Diversity–Multiplexing Tradeoff in Multiple-Access Channels

David N. C. Tse, *Member, IEEE*, Pramod Viswanath, *Member, IEEE*, and Lizhong Zheng, *Member, IEEE*

***Abstract*—In a point-to-point wireless fading channel, multiple transmit and receive antennas can be used to improve the reliability of reception (diversity gain) or increase the rate of communication for a fixed reliability level (multiplexing gain). In a multiple-access situation, multiple receive antennas can also be used to spatially separate signals from different users (multiple-access gain). Recent work has characterized the fundamental tradeoff between diversity and multiplexing gains in the point-to-point scenario. In this paper, we extend the results to a multiple-access fading channel. Our results characterize the fundamental tradeoff between the three types of gain and provide insights on the capabilities of multiple antennas in a network context.**

***Index Terms*—Diversity, multiple input/multiple output (MIMO), multiple access, multiple antennas, space–time codes, spatial multiplexing.**

I. Introduction

THE role of multiple antennas in communication over a wireless channel has been well studied in the point-to-point scenario. The antennas can be used to boost the reliability of reception for a given data rate (providing *diversity* gain) or boost the data rate for a given reliability of reception (providing *multiplexing* or degrees of freedom gain). In a scenario with several users communicating to a common receiver, multiple receive antennas also allow the spatial separation of the signals of different users, thus providing a *multiple-access* gain. This use of multiple antennas is also called space-division multiple access (SDMA). Recent work [12] has characterized the fundamental tradeoff between the diversity and multiplexing gain in the point-to-point context. The objective of this paper is to extend the results to the many-to-one context, thus providing a complete picture on the tradeoff between the three type of gains. This leads to insights on the capabilities of multiple antennas in a network context.

Consider point-to-point wireless communication over a block length of l symbols during which the channel from the m transmit to the n receive antennas is random but not changing over the duration of communication (slow fading scenario). We focus our interest on the high signal-to noise ratio (SNR) scenario and assume that the receiver has full and accurate knowledge of the fading channel. Recognizing that $\log \mathsf{SNR}$ is the capacity of an additive white Gaussian noise (AWGN) channel at high SNR, we define $r := R/\log \mathsf{SNR}$ as the multiplexing gain of a code with data rate R. Consider the behavior of its maximum-likelihood (ML) error probability: if P_e decays as SNR^{-d} for large SNR, then we say that this code has a diversity gain of d. The best decay rate for a given multiplexing gain r is denoted by $d^*_{m,n}(r)$. A complete characterization of this function for independent and identically distributed (i.i.d.) Rayleigh fading is done in [12]: provided that the block length $l \geq m+n-1$

$$d^*_{m,n}(r) = (m-r)(n-r)$$

for every integer $r \leq \min(m,n)$, and the entire curve is piecewise linear joining these points. The inverse of this function, $r^*_{m,n}(d)$, is the largest achievable multiplexing gain for a given diversity gain d. The *maximal* diversity gain is mn, attained when $r \to 0$. The *maximal* multiplexing gain is $\min(m,n)$, the number of degrees of freedom in the channel, attained when $d \to 0$. While the maximal diversity gain is simply the number of independent channel gains between antenna pairs and the maximal multiplexing gain is the dimension of the signal space, the derivation of the entire tradeoff curve requires a more elaborate analysis of channel outage events.

Now consider the i.i.d. Rayleigh-fading *multiple access* channel with K users, with each user having m transmit antennas and the single receiver having n receive antennas. Each user i has a multiplexing gain r_i, i.e., its data rate $R_i = r_i \log \mathsf{SNR}$. The optimal decoder that minimizes the error probability for each user i is the (individual) ML decoder. We require this minimal error probability to decay at least as fast as SNR^{-d}, i.e., each user has a diversity gain of d. In this paper, we characterize exactly the set of multiplexing gain tuples $(r_1,\ldots,r_K)$ that still allow each user to have a diversity gain of d.

In the *symmetric* situation, i.e., the multiplexing gains of all the users are equal (to say r), our characterization takes on a particularly simple form. First, the maximal multiplexing gain achievable by each user is $\min\left(m, \frac{n}{K}\right)$, which can be interpreted as the degrees of freedom per user. This is not

Manuscript received November 12, 2003; revised April 21, 2004. The work of D. N. C. Tse was supported in part by funding from Qualcomm Inc., a matching grant from the California MICRO program, and the National Science Foundation under Grant CCR-0118784. The work of P. Viswanath was supported in part by the National Science Foundation under Grant CCR–0325924 and under a grant from Motorola Inc., as part of the Center for Communication at the University of Illinois at Urbana-Champaign. The material in this paper was presented at the DIMACS Workshop on Multiantenna Channels: Capacity, Coding and Signal Processing, Piscataway, NJ, October 2002 and at the IEEE International Symposium on Information Theory, Yokohama, Japan, June 2003.

D. N. C. Tse is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720-1770 USA (e-mail: dtse@eecs.berkeley.edu).

P. Viswanath is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: pramodv@uiuc.edu).

L. Zheng is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: lizhong@mit.edu).

Communicated by M. Medard, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2004.833347

too surprising as, just like in the point-to-point scenario, this follows from a simple dimension counting argument. More interestingly, we show that within this range of achievable multiplexing gains, the tradeoff performance can be divided into two regimes: the lightly loaded regime and the heavily loaded regime; the corresponding highest achievable diversity gains are, respectively:

- $d^*_{m,n}(r)$, the diversity gain attained as if only one user is in the system, for

$$r \leq \min\left(m, \frac{n}{K+1}\right);$$

- $d^*_{Km,n}(Kr)$, the diversity gain as if the K users pool up their transmit antennas together, for

$$r \in \left(\min\left(m, \frac{n}{K+1}\right), \min\left(m, \frac{n}{K}\right)\right).$$

Thus, there are two fundamental parameters characterizing the performance of each user:

- $\min(m, n/K)$: the degrees of freedom per user, limiting its maximum possible multiplexing gain;
- $\min(m, n/(K+1))$: the threshold on the multiplexing gain below which the error probability of the user is as though it were the only user in the system.

In particular, we note that when the number of transmit antennas m is no more than $\frac{n}{K+1}$, the two parameters coincide. The lightly loaded regime then extends over the entire range of achievable multiplexing gains and the error probability performance of each user is the same as if the other users were not transmitting at all, i.e., *single-user performance*. In this case, the multiple-access gain is obtained for free.

These results are rather surprising: in [11], the authors have shown that under a linear decorrelating receiver, with each additional receive antenna we can either increase the diversity of each user by one, or add an extra user at the same diversity level, but not both. Our results show that this tradeoff is not fundamental and is due to the limitation of a suboptimal receiver structure. Indeed, if we use the ML receiver and we are in the regime $m \leq \frac{n}{K+1}$, one can add an extra user *and* simultaneously increase the diversity of each user if there is an additional receive antenna. We will also see that other strategies, such as successive cancellation and rate splitting, do not significantly close this performance gap between the linear and ML receivers.

Our result also sheds insight into the typical way error occurs in the multiple-access fading channel under optimal decoding. We show the following.

- For $r \leq \min\left(m, \frac{n}{K+1}\right)$, the typical way for error to occur is that just one of the users' message is decoded incorrectly.
- For $r \geq \min\left(m, \frac{n}{K+1}\right)$, the typical way for error to occur is that all the user messages are decoded incorrectly.

This result sheds insight into designing packet retransmit protocols for the fading uplink channel in a cellular wireless system.

The paper is organized as follows. We begin in Section II with notations and the formal statement of the model and the problem studied. Our main result, a characterization of the multiplexing rate tuples of the users as a function of the common diversity gain for each user is in Section III. In Section IV, we go through a few examples to infer the network level impact of multiple antennas in some simple settings. Section V discusses the typical ways in which errors can occur. Section VI deals with the performance of various suboptimal decoders: successive cancellation, time sharing, and rate splitting. Sections VII and VIII contain the proofs of the main results. In this paper, we focus on the equal diversity requirement case: the characterization of the multiplexing–diversity tradeoff when users have different diversity requirements is more difficult and remains an open problem.

II. System Model and Problem Formulation

A. Channel Model

Consider the multiple access channel in Fig. 1. K noncooperating transmitters communicate independent messages to a single receiver. Each of the transmitters has an array of m transmit antennas and the receiver has an array of n receive antennas. Over a block length of time equal to l symbols, the received signal (an element of $\mathbb{C}^{n\times l}$) is

$$\boldsymbol{Y} = \sqrt{\frac{\mathsf{SNR}}{m}} \sum_{i=1}^{K} \boldsymbol{H}_i \boldsymbol{X}_i + \boldsymbol{W}. \tag{1}$$

Here $\boldsymbol{W} \in \mathbb{C}^{n\times l}$ represents additive noise at the receiver. The noise at each of the receive antennas at each time is i.i.d. $\mathcal{CN}(0,1)$; $\mathcal{CN}(0,a)$ denotes a complex Gaussian random variable with i.i.d. zero mean, variance $a/2$, Gaussian random variables as its real and imaginary parts.

The channel between transmitter i and the receiver is represented by the $n \times m$ matrix $\boldsymbol{H}_i$. We assume that the channel stays constant over the entire block length of time l and is known by the receiver, i.e., the slow fading scenario. The transmitter only has a statistical characterization of the channels and is unaware of the actual realizations. We statistically model $\{\boldsymbol{H}_i\}_{i=1\dots K}$ to be i.i.d. with $\mathcal{CN}(0,1)$ entries, the richly scattered Rayleigh-fading environment.

Our focus is on communication by the users over the fixed block of l symbols. A *codebook* of user i (denoted by $\mathcal{C}_i$) comprises of $\lceil 2^{R_i l} \rceil$ codewords, with R_i denoting its rate of communication. We denote the codewords, each an element of $\mathbb{C}^{m\times l}$ as $\left\{\boldsymbol{X}_i(j),\ j = 1 \dots 2^{\lceil R_i l \rceil}\right\}$. There is a constraint on the average unit energy per transmit antenna per symbol per codeword

$$\frac{1}{ml \mid \mathcal{C}_i \mid} \sum_{j=1}^{|\mathcal{C}_i|} \|\boldsymbol{X}_i(j)\|_F^2 \leq 1. \tag{2}$$

Here $\|\cdot\|_F$ is the Frobenius norm on matrices

$$\|A\|_F^2 = \sum_{i,j} \mid A_{ij} \mid^2.$$

B. Diversity and Multiplexing Tradeoff

The receiver makes a decision for each of the users based jointly on the received matrix $\boldsymbol{Y}$ and knowledge of the channel realization. The performance is given by average error probabilities $P_e^{(i)}$, $i = 1, 2 \dots, K$, averaged over the equally likely messages and the channel realizations.

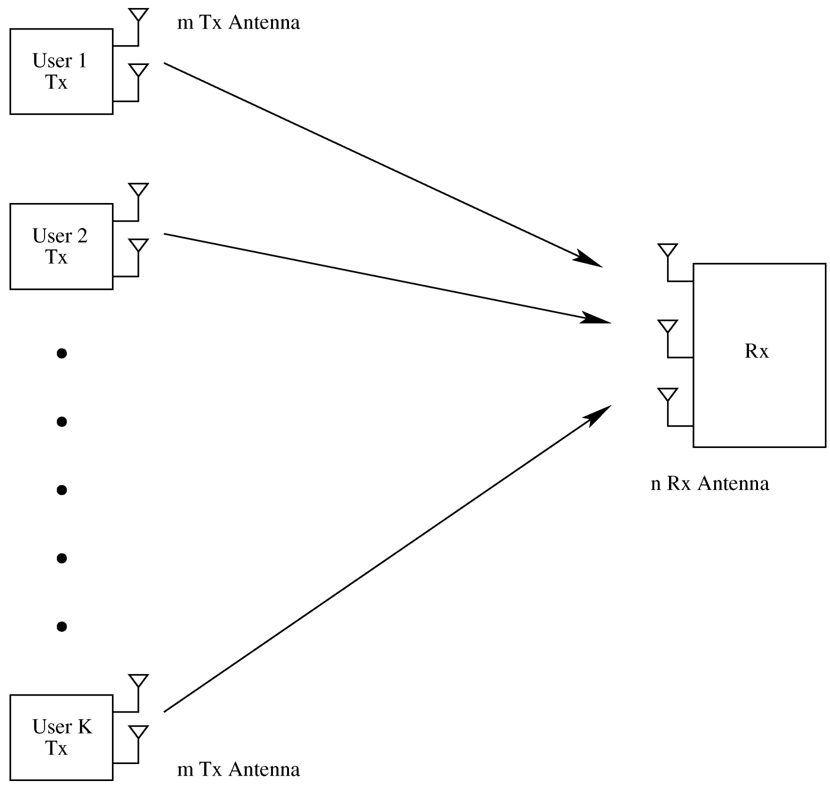

Fig. 1. A multiple-ccess system with K users each with m transmit antennas and a single receiver with n antennas.

Multiple antennas provide two different types of benefits in a fading channel: *diversity* gain and *multiplexing* gain. These gains are well studied in the context of point-to-point communication, i.e., when there is only one transmitting user, and we briefly describe this.

For a fixed rate of transmission R, the error probability can decay with SNR as fast as

$$P_e \sim \frac{1}{\mathsf{SNR}^{mn}}.$$

The factor mn is called the maximal diversity gain, obtained by averaging over the mn independent channels gains between all the antenna pairs. In this context, multiple antennas provide additional reliability over single-antenna systems to *compensate* for the randomness due to fading.

On the other hand, the randomness due to fading can be *taken advantage* of by creating parallel spatial channels. This concept is best motivated by a capacity result: [9], [2] showed that the *ergodic capacity* of the multiple-antenna channel scales like

$$C(\mathsf{SNR}) \sim \min(m,n) \log \mathsf{SNR} \quad \text{(bps/Hz)}$$

at high SNR. The parameter $\min(m,n)$ is the number of degrees of freedom in the channel and yields the maximum amount of spatial multiplexing gain possible.

The ergodic capacity is achieved by averaging over the variation of the channel over time. In the slow fading scenario, no such averaging is possible and one cannot communicate at the capacity $C(\mathsf{SNR})$ reliably. On the other hand, to achieve the maximal diversity gain mn, one needs to communicate at a *fixed* rate R, which becomes very small compared to the capacity at high SNR. This suggests a more interesting formulation of asking what is the largest diversity gain that can be achieved if one wants to communicate at a fixed *fraction* of the capacity. It leads to a formulation of the tradeoff between diversity and multiplexing gains, which we formalize below.

We think of a scheme $\{\mathcal{C}(\mathsf{SNR})\}$ as a family of codes, coding over one single coherence block, one at each SNR level. Let $R(\mathsf{SNR})$ and $P_e(\mathsf{SNR})$ denote their data rate (in bits per symbol period) and the ML probability of detection error, respectively.

Definition 1: A scheme $\{\mathcal{C}(\mathsf{SNR})\}$ is said to achieve spatial multiplexing gain r and *diversity gain* d if the data rate

$$\lim_{\mathsf{SNR}\to\infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} \geq r \tag{3}$$

and the average error probability

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log P_e(\mathsf{SNR})}{\log \mathsf{SNR}} \leq -d. \tag{4}$$

For each r, define $d^*_{m,n}(r)$ to be the supremum of the diversity gain achieved over all schemes. Equivalently, for each d, define $r^*_{m,n}(d)$ to be the supremum of the multiplexing gain achieved over all schemes.

For notational simplicity, we shorten (4) as $P_e(\mathsf{SNR}) \dot{\leq} \mathsf{SNR}^{-d}$; similarly, we say that $P_e(\mathsf{SNR}) \doteq \mathsf{SNR}^{-d}$ if equality holds in the limit.

The fundamental tradeoff between these two types of gains is the subject of [12], where a simple characterization of the diversity–multiplexing tradeoff curve $d^*_{m,n}(r)$ is obtained.

Theorem 1: [12] For block length $l \geq m+n-1$, the diversity–multiplexing tradeoff curve for the i.i.d. Rayleigh point-to-point channel is piecewise linear joining the integer points $(k,(m-k)(n-k)), k=0,\ldots,\min(m,n)$.

The diversity gain decreases from the maximal value mn to zero as the multiplexing gain increases from 0 to the degrees of freedom $\min(m,n)$. Note that the degrees of freedom available in the channel puts a limit on the maximum multiplexing gain achievable.

This formulation naturally generalizes to the multiple-access channel. Given a common diversity requirement d for the users, i.e.,

$$P_e^{(i)} \dot{\leq} \mathsf{SNR}^{-d}, \qquad i=1,\ldots,K$$

we want to characterize the set of the K-tuple multiplexing gains $(r_1,\ldots,r_K)$, i.e.,

$$R_i \sim r_i \log \mathsf{SNR}, \qquad i=1,\ldots,K$$

that can be achieved. This set of multiplexing gains is denoted by $\mathcal{R}(d)$.

In this paper, we focus on the role of antenna arrays in delivering improved diversity and multiplexing gains in multiple-access fading channels. One way to think about a coding scheme for the multiple-access channel is as a point-to-point coding scheme for Km transmit antennas but the signals on the K groups of m antennas each cannot be jointly coded together; independent messages are communicated from the these K groups of transmit antennas. Seen this way, our study here brings to sharp focus the role of joint coding across the transmit antennas in a point-to-point channel. Some related work can be found in [13].

III. Optimal Tradeoff

A. Basic Result

Our first result is an explicit characterization of $\mathcal{R}(d)$ when the block length is large enough.

Theorem 2: If the block length $l \geq Km+n-1$

$$\mathcal{R}(d)=\left\{(r_1,\ldots,r_K): \sum_{s\in S} r_s \leq r^*_{|S|m,n}(d), \forall S\subseteq\{1,\ldots K\}\right\}. \tag{5}$$

where $r^*_{|S|m,n}(\cdot)$ is the multiplexing–diversity tradeoff curve for a point-to-point channel with $|S|\cdot m$ transmit and n receive antennas.

The proof of this result sheds light on the typical way error occurs . We show that for block length $l\geq Km+n-1$, the typical way the error occurs is not by the additive noise being too large but by the channel being bad, i.e., in *outage*, when the target rate tuple does not lie in the multiple-access region defined by the realized channel matrices $\{\boldsymbol{H}_i\}_i$. This is a natural generalization of the concept of outage in point-to-point channel [6], [9]. Our proof technique crucially uses the outage formulation: we calculate the probability of this *outage* event and conditioned on no-outage show that the error probability is no worse than the probability of outage. Thus, the characterization of $\mathcal{R}(d)$ boils down to calculating the probability of outage for a given rate vector. This is easy: there are 2^K-1 constraints in the multiple-access capacity region for a given realization of the channel and for each constraint there is a probability of not meeting it. At the scale of interest, the probability of outage is the worst among all these probabilities. This ensures that we meet the diversity requirements in the 2^K-1 constraints in (5). Details of the proof of Theorem 2 are in Section VII.

B. Symmetric Tradeoff

It turns out that due to the special structure of the functions $r^*_{m,n}(\cdot)$, the tradeoff region can be further simplified. Let us first focus on the largest symmetric multiplexing gain $(r,\ldots,r)$ that can be achieved for a given diversity gain d. From Theorem 2, this symmetric rate is constrained by

$$kr \leq r^*_{km,n}(d), \qquad k=1,\ldots,K \tag{6}$$

and, hence, the largest symmetric multiplexing gain is given by

$$\min_{k=1,\ldots,K} \frac{1}{k} r^*_{km,n}(d). \tag{7}$$

Equivalently, the largest achievable symmetric diversity gain for fixed symmetric multiplexing gains is given by

$$d^*_{\text{sym}}(r) = \min_{k=1,\ldots,K} d^*_{km,n}(kr).$$

We have the following result.

Theorem 3:

$$d^*_{\text{sym}}(r) = \begin{cases} d^*_{m,n}(r), & r\leq \min(m,\frac{n}{K+1}) \\ d^*_{Km,n}(Kr), & r\geq \min(m,\frac{n}{K+1}). \end{cases} \tag{8}$$

Proof: See Section VIII. □

In the multiple-access channel, it is clear that the tradeoff curve cannot be better than the point-to-point single-user tradeoff curve with all but one user absent, namely, $d^*_{m,n}(r)$. The above result says that if the load of the system is sufficiently "light" (r small), the single-user tradeoff can be achieved for every user simultaneously. In particular, if the receiver has enough receive antennas such that $m\leq\frac{n}{K+1}$, then

$$\min\left(m,\frac{n}{K+1}\right)=\min\left(m,\frac{n}{K}\right)=m$$

and single-user performance is achieved for all r: the system is always lightly loaded; see Fig. 2.

On the other hand, if $m\geq\frac{n}{K+1}$, then single-user performance is achieved as long as the users are all transmitting a low enough data rate: $r\leq\frac{n}{K+1}$; see Fig. 3. Moreover, as long as the system operates within the lightly loaded regime, admitting one more user into the system does not degrade the performance of other users, a very desirable property. In this regime, the system provides multiple-access capability without compromising the performance of individual users.

In the heavily loaded regime, i.e., $r>\frac{n}{K+1}$, the symmetric diversity gain is $d^*_{Km,n}(Kr)$. The tradeoff is as though the K users are *pooled* together into a single user with Km antennas and multiplexing gain Kr. In this regime, the performance of each user is affected by the presence of other users. Note that the total number of degrees of freedom in the resulting point-to-point channel is $\min(Km,n)$, and hence,

$$d^*_{\text{sym}}(r)=d^*_{Km,n}(Kr)=0, \quad \forall r\geq\min(m,n/K).$$

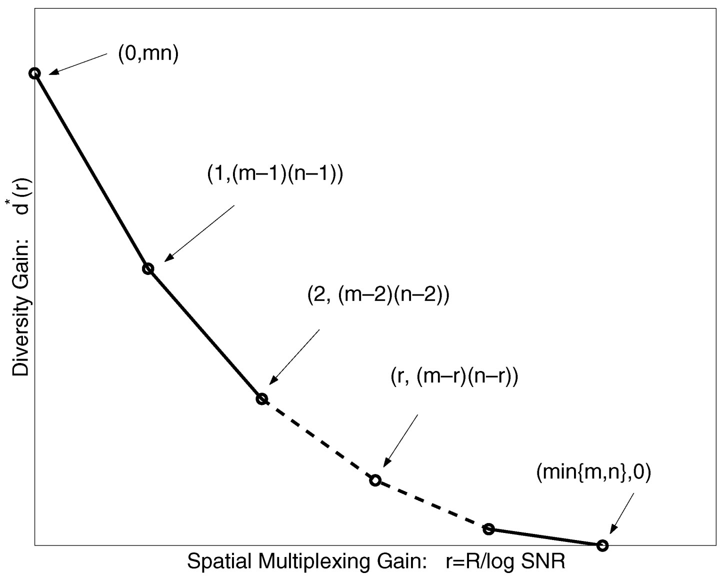

Fig. 2. Symmetric diversity–multiplexing tradeoff for $m \le \frac{n}{K+1}$ is the same as the single-user curve.

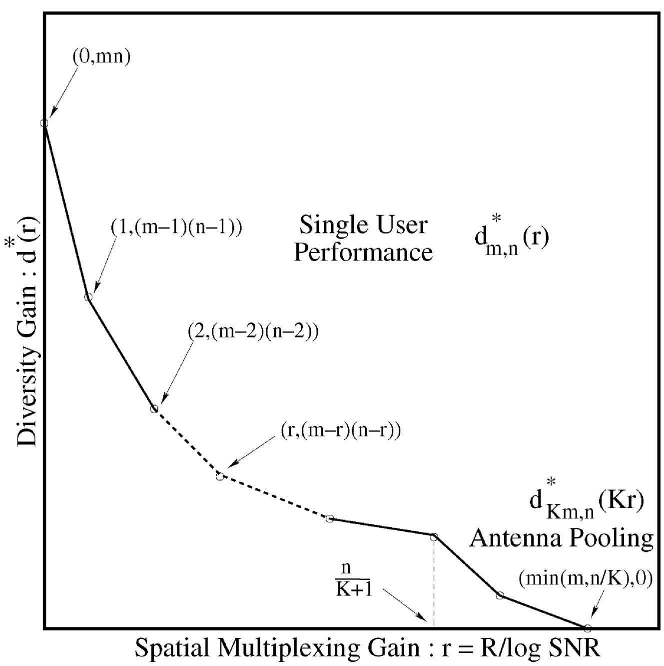

Fig. 3. Symmetric diversity–multiplexing tradeoff for $m > \frac{n}{K+1}$. Same as single-user curve up for $r \le \frac{n}{K+1}$, and switched to the antenna pooled curve for $r > \frac{n}{K+1}$. For $r > \min(m, \frac{n}{K})$, zero diversity gain is achieved. When $\alpha = \frac{n}{K}$ is large, these two thresholds coincide and the multiple-access tradeoff curve is the same as the single-user curve but truncated at $r = \alpha$.

This parameter can be thought of as the number of degrees of freedom *per* user.

Equation (7) says that the symmetric diversity–multiplexing curve is the minimum of K curves. For values of r arbitrarily close to zero, the curve $d^*_{1,n}(r)$ is clearly the smallest one, since $d^*_{km,n}(0) = kmn$ is smallest for $k = 1$. Hence, the single-user curve must determine $d^*_{\rm sym}$ for r sufficiently small. What Theorem 3 says is that no other curve can determine $d^*_{m,n}(r)$ except for $d^*_{Km,n}(Kr)$, and this happens when $r > n/(K+1)$.

In the scenario when the number of users K is much larger than the number of receive antennas n, a particularly simple picture emerges. In this case, $\frac{n}{K+1} \approx \frac{n}{K}$ and

$$\alpha \stackrel{\rm def}{=} \min\left(m, \frac{n}{K}\right) = \frac{n}{K} \tag{9}$$

is the degrees of freedom per user. When $r > \alpha$, $d^*_{\rm sym}(r) = 0$: the multiplexing gain cannot exceed the degrees of freedom per user. When $r < \alpha$, $d^*_{\rm sym}(r) = d^*_{m,n}(r)$, the single-user diversity–multiplexing performance. Thus, the presence of multiple users has the effect of truncating the single-user tradeoff curve at $r = \alpha$; see Fig. 3.

It should be emphasized that, *a priori*, there is no guarantee that transmitting at a multiplexing gain r less than the degrees of freedom $\min(m, n/K)$ per user would yield single-user diversity–multiplexing performance. This condition only guarantees that the multiplexing gain r is achievable with *nonzero* diversity gain: it ensures that the signal space has enough dimensions to linear independently place the spatial signatures of all the users, so that there is a possibility to distinguish between the different users. But when we discuss the diversity–multiplexing tradeoff, we are concerned with the error probability performance itself; even when there are enough dimensions, the random channel-dependent spatial signatures of different users may be closely aligned with each other with some probability, resulting in interference between users and degradation of the single-user error performance. What Theorem 3 says is that, under the *stronger* condition that $r < \min(m, n/(K+1))$, this is *not* a dominating event and single-user error performance is achieved. Somewhat surprisingly, this condition approaches the degree of freedom condition, based on pure dimension counting, when the number of users K is much larger than the number of receive antennas n.

C. Optimal Tradeoff Region Revisited

The structure of $d^*_{\rm sym}$ suggests it is possible to obtain a simpler representation for $\mathcal{R}(d)$ than the one given in Theorem 2.

Indeed, we have the following result, and the proof relegated to Section VIII.

Theorem 4: Suppose $d_1 = 0$ and $d_2, \ldots, d_K$ are defined by

$$d_k \doteq d^*_{m,n}\left(\frac{n}{k+1}\right), \qquad k = 2, \ldots, K. \tag{10}$$

Then for $d \geq d_k$

$$\mathcal{R}(d) = \left\{(r_1, \ldots, r_K) : r_i \leq r^*_{m,n}(d), i = 1, \ldots, K\right\} \tag{11}$$

and for $d \in [d_{l-1}, d_l]$

$$\mathcal{R}(d) = \begin{Bmatrix} (r_1, \ldots, r_K) : \sum_{s \in S} r_s \leq r^*_{(|S|m,n)}(d), \\ \forall S \subseteq \{1, \ldots K\}, | S | \in \{1, l, l+1, \ldots, K\} \end{Bmatrix}. \tag{12}$$

For large enough desired diversity gain $d \geq d_K$, the region of multiplexing gains is a square, i.e., each user achieves single-user performance. This is a direct consequence of the earlier result on the symmetric tradeoff. For smaller diversity gain requirements, other constraints start coming into play. When the diversity gain requirement is small enough, all $2^K - 1$ constraints become relevant.

Furthermore, the tradeoff region has an interesting combinatorial structure.

A polymatroid with rank function f (mapping subsets of $\{1, \ldots, K\}$ to nonnegative reals) is the following polyhedron:

$$\left\{(x_1, \ldots, x_K) \in \mathbb{R}^K_+ : \sum_{i \in S} x_i \leq f(S), \; \forall S \subseteq \{1, \ldots, K\}\right\}. \tag{13}$$

The rank function should be nonnegative, mapping the null set to zero, and

$$f(S \cup \{t\}) \geq f(S) \tag{14}$$

$$f(S \cup T) + f(S \cap T) \leq f(S) + f(T). \tag{15}$$

An important property of polymatroids is a simple characterization of its vertices. In particular, for every permutation π on the set $\{1, \ldots, K\}$ the point x^π

$$x^\pi_{\pi_i} \stackrel{\text{def}}{=} f(\{\pi_1, \ldots, \pi_i\}) - f(\{\pi_1, \ldots \pi_{i-1}\}), \quad i = 1, \ldots, K \tag{16}$$

meets the constraints in (13) and furthermore is a *vertex*. In fact, this can also be taken as a *definition* of a polymatroid. If points defined in (16) satisfy the constraints in (13) for every permutation π, then the function f must have the properties in (14) and (15) and the polyhedron in (13) is a polyhedron. Since the vertices are fully characterized, maximizing linear functions over a polymatroid is easy.

Theorem 5: Given a diversity requirement d, let l satisfy $d \in [d_{l-1}, d_l]$. The tradeoff region $\mathcal{R}(d)$ is a polymatroid, with rank function $f(\cdot)$ given by

$$f(S) = \begin{cases} | S | \, r^*_{m,n}(d), & 0 \leq | S | \leq l-1 \\ r^*_{|S|m,n}(d), & l \leq | S | \leq K. \end{cases}$$

Proof: See Section VIII. □

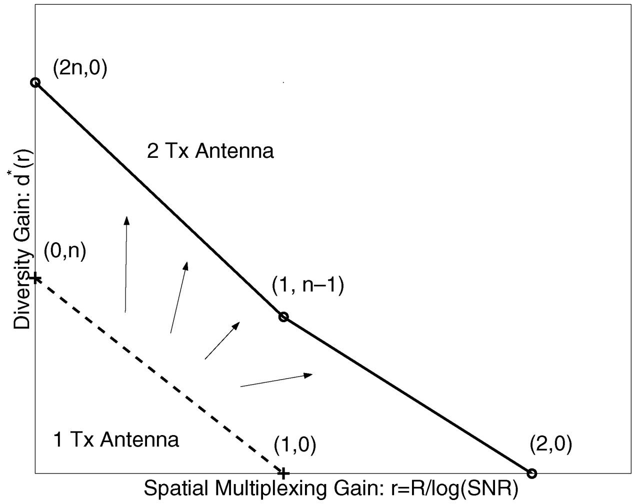

Fig. 4. Improvement in (symmetric) performance by adding a transmit antenna when system is lightly loaded. Increase in both degrees of freedom and diversity is seen.

IV. Examples

In this section we will go through a few examples to explore some implications of the results.

A. Example 1: Adding a Transmit Antenna

Consider a system with a receiver having n antennas and K users each with a single transmit antenna. What is the performance gain from adding an extra transmit antenna for each user? We focus on the symmetric operating point. Consider the following two cases.

Case 1: $K < \frac{n}{2} - 1$.

Here the number of users in the system is relatively small, the system is lightly loaded, and each user attains single-user performance even after adding the extra transmit antenna. The improvement in performance is seen in Fig. 4. In particular, the number of degrees of freedom per user is increased from one to two and the maximal diversity gain increases from n to $2n$.

Case 2: $K \geq n$.

The effect of adding a transmit antenna is seen in Fig. 5. In this case, there is no increase in degrees of freedom per user: it remains at $\frac{n}{K}$. The degrees of freedom is already limited by the number of *receive* antennas. Nevertheless, the diversity gain $d^*_{\text{sym}}(r)$ increases for each $r < \frac{n}{K}$.

This example shows the importance of viewing the multiple-access system as a whole rather than a set of K separate point-to-point links. While the latter view is accurate in the lightly loaded regime where each user attains single-user performance, it can be very misleading in general.

B. Example 2: Adding a Receive Antenna

What is the system-wide benefit of adding a receive antenna at the base station?

This question was asked in [11] in a specific context. The authors considered a multiple-access system with K users, each having one transmit antenna, and a receiver equipped with n antennas, with $n > K$. A simple linear receiver is used to demonstrate the performance improvement due to the use of multiple

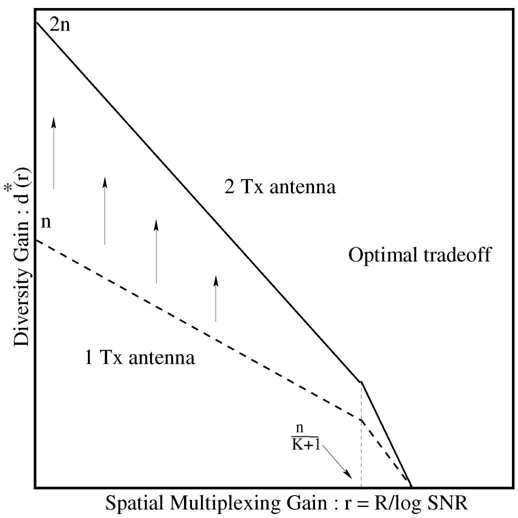

Fig. 5. Improvement in performance by adding a transmit antenna when system is heavily loaded. No increase in degrees of freedom but the tradeoff curve improves.

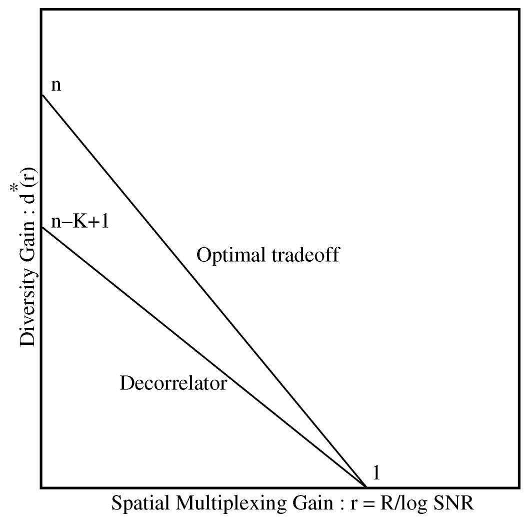

Fig. 6. Comparison of tradeoff curve of the decorrelator with the optimal.

antennas at the receiver. To receive the message from an individual user, the receiver treats the signals from all other users as interference, and uses a decorrelator ([10, Ch. 5]) to null them out. The authors showed that even with this simple receiver, significant performance gain can be obtained by using multiple antennas at the receiver. In particular, for quadrature phase-shift keying (QPSK) modulation, the error probability is of the order

$$P_e \doteq \mathsf{SNR}^{-(n-K+1)}.$$

This means that with n antennas at the receivers, one can null out the interference from $K-1$ users, thus accommodate K users, and provide each of them with interference-free reception with a diversity order $n-(K-1)$. This can be summarized as the following.

> An additional receive antenna can either increase the diversity order of every user by 1, or accommodate one more user at the same diversity order.

Notice that the "diversity order" in this statement corresponds to the maximum diversity gain on the tradeoff curve at $r=0$. In fact, it is easy to compute the entire diversity–multiplexing tradeoff curve under the decorrelator: it is given by $d(r) = (n-K+1)(1-r)$. (See [12, Sec. 7.2] for a derivation of this, in the context of vertical Bell Labs layered space–time (V-BLAST) architecture.)

We can compare this performance with the optimal diversity-multiplexing studied in this paper. For this scenario with K users, each having $m=1$ transmit antenna, and n receive antennas, Theorem 3 specifies the optimal tradeoff performance. Provided that $n>K$, (8) can be rewritten as

$$d^*_{\rm sym}(r) = d^*_{1,n}(r), \qquad r\in[0,1].$$

This is in the lightly loaded regime: each individual user can have *same* tradeoff performance of a point-to-point channel with $m=1$ transmit antenna and n receive antennas: a straight line connecting the maximum diversity gain point $(0,n)$ and the maximum multiplexing gain point $(1,0)$. Adding both an extra receiver and an extra user still maintains the lightly loaded regime. Thus, we can conclude the following.

> An additional receive antenna can increase the diversity order for each user by 1, and ***simultaneously*** accommodate one more user maintaining the tradeoff performance of the existing users.

Under the decorrelator, the additional receive antenna can either provide extra diversity or accommodate one more user, but not both. However, our results show that this tradeoff is not fundamental and is due to the limitation of the decorrelator; with the optimal receiver, you can, in fact, have the cake and eat it too.

More generally, we can compare the diversity-multiplexing tradeoff curve of the decorrelator with the optimal curve; this is shown in Fig. 6

Performance of receiver structures other than the decorrelator will be described in Section VI.

C. Example 3: Implications On Point-to-Point Optimal Codes

We have been analyzing the multiple-access diversity–multiplexing tradeoff in terms of the point-to-point tradeoff curve. But we can turn the table around and use our multiple-access results to shed some light on the point-to-point problem. Consider the point-to-point channel with M transmit and n receive antennas. We ask the question: what part of the tradeoff curve $d^*_{M,n}(r)$ can be achieved without coding across the transmit antennas? This is an interesting question as it potentially simplifies the point-to-point code design problem.

To this end, consider a multiple-access channel with M users and one transmit antenna each. The diversity gain achievable when each user transmits at a multiplexing gain r/M is given by the symmetric diversity–multiplexing tradeoff in Theorem 3

$$d^*_{\rm sym}\left(\frac{r}{M}\right) = \begin{cases} d^*_{1,n}\left(\frac{r}{M}\right), & \frac{r}{M} \le \min\left(1, \frac{n}{M+1}\right) \\ d^*_{M,n}(r), & \frac{r}{M} \ge \min\left(1, \frac{n}{M+1}\right). \end{cases} \tag{17}$$

From this, we observe that if $r \ge \min(M, \frac{nM}{M+1})$ then $d^*_{M,n}(r) = d^*_{\rm sym}(r/M)$. Since there is no coding across the users in the multiple-access channel, this means that for

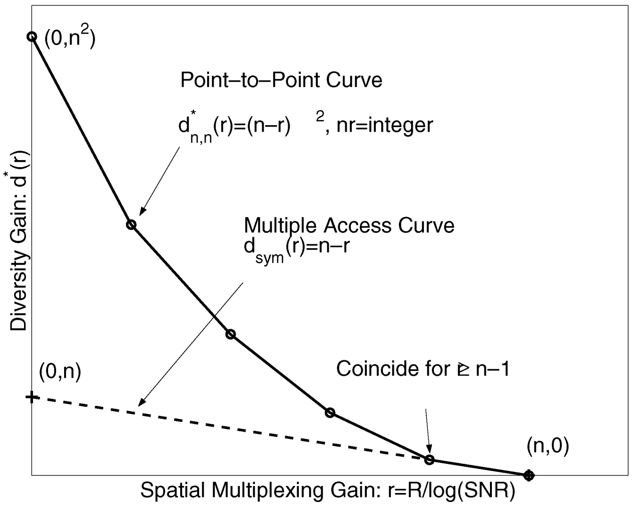

Fig. 7. The $n \times n$ point-to-point tradeoff curve coincides with the multiple-access curve in the high-rate region.

$r \geq \min(M, \frac{nM}{M+1})$, the tradeoff curve in the point-to-point channel $d^*_{M,n}(r)$ can, in fact, be achieved by separate coding at the transmit antennas. On the other hand, if $r \leq \min(M, \frac{nM}{M+1})$, then the symmetric tradeoff is determined by $d^*_{1,n}(r/M)$, which is smaller than or equal to $d^*_{M,n}(r)$. If it is strictly smaller, this implies that coding across the antennas is *necessary* to achieve the point-to-point channel tradeoff curve at those rates.

More specifically, we can consider three cases.

1) $M < n$. In this case, $nM/(M+1)$ is larger than M and further $d^*_{1,n}(r/M) < d^*_{M,n}(r)$ for all values of $r \leq M$, the number of degrees of freedom in the channel. Hence, in this case, without coding across the transmit antennas, one will never achieve the point-to-point tradeoff curve.

2) $M > n$. In this case, the point-to-point tradeoff curve $d^*_{M,n}(r)$ for $r \geq \frac{nM}{M+1}$ can be achieved without coding across the antennas. Further, since $d^*_{1,n}(r/M) < d^*_{M,n}(r)$ for all multiplexing gains $r < \frac{nM}{M+1}$, schemes that do not code across transmit antennas for the point-to-point channel are strictly suboptimal for these rates.

3) $M = n$. In this case, $d^*_{1,n}(r/n) = d^*_{n,n}(r)$ for $r \geq n-1$ and since $(n-1)/n < n/(n+1)$ the symmetric diversity–multiplexing tradeoff in (17) can be simplified to

$$d_{\text{sym}}(r/n) = n - r \tag{18}$$

for $r = 0$ to $r = n$.

This means, that in the point-to-point channel with multiplexing gain larger than $n-1$ the maximal diversity gain can be obtained by coding separately at each of the transmit antennas; see Fig. 7.

V. Typical Error Events

For a multiple-access channel with K users, the detection error event can be decomposed into a collection of disjoint error events, $\mathcal{E}_k, k = 1, \ldots, K$, where $\mathcal{E}_k$ is the event that the message from k users are erroneously decoded, and is referred as a "type-k" error event. An analysis of these error events for the AWGN multiple-access channel is presented in [4].

Now let us turn to the fading multiple-access channel with symmetric multiplexing and diversity gains for each user. We can lower-bound the probability of the type-k error by the probability of outage of the k users considered. From our calculation in Section VII-B, we know that the probability of this outage event is of the order

$$\mathsf{SNR}^{-d^*_{km,n}(kr)}. \tag{19}$$

On the other hand, we know from our discussion in Section VII-C that with a random Gaussian code the average probability of a type-k error event is no more than the same order in (19). We can hence conclude that (19) is the exact order of decay of the probability of type-k error event.

Since the overall error event is the union of the type-k error events, we can write

$$P_e(\mathsf{SNR}) = \sum_{k=1}^{K} P(\mathcal{E}_k) \doteq \sum_{k=1}^{K} \mathsf{SNR}^{-d^*_{km,n}(kr)} \doteq \mathsf{SNR}^{-\min_{k=1,\ldots,K} d^*_{km,n}(kr)}.$$

From Theorem 3, we know that for all rates $r \leq \min(m, n/K+1)$, the type-1 error event dominates all the others and for larger rates, the type-K error event is dominant. Thus, depending on the rates of the users, the typical way errors occur is either one of the users is in error or all the users are in error.

In practical multiple-access systems (such as the uplink of cellular wireless systems), the receiver (base station) uses redundancy in the packet format to check whether it has been correctly decoded (versions of cyclic redundancy check (CRC) codes are commonly used). Then, the base station feeds back to the users whether their packet was successfully received or was in error. This feedback is called automatic repeat request (ARQ) and allows the users to retransmit an erroneously received packet.

Our analysis of the typical way error occurs in the fading uplink channel provides insight into the ARQ protocol design. In particular, one important issue in ARQ protocol design is how much bandwidth has to be allocated to transmit the repeat request. A conservative approach is to reserve enough bandwidth with every packet transmission, to be able to transmit to all the users whether their packet has been correctly received or not; since this resource reservation is continuous (i.e., done with every packet transmission and not just one time), this design costs quite a bit of the downlink bandwidth. On the other hand, when lesser bandwidth is allocated for the repeat request then exceptions (when the number of errors is more than what can be transmitted) will have to be handled separately; if the exceptions happen rarely, then this design is preferable to the conservative one.

For large enough rates, we have identified the dominant error event to be the one where all the users' packets are in error. This suggests that we should allocate just enough resources with every packet transmission to be able to broadcast whether every user has to retransmit (all user packets are received erroneously) or not. On the other hand, for smaller rates, we know that it is most likely that only one of the users' packets is in error. In this case, it makes sense to reserve just enough bandwidth to be able to transmit which of the users' packet is in error (and handle the exceptions separately). In both cases, the insight in the identification of typical error events suggests that we can design the ARQ protocol with minimum reservation of resources to feed back packet errors, thus improving over the conservative resource reservation scheme.

VI. Performance of Some Non-ML Schemes

In Example 2 of Section IV, we have studied the diversity–multiplexing tradeoff performance of suboptimal linear receivers. In this section, we will look at the performance of other receiver structures. The comparison will be restricted to the symmetric scenario where each user attains the same multiplexing gain.

A. Successive Cancellation

The successive cancellation technique is used in multiple-access channels to reduce the joint demodulation of the data from all the users into a sequence of single-user demodulations.

In a system with K users equipped with m transmit antennas each, and n receive antennas, a successive cancellation receiver demodulates the data in K stages. At each stage, the receiver demodulates the data from one user, treating the signals from the uncanceled users as interference. Here, we consider the receiver that nulls out the interference with a decorrelator. After the data symbols from this user are decoded, its contribution is subtracted from the received signals before continuing to the next stage.

We start by studying the case $m = 1$, i.e., each user has only one transmit antenna. The successive cancellation process reduces the multiple-access channel into the following single-user subchannels:

$$\boldsymbol{Y}_i = \sqrt{\mathsf{SNR}}\boldsymbol{g}_i\boldsymbol{x}_i + \boldsymbol{W}_i$$

where $\boldsymbol{x}_i \in \mathcal{C}^{1\times l}$ is the signal transmitted by user i, $\boldsymbol{Y}_i, \boldsymbol{W}_i \in \mathcal{C}^{n\times l}$ are the received signal and noise for user i, respectively. $\boldsymbol{g}_i \in \mathcal{C}^n$ is the effective channel gain, which is the component of $\boldsymbol{h}_i$, the fading coefficients for user i, that is perpendicular to the signal space that needs to be nulled out.

In general, the performance of the successive cancellation receiver depends on the order in which the users are demodulated. We will start with the simple case that the demodulation takes a prescribed order, regardless of the realization of $\boldsymbol{H}_i$'s. Without loss of generality, assume that the data from user 1 is decoded first, and user 2 second, etc. It is clear that the performance of this receiver will be limited by that for the first user, and hence does not provide any improvement over a linear decorrelator without cancellation. The performance of this receiver has already been analyzed in [12]; for ease of generalization to other scenarios, we rederive it here.

Under the decorrelator, $\boldsymbol{g}_i$ is the component of $\boldsymbol{h}_i$ that is perpendicular to the subspace spanned by $\boldsymbol{h}_{i+1}, \ldots, \boldsymbol{h}_k$. Now without loss of optimality, the receiver can project each column

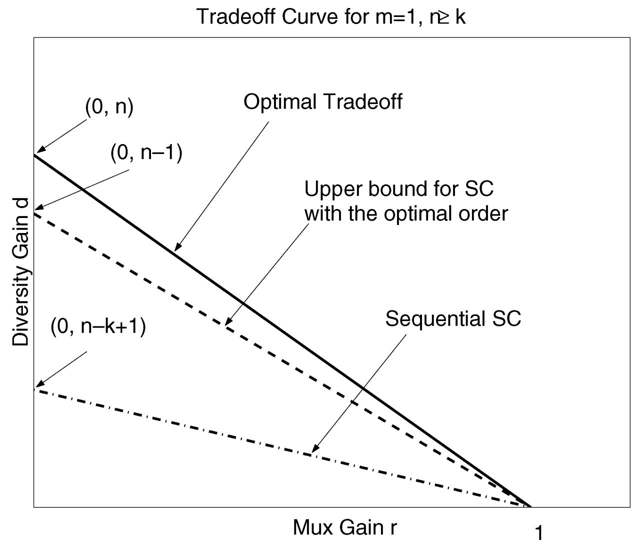

Fig. 8. Tradeoff for successive cancellation schemes with $m = 1$.

vector of $\boldsymbol{Y}_i$ into the direction of $\boldsymbol{g}_i$ and we can rewrite the subchannels as

$$\boldsymbol{y}_i = \sqrt{\mathsf{SNR}}\,\|\boldsymbol{g}_i\|\,\boldsymbol{x}_i + \boldsymbol{w}_i$$

where $\boldsymbol{y}_i = \boldsymbol{Y}_i\boldsymbol{g}_i/\|\boldsymbol{g}_i\| \in \mathcal{C}^{1\times l}$. Moreover, for each $i = 1, \ldots, K$, $\|\boldsymbol{g}_i\|^2$ is chi-square distributed with $n-K+i$ dimensions: $\|\boldsymbol{g}_i\|^2 \sim \chi^2_{2(n-K+i)}$. Clearly, this successive cancellation scheme only works for the case that $K \le n$. It is obvious that the first subchannel, $\boldsymbol{y}_1\sqrt{\mathsf{SNR}}\,\|\boldsymbol{g}_1\|\,\boldsymbol{x}_1 + \boldsymbol{w}_1$ is the bottleneck and hence $P(\mathcal{E}_1)$ dominates the error probability

$$P_e(\mathsf{SNR}) = P\left(\bigcup_i \mathcal{E}_i\right) \doteq P(\mathcal{E}_1).$$

Now we observe that the first subchannel is equivalent to a point-to-point link with one transmit and $n-K+1$ receive antennas, and applying Theorem 1 we have

$$P_e(\mathsf{SNR}) \doteq P(\mathcal{E}_1) \doteq \mathsf{SNR}^{-(n-K+1)(1-r)}.$$

This tradeoff performance is plotted in Fig. 8 in comparison with the optimal tradeoff curve $d^*_{\text{sym}}(r)$ given in (8). We observe that the tradeoff performance is strictly below the optimal. Moreover, with the optimal scheme, each user can achieve a single-user performance as long as the system is not heavily loaded. In the case $m = 1$, this means the performance of a particular user is not affected by the total number of users K in the network, as long as $K \le n$. In contrast, with successive cancellation, adding one user to the network always degrades the performance of all other users.

Now if we allow the receiver to decode for the users in an order that depends on the realization of the channel, the tradeoff performance can be improved. It is shown in [3] that the optimal ordering is to choose the user to decode in each stage such that the effective channel gain $\|\boldsymbol{g}_i\|$ is maximized. The tradeoff performance of this scheme is studied in [12, Sec. 7.2], and it is shown that

$$P_e(\mathsf{SNR}) \dot{\ge} \mathsf{SNR}^{-(n-1)(1-r)}.$$

It is seen that this scheme is still suboptimal.

For the case that each user has $m > 1$ transmit antennas, we can similarly write the single-user subchannels as

$$\boldsymbol{Y}_i = \sqrt{\frac{\mathsf{SNR}}{m}} \boldsymbol{G}_i \boldsymbol{X}_i + \boldsymbol{W}_i, \qquad \text{for } i = 1, \dots, K$$

where $\boldsymbol{X}_i \in \mathcal{C}^{m \times l}$ is the signal transmitted by user i. $\boldsymbol{Y}_i, \boldsymbol{W}_i \in \mathcal{C}^{n \times l}$ are the received signal and noise for user i, respectively. $\boldsymbol{G}_i \in \mathcal{C}^{n \times m}$ is equivalent channel gain for user i. Again, assuming that the users are decoded sequentially, each column vector of $\boldsymbol{G}_i$ is thus the component of the corresponding column vector of $\boldsymbol{H}_i$ that is perpendicular to the subspace spanned by the column vectors of $\boldsymbol{H}_{i+1}, \dots, \boldsymbol{H}_k$. That is, the component of each column vector of $\boldsymbol{H}_i$ in $(K-i)m$ dimensions is nulled out. Consequently, the subchannel for user i is equivalent to a point-to-point channel with m transmit and $n-(K-i)m$ receive antennas, for $i = 1, \dots, K$. Similar to the case that $m = 1$, in order to use successive cancellation, we need an extra constraint that $n \geq Km$.[1]

Under these assumptions, in decoding the ith user, the signals from user $i+1, \dots, K$ that spans a $(K-i)m$-dimensional subspace, has to be nulled out. Effectively, the subchannel for the ith user is a point-to-point link with m transmit and $n-(K-i)m$ receive antennas, and the detection error probability is

$$P(\mathcal{E}_i) \doteq \mathsf{SNR}^{-d^*_{m,n-(K-i)m}(r)}. \tag{20}$$

The performance of the system is thus limited by that of the subchannel for user 1, hence,

$$P_e(\mathsf{SNR}) \doteq P(\mathcal{E}_1) \doteq \mathsf{SNR}^{-d^*_{m,n-(K-1)m}(r)}. \tag{21}$$

Similar to the case with $m = 1$, choosing the ordering in which the users are decoded according to the channel realization helps to improve the performance. The exact tradeoff performance with the optimal ordering is hard to compute. However, we can show that the optimal tradeoff performance is still not achieved.

To see that, we give a simple upper bound of the diversity gain at $r = 0$, and show it is strictly below the optimal. Consider the case that there are only two users (or assume that a genie reveals the data of user $3, \dots, K$ to the receiver). Let Ω_i be the subspace spanned by the column vectors of $\boldsymbol{H}_i$, for $i = 1, 2$. Ω_i's are independently uniformly distributed in the Grassmann manifold $G(n, m)$, which is the set of all m-dimensional subspaces in $\mathcal{C}^n$. The dimensionality of $G(n, m)$ is $m(n-m)$ [8]. Observe that with a high probability, the successive cancellation receiver will make a detection error, if Ω_1 and Ω_2 lie in a small neighborhood of each other, whose size is of the same order as the noise. The probability for that to happen is $\mathsf{SNR}^{-m(n-m)}$. Consequently, the probability of detection error with a successive cancellation receiver is no less than $\mathsf{SNR}^{-m(n-m)}$. In contrast, as discussed in the previous sections, with the optimal ML receiver, the single-user performance of $d^*_{m,n}(0) = mn$ is achieved at

[1]One can actually use fewer receive antennas. For example, if we have $n = 1 + (K-1)m$ receive antennas, after nulling out the other $K-1$ users, user 1 still needs one dimension to communicate. However, the performance of such systems will be severely degraded, since user 1 is the bottleneck of the system. Therefore, we do not consider such cases.

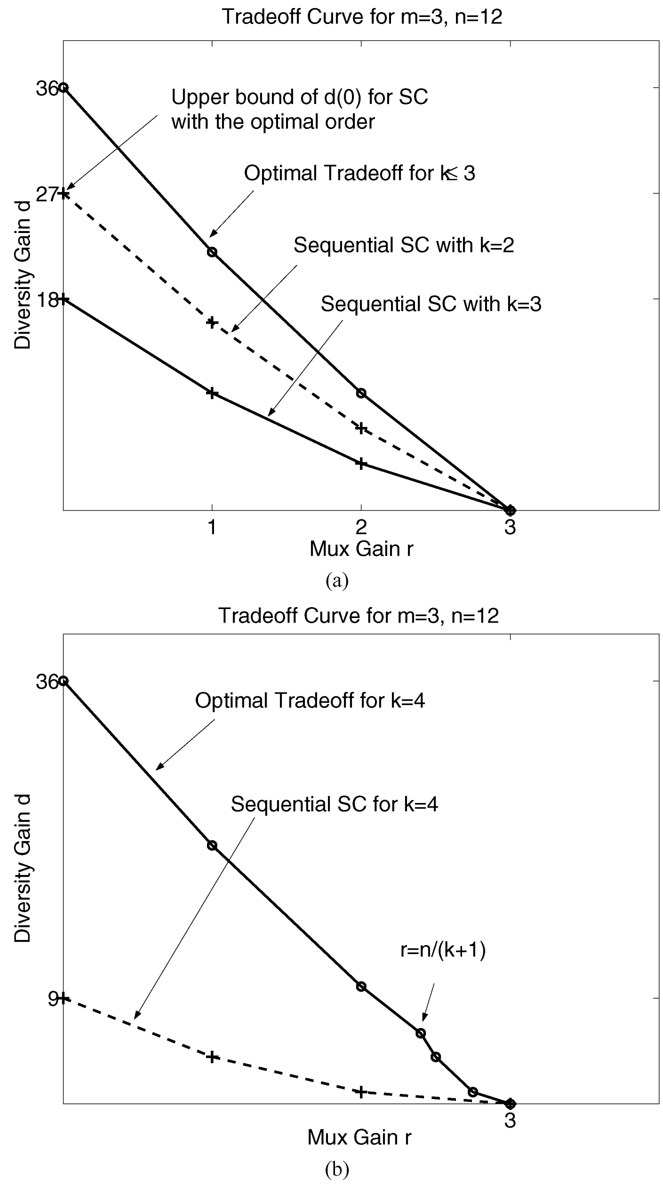

Fig. 9. Tradeoff for successive cancellation schemes with $m > 1$. (a) $m \leq \frac{n}{K+1}$ case; (b) $m > \frac{n}{K+1}$ case.

multiplexing gain $r = 0$. Therefore, the successive cancellation technique is strictly suboptimal. Some examples are plotted in Fig. 9.

To summarize, we have shown in this section that successive cancellation, although simplifies the problem into single-user subchannels and can achieve the maximum sum rate, is strictly suboptimal in terms of the error probability behavior. This is particularly true at low data rates where joint ML detection is significantly better. The successive cancellation technique is biased among the users. For example, the first user that is decoded has the worst channel. In the next two subsections, we study schemes that are symmetric with respect to the users and still achieve the maximal sum rate.

B. Time Sharing

One simple strategy is to time-share and average out the bias. By switching between a set of schemes, we can allow each user to go through the worst channel only for a fraction of time, therefore, potentially improving the average performance.

Suppose we time-share among n_s different schemes. Let the data rate and error probability for user i in the jth scheme be

$$R_j^{(i)} = r_j^{(i)} \log \mathsf{SNR} \qquad P_j^{(i)} \doteq \mathsf{SNR}^{-d_j^{(i)}}$$

respectively, for $i = 1, \dots, K$ and $j = 1, \dots, n_s$. Now by time sharing, we use scheme j with $p_j \le 1$ fraction of the time, where $\sum_{j=1}^{n_s} p_j = 1$. For a fixed choice of $p_j, j = 1, \dots, n_s$, the average data rate and error probability for user i are

$$R^{(i)} = \sum_j p_j r_j^{(i)} \log \mathsf{SNR}$$
$$P^{(i)} \doteq \sum_j p_j \mathsf{SNR}^{-d_j^{(i)}} \doteq \mathsf{SNR}^{-\min_j d_j^{(i)}}. \tag{22}$$

That is, by time sharing, we can achieve the average data rate, but still retain the worst case diversity gain.

Example: Rate Allocation: We consider $n_s = K!$ successive cancellation schemes, one for each of the ordering of the K users. Suppose we want to provide symmetric rate and diversity requirements to the users; without loss of generality, we can compute the performance of user 1. Let p_j be the probability that user 1 is the jth decoded user. By symmetry, we have $p_j = 1/K$ for $j = 1, \dots, K$. Now using (20), the data rate and error probability can be computed as

$$R^{(1)} = \frac{1}{K} \sum_{j=1}^{K} r_j^{(1)} \log \mathsf{SNR}$$
$$P^{(1)} \doteq \mathsf{SNR}^{-d_{\min}}.$$

Here

$$d_{\min} = \min_j d^*_{m,n-(K-j)m}(r_j).$$

If we want to send a data rate of $r \log \mathsf{SNR}$ and set $r_j = r$ for all j, then the performance is still limited by the fraction of time that user 1 goes through the worst (first) subchannel. The probability of error is $\mathsf{SNR}^{-d^*_{m,n-(K-1)m}(r)}$. In order to maximize the data rate at a given diversity requirement $d_{\min} \ge d_{\text{req}}$, the multiplexing gain that should be used when user 1 is the jth decoded user is $r_j^* = r^*_{m,n-(K-j)m}(d_{\text{req}})$. Intuitively, a lower data rate should be transmitted when the user is assigned to a worse channel such that the corresponding diversity is improved.

In Fig. 10, we give an example of the optimal rate allocation and the resulting tradeoff performance for time-sharing schemes. We observe that the tradeoff performance is improved using the optimal rate allocation, but is still strictly below the optimal tradeoff curve with joint ML decoding. This again emphasizes the advantage of using optimal ML decoding in the multiple-access system: when the system is lightly loaded, $r \le \frac{n}{K+1}$, the effect of the interference between different users is completely eliminated by the ML receiver. In comparison, the schemes using a decorrelator to null out interference, as well as the successive cancellation and time-sharing schemes based on that, are strictly suboptimal.

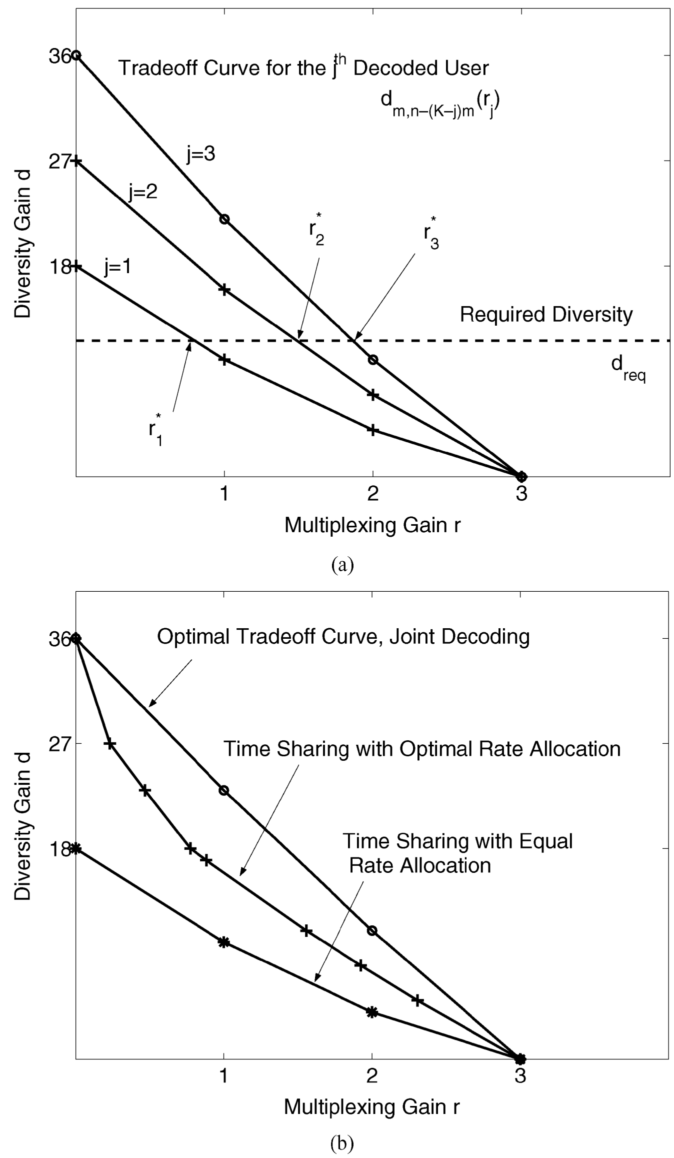

Fig. 10. An example of the rate allocation for the time-sharing scheme: $m = 3, K = 3, n = 12$. (a) The optimal rate allocation r_j^* for a given required diversity gain d_{req} can be read from the tradeoff curves. (b) The resulting performance with the optimal rate allocation. Notice that some transmit antennas need to be shut off ($r_j^* = 0$) to obtain the optimal diversity gain in the low-rate region.

C. Rate Splitting

Another commonly used multiple-access technique is rate splitting [7]. Here, each user is split into virtual users that transmit at different power levels and are decoded in an appropriate order to achieve desired data rates within the capacity region.

In studying rate splitting in multiple-antenna fading channels, we start by treating all the virtual users as independent users, and focus on the power allocation among these users. In our scale of interests, the diversity and multiplexing gains are not changed when scaling the transmitted power of a user by a constant factor that does not depend on SNR. It is only interesting to assign a power of the order $\mathsf{SNR}^{-\alpha}$ to the users. (Notice that in our setup, the transmitted power available for each user is of the order SNR^0.) Unlike the successive cancellation schemes discussed previously (rate splitting with *equal* power allocation for all users), now we can support $K > n$ users.

Example: Single-User Rate Splitting: Consider the simple case with $m = n = 1$ and $K = 2$ users. Let their multiplexing gain be r_1, r_2, respectively. Let user 1 transmit at a power of SNR^0, and user 2 transmit a power $\mathsf{SNR}^{-\alpha}$. The receiver can first decode user 1 treating user 2 as noise, and then cancel its contribution before decoding user 2. Now the effective SNR for user 1 is $\mathsf{SNR}^0/(\mathsf{SNR}^{-\alpha} + \mathsf{SNR}^{-1}) \doteq \mathsf{SNR}^{\alpha}$; and the data rate $R = r\log\mathsf{SNR} = r/\alpha \log(\mathsf{SNR}^{\alpha})$, hence, the effective multiplexing gain is r/α. The error probability it achieves is

$$P(\mathcal{E}_1) \doteq (\mathsf{SNR}^{\alpha})^{d^*_{1,1}(r_1/\alpha)} \doteq \mathsf{SNR}^{-\alpha(1-r_1/\alpha)}.$$

Similarly, the effective SNR for user 2 is $\mathsf{SNR}^{1-\alpha}$, and the probability of error is

$$P(\mathcal{E}_2) \doteq \mathsf{SNR}^{-(1-\alpha)(1-r_2/(1-\alpha))}.$$

Now we can optimize over α to minimize the maximum of two error probabilities, and the resulting overall error probability is

$$P_e \doteq \mathsf{SNR}^{-1/2(1-r_1-r_2)}. \tag{23}$$

Suppose now that these two users are virtual users created by splitting one user with multiplexing gain $r = r_1 + r_2$. From (23), the error performance is $P_e \doteq \mathsf{SNR}^{-(1-r)/2}$. Notice that this is strictly below the single-user performance $\mathsf{SNR}^{-(1-r)}$. Intuitively, since a part of the data rate is transmitted at a lower power level $\mathsf{SNR}^{-\alpha}$, the error probability is increased.

In general, assume that the users transmit at B different power levels

$$\mathsf{SNR}^{-\alpha_1}, \mathsf{SNR}^{-\alpha_2}, \ldots, \mathsf{SNR}^{-\alpha_B}$$

for $0 = \alpha_1 < \cdots, \alpha_B < 1$. Effectively, we have B multiple-access subchannels, with effective SNR

$$\mathsf{SNR}^{\alpha_2-\alpha_1}, \mathsf{SNR}^{\alpha_3-\alpha_2}, \ldots, \mathsf{SNR}^{1-\alpha_B}.$$

For a user i communicating in a subchannels with effective SNR as $\mathsf{SNR}^{-\beta}$, the diversity–multiplexing tradeoff can be computed as in (21), with both the diversity gain and multiplexing gain scaled by β, that is, assuming user i transmitting a rate $r_i \log\mathsf{SNR}$, the error probability is

$$P(\mathcal{E}_i) \doteq \mathsf{SNR}^{-\beta(d^*_{m,n-(K_i-1)m}(r_i/\beta))}$$

where K_i is the number of users sharing the same subchannel.

When a user is split into a number of virtual users, the overall error probability is still dominated by the worst case among the virtual users. The optimal rate splitting and power allocation can be solved as a linear optimization problem. Before this calculation, we can claim that this approach cannot achieve the optimal tradeoff performance. To see this, observe that at a low data rate, Theorem 2 says that single-user tradeoff performance can be achieved. However, as discussed at the end of Section VI-A, with the successive cancellation receiver, whenever there is another user sharing the same subchannel or transmitting at a power that is higher than the noise level, the single-user performance can not achieved. Furthermore, as demonstrated in the example of single-user rate splitting, the rate splitting approach is, in general, not optimal in terms of error exponent.

VII. Proof of Theorem 2

We first prove the lower bound using an outage formulation. Then we prove achievability using a random coding argument.

A. Individual Versus Joint ML Receiver

The receiver that minimizes the error probability for each user i is the *individual ML* receiver. The individual ML receiver for user i treats the other users as *discrete* noise with known structure (codebooks), and makes an ML detection of the message of user i. This is, in general, different from the *joint ML* receiver that jointly detects the messages of all the users ([10, Sec. 4.1.1] has some more discussion on this). But it is easy to relate the error probabilities of the two receivers. Clearly, the joint ML error probability P_e (probability that *any* user is detected incorrectly) is an upper bound to each of the individual ML error probabilities $P_e^{(i)}$. On the other hand, we can consider a joint receiver which uses the individual ML receivers to make a decision on each user's codeword; the performance of this receiver must be an upper bound to P_e. Furthermore, by the union of events bound, the probability of error of this joint receiver is less than the sum of the individual ML probabilities of error. Hence, we conclude that

$$P_e^{(k)} \le P_e \le \sum_{i=1}^{K} P_e^{(i)}, \qquad \text{for all } k.$$

Thus, requiring that each of the $P_e^{(i)}$ to decay like SNR^d is equivalent to requiring the joint ML error probability P_e to decay like SNR^d. Thus, it suffices to work with only the joint ML receiver for the proof below.

B. The Lower Bound: Outage Formulation

In point-to-point channels, the outage is defined as the event that the mutual information of the channel, as a function of the realization of the channel state, does not support the target data rate R, i.e.,

$$\mathcal{O} \triangleq \{H : I(\boldsymbol{X};\boldsymbol{Y} \mid \boldsymbol{H} = H) \le R\}$$

where $I(\boldsymbol{X};\boldsymbol{Y})$ is the mutual information of a point-to-point link with m transmit and n receive antennas.

With the input $\boldsymbol{X}$ having i.i.d. $\mathcal{CN}(0,1)$ entries

$$I(\boldsymbol{X};\boldsymbol{Y} \mid \boldsymbol{H} = H) = \log\det\left(I + \frac{\mathsf{SNR}}{m}HH^{\dagger}\right).$$

It can be shown ([12, Sec. 3.B]) that one can restrict to i.i.d. $\mathcal{CN}$ inputs and the resulting outage probability is characterized in [12, Theorem 4]: at a data rate $R = r\log\mathsf{SNR}$ (bps/Hz)

$$P_{\text{out}}(r\log\mathsf{SNR}) \dot{\le} \mathsf{SNR}^{-d^*_{m,n}(r)} \tag{24}$$

with $d^*_{m,n}(r)$ defined as in Theorem 1 : for integer r, the diversity gain is $(m-r)(n-r)$ and a piecewise-linear function between these integer points. It is shown in [12, Lemma 5] that this outage probability provides a lower bound of the optimal error probability, up to the SNR exponent, i.e., for any coding scheme with a data rate $R = r \log \mathsf{SNR}$ (bps/Hz), the probability of detection error is lower-bounded by

$$P_e(\mathsf{SNR}) \dot{\geq} \mathsf{SNR}^{-d^*_{m,n}(r)}.$$

Intuitively, when an outage occurs, there is a high probability of making a detection error, no matter what coding and decoding techniques are used; therefore, the probability of detection error is lower-bounded by that of outage.

In the multiple-access channel, we can define a corresponding outage event, by which we wish to indicate that the channel is so poor such that the target data rate is not supported, at least for a subset of the users. The definition of outage is given as follows.

Definition 6: Outage Event: For a multiple-access channel with K users, each equipped with m transmit antennas, and a receiver with n receive antennas, the outage event is

$$\mathcal{O} \triangleq \bigcup_S \mathcal{O}_S. \tag{25}$$

The union is taken over all subsets $S \subseteq \{1,\dots,K\}$, and

$$\mathcal{O}_S \triangleq \left\{ H \in \mathbb{C}^{n\times Km} : I(\boldsymbol{X}_S; \boldsymbol{Y} \mid \boldsymbol{X}_{S^c}, \boldsymbol{H} = H) < \sum_{i\in S} R_i \right\}$$

where $\boldsymbol{X}_S$ contains the input signals from the users in S. The significance of this definition is the following: the probability of outage yields a lower bound to the error probability of any scheme. To see that, suppose $\mathcal{O}_S$ occurs for a subset S. Let a genie provide the receiver with the side information of all the correct data symbols $\boldsymbol{X}_{S^c}$ transmitted by users in S^c. But still the sum target rate of the users in S is not supported. Consequently, a detection error (of the users in set S) occurs with a high probability when $\mathcal{O}_S$ occurs.

In the above argument, upon receiving the genie information of the data $\boldsymbol{X}_{S^c}$, the receiver can without loss of optimality, cancel its contribution from the received signals, after which the channel can be written as

$$\begin{aligned}\boldsymbol{Y}_S &= \sqrt{\frac{\mathsf{SNR}}{m}} \sum_{i\in S} \boldsymbol{H}_i \boldsymbol{X}_i + \boldsymbol{W} \\ &= \sqrt{\frac{\mathsf{SNR}}{m}} \boldsymbol{H}_S \boldsymbol{X}_S + \boldsymbol{W}\end{aligned}$$

where $\boldsymbol{H}_S \in \mathcal{C}^{n\times|S|m}$ contains the fading coefficients corresponding to the users in S. By allowing the users in S to cooperate, the problem is reduced to a point-to-point problem with $|S|m$ transmit antennas and n receive antennas, and a fading coefficient matrix $\boldsymbol{H}_S$. Now we can choose the input $\boldsymbol{X}$ to have the i.i.d. Gaussian distribution, such that the $P(\mathcal{O}_S)$ is minimized for all S simultaneously. Let the target data rate of user i be $R_i = r_i \log \mathsf{SNR}$ (bps/Hz) for $i \in \{1,\dots,K\}$, from (24), we have

$$P(\mathcal{O}_S) \doteq \mathsf{SNR}^{-d^*_{|S|m,n}(\sum_{i\in S} r_i)}$$

and

$$P(\mathcal{O}) = P\left(\bigcup_S \mathcal{O}_S\right) \leq \sum_S P(\mathcal{O}_S) \doteq P(\mathcal{O}_{S^*})$$

where S^* be the subset of $\{1,\dots,K\}$ with the slowest decay rate of $P(\mathcal{O}_S)$, i.e.,

$$S^* = \arg\min_S d^*_{|S|m,n}\left(\sum_{i\in S} r_i\right).$$

Combining with the fact that $P(\mathcal{O}) \geq P(\mathcal{O}_{S^*})$, we have

$$P(\mathcal{O}) \doteq P(\mathcal{O}_{S^*}) \doteq \mathsf{SNR}^{-\min_S d^*_{|S|m,n}(\sum_{i\in S} r_i)}$$

as summarized below.

Lemma 7: For a multiple-access system with K users, each equipped with m transmit antennas and a receiver with n receive antennas, let the data rate of user i be $R_i = r_i \log \mathsf{SNR}$ (bps/Hz), for $i = 1,\dots,K$. The detection error probability of any coding scheme is lower-bounded

$$P_e(\mathsf{SNR}) \dot{\geq} P(\mathcal{O}) \doteq \mathsf{SNR}^{-d_{\text{out}}(r_1,\dots,r_K)}$$

where

$$d_{\text{out}}(r_1,\dots,r_K) = \min_S d^*_{|S|m,n}\left(\sum_{i\in S} r_i\right),$$

with $d^*_{m,n}(r)$ as given in Theorem 1.

Consequently, to meet a diversity requirement of d for every user, the transmitted data rates must satisfy

$$d^*_{|S|m,n}\left(\sum_{i\in S} r_i\right) \geq d$$

or equivalently

$$\sum_{i\in S} r_i \leq r^*_{|S|m,n}(d) \tag{26}$$

for all $S \subseteq \{1,\dots,K\}$.

C. The Upper Bound: Random Coding

Lemma 7 gives a lower bound of the optimal error probability. In this subsection, we complete the proof of Theorem 2 by showing that this bound is actually tight, up to the scale of the SNR exponent, provided that the block length $l \geq Km + n - 1$. We show that for any $(r_1,\dots,r_K)$ satisfying (26), there exists a coding scheme that achieves the common diversity d.

To do this, we consider the ensemble of i.i.d. $\mathcal{CN}$ random codes. Specifically, each user generates a codebook $\mathcal{C}^{(i)}$ containing $\mathsf{SNR}^{r_i\times l}$ codewords, denoted as $X_1^{(i)}, X_2^{(i)}, \dots, X^{(i)}_{\mathsf{SNR}^{r_i l}}$. Each codeword is an $m\times l$ matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. Once picked, the codebooks are revealed to the receiver. In each block period, the transmitted signals of user i is simply chosen from the corresponding codebook $\mathcal{C}^{(i)}$ equiprobably according to the message to be transmitted.

Consider the detection error probability of the joint ML receiver. We first define for each nonempty set $S \subseteq \{1,\dots,K\}$ an error event (referred to as a "type S error")

$$\mathcal{E}^S \triangleq \{\hat{m}_i = m_i, \forall i \in S^c \text{ and } \hat{m}_i \neq m_i, \forall i \in S\}$$

where $\hat{m}_i$ is the decoded message for user i. Thus, $\mathcal{E}^S$ is the event that the receiver makes wrong decisions on the messages of all the users in set S, and makes correct decisions for the rest. Clearly, we have

$$P_e(\mathsf{SNR}) = P\left(\bigcup_S \mathcal{E}^S\right) \le \sum_S P(\mathcal{E}^S).$$

In the following, we study $P(\mathcal{E}^S)$, assuming without loss of generality that $S = \{1, 2, \ldots, |S|\}$. Let

$$\boldsymbol{X}_0 = (\boldsymbol{X}_0^{(1)}, \boldsymbol{X}_0^{(2)}, \ldots, \boldsymbol{X}_0^{(K)})$$

be transmitted, where $\boldsymbol{X}_0^{(i)} \in \mathcal{C}^{(i)}$ is the codeword transmitted by user i. Denote $\boldsymbol{X}_1$ to be another codeword which differs from $\boldsymbol{X}_0$ on the symbols transmitted by all the users in S but coincides on those transmitted by the other users, that is,

$$\boldsymbol{X}_1 = (\boldsymbol{X}_1^{(1)}, \boldsymbol{X}_1^{(2)}, \ldots, \boldsymbol{X}_1^{(|S|)}, \boldsymbol{X}_0^{(|S|+1)}, \ldots, \boldsymbol{X}_0^{(K)})$$

where $\boldsymbol{X}_1^{(i)} \ne \boldsymbol{X}_0^{(i)},\ \forall i = 1, \ldots, |S|$.

Now a type-S error occurs if the receiver makes a (wrong) decision in favor of one of such codewords $\boldsymbol{X}_1$. This occurs exactly when

$$\begin{aligned}\|\boldsymbol{W}\|^2 &\ge \left\| \frac{1}{2}\sqrt{\frac{\mathsf{SNR}}{m}} \boldsymbol{H}(\boldsymbol{X}_1 - \boldsymbol{X}_0) \right\|^2 \\ &= \left\| \frac{1}{2}\sqrt{\frac{\mathsf{SNR}}{m}} \boldsymbol{H}_S(\boldsymbol{X}_1^S - \boldsymbol{X}_0^S) \right\|^2. \end{aligned} \tag{27}$$

Here $\boldsymbol{H}_S = \left[\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_{|S|}\right]$, the first $|S|m$ columns of $\boldsymbol{H}$ and $\boldsymbol{X}_i^S = \left[\boldsymbol{X}_i^{(1)}, \boldsymbol{X}_i^{(2)}, \ldots, \boldsymbol{X}_i^{(|S|)}\right]$ for $i = 0, 1$.

Now the computation of $P(\mathcal{E}^S)$ is reduced to finding the probability, averaged over $\boldsymbol{H}$ and $\boldsymbol{W}$, that there exists a codeword $\boldsymbol{X}_1^S \ne \boldsymbol{X}_0^S$ such that (27) is satisfied.

Since the codewords are all i.i.d. $\mathcal{CN}(0,1)$, this computation is the same as that for the error probability of a point-to-point link with $|S|m$ transmit and n receive antennas, with i.i.d. $\mathcal{CN}(0,1)$ random code as the input, and an overall data rate of $\sum_{i=1}^{|S|} r_i \log \mathsf{SNR}$. In [12, Sec. 3.3], it is shown that for the point-to-point channel described above, provided that the block length $l \ge |S|m + n - 1$, the error probability, averaged over the $\mathcal{CN}(0,1)$ random code ensemble, has diversity

$$d^*_{|S|m,n}\left(\sum_{i=1}^{|S|} r_i\right). \tag{28}$$

Now the error probability coincides with the lower bound from the outage formulation

$$P_e(\mathsf{SNR}) \doteq \mathsf{SNR}^{-d^*_{|S|m,n}(r_S)}$$

for $d^*_{m,n}(r)$ defined in Theorem 1.

The proof of this statement is based on the computation of the conditional pairwise error probability $P(\boldsymbol{X}_0 \to \boldsymbol{X}_1 \mid \boldsymbol{H} = H)$ as in [12, eq. (19)], averaged over the ensemble of the codes. In other words, we only used the pairwise independent property of the codebook, i.e., for any pair of distinct codewords $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$, all the entries are generated independently from the Gaussian ensemble.

In computing $P(\mathcal{E}^S)$ for the multiple-access channel, we make the key observation that $\boldsymbol{X}_0^S$ and $\boldsymbol{X}_1^S$ in (27) are pairwise independent. Consequently, the proof in [12] can be used to show

$$P(\mathcal{E}^S) \dot{\le} \mathsf{SNR}^{-d^*_{|S|m,n}(r_s)} \tag{29}$$

where $r_S = \sum_{i \in S} r_i$ is the sum multiplexing gain of the users in S.

The overall error probability is

$$\begin{aligned} P_e(\mathsf{SNR}) &\le \sum_S P(\mathcal{E}^S) \\ &\doteq P(\mathcal{E}^{S^*}) \end{aligned}$$

where S^* maximizes the SNR exponent of $P(\mathcal{E}^S)$, i.e.,

$$S^* = \arg\min_S \qquad d^*_{|S|m,n}(r_S).$$

This completes the proof of our main result.

VIII. Proofs of Theorems 3, 4, and 5

A. Proof of Theorem 3

Recall that

$$d^*_{\text{sym}}(r) = \min_{k=1\ldots K} d^*_{km,n}(kr).$$

To prove the result we have to show that for

$$r \le \min(m, n/(K+1))$$

$d^*_{m,n}(r)$ is the smallest and, otherwise, $d^*_{Km,n}(Kr)$ is the smallest.

Fix $1 \le k_2 < k_1 \le K$ and consider the following key observation:

$$d^*_{k_1 m,n}(k_1 r) - d^*_{k_2 m,n}(k_2 r) \begin{cases} \ge 0, & r \in \left[0, \min\left(m, \frac{n}{k_1+k_2}\right)\right] \\ \le 0, & r \in \left[\min\left(m, \frac{n}{k_1+k_2}\right), m\right]. \end{cases} \tag{30}$$

Suppose this is true. It can be directly seen from the definition of d^* that

$$\begin{aligned} d^*_{k_1 m,n}(0) &> d^*_{k_2 m,n}(0), \qquad k_1 > k_2 \\ d^*_{km,n}(r_1) &> d^*_{km,n}(r_2), \qquad r_1 < r_2 \\ d^*_{km,n}(r) &= 0, \qquad r \ge \min(km, n). \end{aligned}$$

In the case $m \le n/(K+1)$, we complete the proof by observing from (30) that $d^*_{1,n}(r)$ is below every other curve. If this is not the case, then $d^*_{1,n}(r)$ is still below every other curve up to $r \le n/(K+1)$ at which point the curve $d^*_{Km,n}(r)$ intersects it. Since the curve $d^*_{Km,n}(r)$ must have intersected all the other curves by $r \le n/(K+1)$, it is now below all the other curves for $r \in [n/(K+1), n/K]$. This completes the proof of the proposition.

We now show (30). Fix $1 \le k \le K$. Consider the following parabola:

$$g_k(r) \stackrel{\text{def}}{=} k(m-r)(n-kr), \qquad r \in [0, \min(m, n/k)].$$

This parabola is below the corresponding single-user tradeoff curve $d^*_{km,n}(kr)$ for all values of r (since this tradeoff curve is piecewise linear) and equal only when r is such that kr is an integer. It follows that the two tradeoff curves $d^*_{k_i m,n}(r),\ i = 1, 2$ cross over if and only if the corresponding parabolas $g_{k_i}(r),\ i = 1, 2$ intersect. A simple calculation shows that the two parabolas intersect at a point r exactly when r satisfies the quadratic equation

$$r^2(k_1+k_2) - r(n+(k_1+k_2)m) + mn = 0.$$

There are two solutions: m and $n/(k_1+k_2)$. The interesting range of intersection of the parabolas is restricted to $r < \min(m, n/k_1, n/k_2) \le m$; at least one of the tradeoff curves is identically zero for r above this value. Thus, we have now shown (30) for the case $m \le n/(k_1+k_2)$ and will henceforth assume otherwise. In this regime, we conclude that the tradeoff curves cross over exactly once in the range $[0, \min(m, n/k_1, n/k_2))$ and only need to determine the crossover point of the tradeoff curves.

While the intersection of the two parabolas occurs at $n/(k_1+k_2)$, this might not be the same as the crossover point between the tradeoff curves. In general, the parabolas are below the corresponding tradeoff curves, but if $nk_1/(k_1+k_2)$ is an integer (observe that in this case it must be that $nk_2/(k_1+k_2)$ is also an integer) then we have found the crossover point of the tradeoff curves as well to be $n/(k_1+k_2)$. We are hence only left with the case when $m > n/(k_1+k_2)$ and $nk_i/(k_1+k_2),\ i=1,2$ are not integers. We show that even in this case, somewhat surprisingly, the crossover point of the tradeoff curves is still the same as the intersection point between the parabolas.

Since the tradeoff curve is piecewise linear, the crossover point can be found as the intersection of the line segments of $d^*_{k_i m,n}(k_i r)$ passing through the two points

$$(n_i, (k_i m - n_i)(n-n_i)) \quad \text{and} \quad (n_i+1, (k_i m - n_i - 1)(n - n_i - 1))$$

for $i = 1, 2$. Here we have written

$$n_i \stackrel{\text{def}}{=} \left\lfloor \frac{nk_i}{k_1+k_2} \right\rfloor, \qquad i = 1, 2.$$

Hence, the intersection point r satisfies the linear equation $(k_1 a_1 - k_2 a_2) r + b = 0$ where

$$\begin{aligned} a_i &= (k_i m + n - 2n_i - 1), \quad i = 1, 2, \\ b &= (k_2 m - n_2)(n - n_2) - (k_1 m - n_1)(n - n_1) \\ &\quad + n_2 a_2 - n_1 a_1. \end{aligned}$$

Observe that since $nk_i/(k_1+k_2),\ i=1,2$ are not integers we must have

$$n_1 + n_2 = n - 1. \tag{31}$$

Using (31) it can be easily verifiedy that

$$\begin{aligned} k_1 a_1 + k_2 a_2 &= (k_1+k_2)(m(k_1-k_2) - (n_1-n_2)) \\ b &= n((n_1-n_2) - m(k_1-k_2)). \end{aligned}$$

It now follows that the intersection point between the line segments, and hence that between the tradeoff curves, is $r = n/(k_1+k_2)$. This completes the proof.

B. Proof of Theorem 4

From the proof of Theorem 3 (in particular from (30)), it follows that the single-user tradeoff curve $d^*_{m,n}(r)$ is below all the other curves $d^*_{km,n}(kr)$ for $k = 2, \ldots K$ for $r \le n/(K+1)$. Recall that

$$d_K \stackrel{\text{def}}{=} d^*_{m,n}\left(\frac{n}{K+1}\right) \tag{32}$$

and $r^*_{m,n}(d)$ is the multiplexing–tradeoff curve (inverse of $d^*_{m,n}(r)$). Since the tradeoff curves are monotonically decreasing, (30) means that

$$r^*_{m,n}(d) \le \frac{1}{k} r^*_{km,n}(d), \qquad d \ge d_K.$$

From the characterization of $\mathcal{R}(d)$ in Theorem 2, it now follows that, for $d \ge d_K$

$$\mathcal{R}(d) = \left\{(r_1, \ldots, r_K) : 0 \le r_i \le d^*_{m,n}(r_i), \ \ i = 1, \ldots, K\right\}$$

i.e., the optimal tradeoff region is a cube.

Toward generalizing this observation, define (analogous to (32))

$$d_k \stackrel{\text{def}}{=} d^*_{m,n}\left(\frac{n}{k+1}\right), \qquad k = 2, \ldots, K. \tag{33}$$

From (30) it follows that for $d \in [d_{l-1}, d_l]$

$$\begin{aligned} \frac{1}{K} r^*_{Km,n}(d) &\le \frac{1}{K-1} r^*_{(K-1)m,n}(d) \\ &\le \cdots \le \frac{1}{l} r^*_{lm,n}(d) \le r^*_{m,n}(d) \end{aligned} \tag{34}$$

$$r^*_{m,n}(d) \ge \frac{1}{k} r^*_{km,n}(d), \qquad k = 2, \ldots, l-1. \tag{35}$$

It follows that the constraint

$$r_i \le r^*_{m,n}(d), \qquad i = 1, \ldots, K$$

implies the constraints

$$\sum_{s \in S} r_s \le r^*_{|S|m,n}(d)$$

for any subset S with $|S| = 2, \ldots, l-1$. This proves the simplification of $\mathcal{R}(d)$ from (5) to (12).

C. Proof of Theorem 5

Observe that the characterization of (12) for $d \in [d_{l-1}, d_l]$ can be rewritten as

$$\mathcal{R}(d) = \left\{(r_1, \ldots, r_K) : \sum_{s \in S} r_s \le f(|S|), S \subseteq \{1, \ldots, K\}\right\}. \tag{36}$$

Here we have written the rank function

$$f : |S| \mapsto \begin{cases} |S|\, r^*_{m,n}(d), & 0 \le |S| \le l-1 \\ r^*_{|S|m,n}(d), & l \le |S| \le K. \end{cases}$$

Fix an ordering of the users π, a permutation of $\{1, \ldots, K\}$. Using (34), it follows that the multiplexing gain vector $(r_1^\pi, \ldots, r_K^\pi)$ with

$$r^\pi_{\pi(i)} \stackrel{\text{def}}{=} f(i) - f(i-1), \qquad i = 1, \ldots, K$$

is contained in the region $\mathcal{R}(d)$ in (36). Since this is true for every permutation π, and for every l, we have shown that $\mathcal{R}(d)$ is indeed a polymatroid.

REFERENCES

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

[2] G. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

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

[4] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, Mar. 1985.

[5] T. Guess and M. Varanasi, "Error exponents for the Gaussian multiple-access channel," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 214.

[6] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information-theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[7] B. Rimoldi and R. Urbanke, "A rate splitting approach to the Gaussian multiple access channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 364–75, Mar. 1996.

[8] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, 3rd ed: Publish or Perish, Inc., 1999, vol. 1.

[9] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999.

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

[11] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, pp. 1740–1750, Feb./Mar./Apr. 1994.

[12] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, May 2003.

[13] N. Prasad and M. K. Varanasi. Analysis and optimization for multi-access/V-BLAST architecture over MIMO Raleigh fading channels. presented at 41st Annual Allerton Conference on Communications, Control and Computing, Monticello, IL Oct. 2003. [Online] Available: http://ece-www.colorado.edu/~varanasi/pprasaller2003-10-f.htm