square of the projection height[3] from $\mathbf{h}_k^{(j)}$ to the range of $\mathbf{H}_k^{(j)}$, or alternatively, the square of the norm of the projection of $\mathbf{h}_k^{(j)}$ onto the null space of $(\mathbf{H}_k^{(j)})^T$. That is

$$\rho_{k,\mathrm{ZF}}^{(j)} = \frac{\rho_0}{L}R_k^{(j)} = \frac{\rho_0}{L}\left\|\mathbf{h}_k^{(j)}\right\|^2\sin^2\theta_k^{(j)} \tag{6}$$

where $\|\mathbf{h}_k^j\|$ is the norm of $\mathbf{h}_k^{(j)}$, while $\theta_k^{(j)}$ is the angle between $\mathbf{h}_k^{(j)}$ and its projection on the range of $\mathbf{H}_k^{(j)}$, defined as

$$\theta_k^{(j)} = \sin^{-1}\frac{\sqrt{R_k^{(j)}}}{\left\|\mathbf{h}_k^{(j)}\right\|}, \quad 0 < \theta_k^{(j)} < \frac{\pi}{2}.$$

[1]Without loss of generality, it is assumed that in each subset the selected columns are ordered increasingly by index.

[2]We are mainly interested in the maximum diversity gain here, corresponding to a zero multiplexing gain scenario in [7]. An exception is in Section III-C, where a family of block codes with multiplexing gain r is considered for each stream.

[3]Projection height refers to the norm of the error vector, i.e., the difference between a vector and its projection onto a subspace.

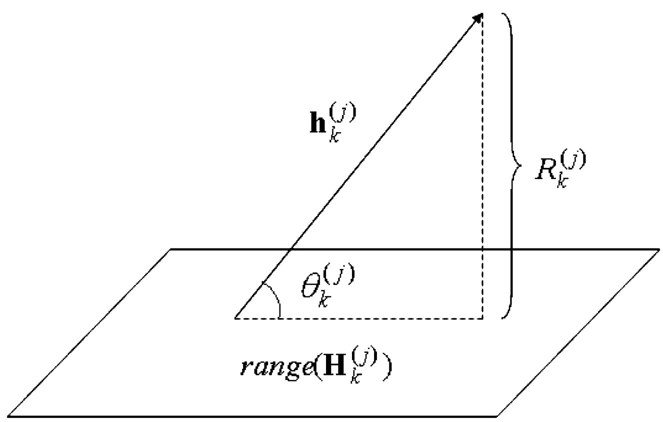

Fig. 1. Illustration of (6).

Fig. 1 gives a geometrical view of the relationship in (6).
We further define

$$R_{\min}^{(j)} = \min_{k=1\ldots L} \left\{ R_k^{(j)} \right\}, \quad 1 \le j \le N_U \tag{7}$$

for each subset U_j, which essentially determines the system performance at high SNR [10], [17].

For a linear MMSE receiver, the spatial equalizer is given by $\mathbf{G}_{\mathrm{MMSE}} = [\mathbf{H}_j^H \mathbf{H}_j + L/\rho_0 \mathbf{I}]^{-1} \mathbf{H}_j^H$, and the postprocessing SNRs are given by [10], [22]

$$\begin{aligned} \rho_{k,\mathrm{MMSE}}^{(j)} &= \frac{\rho_0}{L \left[\mathbf{H}_j^H \mathbf{H}_j + L/\rho_0 \mathbf{I} \right]_{kk}^{-1}} - 1, \quad 1 \le k \le L, \\ &= \frac{\rho_0 \Gamma_k^{(j)}}{L} - 1 \end{aligned} \tag{8}$$

where $\Gamma_k^{(j)} = \frac{1}{[\mathbf{H}_j^H \mathbf{H}_j + L/\rho_0 \mathbf{I}]_{kk}^{-1}}$. Similarly, we define

$$\Gamma_{\min}^{(j)} = \min_{k=1\ldots L} \left\{ \Gamma_k^{(j)} \right\}, \quad l \le j \le N_U. \tag{9}$$

In this paper, we adopt the antenna selection rule that maximizes the postprocessing SNR of the worst data stream (as in [1]). That is, we choose the subset among (1) such that $R_{\min}^{(j)}$ in (7) or $\Gamma_{\min}^{(j)}$ in (9) is maximized, and we denote

$$R_{\mathrm{SL}} = \max_{1 \le j \le N_U} \left\{ R_{\min}^{(j)} \right\} \text{ and } \Gamma_{\mathrm{SL}} = \max_{1 \le j \le N_U} \left\{ \Gamma_{\min}^{(j)} \right\}. \tag{10}$$

We will show that this selection rule is optimal for linear receivers with respect to diversity order.

Before we proceed, we first introduce some notations. The diversity order is defined as the slope of the average frame error probability $P_e(\rho_0)$[4] in log-scale in the high-SNR regime [7]

$$d = -\lim_{\rho_0 \to \infty} \frac{\log \mathrm{P}_e(\rho_0)}{\log(\rho_0)} = \lim_{\rho_0 \to \infty} \frac{\log \mathrm{P}_e(\rho_0)}{\log(1/\rho_0)}. \tag{11}$$

[4]By frame error probability, we mean an error is declared if any substream is decoded unsuccessfully. For each substream, assuming coding over a single block with constant fading, this is the true error probability of a code averaged over the transmitted codewords, channel fading, and additive noise.

We adopt the operator $\doteq$ as defined in [7], to denote exponential equality, i.e., we write $f(\rho_0) \doteq \rho_0^{-b}$ to represent

$$-\lim_{\rho_0 \to \infty} \frac{\log f(\rho_0)}{\log(\rho_0)} = b.$$

Equivalently (for the convenience of analysis in this paper), we use $f(x) \doteq x^b$ to represent

$$\lim_{x \to 0} \frac{\log f(x)}{\log x} = b. \tag{12}$$

The operators $\dot{\le}, \dot{\ge}, \dot{<}, \dot{>}$ are similarly defined. Note that according to our notation, $f(x) \dot{\le} g(x)$ indicates $f(x) \ge g(x)$ for sufficiently small x.

Lemma I: For separately encoded spatial multiplexing systems with linear ZF/MMSE receivers, the antenna selection rule that chooses the antenna subset with the strongest weakest data link achieves the optimal diversity order among all antenna selection rules. Furthermore, the optimal diversity order can be evaluated as

$$d_{\mathrm{opt}}^L = \lim_{x \to 0} \frac{\log \Pr(R_{\mathrm{SL}} \le x)}{\log(x)} \tag{13}$$

for both receivers.

Proof: See Appendix A. □

Remarks: Lemma I indicates that we only need to evaluate (13) for the optimal diversity order of separately encoded SM systems with transmit antenna selection and linear ZF/MMSE receivers. We also make no claims on the practicality of this selection algorithm, as the main focus of this paper is on theoretical analysis. Note that efficient antenna selection algorithms exist in literature (e.g., [2], [15] and references therein).

III. Diversity Order and Diversity–Multiplexing Tradeoff When $L = 2$

In this section, we discuss the main idea of our geometrical approach for $L = 2$, which also admits an exact result. Extension to the general L scenario is not trivial, as will be seen in Section IV.

Theorem I: In an $N_R \times N_T$ spatial multiplexing system with linear ZF/MMSE or ZF/MMSE decision-feedback receivers satisfying $N_T \ge 2$ and $N_R \ge 2$, if separately encoded data streams are transmitted from two selected antennas, the optimal achievable diversity order is

$$d_{\mathrm{opt}} = (N_T - 1)(N_R - 1). \tag{14}$$

The optimal diversity-multiplexing tradeoff curve is the piecewise function of r connecting the points $(r, d_{\mathrm{opt}}(r))$ for $0 \le r \le 2$, with

$$d_{\mathrm{opt}}(r) = (N_T - 1)(N_R - 1)(1 - r/2)^+ \tag{15}$$

where $(x)^+ = \max(x, 0)$.

This theorem will be proved in three steps: we start with linear ZF/MMSE receivers by evaluating (13) in Section III-A. Then we extend the analysis to ZF/MMSE-DF receivers in Section III-B. Finally we analyze the corresponding diversity-multiplexing tradeoff curves in Section III-C.

As $L = 2$, each possible antenna subset in (1) contains two selected column vectors from the original channel matrix $\mathbf{H}$. For convenience, we will use R_{kj} to denote the squared projection height from $\mathbf{h}_k$ to $\mathbf{h}_j$. In another word, given that $U_s = \{\mathbf{h}_k, \mathbf{h}_j\}$ is selected, $R_{kj} \triangleq R_1^{(s)}$ and $R_{jk} \triangleq R_2^{(s)}$ (cf. (6) in Section II). Similar notation transforms are adopted for $\theta_k^{(j)}$. Clearly

$$R_{kj} = \|\mathbf{h}_k\|^2 \sin^2 \theta_{kj} \tag{16}$$

where it can be shown that [12], [18], [26], [27] $\|\mathbf{h}_k\|^2$ is $\chi^2(2N_R)$ distributed with probability density function (pdf)[5]

$$f_{\|\mathbf{h}_k\|^2}(x) = \frac{1}{\Gamma(N_R)} x^{N_R-1} e^{-x} = \frac{1}{(N_R-1)!} x^{N_R-1} e^{-x} \tag{17}$$

and θ_{kj} assumes a pdf of

$$f_{\theta_{kj}}(\theta) = (N_R - 1)\sin 2\theta (\sin\theta)^{2N_R-4}, \theta \in \left(0, \frac{\pi}{2}\right). \tag{18}$$

Furthermore, $\|\mathbf{h}_k\|^2, \|\mathbf{h}_j\|^2$ and θ_{kj} are mutually independent. Also, R_{kj} is a $\chi^2(2(N_R-1))$ distributed random variable, which leads to

$$\Pr(R_{kj} \le x) \doteq x^{N_R-1}. \tag{19}$$

From (19) we can see that without antenna selection ($N_T = L$), any data stream with a ZF/MMSE receiver bears a diversity order of $N_R - 1$, so does the overall error probability. When $N_T > L$, through the diversity-maximization transmit antenna selection (cf. (7) and (10))

$$R_{\rm SL} = \max_{k \ne j \in \{1,\ldots,N_T\}} \{\min\{R_{kj}, R_{jk}\}\} \tag{20}$$

we will show that a product gain of $N_T - 1$ on the diversity order can be achieved.

A. Linear ZF/MMSE Receiver

In the following, the optimal diversity order (13) for linear ZF/MMSE receivers $d_{\rm opt}^L$ will be explicitly explored. Note that neither the exact pdf of $R_{\rm SL}$ nor its polynomial expansion near zero seems tractable, which motivates us to solve the problem through tight upper and lower bounds.

By definition

$$\begin{aligned}\Pr(R_{\rm SL} \le x) &= \Pr(\min(R_{12}, R_{21}) \le x, \min(R_{13}, R_{31}) \le x \\ &\quad \ldots, \min(R_{(N_T-1)N_T}, R_{N_T(N_T-1)}) \le x) \\ &= \Pr\left(\bigcup_{i=1}^{N} A_i\right)\end{aligned} \tag{21}$$

where the $N = 2^{N_U}$ events $\{A_i\}$ are defined as

$$\begin{aligned}A_1 &= \{R_{12} \le x, R_{13} \le x, \ldots, R_{1N_T} \le x, \\ &\qquad R_{23} \le x, \ldots, R_{(N_T-1)N_T} \le x\} \\ A_2 &= \{R_{12} \le x, R_{13} \le x, \ldots, R_{1N_T} \le x, \\ &\qquad R_{23} \le x, \ldots, R_{N_T(N_T-1)} \le x\} \\ &\vdots \\ A_N &= \{R_{21} \le x, R_{31} \le x, \ldots, R_{N_T1} \le x \\ &\qquad R_{32} \le x, \ldots, R_{N_T(N_T-1)} \le x\}\end{aligned} \tag{22}$$

each of which corresponds to the case that one data stream from each of the N_U subsets is in outage (i.e., close to zero for sufficiently small x). For example, event A_1 picks the first element from each possible subset $U_1 \sim U_{N_U}$, i.e.,

$$A_1 = \left\{\bigcap_{1\le k<j\le N_T} \{R_{kj} \le x\}\right\}.$$

Then from (21) we have

$$\Pr(A_1) \le \Pr(R_{\rm SL} \le x) \le \sum_{i=1}^{N} \Pr(A_i) \tag{23}$$

which indicates

$$\lim_{x\to 0} \frac{\log \max_i \Pr(A_i)}{\log(x)} \le d_{\rm opt}^L \le \lim_{x\to 0} \frac{\log \Pr(A_1)}{\log(x)}. \tag{24}$$

Clearly the lower bound in (24) is actually tight. However, it is generally difficult to identify the dominant terms at high SNR from (22), since the cardinality grows exponentially with N_T^2. Alternatively, we take the following approach. First, we find a common upper bound for $\Pr(A_i), \forall i$, i.e., obtaining such a P_U that

$$\max_i \Pr(A_i) \le P_U \tag{25}$$

which determines a lower bound for $d_{\rm opt}^L$ (cf. the lower bound in (24)).[6] We then obtain a lower bound for $\Pr(A_1)$

$$\Pr(A_1) \ge P_L \tag{26}$$

and evaluate its error exponential, which gives an upper bound for $d_{\rm opt}^L$ (cf. the upper bound in (24)). It turns out that these two bounds coincide and represent the best achievable diversity order, given by $(N_T-1)(N_R-1)$.

Diversity Lower Bound

Proposition I: $d_{\rm opt}^L \ge (N_T-1)(N_R-1)$.

Proof: Define S_i as the set consisting of the $N_U = \binom{N_T}{2}$ random variables in A_i (see (22)). For example

$$\begin{aligned}S_1 &= \{R_{12}, R_{13}, \ldots, R_{1N_T}, R_{23}, \ldots, R_{(N_T-1)N_T}\} \\ &= \{R_{kj}\}_{1\le k<j\le N_T}.\end{aligned}$$

First we claim that, in any S_i we can always find a subset of $N_T - 1$ random variables in the form of

$$S_{i\_\rm indep} = \left\{R_{k_1k_2}, R_{k_2k_3}, \ldots, R_{k_{N_T-1}k_{N_T}}\right\}$$

[5] $\chi^2(2N_R)$ denotes Chi-square distribution with $2N_R$ degrees of freedom and mean N_R. $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt$ is the Gamma function.

[6] Recall that an error probability with lower diversity order corresponds to an exponentially larger one, and vice versa.

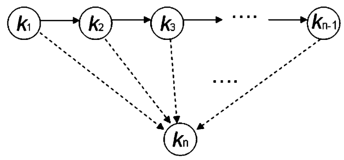

Fig. 2. Illustration of the Proof for Proposition I: A graph with size n.

where $k_1 \sim k_{N_r}$ is some permutation of the integer array $1 \sim N_T$. For example

$$S_{1\_\text{indep}} = \{R_{12}, R_{23}, \ldots, R_{(N_T-1)N_T}\}$$

i.e., $k_i = i, 1 \le i \le N_T$. By Lemma II given in Appendix B, we obtain the following common upper bound for $\Pr(A_i)$:

$$\Pr(A_i) \le P_U = [\Pr(R_{kj} \le x)]^{(N_T-1)}, \quad k \ne j, \forall A_i. \tag{27}$$

With (19) we have

$$\max_i \Pr(A_i) \ge x^{(N_T-1)(N_R-1)} \tag{28}$$

and from (24) Proposition I follows.

The claim is proved as follows. A graph is drawn in Fig. 2 to visualize S_i, where the nodes represent the transmit antenna elements and an arrow directed from node k_1 to k_2 appears in the graph if $R_{k_1k_2}$ exists in S_i. Based on the definition of S_i, there is one and only one arrow between any two nodes. For such a graph of size n, a *complete path* is defined as one that goes through a series of $n-1$ arrows, passing each of the n nodes one and only one time. Therefore we can find a subset S_{i_indep} in S_i if at least one complete path exists in the graph of size N_T, passing the nodes with the order $k_1 \rightarrow k_2 \rightarrow \cdots \rightarrow k_{N_T}$. It is easily shown that all graphs of size 3 contain one complete path. Now suppose that all graphs of size $n-1$ contain at least one complete path. Then in any graph of size n, we can always find a path passing $n-1$ nodes one and only one time, assumed in the order $k_1 \rightarrow k_2 \rightarrow \cdots \rightarrow k_{n-1}$. From Fig. 2, we see that if there is no complete path existing in the graph of size n, the arrow between k_1 and k_n should be in the direction of $k_1 \rightarrow k_n$, otherwise $k_n \rightarrow k_1 \rightarrow \cdots \rightarrow k_{n-1}$ forms a complete path. For the same reason, the arrow between k_2 and k_n should be in the direction of $k_2 \rightarrow k_n$, otherwise $k_1 \rightarrow k_n \rightarrow k_2 \cdots \rightarrow k_{n-1}$ forms a complete path. By repeating the same reasoning, we finally reach $k_{n-1} \rightarrow k_n$. However, in this case $k_1 \rightarrow k_2 \rightarrow \cdots \rightarrow k_n$ is a complete path, and we conclude that a complete path always exists for any graph of size n. Thus the claim is proved for any N_T by induction. $\square$

Diversity Upper Bound

Proposition II: $d_{\text{opt}}^L \le (N_T-1)(N_R-1)$.

Proof: This proof is technically involved, so we divide it into three parts for ease of illustration: (1) by some geometrical analysis $\Pr(A_1)$ is further lower bounded (see (31)); (2) by some lemmas in Appendix C, the probability lower bound is transformed into an exponentially equivalent form with known statistics (see (35)); (3) the diversity upper bound is explicitly evaluated (see (38)).

Part 1: Given

$$\begin{aligned}\Pr(A_1) &= \Pr\left(\bigcap_{1\le k<j\le N_T} R_{kj} \le x\right)\\ &= \Pr\left(\bigcap_{1\le k<j\le N_T} \|\mathbf{h}_k\|^2 \sin^2\theta_{kj} \le x\right)\end{aligned}$$

and by defining the summation $z = \sum_{k=1}^{N_T-1} \|\mathbf{h}_k\|^2$, distributed as $\chi^2(2(N_T-1)N_R)$, and $\psi_0 = (\pi/2)/(N_T-1)$, we have

$$\begin{aligned}&\Pr(A_1)\\ &\ge \Pr\left(\bigcap_{1\le k<j\le N_T} z\sin^2\theta_{kj} \le x\right)\\ &\ge \Pr\left(\bigcap_{1\le k<j\le N_T} z\sin^2\theta_{kj} \le x, \max_{2\le k\le N_T} \theta_{1k} < \psi_0\right)\end{aligned} \tag{29}$$

where for the second inequality we have further restricted the ranges of the N_T-1 i.i.d. random variables $\{\theta_{1k}\}_{k=2}^{N_T}$ within $(0,\psi_0)$ (cf. Corollary I in Appendix B).

Based on the geometric structure involved, given θ_{1k} and $\theta_{1j}, 2 \le k < j \le N_T, \theta_{kj}$ is constrained as $\theta_{kj} \le \theta_{1k} + \theta_{1j}$, where the equality holds only when $\mathbf{h}_1, \mathbf{h}_k, \mathbf{h}_j$ are linearly dependent (located in the same subspace with a dimension less than 3) [11]. Then within the range $(0,\psi_0)$, we have

$$\begin{aligned}\sin^2\theta_{kj} &\le \sin^2(\theta_{1k}+\theta_{1j})\\ &\le \sin^2(\theta_{12}+\theta_{13}+\cdots+\theta_{1N_T})\\ &= \sin^2\theta_\Sigma\end{aligned} \tag{30}$$

where $\theta_\Sigma = \sum_{k=2}^{N_T} \theta_{1k}$ is still in the range of $(0,\pi/2)$. Therefore (29) can be further lower bounded as:

$$\begin{aligned}\Pr(A_1) &\ge P_L\\ &= \Pr\left(z\sin^2\theta_\Sigma \le x, \max_{2\le k\le N_T} \theta_{1k} < \psi_0\right)\\ &\doteq \Pr(z\sin^2\theta'_\Sigma \le x)\end{aligned} \tag{31}$$

where we define a new set of i.i.d. random variables $\theta'_{12} \sim \theta'_{1N_T}$ with pdf of

$$\begin{aligned}f_{\theta'_{1i}}(x) &= \frac{f_{\theta_{1i}}(x)}{\int_0^{\psi_0} f_{\theta_{1i}}(x)\,dx}\\ &= \frac{f_{\theta_{1i}}(x)}{C}, 0 < x < \psi_0, 2 \le i \le N_T\end{aligned} \tag{32}$$

i.e., the restriction of $\theta_{12} \sim \theta_{1N_T}$ in the range of $(0,\psi_0)$, and $\theta'_\Sigma = \sum_{i=2}^{N_T} \theta'_{1i}$.

Part 2:

With Lemma III–VI on exponential equivalence given in Appendix C, the evaluation of (31) can be further simplified. Specifically, by defining $\theta'_0 = \max_k \theta'_{1k}$ and $m(x) = \sin^2(x)$

for $x \in (0, \psi_0)$, from Lemma IV (whose proof requires Lemma III) we have

$$\Pr(\sin^2\theta'_\Sigma \le x) \doteq \Pr(\sin^2\theta'_0 \le x) \tag{33}$$

i.e., with respect to monotonic functions, replacing the sum of a set of independent random variables with the maximum of them does not change the associated exponential behavior.

Further by Lemma V, we have

$$\Pr(z\sin^2\theta'_\Sigma \le x) \doteq \Pr(z\sin^2\theta'_0 \le x) \tag{34}$$

i.e., the exponential behavior is still not changed by multiplication of an independent random variable.

Finally by Lemma VI, we have

$$\Pr(z\sin^2\theta'_0 \le x) \doteq \Pr(z\sin^2\theta_0 \le x) \tag{35}$$

where $\theta_0 = \max_k \theta_{1k}$, i.e., without the constraint of $(0, \psi_0)$ on $\theta_{12} \sim \theta_{1N_T}$.

Part 3:

We are then left to evaluate the smallest exponential in $\Pr(z\sin^2\theta_0 \le x)$ in (35). Note that z is a $\chi^2(2(N_T-1)N_R)$ distributed random variable with cumulative distribution function (cdf)

$$F_z(x) = 1 - e^{-x}\sum_{k=0}^{M+N_T-2}\frac{x^k}{k!}, M = (N_T-1)(N_R-1) \tag{36}$$

while from (18) the pdf of θ_0 can be easily derived as:

$$f_{\theta_0}(\theta) = M[\sin^2\theta]^{M-1}\sin 2\theta,\ \theta \in \left(0, \frac{\pi}{2}\right). \tag{37}$$

After some algebra presented in Appendix D, we can get the following equivalent polynomial form as $x \to 0$:

$$\Pr(z\sin^2\theta_0 \le x) = \left(\frac{1}{M!} - M\sum_{k=0}^{N_T-3}\frac{k!}{(M+k+1)!}\right)x^M + o(x^M) \tag{38}$$

where the coefficient of x^M in (38) is always positive (also shown in Appendix D), which completes the proof. □

For example, when $N_T = 3, N_R = 3$, and $L = 2$, from (38) we get as $x \to 0$

$$\Pr(z\sin^2\theta_0 \le x) = \frac{x^4}{120} + o(x^5).$$

To illustrate the soundness of our approach, Fig. 3 presents by simulations the exponential behavior of $\Pr(z\sin^2\theta_0 \le x)$, together with the outage probability in (21) and the probabilities representing the diversity bounds as in (23), (24). The results show a diversity order of 4, verifying our derivations above.

With Propositions I and II, (14) is proved for the linear receiver case.

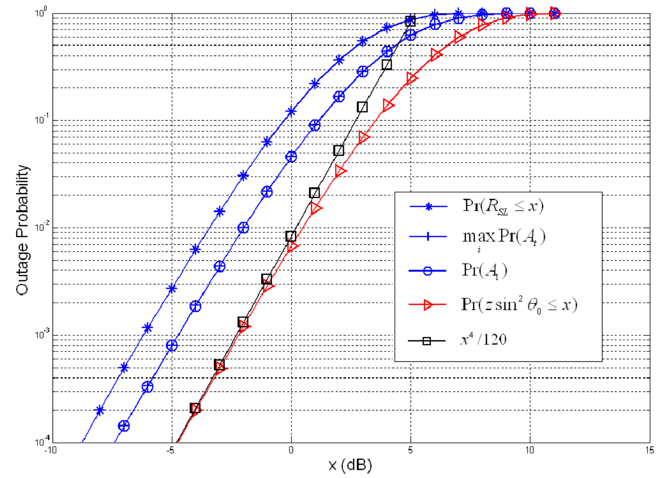

Fig. 3. Exponential behaviors of the outage probabilities for the $N_T = N_R = 3, L = 2$ scenario.

B. Decision-Feedback Receivers

In this subsection we continue to explore the optimal achievable diversity order for separately encoded SM systems with transmit antenna selection and DF receivers, denoted as $d_{\rm opt}^{\rm DF}$. As expected, the performance analysis for DF receivers will rely heavily on that for their linear counterparts in Section III-A.

For DF receivers with $L = 2$, the frame error probability is given by

$$P_e = P_{e1} + P_{e2}(1 - P_{e1}) \tag{39}$$

where P_{e1} is the error probability of the first decoded data stream, and P_{e2} is that of the second stream *assuming perfect feedback*. Therefore $P_e \doteq \max\{P_{e1}, P_{e2}\}$. For a fixed-order ZF/MMSE DF receiver without antenna selection, $P_{e1} \ge P_{e2}$ is always fulfilled, so that we can investigate its diversity order solely from the first decoded data stream, which is processed by just a linear ZF/MMSE receiver (see, e.g., [7], [17]). This may entice one to consider an antenna selection rule that maximizes the performance of the first decoded data stream. However, as will be shown, this antenna selection rule is generally not optimal.

Instead of seeking an optimal antenna selection rule for DF receivers like Lemma I, we take the following simpler approach to evaluate $d_{\rm opt}^{\rm DF}$, thanks to the close connection between DF receivers and their linear counterparts. First, since for any antenna selection rule $j, P_e^{(j)} \dot{\le} P_{e1}^{(j)}$, we have $d_{(j)}^{\rm DF} \le d_{1,(j)}^{\rm DF}$, where $d_{1,(j)}^{\rm DF}$ is the diversity order of the corresponding first decoded data stream. Consequently

$$d_{\rm opt}^{\rm DF} \le d_{1,{\rm opt}}^{\rm DF} \tag{40}$$

where $d_{1,{\rm opt}}^{\rm DF}$ is the best achievable diversity order of the first decoded data stream. Note that there may be multiple choices of antenna selection rules to achieve the optimum for both sides of (40), and a selection rule that is best for one may fail for the other. Nonetheless, following a procedure similar to what we have discussed above for linear receivers, we will explicitly evaluate $d_{1,\mathrm{opt}}^{\mathrm{DF}}$ and derive a diversity upper bound for $d_{\mathrm{opt}}^{\mathrm{DF}}$. We then continue by constructing a specific antenna selection algorithm whose diversity order is easy to assess and serve as a lower bound for $d_{\mathrm{opt}}^{\mathrm{DF}}$.

Diversity Upper Bound—Maximizing the SNR of the First Decoded Data Stream

We investigate an antenna selection algorithm that achieves $d_{1,\mathrm{opt}}^{\mathrm{DF}}$: selecting the antenna subset that maximizes the postprocessing SNR of the first decoded data stream. We distinguish two scenarios with respect to detection order: arbitrary but fixed ordering and optimal ordering [9]. Our main result is summarized next.

Proposition III: $d_{\mathrm{opt}}^{\mathrm{DF}} \le d_{1,\mathrm{opt}}^{\mathrm{DF}} = (N_T-1)(N_R-1)$.

Proof: For fixed ordering, without loss of generality, we assume that decoding starts from the signal transmitted from the antenna with the smallest index number in the selected antenna subset. The antenna selection rule indicates that the postprocessing SNR of the first decoded data stream is given by (cf. (20))

$$R_{\mathrm{SL1}} = \max_{1\le k<j\le N_T} \{R_{kj}\}. \tag{41}$$

Clearly we have (cf. (22))

$$\Pr(R_{\mathrm{SL1}} \le x) = \Pr(A_1) \tag{42}$$

whose upper bound can be derived from (27), while its lower bound exponential behavior evaluation directly follows (29)–(38). Therefore for arbitrary but fixed ordering

$$d_{1,\mathrm{opt}}^{\mathrm{DF}} = (N_T-1)(N_R-1). \tag{43}$$

For optimal ordering, our selection rule is reformulated with

$$R_{\mathrm{SL2}} = \max_{k\ne j\in\{1,\ldots,N_T\}} \{\max\{R_{kj}, R_{jk}\}\} \tag{44}$$

and we have

$$\begin{aligned}\Pr(R_{\mathrm{SL2}} \le x) &= \Pr\left(\bigcap_{1\le k<j\le N_T} \max(R_{kj}, R_{\mathrm{jk}}) \le x\right)\\ &= \Pr\left(\bigcap_{1\le k<j\le N_T} \max\left(\|\mathbf{h}_k\|^2, \|\mathbf{h}_j\|^2\right)\sin^2\theta_{kj} \le x\right).\end{aligned} \tag{45}$$

It is straightforward to upper and lower bound (45) as (cf. (27) and (31))

$$\begin{aligned}\Pr(R_{\mathrm{SL2}} \le x) &\le \Pr(A_1)\\ &\le [\Pr(R_{kj}\le x)]^{N_T-1}, \quad \forall k\ne j\end{aligned} \tag{46}$$

$$\begin{aligned}\Pr(R_{\mathrm{SL2}} \le x) &\ge \Pr\left(\max\left(\|\mathbf{h}_1\|^2,\cdots,\|\mathbf{h}_{N_T}\|^2\right)\sin^2\theta'_\Sigma \le x\right)\\ &\ge \Pr\left(\left(\sum_{i=1}^{N_T}\|\mathbf{h}_i\|^2\right)\sin^2\theta'_\Sigma \le x\right).\end{aligned} \tag{47}$$

The evaluation of the lower bound in (47) is similar as in (31), except that z is redefined as $z = \sum_{k=1}^{N_T}\|\mathbf{h}_k\|^2$. Following a similar approach as (31)–(38), we get

$$\begin{aligned}&\Pr(R_{\mathrm{SL2}} \le x)\\ &\quad \dot{\le} \left(\frac{1}{M!} - M\sum_{k=0}^{N_T+N_R-3}\frac{k!}{(M+k+1)!}\right)x^M\\ &\qquad + o(x^{M+1}).\end{aligned} \tag{48}$$

So the same result as in (43) is obtained with optimal ordering. □

Remarks: Even though $R_{\mathrm{SL2}} \ge R_{\mathrm{SL1}} \ge_{\mathrm{SL1}}$ with probability 1, no advantage in diversity order can be achieved. Of course, some coding gain in SNR is naturally expected, which is beyond the scope of this paper. So $d_{\mathrm{opt}}^{\mathrm{DF}} \le d_{1,\mathrm{opt}}^{\mathrm{DF}} = d_{\mathrm{opt}}^{L}$. On the other hand, we anticipate the performance of DF receivers is no worse than their linear counterparts at high SNR, which will be verified below.

In [7], [12], the authors have shown that optimal ordering will not increase the diversity order in the first decoded data stream of a separately encoded SM system with DF receivers in the $L=2$ case.[7] As a side product, here we reach the same conclusion in the antenna selection context.

Finally we note that this antenna selection rule is in general suboptimal with respect to the diversity order for the whole system. The reason is that although P_{e1} in (39) is minimized, P_{e2} is not affected by the selection process. Rather, it behaves the same as in a nonselection scheme with N_R -order diversity. Therefore $P_e = \max\{P_{e1}, P_{e2}\}$ is mostly dominated by the second stream, and the diversity order is given by $d = \min\{(N_T-1)(N_R-1), N_R\} = N_R$, for $N_T \ge 3, N_R \ge 2$.

Diversity Lower Bound—QR Decomposition-Based Antenna Selection Algorithm

Here we analyze a simple yet effective antenna selection algorithm, for which the first decoded data stream achieves a diversity order of $(N_T-1)(N_R-1)$, while the second stream performs better than the first one. This antenna selection algorithm is based on QR decomposition, which was originally proposed for capacity maximization [14], [15]. Compared with brute force methods, this algorithm greatly reduces the computational complexity while achieving a near optimal performance with respect to channel capacity. Here we apply it in SM systems with DF receivers with the goal of minimizing the error rate, and show that it also maximizes the diversity order, therefore verifying our observations in [15] and revealing its great potential.

Proposition IV:

$$d_{\mathrm{opt}}^{\mathrm{DF}} \ge (N_T-1)(N_R-1).$$

Proof: For simplicity we mainly focus on ZF-DF receivers. From a geometrical viewpoint, this incremental antenna selection procedure starts by seeking the column vector with the largest norm (or the largest projection height to the null-space); and in each of the following steps, one column with the largest

[7]Extension to the general L scenario was done recently in [24], [25].

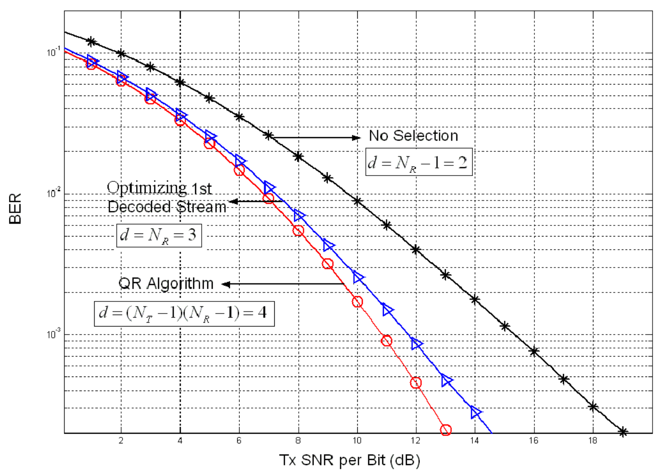

Fig. 4. BER performance of the ZF-DF receiver with $N_T = N_R = 3, L = 2$.

projection height to the space spanned by the selected column vectors is chosen until the Lth antenna is selected. The detection order is the reverse of the selection order, i.e., the stream decoded at the lth step is transmitted from the antenna selected at the $(L-l+1)$th step. The corresponding postprocessing SNR for the stream decoded at the lth step is then proportional to the maximum value among $N_T - L + l$ independent random variables distributed as $\chi^2(2(N_R - L + 1))$, resulting in a diversity order of $(N_T - L + l)(N_R - L + l)$. Therefore for $L = 2$, the first decoded stream achieves a diversity order of $d_1 = (N_T - 1)(N_R - 1)$ and the second decoded stream achieves $d_2 = N_T N_R$, and the diversity order for the whole system equals $(N_T - l)(N_R - l)$.

For MMSE-DF receivers, the algorithm is slightly modified: at the lth step, a column $\mathbf{h}_{(l)}$ is selected such that

$$l = \arg\max_{l'} \frac{1}{\left[\left[\mathbf{H}_{(l-1)}, \mathbf{h}_{l'}\right]^H \left[\mathbf{H}_{(l-1)}, \mathbf{h}_{l'}\right] + L/\rho_0 \mathbf{I}\right]^{-1}_{ll}}$$

where $\mathbf{H}_{(l-1)}$ is the matrix formed by the previous selected $l-1$ column vectors, and $[\mathbf{H}_{(l-1)}, \mathbf{h}_{l'}]$ means appending $\mathbf{h}_{l'}$ to the last column. Following the proof of Lemma I (in Appendix A), the same diversity lower bound is achieved. □

Simulation results are presented in Fig. 4 for the $N_T = 3$, $N_R = 3, L = 2$ scenario, where the bit error rates (BER) of the ZF-DF receiver with the above two selection rules are shown, together with that for the DF system without antenna selection. The achieved diversity orders of different schemes are also provided for a clearer comparison. We see that although simpler, the QR based method (bearing a diversity of 4) outperforms the algorithm that optimizes only the first decoded data stream, which only achieves a diversity order of 3, due to the unaffected second data stream, as discussed above.

With Propositions III and IV, (14) is also proved for ZF/MMSE-DF receivers.

C. The Diversity–Multiplexing Tradeoff

Using the same geometrical approach, we can also obtain the diversity-multiplexing tradeoff curve introduced in [7] for the separately encoded SM systems with antenna selection. With quasistatic fading assumption, a family of codes $\{\zeta(\rho_0)\}$ over a block length shorter than fading coherence time is employed, one at each SNR level. We further assume that the rate of the code increases with SNR, so a scheme achieves a multiplexing gain r if the rate $R(\rho_0) = r \log \rho_0$. Based on the diversity order analysis in Sections III-A and -B (see especially proof of Lemma I in Appendix A), we get the following equation:

$$P_e \doteq P_{e\_\max} \doteq P_{\text{out}\_\max}$$

where $P_{\text{out}\_\max}$ can be viewed as the outage probability of the worst stream for linear receivers, and that of the first decoded data stream for DF receivers, respectively. The diversity-multiplexing tradeoff curve $d(r)$ can be evaluated directly from $P_{\text{out}\_\max}$ as $\rho_0 \to \infty$[8]

$$\begin{aligned} P_{\text{out}\_\max} &= \Pr\left[\log\left(1 + \frac{\rho_0}{L} R_{\text{SL}}\right) \le \frac{r}{L}\log\rho_0\right] \\ &\doteq \Pr\left[R_{\text{SL}} \le L\rho_0^{-(1-\frac{r}{L})}\right] \\ &\doteq \left[\rho_0^{-(1-\frac{r}{L})}\right]^M = \rho_0^{-M(1-\frac{r}{L})} \end{aligned}$$

where $M = (N_T - 1)(N_R - 1)$.

Therefore, (15) is proved and we finalize the proof of Theorem I.

IV. Extension to General L

Analyzes in Section III can be partially extended to the general scenario. In particular, the diversity lower bound can be obtained following a similar approach. In contrast, a tight diversity upper bound could not be achieved with straightforward extension. Evaluations become more involved as now the postprocessing SNR is proportional to the squared projection height from a column vector to a *nondegenerated* space. In the following, we present our results for general L , highlight the key challenges, and finally give our conjecture and future direction.

Our main result for general L is given below, which should be compared with Theorem I.

Theorem II: In an $N_R \times N_T$ spatial multiplexing system with linear ZF/MMSE or ZF/MMSE decision-feedback receivers satisfying $N_T \ge L$ and $N_R \ge L$, if separately encoded data streams are transmitted from L selected antennas, the optimal achievable diversity order is bounded as

$$M_L \le d_{\text{opt}} \le M_U \tag{49}$$

where

$$\begin{aligned} M_L &= (N_T - L + 1)(N_R - L + 1) \\ M_U &= (N_T - L + 1)(N_R - 1). \end{aligned}$$

The optimal diversity-multiplexing tradeoff curve is bounded as

$$M_L\left(1 - \frac{r}{L}\right)^+ \le d_{\text{opt}}(r) \le M_U\left(1 - \frac{r}{L}\right)^+.$$

We will mainly focus on the extension for linear receivers; those for diversity order of DF receivers and diversity-multiplexing tradeoff are relatively straightforward and will be briefly discussed.

[8] Replace R_{SL} with R_{SL1} or R_{SL2} for DF receivers.

Employing the antenna selection method (10), we can then derive a similar outage probability expression as (21):

$$\begin{aligned}\Pr(R_{\mathrm{SL}} \leq x) &= \Pr\left(\min_{U_1}\{R_k^{(1)}\} \leq x, \min_{U_2}\{R_k^{(2)}\}\right.\\ &\qquad \left.\leq x, \ldots, \min_{U_{N_U}}\{R_k^{(N_U)}\} \leq x\right)\\ &= \Pr\left(\bigcup_{i=1}^{N} A_i\right) \qquad (51)\end{aligned}$$

where $U_1 \sim U_{N_U}$ are defined in (1), $R_k^{(j)}$ in (6), and the $N = L^{\binom{N_T}{L}}$ events $\{A_i\}$ are similarly defined with a generic form (cf. (22))

$$A_j = \left\{\bigcap_{l \leq i \leq N_U} \left\{R_{j(i)}^{(i)} \leq x\right\}\right\}, \quad 1 \leq j \leq N$$

where $1 \leq j(i) \leq L$ denotes a pick from subset U_i. Clearly, (23) still holds and we proceed with the evaluation of upper and lower bounds.

Proposition V: For general L, the optimal diversity order for linear receivers can be lower bounded as $d_{\mathrm{opt}}^L \geq (N_T - L + 1)(N_R - L + 1)$.

Proof: From the definition of A_i and $U_l \sim U_{N_U}$, in any S_i (the set of random variables in A_i) we can always find a subset bearing the form

$$S_{i\_\mathrm{indep}} = \{R_{k_1}^{(j_1)'}, R_{k_2}^{(j_2)'}, \ldots, R_{k_{N_T-L+1}}^{(j_{N_T-L+1})'}\}$$

with $N_T - L + 1$ random variables. In this form, $R_k^{(j)}$ is slightly modified from $R_k^{(j)}$ with k indicating the index of column vector in the original channel matrix $\mathbf{H}$, instead of $\mathbf{H}_j$, i.e., the squared projection height from $\mathbf{h}_k$ (instead of $\mathbf{h}_k^{(j)}$) to the subspace spanned by the remaining column vectors in subset U_j , with the implication that $\mathbf{h}_k \in U_j$. Here $k_1 \sim k_{N_T-L+1}$ are some $N_T - L + 1$ distinct integers within $[1, N_T]$, and $U_{j1} \sim U_{j_{N_T-L+1}}$ is an ordered list of $N_T - L + 1$ different subsets from $U_1 \sim U_{N_U}$. It is further required that $\mathbf{h}_{k_i} \in U_{j_i}$, and $\mathbf{h}_{k_i} \notin U_{j_1}$ for $l > i$, i.e., $\mathbf{h}_{k_i}$ belongs to subset U_{j_i} but not those after. Therefore by Lemma VII in Appendix E we can get the following upper bound for any $\Pr(A_i)$:

$$\Pr(A_i) \leq \left[\Pr\left(R_{k_1}^{(j_1)'} \leq x\right)\right]^{N_T-L+1} \qquad (52).$$

Furthermore, since $R_k^{(j)'}$ is $\chi^2(2(N_R - L + 1))$ distributed, we have

$$\Pr(A_i) \dot{\geq} x^{(N_T-L+1)(N_R-L+1)} \qquad (53)$$

and Proposition V is proved. □

On the other hand, for $L > 2$ scenarios, the derivation of a tight lower bound for $\Pr(A_1)$ is much more involved as compared to the $L = 2$ case, because the angles $\{\theta_k^{(l)}\}$ are correlated in a complicated manner, and a general form of their joint pdf expressions is not accessible. Nonetheless, we have the following result.

Proposition VI: For general L, the optimal diversity order for linear receivers can be upper bounded as

$$d_{\mathrm{opt}}^L \leq (N_T - L + 1)(N_R - 1).$$

Proof: Since the projection height from a vector to a subspace represents the shortest distance from the vector to any point in the subspace, we have $R_k^{(j)} \leq R_{kl}^{(j)}$, for any $\mathbf{h}_k^{(j)}$, $\mathbf{h}_l^j \in U_j$, where $R_{kl}^{(j)}$ denotes the squared projection height from $\mathbf{h}_k^{(j)}$ to $\mathbf{h}_l^{(j)}$. It is then not difficult to build up the following lower bound:

$$\begin{aligned}\Pr(A_1) &= \Pr\left(\bigcap_{1 \leq i \leq N_U} \{R_1^i \leq x\}\right)\\ &\geq \Pr\left(\bigcap_{1 \leq i \leq N_U} \{R_{12}^i \leq x\}\right). \qquad (54)\end{aligned}$$

Carefully examining the first two elements in all subsets (see (1)) reveals that the last term in (54) bears a similar form as the $\Pr(A_1)$ in $L = 2$ case (see (22)), replacing $N_T - 1$ with $N_T - L + 1$. Following the same lines as in Section III, we have for general L

$$\Pr(R_{\mathrm{SL}} \leq x) \dot{\leq} \Pr(A_1) \dot{\leq} x^{(N_T-L+1)(N_R-1)}. \qquad (55)$$

□

Combining the above two propositions, (49) is proved for linear receivers. Also, as the $L = 2$ case it is straightforward to show that (49) applies for the first decoded data stream of DF receivers. Since the diversity order of the first decoded data stream is upper bounded by $(N_T - L + 1)(N_R - 1)$ and $P_e \dot{\leq} P_{e1}$, the upper bound $(N_T - L + 1)(N_R - 1)$ also applies for P_e with DF receivers. On the other hand, by employing the QR based selection algorithm for DF receivers, a diversity order lower bound $(N_T - L + 1)(N_R - L + 1)$ is achieved. Therefore (49) also holds for DF receivers. Finally the derivation of (50) follows the same method as given in Section III-C, and Theorem II follows.

Remarks: Note that when $L = 2$, the two bounds in (49) and (50) coincide and conform to the results obtained in Section III. A conjecture on the diversity order of separately encoded SM systems with transmit antenna selection and linear ZF receivers was made in [1] based on numerical results, which actually has motivated our research: *for linear ZF receivers, when $N_R = L$, the achievable diversity order is $N_T - L + 1$.* Our results prove its correctness and further extend it to general N_R.

In Fig. 5 we investigate the exponential behavior of $\Pr(R_{\mathrm{SL}} \leq x)$ and $\Pr(A_1)$ for an exemplary scenario $N_T = N_R = 4, L = 3$ through simulations. By comparing with the reference curve x^4, we find that both of them present a diversity of $M_L = 4$, verifying its achievability as we claim in Theorem II. On the other hand, it shows that in this scenario the probability lower bound $\Pr(A_1)$ also achieves the diversity lower bound M_L, revealing that the diversity orders between $M_L + 1$ and M_U may actually not be achievable. We thus have the following conjecture, which constitutes part of our future

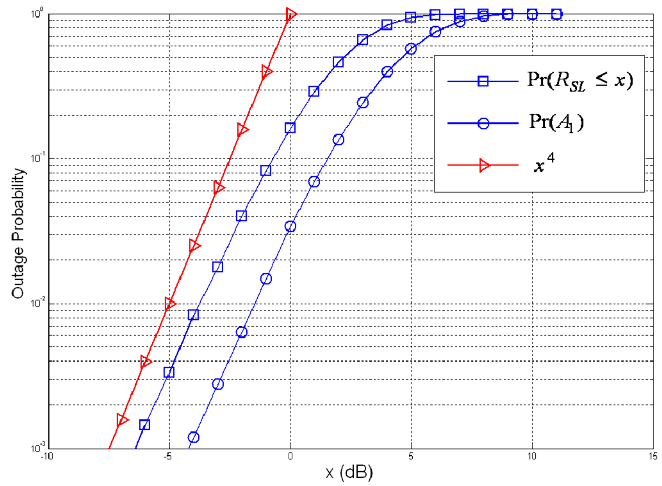

Fig. 5. Exponential Behaviors of the Outage Probabilities for the $N_T = N_R = 4, L = 3$ Scenario.

work. The key seems to lie on finding a better lower bound of $\Pr(A_1)$ than (54).

Conjecture: for general L, the optimal transmit antenna selection achieves a diversity order exactly equal to M_L.

V. Conclusion

In this paper, we have analyzed the diversity order achieved by transmit antenna selection for separately encoded SM systems with linear and decision-feedback receivers. Using a geometric approach, we have rigorously derived their achievable diversity order for the $L = 2$ scenario. We have also used the same geometrical approach to obtain bounds on the achievable diversity order for general L. Our results prove and extend the previous conjectures in literature drawn from simulations, and verify the predicted potential of antenna selection for practical spatial multiplexing systems. Furthermore, the proposed geometrical approach may also be used to solve other open problems related to MIMO.

Besides verifying the conjecture for general L, the analysis for maximum-likelihood receivers, joint transmitter and receiver selection, and multiuser MIMO scenarios also direct our future research.

Appendix

A. Proof of Lemma I

Proof: Suppose subset U_j is selected following an arbitrary antenna selection rule, which may be channel dependent. Given any channel realization $\mathbf{H}$, let the corresponding selected antenna subset be $\mathbf{H}_j$ (as in (2)). Let random variable $R_{\min}^{(j)}$ denote the minimum squared projection height for this antenna selection rule (cf. (7)). We claim that the maximum diversity order achieved for this antenna selection rule is given by

$$d_{(j)}^{L} = \lim_{x\to 0} \frac{\log \Pr\left(R_{\min}^{(j)} \le x\right)}{\log(x)}. \tag{56}$$

Since our antenna selection rule (10) dictates $R_{\mathrm{SL}} \ge R_{\min}^{(j)}$ with probability 1, the lemma readily follows.

The claim of (56) is proved as follows. Let us consider a linear ZF receiver first. For block fading channels, given any channel realization $\mathbf{H}$, let $P_e^{(j)}(\mathbf{H})$ and $P_{el}^{(j)}(\mathbf{H})$ denote the conditional error probability of the whole frame and the lth substream, respectively, assuming optimal one-dimensional coding for each substream. The worst link, with conditional error probability $P_{e\_\max}^{(j)}(\mathbf{H}) = \max_l P_{el}^{(j)}(\mathbf{H})$, has instantaneous postdetection SNR $\frac{\rho_0}{L} R_{\min}^{(j)}(\mathbf{H})$.[9] The corresponding average error probabilities are denoted as $P_e^{(j)}$, $P_{el}^{(j)}$, and $P_{e\_\max}^{(j)}$. Clearly, we have

$$P_{e\_\max}^{(j)}(\mathbf{H}) \le P_e^{(j)}(\mathbf{H}) \le \sum_{l=1}^{L} P_{el}^{(j)}(\mathbf{H}) \tag{57}$$

which leads to

$$\max_l P_{el}^{(j)} \le P_{e\_\max}^{(j)} \le P_e^{(j)} \le \sum_{l=1}^{L} P_{el}^{(j)}. \tag{58}$$

Since the upper and lower bounds above bear the same diversity order [7], [10], we have

$$P_e^{(j)} \doteq P_{e\_\max}^{(j)} \triangleq E\left\{P_{e\_\max}^{(j)}(\mathbf{H})\right\}. \tag{59}$$

It is known from Lemma 5 of [7] that the error probability $P_{e\_\max}^{(j)}$ is (uniformly) lower bounded by the corresponding outage probability, given by $\Pr(\log(1 + \frac{\rho_0}{L} R_{\min}^{(j)}) \le R_0)$, where R_0 is the fixed data rate (see Section II). Therefore the diversity order is upper-bounded by

$$\begin{aligned} d_{(j)}^{\mathrm{ZF}} &\le -\lim_{\rho_0\to\infty} \frac{\log \Pr\left(R_{\min}^{(j)} \le \frac{L}{\rho_0}(2^{R_0} - 1)\right)}{\log(\rho_0)} \\ &= \lim_{x\to 0} \frac{\log \Pr\left(R_{\min}^{(j)} \le x\right)}{\log(x)}. \end{aligned} \tag{60}$$

The above error probability lower bound is obtained without specifying any coding scheme. On the other hand, the error probability with uncoded square M-QAM signaling can serve as an upper bound [7], [21]

$$P_{e\_\max}^{(j)} \le P_{e\_\max\_\mathrm{QAM}}^{(j)} \le 4P(S_0 \to S_1) \tag{61}$$

where $P(S_0 \to S_1)$ denotes the pairwise error probability between two closest QAM constellation points. By (21) in [7], we have

$$P(S_0 \to S_1) \doteq P\left(\frac{\rho_0}{L} R_{\min}^{(j)} \cdot \frac{3}{M-1} < 1\right). \tag{62}$$

We, therefore, obtain

$$d_{(j)}^{\mathrm{ZF}} \ge \lim_{x\to 0} \frac{\log \Pr\left(R_{\min}^{(j)} \le x\right)}{\log(x)} \tag{63}$$

and thus have proved the claim of (56) for a linear ZF receiver. An immediate conclusion from analysis above is that, one-dimensional channel coding has no impact on system diversity order.

[9] $R_{\min}^{(j)}(\mathbf{H})$ refers to the realization of $R_{\min}^{(j)}$ with respect to $\mathbf{H}$.

Note that the above analysis (see (58)) also indicates that

$$d_{(j)}^{\rm ZF} = \lim_{x\to 0} \frac{\log \Pr\left(R_{\min}^{(j)} \le x\right)}{\log(x)} = \lim_{\rho_0\to\infty} \frac{\log \Pr\left(\min_{k=1\cdots L}\left\{\rho_{k,{\rm ZF}}^{(j)}\right\} \le 1\right)}{\log(1/\rho_0)} = \min_{k=1\cdots L} \lim_{\rho_0\to\infty} \frac{\log \Pr\left(\rho_{k,{\rm ZF}}^{(j)} \le 1\right)}{\log(1/\rho_0)}. \tag{64}$$

As for a linear MMSE receiver, following a similar approach, we have

$$d_{(j)}^{\rm MMSE} = \lim_{\rho_0\to\infty} \frac{\log \Pr\left(\min_{k=1\cdots L}\left\{\rho_{k,{\rm MMSE}}^{(j)}\right\} \le 1\right)}{\log(1/\rho_0)} = \min_{k=1\cdots L} \lim_{\rho_0\to\infty} \frac{\log \Pr\left(\rho_{k,{\rm MMSE}}^{(j)} \le 1\right)}{\log(1/\rho_0)}. \tag{65}$$

In the following, we complete the proof by showing that

$$\Pr\left(\rho_{k,{\rm ZF}}^{(j)} \le 1\right) \doteq \Pr\left(\rho_{k,{\rm MMSE}}^{(j)} \le 1\right), \quad \forall k. \tag{66}$$

We alternatively have expressions of postprocessing SNRs for ZF and MMSE receivers as (see, e.g., [10])

$$\rho_{k,{\rm ZF}}^{(j)} = \frac{\rho_0}{L}\left(\mathbf{h}_k^{(j)}\right)^H \cdot \left[\mathbf{I} - \mathbf{H}_k^{(j)}\left(\left(\mathbf{H}_k^{(j)}\right)^H \mathbf{H}_k^{(j)}\right)^{-1}\left(\mathbf{H}_k^{(j)}\right)^H\right]\mathbf{h}_k^{(j)} \tag{67}$$

and

$$\rho_{k,{\rm MMSE}}^{(j)} = \frac{\rho_0}{L}\left(\mathbf{h}_k^{(j)}\right)^H \cdot \left[\mathbf{I} - \mathbf{H}_k^{(j)}\left(\left(\mathbf{H}_k^{(j)}\right)^H \mathbf{H}_k^{(j)} + \frac{L}{\rho_0}\mathbf{I}\right)^{-1}\left(\mathbf{H}_k^{(j)}\right)^H\right]\mathbf{h}_k^{(j)}. \tag{68}$$

By matrix inversion Lemma [20], with

$$\mathbf{A}_k^{(j)} = \left((\mathbf{H}_k^{(j)})^H \mathbf{H}_k^{(j)}\right)^{-1}$$

we have

$$\rho_{k,{\rm MMSE}}^{(j)} - \rho_{k,{\rm ZF}}^{(j)} = \left(\mathbf{h}_k^{(j)}\right)^H \mathbf{H}_k^{(j)}\mathbf{A}_k^{(j)} \times \left(\frac{L}{\rho_0}\mathbf{A}_k^{(j)} + \mathbf{I}\right)^{-1}\mathbf{A}_k^{(j)}\left(\mathbf{H}_k^{(j)}\right)^H \mathbf{h}_k^{(j)} \triangleq \eta_k^{(j)}(\rho_0). \tag{69}$$

Clearly, $\eta_k^{(j)}(\rho_0) \ge 0$ and thus

$$\Pr\left(\rho_{k,{\rm MMSE}}^{(j)} \le 1\right) \dot{\ge} \Pr\left(\rho_{k,{\rm ZF}}^{(j)} \le 1\right). \tag{70}$$

Another key observation is that, $\eta_k^{(j)}(\rho_0)$ is statistically independent with $\rho_{k,{\rm ZF}}^{(j)}$, as the latter is proportional to the squared norm of the projection of $\mathbf{h}_k^{(j)}$ onto the null space of $(\mathbf{H}_k^{(j)})^T$, while the former is the correlation of $\mathbf{h}_k^{(j)}$ with a vector in the range of $\mathbf{H}_k^{(j)}$ [18], [26]. Furthermore, as $\rho_0 \to \infty$, it can be shown that

$$\eta_k^{(j)}(\rho_0) \xrightarrow{a.s.} \eta_k^{(j)} \tag{71}$$

i.e., it converges almost surely to a positive random variable with finite mean. This observation was also made in [24] where the pdf of $\eta_k^{(j)}$ was explicitly given. For our purpose, it is sufficient to know that $\Pr(\eta_k^{(j)} \le 1/2)$ is a positive probability of $O(1)$ (the same order as 1). We thus have

$$\frac{\log \Pr\left(\rho_{k,{\rm MMSE}}^{(j)} \le 1\right)}{\log(1/\rho_0)} = \frac{\log \Pr\left(\rho_{k,{\rm ZF}}^{(j)} + \eta_k^{(j)}(\rho_0) \le 1\right)}{\log(1/\rho_0)} \ge \frac{\log \Pr\left(\rho_{k,{\rm ZF}}^{(j)} \le 1/2\right)}{\log(1/\rho_0)} + \frac{\log \Pr\left(\eta_k^{(j)}(\rho_0) \le 1/2\right)}{\log(1/\rho_0)} \tag{72}$$

with

$$\lim_{\rho_0\to\infty} \frac{\log \Pr\left(\eta_k^{(j)}(\rho_0) \le 1/2\right)}{\log(1/\rho_0)} = \lim_{\rho_0\to\infty} \frac{\log \Pr\left(\eta_k^{(j)} \le 1/2\right)}{\log(1/\rho_0)} = 0 \tag{73}$$

where (73) is due to the fact that almost sure convergence always leads to convergence in distribution. (72) and (73) together gives rise to

$$\Pr\left(\rho_{k,{\rm MMSE}}^{(j)} \le 1\right) \dot{\le} \Pr\left(\rho_{k,{\rm ZF}}^{(j)} \le 1\right) \tag{74}$$

and (66) follows with (70) and (74). □

B. Lemma II and Corollary I

Lemma II: For any permutation of the integer array $1 \sim N_T$, denoted as $k_1 \sim k_{N_T}$, we have

$$\Pr\left(R_{k_1k_2} \le x, R_{k_2k_3} \le x, \ldots, R_{k_{(N_T-1)}k_{N_T}} \le x\right) = [\Pr(R_{k_1k_2} \le x)]^{(N_T-1)} \tag{75}$$

i.e., random variables $R_{k_1k_2}, R_{k_2k_3}, \ldots, R_{k_{(N_T-1)}k_{N_T}}$ are jointly independent.

Proof: Essentially $R_{k_ik_{i+1}}$ is only a function of $\mathbf{h}_{k_i}$ and $\mathbf{h}_{k_{i+1}}$, denoted as $R_{k_ik_{i+1}} = g(\mathbf{h}_{k_i}, \mathbf{h}_{k_{i+1}})$, therefore the conditional pdf of $R_{k_ik_{i+1}}$ given those variables appearing earlier in the sequence admits

$$\begin{aligned} f&\left(R_{k_ik_{i+1}} \,|\, R_{k_{i-1}k_i}, \ldots, R_{k_1k_2}\right) \\ &= f\left(g\left(\mathbf{h}_{k_i}, \mathbf{h}_{k_{i+1}}\right) \,|\, g\left(\mathbf{h}_{k_{i-1}}, \mathbf{h}_{k_i}\right), \ldots, g\left(\mathbf{h}_{k_1}, \mathbf{h}_{k_2}\right)\right) \\ &= f\left(g\left(\mathbf{h}_{k_i}, \mathbf{h}_{k_{i+1}}\right) \,|\, g\left(\mathbf{h}_{k_{i-1}}, \mathbf{h}_{k_i}\right)\right) \\ &= f\left(R_{k_ik_{i+1}} \,|\, R_{k_{i-1}k_i}\right) \end{aligned}$$

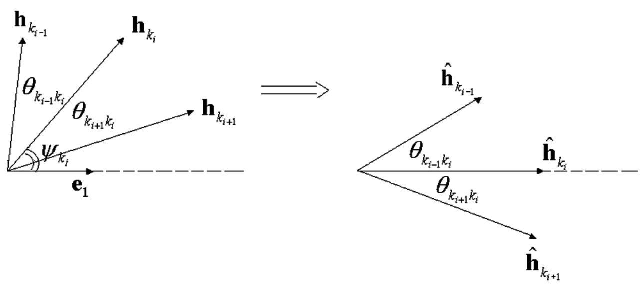

Fig. 6. Independence of $\theta_{k_{i-1}k_i}$ and $\theta_{k_{i+1}k_i}$.

where the second equality holds because the states of $\mathbf{h}_{k_1} \sim \mathbf{h}_{k_{i-1}}$ do not affect $\mathbf{h}_{k_i}$ and $\mathbf{h}_{k_{i+1}}$. Consequently, the above sequence forms a Markov chain.

We are left to prove the independence between $R_{k_ik_{i+1}}$ and $R_{k_{i-1}k_i}$. Given

$$R_{k_ik_{i+1}} = \|\mathbf{h}_{k_i}\|^2 \sin^2\theta_{k_{i+1}k_i} \quad \text{and}$$
$$R_{k_{i-1}k_i} = \left\|\mathbf{h}_{k_{i-1}}\right\|^2 \sin^2\theta_{k_{i-1}k_i}$$

together with the independence between $\|\mathbf{h}_{k_i}\|^2$ and $\|\mathbf{h}_{k_{i-1}}\|^2$, and between vector norms and directions (angles) [18], we only need to show that $\theta_{k_{i+1}k_i}$ and $\theta_{k_{i-1}k_i}$ are independent.

Following a similar rotation approach as in [12], we define $\mathbf{e}_1 \sim \mathbf{e}_{N_R}$ as a *fixed* orthonormal basis (e.g., Cartesian coordinates) of the vector space $\mathbb{C}^{N_R}$. We rotate $[\mathbf{h}_{k_{i-1}}, \mathbf{h}_{k_i}, \mathbf{h}_{k_{i+1}}]$ as a whole so that $\mathbf{h}_{k_i}$ is parallel to $\mathbf{e}_1$, denoted as $[\tilde{\mathbf{h}}_{k_{i-1}}, \tilde{\mathbf{h}}_{k_{i+1}}] = \mathbf{Q}(\psi_{k_i})[\mathbf{h}_{k_{i-1}}, \mathbf{h}_{k_{i+1}}]$, where ψ_{k_i} is the angle between $\mathbf{h}_{k_i}$ and $\mathbf{e}_1$, and $\mathbf{Q}(\psi_{k_i})$ is the corresponding unitary rotation matrix. Since $[\mathbf{h}_{k_{i-1}}, \mathbf{h}_{k_{i+1}}]$ is an i.i.d. Gaussian matrix (therefore the joint distribution is rotationally invariant) and is independent with ψ_{k_i}, $[\tilde{\mathbf{h}}_{k_{i-1}}, \tilde{\mathbf{h}}_{k_{i+1}}]$ is still i.i.d. Gaussian. Because $\theta_{k_{i+1}k_i}$ and $\theta_{k_{i-1}k_i}$ are unchanged after the rotation, and equal to the angles between $\tilde{\mathbf{h}}_{k_{i+1}}$ and $\mathbf{e}_1$, and between $\tilde{\mathbf{h}}_{k_{i-1}}$ and $\mathbf{e}_1$, respectively (see Fig. 6), given the fact that $\tilde{\mathbf{h}}_{k_{i+1}}$ and $\tilde{\mathbf{h}}_{k_{i-1}}$ are independent, it is straightforward to show that $\theta_{k_{i+1}k_i}$ and $\theta_{k_{i-1}k_i}$ are independent, so are $R_{k_ik_{i+1}}$ and $R_{k_{i-1}k_i}$, and Lemma II follows. □

Corollary I: $\theta_{12}, \theta_{13}, \ldots, \theta_{1N_T}$ are jointly independent.

Proof: Using the same rotation approach as above, if we rotate $[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_T}]$ as a whole such that $\mathbf{h}_1$ is parallel to $\mathbf{e}_1, \tilde{\mathbf{h}}_2, \ldots, \tilde{\mathbf{h}}_{N_T}$ are jointly independent vectors; their angles with $\mathbf{e}_1$, which are equal to $\theta_{12}, \theta_{13}, \ldots, \theta_{1N_T}$, respectively, are also jointly independent. □

C. Some Lemmas Used in the Derivation of Proposition II

Lemma III: If $\theta_1, \ldots, \theta_K$ are independent positive random variables, whose cumulative distribution functions admit $F_{\theta_k}(x) \doteq x^{n_k}$, we have

$$\Pr\left[\sum_{k=1}^{K}\theta_k \le x\right] \doteq x^{\sum_{k=1}^{K} n_k}. \tag{76}$$

Proof: At first we evaluate the exponential behavior of $\Pr(\theta_1 + \theta_2 \le x)$. Note that the following expression assumes $x \to 0$:

$$\begin{aligned}
\Pr(\theta_1+\theta_2 \le x) &= \int_0^x \left[\int_0^{x-\theta_1} f_{\theta_2}(\theta_2)d\theta_2\right] f_{\theta_1}(\theta_1)d\theta_1 \\
&= \int_0^x f_{\theta_1}(\theta_1)F_{\theta_2}(x-\theta_1)d\theta_1 \\
&\doteq \int_0^x \left[\theta_1^{n_1-1} + o\left(\theta_1^{n_1-1}\right)\right] \\
&\quad \cdot [(x-\theta_1)^{n_2} + o((x-\theta_1)^{n_2})]d\theta_1 \\
&\doteq x^{n_2}\int_0^x \left[\theta_1^{n_1-1} + o\left(\theta_1^{n_1-1}\right)\right] d\theta_1 \\
&\doteq x^{n_1+n_2}
\end{aligned}$$

where $o(x)$ denotes a higher-order function of x such that $\lim_{x\to 0} o(x)/x = 0$. Next, we can treat $\theta_1 + \theta_2$ as a new random variable exponentially equivalent to $x^{n_1+n_2}$ and evaluate $(\theta_1 + \theta_2) + \theta_2$ following the same approach, whose cdf asymptotically behaves as $x^{(n_1+n_2)+n_3}$. Lemma IV follows after such repeated operations. □

Lemma IV: Let $m(\theta)$ be a positive function of θ, which monotonically increases with θ, satisfying $m(0) = 0$ and $m^{-1}(x) \doteq x^{n_0}$. If $\theta_1, \ldots, \theta_k$ are independent positive random variables, whose cumulative distribution functions admit $F_{\theta_k}(x) \doteq x^{n_k}$, we have

$$\Pr\left[m\left(\sum_{k=1}^{K}\theta_k\right) \le x\right] \doteq \Pr\left[m\left(\max_k \theta_k\right) \le x\right] \doteq x^{n_0\sum_{k=1}^{K} n_k}. \tag{77}$$

Proof: Since $\theta_1, \ldots, \theta_k$ are independent, we get

$$\Pr\left[\max_k \theta_k \le x\right] = \prod_{k=1}^{K} F_{\theta_1}(x) \doteq x^{\sum_{k=1}^{K} n_k}. \tag{78}$$

By Lemma III, we have

$$\Pr\left[\sum_{k=1}^{K}\theta_k \le x\right] \doteq x^{\sum_{k=1}^{K} n_k}. \tag{79}$$

Because both θ and $m(\theta)$ are positive, and $m(\theta)$ monotonically increases with θ, $m^{-1}(x)$ is also a positive function of x, which monotonically increases with x. Therefore,

$$\begin{aligned}
&\Pr\left[m\left(\max_k \theta_k\right) \le x\right] \\
&= \Pr\left[\max_k \theta_k \le m^{-1}(x)\right] \\
&\doteq [m^{-1}(x)]^{\sum_{k=1}^{K} n_k} \doteq x^{n_0\sum_{k=1}^{K} n_k}.
\end{aligned} \tag{80}$$

Similarly

$$\Pr\left[m\left(\sum_{k=1}^{K}\theta_k\right)\le x\right]\doteq x^{n_0\sum_{k=1}^{K}n_k} \tag{81}$$

and Lemma V follows. □

Lemma V: For independent continuous random variables a, b_1 and b_2 satisfying $a \ge 0$ with $\Pr(a \le x)\doteq x^{n_a}$, and $0 \le b_1, b_2 \le 1$ with $\Pr(b_1 \le x)\doteq\Pr(b_2 \le x)\doteq x^{n_b}$, we have

$$\Pr(ab_1 \le x)\doteq\Pr(ab_2 \le x)\dot{\le}x^{n_a}. \tag{82}$$

Proof: We have

$$\begin{aligned}\Pr(ab_1 \le x) &= \int_0^1 \Pr\left(a \le \frac{x}{y}\right) f_{b_1}(y)\cdot dy\\ &= \int_0^{\varepsilon} \Pr\left(a \le \frac{x}{y}\right) f_{b_1}(y)\cdot dy\\ &\quad + \int_{\varepsilon}^1 \Pr\left(a \le \frac{x}{y}\right) f_{b_1}(y)\cdot dy\\ &= t_{11}(x)+t_{12}(x)\end{aligned} \tag{83}$$

and similarly

$$\begin{aligned}\Pr(ab_2 \le x) &= \int_0^{\varepsilon} \Pr\left(a \le \frac{x}{y}\right) f_{b_2}(y)\cdot dy\\ &\quad + \int_{\varepsilon}^1 \Pr\left(a \le \frac{x}{y}\right) f_{b_2}(y)\cdot dy\\ &= t_{21}(x)+t_{22}(x)\end{aligned} \tag{84}$$

where $\varepsilon > 0$ is a fixed small positive number, small enough to make the pdf approximations for b_1 and b_2 :

$$\begin{aligned}f_{b_1}(x) &= c_1 x^{n_b-1} + o\left(x^{n_b-1}\right)\\ f_{b_2}(x) &= c_2 x^{n_b-1} + o\left(x^{n_b-1}\right)\end{aligned} \tag{85}$$

hold true.

Since $\Pr(a \le \frac{x}{y})$ is positive, and is a decreasing function of y, when $x \to 0$ we have

$$t_{12}(x) \le \Pr\left(a \le \frac{x}{\varepsilon}\right)\int_{\varepsilon}^1 f_{b_1}(y)\cdot dy \le \Pr\left(a \le \frac{x}{\varepsilon}\right)\doteq x^{n_a}. \tag{86}$$

In other words, $t_{12}(x)\dot{\ge}x^{n_a}$.

On the other hand, since $0 \le b_1 \le 1$, we have

$$\Pr(ab_1 \le x) \ge \Pr(a \le x)\doteq x^{n_a} \tag{87}$$

which means that $\Pr(ab_1 \le x)\dot{\le}x^{n_a}$.

The same inequalities as in (86), (87) also hold true for $t_{22}(x)$ and $\Pr(ab_2 \le x)$, respectively.

Meanwhile, since ε is small enough, by applying (85), we get that

$$t_{11}(x)\doteq t_{21}(x)\doteq\int_0^{\varepsilon}\Pr\left(a \le \frac{x}{y}\right)y^{n_b-1}\cdot dy. \tag{88}$$

If $t_{12}(x)\dot{>}x^{n_a}$ or $t_{22}\dot{>}x^{n_a}$, $t_{11}(x)\doteq t_{21}(x)\dot{\le}x^{n_a}$ dominate and (82) follows. We are left to check the case $t_{12}(x)\doteq t_{22}(x)\doteq x^{n_a}$, which together with (88) leads to (82) as well. □

Lemma VI: With z, θ_0', and θ_0 as defined in Section III-A, we have

$$\Pr\left(z\sin^2\theta_0' \le x\right)\doteq\Pr\left(z\sin^2\theta_0 \le x\right).$$

Proof: From (18), (32) the pdf of $\theta_0' = \max_k \theta_{1k}'$ is derived through results in order statistics

$$\begin{aligned}f_{\theta_0'}(\theta) &= \frac{M}{C^{N_T-1}}[\sin^2\theta]^{M-1}\sin 2\theta,\\ M &= (N_T-1)(N_R-1), \quad \theta\in(0,\psi_0)\end{aligned} \tag{89}$$

where $C = \int_0^{\psi_0} f_{\theta_{li}}(x)dx,\ 2 \le i \le N_T$. Therefore, we have

$$\begin{aligned}&\Pr\left(z\sin^2\theta_0' \le x\right)\\ &= \int_0^{\psi_0} F_z\left(\frac{x}{\sin^2\theta}\right) f_{\theta_0'}(\theta)\cdot d\theta\\ &= \frac{M}{C^{N_T-1}}\int_0^{\psi_0} F_z\left(\frac{x}{\sin^2\theta}\right)[\sin^2\theta]^{M-1}\sin 2\theta\cdot d\theta\\ &\overset{t=\sin^2\theta}{=} \frac{M}{C^{N_T-1}}\int_0^{a_0} F_z\left(\frac{x}{t}\right)t^{M-1}\cdot dt\\ &\overset{t_1=\frac{t}{a_0}}{=} \frac{Ma_0^M}{C^{N_T-1}}\int_0^1 F_z\left(\frac{(x/a_0)}{t_1}\right)t_1^{M-1}\cdot dt_1\end{aligned} \tag{90}$$

where $a_0 = \sin^2\psi_0$ is a positive real number. On the other hand, by applying (37) we have

$$\begin{aligned}\Pr\left(z\sin^2\theta_0 \le x\right) &= \int_0^{\pi/2} F_z\left(\frac{x}{\sin^2\theta}\right) f_{\theta_0}(\theta)\cdot d\theta\\ &= M\int_0^{\pi/2} F_z\left(\frac{x}{\sin^2\theta}\right)[\sin^2\theta]^{M-1}\sin 2\theta\cdot d\theta\\ &\overset{t=\sin^2\theta}{=} M\int_0^1 F_z\left(\frac{x}{t}\right)t^{M-1}\cdot dt.\end{aligned} \tag{91}$$

By comparing (90) and (91), we get

$$\Pr\left(z\sin^2\theta_0' \le x\right)\doteq\Pr\left(z\sin^2\theta_0 \le x\right).$$ □

D. The Derivation of (38)

We continue the evaluation of (91) to derive the polynomial expansion of $\Pr(z\sin^2\theta_0' \le x)$

$$\begin{aligned}&\Pr\left(z\sin^2\theta_0 \le x\right)\\ &= M\int_0^1 F_z\left(\frac{x}{t}\right)t^{M-1}\cdot dt\end{aligned}$$

$$\begin{aligned}&\overset{m=1/t}{=} M\int_1^\infty F_z(mx)\frac{1}{m^{M+1}}\cdot dm\\&= M\int_1^\infty\left[1-e^{-mx}\sum_{k=0}^{M+N_T-2}\frac{m^kx^k}{k!}\right]\frac{1}{m^{M+1}}\cdot dm\\&=1-P_1(x)-P_2(x)\end{aligned}\tag{92}$$

where

$$P_1(x)=M\sum_{k=0}^{M}\frac{x^k}{k!}E_{M+1-k}(x),$$
$$P_2(x)=M\sum_{k=M+1}^{M+N_T-2}\frac{x^k}{k!}E_{-(k-M-1)}(x)$$

with $E_k(x)=\int_1^\infty\frac{e^{-xm}}{m^k}dm$ the integral exponential function. From [13], we have the following recursive rules:

$$E_{k+1}(x)=\frac{1}{k}\left(e^{-x}-xE_k(x)\right)$$
$$E_{k+1}(x)=\frac{e^{-x}}{k!}\sum_{i=0}^{k-1}(-1)^i(k-i-1)!x^i+\frac{(-1)^k}{k!}x^kE_1(x).$$

Therefore, after some involved mathematical manipulations, we have

$$P_1(x)=Me^{-x}\sum_{k=0}^{M-1}c_kx^k$$

where

$$c_k=\sum_{i=0}^{k}\frac{(-k)^{k-i}(M-k-1)!}{(M-i)!i!}$$

for $k\le M-1$. We can expand e^{-x} by its Taylor's series and get the polynomial equivalent form: $P_1(x)=\sum_{n=0}^\infty a_nx^n$. For $n\le M-1$

$$\begin{aligned}a_n&=M\sum_{k=0}^{n}\frac{(-1)^{n-k}}{(n-k)!}\sum_{i=0}^{k}\frac{(-1)^{k-i}(M-k-1)!}{(M-i)!i!}\\&=M\frac{1}{n!}\sum_{k=0}^{n}\frac{(-1)^n n!(M-k-1)!}{(n-k)!M!}\sum_{i=0}^{k}\frac{(-1)^iM!}{(M-i)!i!}\\&=\frac{1}{n!}\sum_{k=0}^{n}\frac{(-1)^n\binom{n}{k}}{\binom{M-1}{k}}\sum_{i=0}^{k}(-1)^i\binom{M}{i}\\&=\frac{1}{n!}\sum_{k=0}^{n}\frac{(-1)^n\binom{n}{k}}{\binom{M-1}{k}}(-1)^k\binom{M-1}{k}\\&=\frac{(-1)^n}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}=\begin{cases}1,&n=0\\0,&0<n\le M-1\end{cases}\end{aligned}$$

where we use the equality [13]

$$\sum_{i=0}^{k}(-1)^i\binom{M}{i}=(-1)^k\binom{M-1}{k}.$$

With a similar approach, we can obtain $a_M=-\frac{1}{M!}$, therefore

$$P_1(x)=1-\frac{1}{M!}x^M+o(x^M).\tag{93}$$

On the other hand, $P_2(x)$ can be represented as

$$P_2(x)=Me^{-x}\times\sum_{k=M+1}^{M+N_T-2}\left\{\frac{(k-M-1)!}{k!}\left[\sum_{i=0}^{k-M-1}\frac{x^{l+M}}{i!}\right]\right\}$$

and after some manipulations, we obtain

$$P_2(x)=b_Mx^M+o(x^M)\tag{94}$$

where $b_M=M\sum_{k=0}^{N_T-3}\frac{k!}{(M+k+1)!}$. By combining (92), (93) and (94), we can derive (38). Furthermore, from [13], we have

$$\sum_{k=1}^{\infty}\frac{k!}{(M+k+1)!}=\frac{1}{M(M+1)!}$$

therefore

$$\begin{aligned}\sum_{k=0}^{N_T-3}\frac{k!}{(M+k+1)!}&<\sum_{k=0}^{\infty}\frac{k!}{(M+k+1)!}\\&=\sum_{k=1}^{\infty}\frac{k!}{(M+k+1)!}+\frac{1}{(M+1)!}\\&=\frac{1}{M(M+1)!}+\frac{1}{(M+1)!}=\frac{1}{M.M!}\end{aligned}$$

so the coefficient of x^M in (38) is always positive.

E. Lemma VII

Lemma VII: For any N_T-L+1 distinct integers within $[1,N_T]$, denoted as an ordered list $k_1\sim k_{N_T-L+1}$, and any N_T-L+1 different subsets from $U_1\sim U_{N_U}$ denoted as an ordered list $U_{j_1}\sim U_{j_{N_T-L+1}}$, suppose $\mathbf{h}_{k_i}\in U_{j_i}$, and $\mathbf{h}_{k_i}\notin U_{j_l}$ for $l>i$, i.e., $\mathbf{h}_{k_i}$ belongs to subset U_{j_i} but not those after. We have

$$\Pr\left(R_{k_1}^{(j_1)'}\le x,R_{k_2}^{(j_2)'}\le x,\ldots,R_{k_{N_T-L+1}}^{(j_{N_T-L+1})'}\le x\right)=\left[\Pr(R_{k_1}^{(j_1)'}\le x)\right]^{(N_T-L+1)}\tag{95}$$

where $R_k^{(j)'}$ is modified from $R_k^{(j)}$ with k indicating the index of the column vector in the original channel matrix $\mathbf{H}$, instead of $\mathbf{H}_j$.[10]

[10]Lemma II is a special case of Lemma VII. However, the proof of Lemma II bears some interesting geometric elements, and is needed for Corollary I and Proposition II.

Proof: From the definition, $R_{k_1}^{(j_1)'} = \|\mathbf{P}\mathbf{h}_{k_1}\|^2$, where $\mathbf{P} = \mathbf{I} - \mathbf{B}\mathbf{B}^\dagger$ is the projection matrix to the null space of $\mathrm{range}\{U_{j_1}\backslash\mathbf{h}_{k_1}\}$, the subspace spanned by the remaining column vectors in subset U_{j_1} except $\mathbf{h}_{k_1}$ [20], and $\mathbf{B}$ is composed of any basis of this subspace. Since any projection matrix is idempotent, i.e., , $\mathbf{P}^2 = \mathbf{P}$, its eigenvalues are either 1 or 0. Noting that $\mathbf{P}$ is Hermitian, we can write the eigenvalue decomposition of $\mathbf{P}$ as $\mathbf{P} = \mathbf{V}\Lambda\mathbf{V}^H$, where $\mathbf{V}$ is unitary, and $\Lambda = \mathrm{diag}(1^{N_R-L+1}, 0^{L-1})$. Therefore

$$R_{k_1}^{(j_1)'} = \left\|\mathbf{V}\Lambda\mathbf{V}^H\mathbf{h}_{k_1}\right\|^2 = \left\|\Lambda\mathbf{V}^H\mathbf{h}_{k_1}\right\|^2 \tag{96}$$

where the second equality follows by the fact that a unitary transformation preserves length. From the definition of $\mathbf{P}$, the unitary matrix $\mathbf{V}^H$ is independent of $\mathbf{h}_{k_1}$. Therefore, the conditional pdf

$$\begin{aligned} f\left(R_{k_1}^{(j_1)'} \mid \{U_{j_1}\backslash\mathbf{h}_{k_1}\}\right) &= f\left(\left\|\mathbf{\Lambda}\mathbf{V}_0^H\mathbf{h}_{k_1}\right\|^2\right) \\ &= f\left(\|\mathbf{\Lambda}\mathbf{h}_{k_1}\|^2\right) = f\left(R_{k_1}^{(j_1)'}\right) \end{aligned} \tag{97}$$

where $\mathbf{V}_0^H$ is a fixed unitary matrix dependent on the given realizations of $\{U_{j_1}\backslash\mathbf{h}_{k_1}\}$, and the second equality comes from the rotationally invariant property of the i.i.d. Gaussian vector $\mathbf{h}_{k_1}$ [18], [26]. That is, $R_{k_1}^{(j_1)'}$ is independent of any vector in $\mathrm{range}(U_{j_1}\backslash\mathbf{h}_{k_1})$. It is therefore straightforward to show that $R_{k_1}^{(j_1)'} \sim R_{k_{N_T-L+1}}^{(j_{N_T-L+1})'}$ are jointly independent and identically distributed, and Lemma VII holds. □

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments.

References

[1] R. W. Heath and A. Paulraj, "Antenna selection for spatial multiplexing systems based on minimum error rate," in *Proc. 2001 IEEE Int. Conf. Commun. (ICC'01)*, June 2001, vol. 7, pp. 2276–2280.

[2] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection-an overview," *IEEE Microw. Mag.*, vol. 5, no. 1, pp. 46–56, Mar. 2004.

[3] M. K. Simon and M. S. Alouini, *Digital Communications over Fading Channels*. New York: Wiley, 2000.

[4] D. A. Gore and A. J. Paulraj, "MIMO antenna subset selection with space-time coding," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2580–2588, Oct. 2002.

[5] I. Bahceci, T. M. Duman, and Y. Altunbasak, "Antenna selection for multiple-antenna transmission systems: Performance analysis and code construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2669–2681, Oct. 2003.

[6] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: Theory and algorithms," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2796–2807, Nov. 2003.

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

[8] A. Gorokhov, D. Gore, and A. Paulraj, "Diversity versus multiplexing in MIMO systems with antenna selection," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2003.

[9] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1841–1852, 1999.

[10] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge University Press, 1998.

[11] D. Pedoe, *A Course of Geometry for Colleges and Universities*. Cambridge, U.K.: University Press, 1970.

[12] S. Loyka and F. Gagnon, "Performance analysis of the V-BLAST algorithm: An analytical approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1326–1337, Jul. 2004.

[13] I. S. Gradsbteyn and I. M. Ryzbik, *Table of Integrals, Series, and Products*, 6th ed. New York: Academic, 2000.

[14] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 339–347, Feb. 2004.

[15] H. Zhang and H. Dai, "Fast transmit antenna selection algorithms for MIMO systems with fading correlation," in *Proc. Vehicular Technology Conference, Fall 2004, VTC Fall 04*, Sep. 2004.

[16] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1397, Aug. 2003.

[17] N. Prasad and M. K. Varanasi, "Analysis of decision feedback detection for MIMO Rayleigh fading channels and optimization of power and rate allocations," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1009–1025, June 2004.

[18] R. J. Muirhead, *Aspect of Multivariate Statistics Theory*. New York: Wiley, 1982.

[19] H. A. David, *Order Statistics*, 3rd ed. New York: Wiley, 2003.

[20] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Singapore: SIAM, 2000.

[21] J. G. Proakies, *Digital Communications*, 3 ed. New York: McGraw-Hill, 1995.

[22] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2003.

[23] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. Boston, MA: McGraw-Hill, 2002.

[24] Y. Jiang, X. Zheng, and J. Li, "Asymptotic performance analysis of V-BLAST," in *Proc. 2005 IEEE Global Communications Conference (GLOBECOM)*, St. Louis, MO, Nov. 2005.

[25] H. Zhang, H. Dai, and B. L. Hughes, "On the diversity-multiplexing tradeoff for ordered SIC receivers over MIMO channels," in *2006 IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, Jun. 2006.

[26] B. Hassibi, "Random matrices, integrals and space-time systems," in *DIMACS Workshop on Algebraic Coding and Information Theory*, Dec. 2003 [Online]. Available: http://dimacs.rutgers.edu/Workshops/CodingTheory/slides/hassibi.ppt

[27] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," Ph.D. Dissertation, MIT, , 1989 [Online]. Available: http://math.mit.edu/~edelman/thesis/thesis.ps