

Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna Channels

Lizhong Zheng, *Member, IEEE* David N.C. Tse, *Member, IEEE*

Abstract—Multiple antennas can be used for increasing the amount of diversity or the number of degrees of freedom in wireless communication systems. In this paper, we propose the point of view that both types of gains can be simultaneously obtained for a given multiple antenna channel, but there is a fundamental tradeoff between how much of each any coding scheme can get. For the richly scattered Rayleigh fading channel, we give a simple characterization of the optimal tradeoff curve and use it to evaluate the performance of existing multiple antenna schemes.

Index Terms—diversity, multiple antennas, MIMO, spatial multiplexing, space-time codes .

I. INTRODUCTION

Multiple antennas are an important means to improve the performance of wireless systems. It is widely understood that in a system with multiple transmit and receive antennas (MIMO channel), the spectral efficiency is much higher than that of the conventional single antenna channels. Recent research on multiple antenna channels, including the study of channel capacity [1], [2] and the design of communication schemes [3], [4], [5], demonstrates a great improvement of performance.

Traditionally, multiple antennas have been used to increase *diversity* to combat channel fading. Each pair of transmit and receive antennas provides a signal path from the transmitter to the receiver. By sending signals that carry the same information through different paths, multiple independently faded replicas of the data symbol can be obtained at the receiver end; hence more reliable reception is achieved. For example, in a slow Rayleigh fading environment with 1 transmit and n receive antennas, the transmitted signal is passed through n different paths. It is well known that if the fading is independent across antenna pairs, a maximal diversity gain (advantage) of n can be achieved: the average error probability can be made to decay like $1/\text{SNR}^n$ at high SNR, in contrast to the SNR^{-1} for the single antenna fading channel. More recent work has concentrated on using multiple *transmit* antennas to get diversity (some examples are trellis-based space-time codes [6], [7] and orthogonal designs [8], [3]). However, the underlying idea is still averaging over multiple path gains

(fading coefficients) to increase the reliability. In a system with m transmit and n receive antennas, assuming the path gains between individual antenna pairs are i.i.d. Rayleigh faded, the maximal diversity gain is mn , which is the total number of fading gains that one can average over.

Transmit or receive diversity is a means to *combat* fading. A different line of thought suggests that in a MIMO channel, fading can in fact be *beneficial* through increasing the *degrees of freedom* available for communication [2], [1]. Essentially, if the path gains between individual transmit-receive antenna pairs fade independently, the channel matrix is well-conditioned with high probability, in which case multiple parallel *spatial channels* are created. By transmitting independent information streams in parallel through the spatial channels, the data rate can be increased. This effect is also called *spatial multiplexing* [5], and is particularly important in the high signal-to-noise ratio (SNR) regime where the system is degree-of-freedom-limited (as opposed to power-limited). Foschini [2] has shown that in the high SNR regime, the capacity of a channel with m transmit, n receive antennas and i.i.d. Rayleigh faded gains between each antenna pair is given by:

$$C(\text{SNR}) = \min\{m, n\} \log \text{SNR} + O(1).$$

The number of degrees of freedom is thus the minimum of m and n . In recent years, several schemes have been proposed to exploit the spatial multiplexing phenomenon (for example BLAST [2]).

In summary, a MIMO system can provide two types of gains: diversity gain and spatial multiplexing gain. Most of current research focuses on designing schemes to extract either maximal diversity gain *or* maximal spatial multiplexing gain. (There are also schemes which switch between the two modes, depending on the instantaneous channel condition [5].) However, maximizing one type of gain may not necessarily maximize the other. For example, it was observed in [9] that the coding structure from the orthogonal designs [3], while achieving the full diversity gain, reduces the achievable spatial multiplexing gain. In fact, each of the two design goals addresses only one aspect of the problem. This makes it difficult to compare the performance between diversity-based and multiplexing-based schemes

In this paper, we put forth a different viewpoint: given a MIMO channel, both gains can in fact be *simultaneously* obtained, but there is a *fundamental tradeoff* between how much of each type of gain any coding scheme can extract: higher

both authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

This research is supported by a National Science Foundation Early Faculty CAREER Award, with matching grants from A.T.&T., Lucent Technologies and Qualcomm Inc., and by the National Science Foundation under grant CCR-01-18784

spatial multiplexing gain comes at the price of sacrificing diversity. Our main result is a simple characterization of the optimal tradeoff curve achievable by *any* scheme. To be more specific, we focus on the high SNR regime, and think of a *scheme* as a family of codes, one for each SNR level. A scheme is said to have a spatial multiplexing gain r and a diversity advantage d if the rate of the scheme scales like $r \log \text{SNR}$ and the average error probability decays like $1/\text{SNR}^d$. The optimal tradeoff curve yields for each multiplexing gain r the optimal diversity advantage $d^*(r)$ achievable by *any* scheme. Clearly, r cannot exceed the total number of degrees of freedom $\min\{m, n\}$ provided by the channel; and $d^*(r)$ cannot exceed the maximal diversity gain mn of the channel. The tradeoff curve bridges between these two extremes. By studying the optimal tradeoff, we reveal the relation between the two types of gains, and obtain insights to understand the overall resources provided by multiple antenna channels.

For the i.i.d. Rayleigh flat fading channel, the optimal tradeoff turns out to be very simple for most system parameters of interest. Consider a slow fading environment in which the channel gain is random but remains constant for a duration of l symbols. We show that as long as the block length $l \geq m + n - 1$, the optimal diversity gain $d^*(r)$ achievable by any coding scheme of block length l and multiplexing gain r (r integer) is precisely $(m - r)(n - r)$. This suggests an appealing interpretation: out of the total resource of m transmit and n receive antennas, it is *as though* r transmit and r receive antennas were used for multiplexing and the remaining $m - r$ transmit and $n - r$ receive antennas provided the diversity. Thus, by adding one transmit and one receive antenna, the spatial multiplexing gain can be increased by one while maintaining the *same* diversity level. It should also be observed that this optimal tradeoff does not depend on l as long as $l \geq m + n - 1$; hence, no more diversity gain can be extracted by coding over block lengths greater than $m + n - 1$ than using a block length equal to $m + n - 1$.

The tradeoff curve can be used as a unified framework to compare the performance of many existing diversity-based and multiplexing-based schemes. For several well-known schemes, we compute the achieved tradeoff curves $d(r)$ and compare it to the optimal tradeoff curve. That is, the performance of a scheme is evaluated by the tradeoff it achieves. By doing this, we take into consideration not only the capability of the scheme to combat against fading, but also its ability to accommodate higher data rate as SNR increases, and therefore provide a more complete view.

The diversity-multiplexing tradeoff is essentially the tradeoff between the error probability and the data rate of a system. A common way to study this tradeoff is to compute the *reliability function* from the theory of *error exponents* [10]. However, there is a basic difference between the two formulations: while the traditional reliability function approach focuses on the asymptotics of *large block lengths*, our formulation is based on the asymptotics of *high SNR* (but fixed block length). Thus, instead of using the machinery of the error exponent theory, we exploit the special properties of fading channels and develop a simple approach, based on the outage capacity formulation [11], to analyze the diversity-

multiplexing tradeoff in the high SNR regime. On the other hand, even though the asymptotic regime is different, we do conjecture an intimate connection between our results and the theory of error exponents.

The rest of the paper is outlined as follows. Section II presents the system model and the precise problem formulation. The main result on the optimal diversity-multiplexing tradeoff curve is given in Section III, for block length $l \geq m + n - 1$. In Section IV, we derive bounds on the tradeoff curve when the block length is less than $m + n - 1$. While the analysis in this section is more technical in nature, it provides more insights to the problem. Section V studies the case when spatial diversity is combined with other forms of diversity. Section VI discusses the connection between our results and the theory of error exponents. We compare the performance of several schemes with the optimal tradeoff curve in Section VII. Section VIII contains the conclusions.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Channel Model

We consider the wireless link with m transmit and n receive antennas. The fading coefficient \mathbf{h}_{ij} is the complex path gain from transmit antenna j to receive antenna i . We assume that the coefficients are independently Rayleigh distributed with unit variance, and write $\mathbf{H} = [\mathbf{h}_{ij}] \in \mathcal{C}^{n \times m}$. \mathbf{H} is assumed to be known to the receiver, but not at the transmitter. We also assume that the channel matrix \mathbf{H} remains constant within a block of l symbols, i.e. the block length is much small than the channel coherence time. Under these assumptions, the channel, within one block, can be written as:

$$\mathbf{Y} = \sqrt{\frac{\text{SNR}}{m}} \mathbf{H} \mathbf{X} + \mathbf{W} \quad (1)$$

where $\mathbf{X} \in \mathcal{C}^{m \times l}$ has entries $\mathbf{x}_{ij}, i = 1, \dots, m, j = 1, \dots, l$ being the signals transmitted from antenna i at time j ; $\mathbf{Y} \in \mathcal{C}^{n \times l}$ has entries $\mathbf{y}_{ij}, i = 1, \dots, n, j = 1, \dots, l$ being the signals received from antenna i at time j ; the additive noise \mathbf{W} has i.i.d. entries $\mathbf{w}_{ij} \sim \mathcal{CN}(0, 1)$; SNR is the average signal to noise ratio at each receive antenna.

We will first focus on studying the channel within this single block of l symbol times. In section V, our results are generalized to the case when there is a multiple of such blocks, each of which experiences independent fading.

A rate R bps/Hz codebook \mathcal{C} has $|\mathcal{C}| = \lfloor 2^{Rl} \rfloor$ codewords $\{X(1), \dots, X(|\mathcal{C}|)\}$, each of which is an $m \times l$ matrix. The transmitted signal \mathbf{X} is normalized to have the average transmit power at each antenna in each symbol period to be 1. We interpret this as an overall power constraint on the codebook \mathcal{C} :

$$\frac{1}{|\mathcal{C}|} \sum_{i=1}^{|\mathcal{C}|} \|\mathbf{X}(i)\|_F^2 \leq ml. \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix: $\|\mathbf{R}\|_F^2 \triangleq \sum_{ij} \|R_{ij}\|^2 = \text{trace}(\mathbf{R}\mathbf{R}^\dagger)$.

B. Diversity and Multiplexing

Multiple antenna channels provide *spatial diversity*, which can be used to improve the reliability of the link. The basic idea is to supply to the receiver multiple independently faded replicas of the same information symbol, so that the probability that all the signal components fade simultaneously is reduced.

As an example, consider uncoded binary PSK signals over a single antenna fading channel ($m = n = l = 1$ in the above model). It is well known [12] that the probability of error at high SNR (averaged over the fading gain \mathbf{H} as well as the additive noise) is

$$P_e(\text{SNR}) \approx \frac{1}{4} \text{SNR}^{-1}.$$

In contrast, transmitting the same signal to a receiver equipped with 2 antennas, the error probability is

$$P_e(\text{SNR}) \approx \frac{3}{16} \text{SNR}^{-2}.$$

Here we observe that by having the extra receive antenna, the error probability decreases with SNR at a faster speed of SNR^{-2} . This phenomenon implies that at high SNR, the error probability is much smaller. Similar results can be obtained if we change the binary PSK signals to other constellations. Since the performance gain at high SNR is dictated by the SNR exponent of the error probability, this exponent is called the *diversity gain*. Intuitively, it corresponds to the number of independently faded paths that a symbol passes through; in other words, the number of independent fading coefficients that can be averaged over to detect the symbol. In a general system with m transmit and n receive antennas, there are in total $m \times n$ random fading coefficients to be averaged over; hence the *maximal (full) diversity gain* provided by the channel is mn .

Besides providing diversity to improve reliability, multiple antenna channels can also support a higher data rate than single antenna channels. As an evidence of this, consider an ergodic block fading channel in which each block is as in (1) and the channel matrix is independent and identically distributed across blocks. The ergodic capacity (bps/Hz) of this channel is well-known [1], [2]:

$$C(\text{SNR}) = \mathcal{E} \left[\log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{H}^\dagger \right) \right]$$

At high SNR

$$C(\text{SNR}) = \min\{m, n\} \log \frac{\text{SNR}}{m} + \sum_{i=|\min\{m, n\}|+1}^{\max\{m, n\}} \mathcal{E}[\log \chi_{2i}^2] + o(1),$$

where χ_{2i}^2 is Chi-square distributed with $2i$ degrees of freedom. We observe that at high SNR, the channel capacity increases with SNR as $\min\{m, n\} \log \text{SNR}$ (bps/Hz), in contrast to $\log \text{SNR}$ for single antenna channels. This result suggests that the multiple antenna channel can be viewed as $\min\{m, n\}$ parallel *spatial channels*; hence the number $\min\{m, n\}$ is the total *number of degrees of freedom* to communicate. Now

one can transmit independent information symbols in parallel through the spatial channels. This idea is also called *spatial multiplexing*.

Reliable communication at rates arbitrarily close to the ergodic capacity requires averaging across many independent realizations of the channel gains over time. Since we are considering coding over only a single block, we must lower the data rate and step back from the ergodic capacity to cater for the randomness of the channel \mathbf{H} . Since the channel capacity increases linearly with $\log \text{SNR}$, in order to achieve a certain fraction of the capacity at high SNR, we should consider schemes that support a data rate which also increases with SNR. Here, we think of a *scheme* as a family of codes $\{\mathcal{C}(\text{SNR})\}$ of block length l , one at each SNR level. Let $R(\text{SNR})$ (bits/symbol) be the rate of the code $\mathcal{C}(\text{SNR})$. We say that a scheme achieves a *spatial multiplexing gain* of r if the supported data rate

$$R(\text{SNR}) \approx r \log \text{SNR} \text{ (bps/Hz)}$$

One can think of spatial multiplexing as achieving a *non-vanishing* fraction of the degrees of freedom in the channel. According to this definition, any fixed-rate scheme has a zero multiplexing gain, since eventually at high SNR, any fixed data rate is only a vanishing fraction of the capacity.

Now to formalize, we have the following definition.

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

$$\lim_{\text{SNR} \rightarrow \infty} \frac{R(\text{SNR})}{\log \text{SNR}} = r$$

and the average error probability

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\log P_e(\text{SNR})}{\log \text{SNR}} = -d \quad (3)$$

For each r , define $d^*(r)$ to be the supremum of the diversity advantage achieved over all schemes. We also define

$$\begin{aligned} d_{max}^* &\triangleq d^*(0) \\ r_{max}^* &\triangleq \sup\{r : d^*(r) > 0\} \end{aligned}$$

which are respectively the maximal diversity gain and the maximal spatial multiplexing gain in the channel.

Throughout the rest of the paper, we will use the special symbol \doteq to denote *exponential equality*, i.e., we write $f(\text{SNR}) \doteq \text{SNR}^b$ to denote

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\log f(\text{SNR})}{\log \text{SNR}} = b$$

and $\dot{\geq}, \dot{\leq}$ are similarly defined. (3) can thus be written as

$$P_e(\text{SNR}) \doteq \text{SNR}^{-d}.$$

The error probability $P_e(\text{SNR})$ is averaged over the additive noise \mathbf{W} , the channel matrix \mathbf{H} and the transmitted codewords (assumed equally likely). The definition of diversity gain here differs from the standard definition in the space-time coding literature (see for example [7]) in two important ways:

- This is the *actual* error probability of a code, and not the *pairwise* error probability between two codewords as

is commonly used as a diversity criterion in space-time code design.

- In the standard formulation, diversity gain is an asymptotic performance metric of one *fixed* code. To be specific, the input of the fading channel is fixed to be a particular code, while SNR increases. The speed that the error probability (of a ML detector) decays as SNR increases is called the diversity gain. In our formulation, we notice that the channel capacity increases linearly with \log SNR. Hence in order to achieve a non-trivial fraction of the capacity at high SNR, the input data rate must also *increase* with SNR, which requires a sequence of codebooks with increasing size. The diversity gain here is use as a performance metric of such a sequence of codes, which is formulated as a "scheme". Under this formulation, any fixed code has 0 spatial multiplexing gain. *Allowing both the data rate and the error probability scale with the SNR is the crucial element of our formulation and, as we will see, allows us to talk about their tradeoff in a meaningful way.*

The spatial multiplexing gain can also be thought as the data rate normalized with respect to the SNR level. A common way to characterize the performance of a communication scheme is to compute the error probability as a function of SNR for a fixed data rate. However, different designs may support different data rate. In order to compare these schemes fairly, Forney [13] proposed to plot the error probability against the *normalized* SNR:

$$\text{SNR}_{\text{norm}} \triangleq \frac{\text{SNR}}{C^{-1}(R)},$$

where $C(\text{SNR})$ is the capacity of the channel as a function of SNR. That is, SNR_{norm} measures how far the SNR is above the minimal required to support the target data rate.

A dual way to characterize the performance is to plot the error probability as a function of the data rate, for a fixed SNR level. Analogous to Forney's formulation, to take into consideration the effect of the SNR, one should use the *normalized data rate* R_{norm} instead of R :

$$R_{\text{norm}} \triangleq \frac{R}{C(\text{SNR})}$$

which indicates how far a system is operating from the Shannon limit. Notice that at high SNR, the capacity of the multiple antenna channel is $C(\text{SNR}) \approx \min\{m, n\} \log \text{SNR}$; hence the spatial multiplexing gain

$$r = \frac{R}{\log \text{SNR}} \approx \min\{m, n\} R_{\text{norm}}$$

is just a constant multiple of R_{norm} .

III. OPTIMAL TRADEOFF: $l \geq m + n - 1$ CASE

In this section, we will derive the optimal tradeoff between the diversity gain and the spatial multiplexing gain that any scheme can achieve in the Rayleigh fading multiple antenna channel. We will first focus on the case that the block length $l \geq m + n - 1$, and discuss the other cases in section IV.

A. Optimal Tradeoff Curve

The main result is given in the following theorem.

Theorem 2: Assume $l \geq m + n - 1$. The optimal tradeoff curve $d^*(r)$ is given by the piecewise linear function connecting the points $(k, d^*(k))$, $k = 0, 1, \dots, \min\{m, n\}$, where

$$d^*(k) = (m - k)(n - k) \quad (4)$$

In particular, $d_{\text{max}}^* = mn$, and $r_{\text{max}}^* = \min\{m, n\}$.

The function $d^*(r)$ is plotted in Figure 1.

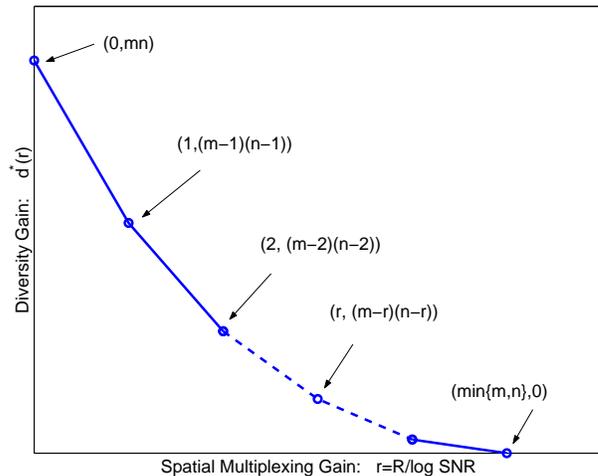


Fig. 1. Diversity-multiplexing tradeoff, $d^*(r)$ for general m, n and $l \geq m + n - 1$.

The optimal tradeoff curve intersects the r axis at $\min\{m, n\}$. This means that the maximum achievable spatial multiplexing gain r_{max}^* is the total number of degrees of freedom provided by the channel as suggested by the ergodic capacity result in (3). Theorem 2 says that at this point, however, no positive diversity gain can be achieved. Intuitively, as $r \rightarrow r_{\text{max}}^*$, the data rate approaches the ergodic capacity and there is no protection against the randomness in the fading channel.

On the other hand, the curve intersects the d axis at the maximal diversity gain $d_{\text{max}}^* = mn$, corresponding to the total number of random fading coefficients that a scheme can average over. There are known designs that achieve the maximal diversity gain at a fixed data rate [8]. Theorem 2 says that in order to achieve the maximal diversity gain, no positive spatial multiplexing gain can be obtained at the same time.

The optimal tradeoff curve $d^*(r)$ bridges the gap between the above two design criteria, by connecting the two extreme points: $(0, d_{\text{max}}^*)$ and $(r_{\text{max}}^*, 0)$. This result says that positive diversity gain and spatial multiplexing gain can be achieved simultaneously. However, increasing the diversity advantage comes at a price of decreasing the spatial multiplexing gain, and vice versa. The tradeoff curve provides a more complete picture of the achievable performance over multiple antenna channels than the two extreme points corresponding to the maximum diversity gain and multiplexing gain. For example, the ergodic capacity result suggests that by increasing the minimum of the number of transmit and receive antennas, $\min\{m, n\}$, by one, the channel gains one more degree of

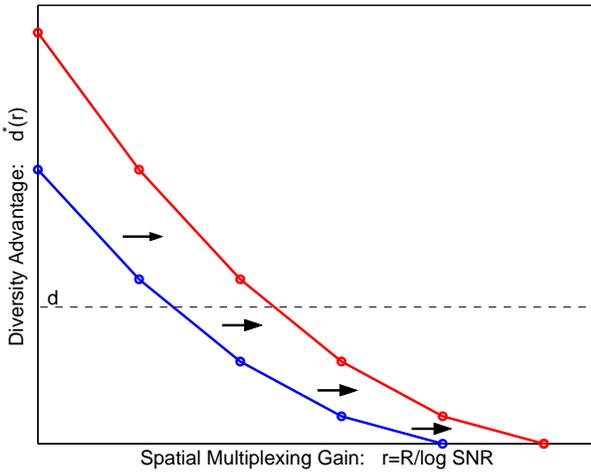


Fig. 2. Adding one transmit and one receive antenna increases spatial multiplexing gain by 1 at each diversity level.

freedom; this corresponds to r_{max}^* being increased by 1. Theorem 2 makes a more informative statement: if we increase both m and n by 1, the entire tradeoff curve is shifted to the right by 1, as shown in Figure 2; i.e., for any given diversity gain requirement d , the supported spatial multiplexing gain is increased by 1.

To understand the operational meaning of the tradeoff curve, we will first use the following example to study the tradeoff performance achieved by some simple schemes.

Example: 2×2 system

Consider the multiple antenna channel with 2 transmit and 2 receive antennas. Assume $l \geq m + n - 1 = 3$. The optimal tradeoff for this channel is plotted in Figure 3-(a). The maximum diversity gain for this channel is $d_{max}^* = 4$, and the total number of degrees of freedom in the channel is $r_{max}^* = 2$.

In order to get the maximal diversity gain, d_{max}^* , each information bit needs to pass through all the 4 paths from the transmitter to the receiver. The simplest way of achieving this is to repeat the same symbol on the two transmit antennas in two consecutive symbol times:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & 0 \\ 0 & \mathbf{x}_1 \end{bmatrix}. \quad (5)$$

d_{max}^* can only be achieved with a multiplexing gain $r = 0$. If we increase the size of the constellation for the symbol \mathbf{x}_1 as SNR increases to support a data rate $R = r \log \text{SNR} (bps/Hz)$ for some $r > 0$, the distance between constellation points shrinks with the SNR and the achievable diversity gain is decreased. The tradeoff achieved by this repetition scheme is plotted in Figure 3-(b)¹. Notice the maximal spatial multiplexing gain achieved by this scheme is 1/2, corresponding to the point (1/2, 0), since only one symbol is transmitted in two symbol times.

The reader should distinguish between the notion of the maximal diversity gain achieved by a scheme, $d(0)$, and the maximal diversity provided by the channel d_{max}^* . For the

above example, $d(0) = d_{max}^*$ but for some other schemes $d(0) < d_{max}^*$ strictly. Similarly, the maximal spatial multiplexing gain achieved by a scheme is in general different from the degrees of freedom r_{max}^* in the channel.

Consider now the Alamouti scheme as an alternative to the repetition scheme in (5). Here, two data symbols are transmitted in every block of length 2 in the form:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & -\mathbf{x}_2^\dagger \\ \mathbf{x}_2 & \mathbf{x}_1^\dagger \end{bmatrix} \quad (6)$$

It is well known that the Alamouti scheme can also achieve the full diversity gain, d_{max}^* , just like the repetition scheme. However, in terms of the tradeoff achieved by the two schemes, as plotted in Figure 3-(b), the Alamouti scheme is strictly better than the repetition scheme, since it yields a strictly higher diversity gain for any positive spatial multiplexing gain. The maximal multiplexing gain achieved by the Alamouti scheme is 1, since one symbol is transmitted per symbol time. This is twice as much as that for the repetition scheme. However, the tradeoff curve for Alamouti scheme is still below the optimal for any $r > 0$.

In the literature on space-time codes, the diversity gain of a scheme is usually discussed for a fixed data rate, corresponding to a multiplexing gain $r = 0$. This is, in fact, the *maximal diversity gain* $d(0)$ achieved by the given scheme. We observe that if the performance of a scheme is only evaluated by the maximal diversity gain $d(0)$, one cannot distinguish the performance of the repetition scheme in (5) and the Alamouti scheme. More generally, the problem of finding a code with the highest (fixed) rate that achieves a given diversity gain is not a well-posed one: any code satisfying a mild non-degenerate condition (essentially a full-rank condition like the one in [7]) will have full diversity gain, no matter how dense the symbol constellation is. This is because diversity gain is an asymptotic concept, while for any fixed code the minimum distance is fixed and does not depend on the SNR. (Of course, the higher the rate, the higher the SNR needs to be for the asymptotics to be meaningful.) In the space-time coding literature, a common way to get around this problem is to put further constraints on the class of codes. In [7], for example, each codeword symbol \mathbf{x}_{ij} is constrained to come from the same fixed constellation. (c.f. Theorem 3.31 there) These constraints are however not fundamental. In contrast, by defining the multiplexing gain as the data rate *normalized* by the capacity, the question of finding schemes that achieves the maximal multiplexing gain for a given diversity gain becomes meaningful.

B. Outage Formulation

As a step to prove Theorem 2, we will first discuss another commonly used concept for multiple antenna channels: the outage capacity formulation, proposed in [11] for fading channels and applied to multi-antenna channels in [1].

Channel outage is usually discussed for non-ergodic fading channels, i.e., the channel matrix \mathbf{H} is chosen randomly but is held fixed for all time. This non-ergodic channel can be written as:

$$\mathbf{y}_t = \sqrt{\frac{\text{SNR}}{m}} \mathbf{H} \mathbf{x}_t + \mathbf{w}_t, \text{ for } t = 1, 2, \dots, \infty \quad (7)$$

¹How these curves are computed will become evident in Section VII.

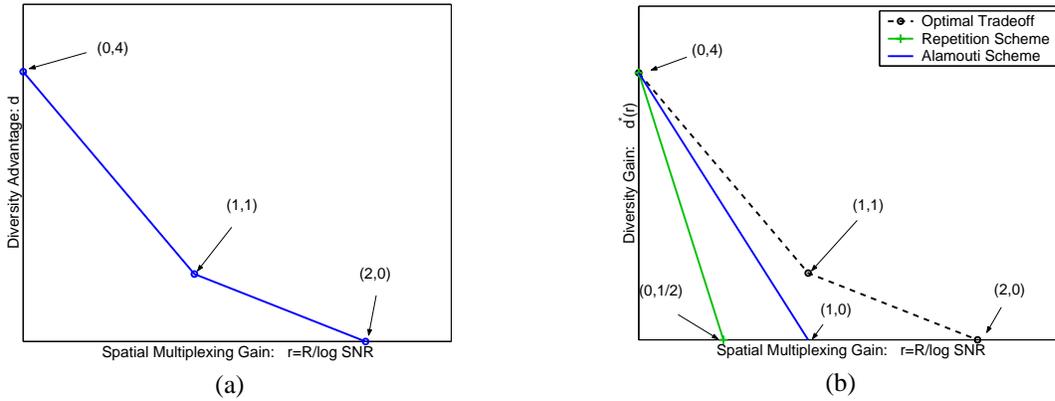


Fig. 3. Diversity-multiplexing tradeoff for (a): $m = n = 2, l \geq 3$; (b): Comparison between two schemes.

where $\mathbf{x}_t \in \mathcal{C}^m, \mathbf{y}_t \in \mathcal{C}^n$ are the transmitted and received signals at time t , and $\mathbf{w}_t \in \mathcal{C}^n$ is the additive Gaussian noise. An outage is defined as the event that the mutual information of this channel does not support a target data rate :

$$\{H : I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{H} = H) < R\}$$

The mutual information is a function of the input distribution $P(x_t)$, and the channel realization. Without loss of optimality, the input distribution can be taken to be Gaussian with a covariance matrix Q , in which case:

$$I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{H} = H) = \log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{Q} \mathbf{H}^\dagger \right)$$

Optimizing over all input distributions, the outage probability is

$$P_{out}(R) = \inf_{Q \geq 0, \text{trace}(Q) \leq m} P \left[\log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{Q} \mathbf{H}^\dagger \right) < R \right]$$

where the probability is taken over the random channel matrix \mathbf{H} . We can simply pick $Q = I$ to get an upper bound on the outage probability.

On the other hand, Q satisfies the power constraint, $\text{trace}(Q) \leq m$, and hence $mI - Q$ is a positive-semidefinite matrix. Notice that $\log \det(\cdot)$ is an increasing function on the cone of positive-definite Hermitian matrices, i.e., if A and B are both positive-semidefinite Hermitian matrices, written as $A \geq 0$ and $B \geq 0$, then

$$A - B \geq 0 \implies \log \det A \geq \log \det B.$$

Therefore, if we replace Q by mI_m , the mutual information is increased:

$$\log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{Q} \mathbf{H}^\dagger \right) \leq \log \det \left(I + \text{SNR} \mathbf{H} \mathbf{H}^\dagger \right);$$

hence the outage probability satisfies

$$\begin{aligned} & P \left[\log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{H}^\dagger \right) < R \right] \\ & \geq P_{out}(R) \\ & \geq P \left[\log \det \left(I + \text{SNR} \mathbf{H} \mathbf{H}^\dagger \right) < R \right] \end{aligned} \quad (8)$$

At high SNR,

$$\begin{aligned} & \lim_{\text{SNR} \rightarrow \infty} \frac{\log P[\log \det(I + \text{SNR} \mathbf{H} \mathbf{H}^\dagger) < R]}{\log \text{SNR}} \\ & = \lim_{\text{SNR} \rightarrow \infty} \frac{\log P[\log \det(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{H}^\dagger) < R]}{\log \frac{\text{SNR}}{m}} \\ & = \lim_{\text{SNR} \rightarrow \infty} \frac{\log P[\log \det(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{H}^\dagger) < R]}{\log \text{SNR}} \end{aligned}$$

Therefore on the scale of interest, the bounds are tight, and we have

$$P_{out}(R) \doteq P[\log \det(I + \text{SNR} \mathbf{H} \mathbf{H}^\dagger) < R]. \quad (9)$$

and we can without loss of generality assume the input (Gaussian) distribution to have covariance matrix $Q = I$.

In the outage capacity formulation, we can ask an analogous question as in our diversity-tradeoff formulation: given a target rate R which scales with SNR as $r \log \text{SNR}$, how does the outage probability decrease with the SNR? To perform this analysis, we can assume without loss of generality that $m \geq n$. This is because

$$\log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{H}^\dagger \right) = \log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H}^\dagger \mathbf{H} \right),$$

hence swapping m and n has no effect on the mutual information, except a scaling factor of m/n on the SNR, which can be ignored on the scale of interest.

We start with the following example.

Example: Single Antenna Channel

Consider the single antenna fading channel

$$\mathbf{y} = \sqrt{\text{SNR}} \mathbf{h} \mathbf{x} + \mathbf{w}$$

where $\mathbf{h} \in \mathcal{C}$ is Rayleigh distributed, and $\mathbf{y}, \mathbf{x}, \mathbf{w} \in \mathcal{C}$. To achieve a spatial multiplexing gain of r , we set the input data rate be $R = r \log \text{SNR}$ for $0 \leq r \leq 1$. The outage probability for this target rate is

$$\begin{aligned} P_{out}(r \log \text{SNR}) & = P(\log(1 + \text{SNR} \|\mathbf{h}\|^2) \leq r \log \text{SNR}) \\ & = P(1 + \text{SNR} \|\mathbf{h}\|^2 \leq \text{SNR}^r) \\ & \approx P(\|\mathbf{h}\|^2 \leq \text{SNR}^{-(1-r)}) \end{aligned}$$

Notice $\|\mathbf{h}\|^2$ is exponentially distributed, with density $p_{\|\mathbf{h}\|^2}(t) = e^{-t}$; hence

$$\begin{aligned} P_{out}(r \log \text{SNR}) &\approx P(\|\mathbf{h}\|^2 \leq \text{SNR}^{-(1-r)}) \\ &= 1 - \exp(-\text{SNR}^{-(1-r)}) \\ &\doteq \text{SNR}^{-(1-r)} \end{aligned}$$

This simple example shows the relation between the data rate and the SNR exponent of the outage probability. The result depends on the Rayleigh distribution of \mathbf{h} only through the near zero behavior: $P(\|\mathbf{h}\|^2 \leq \epsilon) \sim \epsilon$; hence is applicable to any fading distribution with a non-zero finite density near 0. We can also generalize to the case that the fading distribution has $P(\|\mathbf{h}\|^2 \leq \epsilon) \sim \epsilon^k$, in which case the resulting SNR exponent is $k(1-r)$ instead of $1-r$.

In a general $m \times n$ system, an outage occurs when the channel matrix \mathbf{H} is ‘‘near singular’’. The key step in computing the outage probability is to explicitly quantify how singular \mathbf{H} needs to be for outage to occur, in terms of the target data rate and the SNR. In the above example with a data rate $R = r \log \text{SNR}$, outage occurs when $\|\mathbf{h}\|^2 \leq \text{SNR}^{-(1-r)}$, with a probability $\text{SNR}^{-(1-r)}$. To generalize this idea to multiple antenna systems, we need to study the probability that the singular values of \mathbf{H} are close to zero. We quote the joint probability density function (pdf.) of these singular values [14].

Lemma 3: Let \mathbf{R} be an $m \times n$ random matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries. Suppose $m \geq n$, $\mu_1 \leq \mu_2 \leq \dots \leq \mu_n$ be the ordered non-zero eigenvalues of $\mathbf{R}^\dagger \mathbf{R}$, then the joint pdf. of μ_i 's is

$$p(\mu_1, \dots, \mu_n) = K_{m,n}^{-1} \prod_{i=1}^n \mu_i^{m-n} \prod_{i < j} (\mu_i - \mu_j)^2 e^{-\sum_i \mu_i} \quad (10)$$

where $K_{m,n}$ is a normalizing constant. Define $\alpha_i := -\log \mu_i / \log \text{SNR}$ for all i . The joint pdf. of the random vector $\alpha = [\alpha_1, \dots, \alpha_n]$ is

$$p(\alpha) = K_{m,n}^{-1} (\log \text{SNR})^n \prod_{i=1}^n \text{SNR}^{-(m-n+1)\alpha_i} \prod_{i < j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 \exp \left[-\sum_{i=1}^n \text{SNR}^{-\alpha_i} \right]$$

This can be obtained from (10) by the change of variables $\mu_i = \text{SNR}^{-\alpha_i}$.

Now consider (9) with $R = r \log \text{SNR}$, let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the non-zero eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$, we have

$$\begin{aligned} P_{out}(R) &\doteq P[\log \det(I + \text{SNR}\mathbf{H}\mathbf{H}^\dagger) < R] \\ &= P \left[\prod_{i=1}^n (1 + \text{SNR}\lambda_i) < \text{SNR}^r \right] \end{aligned}$$

Let $\lambda_i = \text{SNR}^{-\alpha_i}$. At high SNR, we have $(1 + \text{SNR}\lambda_i) \doteq \text{SNR}^{(1-\alpha_i)^+}$, where $(x)^+$ denotes $\max\{0, x\}$. The above can thus be written as

$$\begin{aligned} P_{out}(R) &\doteq P \left[\prod_i \text{SNR}^{(1-\alpha_i)^+} < \text{SNR}^r \right] \\ &= P \left[\sum_i (1 - \alpha_i)^+ < r \right] \end{aligned}$$

Here, the random vector α indicates the level of singularity of the channel matrix \mathbf{H} . The larger α_i 's are, the more singular \mathbf{H} is. The set $\mathcal{A} = \{\alpha : \sum_i (1 - \alpha_i)^+ < r\}$ describes the outage event in terms of the singularity level. With the distribution of α given in Lemma 3, we can simply compute the probability that $\alpha \in \mathcal{A}$ to get the outage probability:

$$\begin{aligned} &P_{out}(r \log \text{SNR}) \\ &\doteq \int_{\mathcal{A}} p(\alpha) d\alpha \\ &= \int_{\mathcal{A}} K_{m,n}^{-1} (\log \text{SNR})^n \prod_{i=1}^n \text{SNR}^{-(m-n+1)\alpha_i} \\ &\quad \prod_{i < j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 \exp \left[-\sum_i \text{SNR}^{-\alpha_i} \right] d\alpha \end{aligned}$$

Since we are only interested in the SNR exponent of P_{out} , i.e.,

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\log P_{out}(r \log \text{SNR})}{\log \text{SNR}},$$

we can make some approximations to simplify the integral. First, the term $K_{m,n}^{-1} (\log \text{SNR})^n$ has no effect on the SNR exponent, since

$$\frac{\log(K_{m,n}^{-1} (\log \text{SNR})^n)}{\log \text{SNR}} \rightarrow 0$$

Secondly, for any $\alpha_i < 0$, the term $\exp(-\text{SNR}^{-\alpha_i})$ decays with SNR exponentially. At high SNR, we can therefore ignore the integral over the range with any $\alpha_i < 0$; and replace the above integral range \mathcal{A} with $\mathcal{A}' = \mathcal{A} \cap \mathcal{R}^{n+}$ (\mathcal{R}^{n+} is the set of real n -vectors with non-negative elements). Moreover, within \mathcal{A}' , $\exp(-\text{SNR}^{-\alpha_i})$ approaches to 1 for $\alpha_i > 0$ and e for $\alpha_i = 0$ and thus have no effect on the SNR exponent, and

$$P_{out}(r \log \text{SNR}) \doteq \int_{\mathcal{A}'} \prod_{i=1}^n \text{SNR}^{-(m-n+1)\alpha_i} \prod_{i < j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 d\alpha \quad (11)$$

By definition, $\alpha_i \geq \alpha_j$ for any $i < j$. We only need to consider the case that α_i 's are distinct, since otherwise the integrand is zero. In this case, the term $|\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j}|$ is dominated by $\text{SNR}^{-\alpha_j}$ for any $i < j$, therefore

$$P_{out}(r \log \text{SNR}) \doteq \int_{\mathcal{A}'} \prod_{i=1}^n \text{SNR}^{-(2i-1+m-n)\alpha_i} d\alpha \quad (12)$$

Finally, as $\text{SNR} \rightarrow \infty$, the integral is dominated by the term with the largest SNR exponent. This heuristic calculation is made rigorous in Appendix and the result stated precisely in the following theorem.

Theorem 4: Outage Probability

For the multiple antenna channel (1), let the data rate be $R = r \log \text{SNR}$, with $r \leq \min\{m, n\}$. The outage probability satisfies

$$P_{out}(r \log \text{SNR}) \doteq \text{SNR}^{-d_{out}(r)} \quad (13)$$

where

$$d_{out}(r) = \inf_{\alpha \in \mathcal{A}'} \sum_{i=1}^{\min\{m,n\}} (2i-1 + |m-n|) \alpha_i \quad (14)$$

and

$$\mathcal{A}' = \{ \alpha \in \mathcal{R}^{\min\{m,n\}+} \mid \alpha_1 \geq \dots \geq \alpha_{\min\{m,n\}} \geq 0, \\ \text{and } \sum_i (1 - \alpha_i)^+ < r \}$$

$d_{out}(r)$ can be explicitly computed. The resulting $d_{out}(r)$ coincides with $d^*(r)$ given in (4) for all r .

Proof: See Appendix \blacksquare

The analysis of the outage probability provides useful insights to the problem at hand. Again assuming $m \geq n$, (12) can in fact be generalized to any set $\mathcal{B} \subset \mathcal{R}^{n+}$,

$$P(\alpha \in \mathcal{B}) \doteq \int_{\mathcal{B}} \prod_{i=1}^n \text{SNR}^{-(m-n+2i-1)\alpha_i} d\alpha \\ \doteq \text{SNR}^{-\min_{\alpha \in \mathcal{B}} \sum_i (m-n+2i-1)\alpha_i}$$

In particular, we consider for any $b = [b_1, \dots, b_n] \in \mathcal{R}^{n+}$ the set $\mathcal{B}_b = \{ \alpha : \alpha_i \geq b_i \}$. Now

$$P(\alpha \in \mathcal{B}_b) = P(\lambda_i \leq \text{SNR}^{-b_i}, \forall i) \\ \doteq \text{SNR}^{-\sum_i (m-n+2i-1)b_i}$$

Notice that SNR is a dummy variable, this result can also be written as

$$\lim_{\epsilon \rightarrow 0} \frac{\log P(\lambda_i \leq \epsilon^{b_i}, \forall i)}{\log \epsilon} \\ = \sum_{i=1}^n (m-n+2i-1)b_i, \quad \text{for } m \geq n \\ = \sum_{i=1}^{\min\{m,n\}} (|m-n| + 2i-1)b_i, \quad \text{for general } m, n \quad (15)$$

which characterizes the near singular distribution of the channel matrix \mathbf{H} .

This result has a geometric interpretation as follows. For $k = 0, 1, \dots, n$, define

$$\mathcal{R}_k \triangleq \{ X \in \mathcal{C}^{n \times m} : \text{rank}(X) = k \} \\ \mathcal{U}_k \triangleq \{ X \in \mathcal{C}^{n \times m} : \text{rank}(X) \leq k \} \\ = \bigcup_{j=0}^k \mathcal{R}_j$$

It can be shown that \mathcal{R}_k is a differentiable manifold; hence the dimensionality of \mathcal{R}_k is well-defined. Intuitively, we observe that in order to specify a rank k matrix in $\mathcal{C}^{n \times m}$, one needs to specify k linearly independent row vectors of dimension m , and the rest $n-k$ rows as a linear combination of them. These add up to

$$d_k \triangleq mk + (n-k)k = mn - (m-k)(n-k),$$

which is the dimensionality of \mathcal{R}_k .

We also observe that the closure of \mathcal{R}_k is

$$\bar{\mathcal{R}}_k = \mathcal{U}_k = \mathcal{R}_k \cup \mathcal{U}_{k-1}$$

which means that \mathcal{U}_{k-1} , the set of matrices with rank less than k , is the boundary of \mathcal{R}_k , and is the union of some lower dimensional manifolds. Now consider any point X_k in \mathcal{R}_k , we say X_k is near singular if it is close to the boundary \mathcal{U}_{k-1} . Intuitively, we can find X_k 's projection X_{k-1} in \mathcal{U}_{k-1} , and the difference $X_k - X_{k-1}$ has at least $d_k - d_{k-1}$ dimensions. Now X_k being near singular requires that its components in these $d_k - d_{k-1}$ dimensions to be small.

Consider the i.i.d. Gaussian distributed channel matrix $\mathbf{H} \in \mathcal{C}^{n \times m} = \mathcal{U}_n$. The event that the smallest singular value of \mathbf{H} is close to 0, $\lambda_1 \leq \epsilon^{b_1}$, occurs when \mathbf{H} is close to its projection, \mathbf{H}' in \mathcal{U}_{n-1} . This means that the component of \mathbf{H} in $d_n - d_{n-1}$ dimensions is of order $\epsilon^{b_1/2}$, with a probability $\epsilon^{(d_n - d_{n-1})b_1}$. Conditioned on this event, the second smallest singular value of \mathbf{H} being small, $\lambda_2 \leq \epsilon^{b_2}$, means that $\mathbf{H}' \in \mathcal{U}_{n-1}$ is close to its boundary, with a probability $\epsilon^{(d_{n-1} - d_{n-2})b_2}$. By induction, (15) is obtained.

Now the outage event at multiplexing gain r is $\{ \sum_i (1 - \alpha_i)^+ < r \}$. There are many choices of α that satisfy this singularity condition. According to (15), for each of these α 's, the probability $P(\lambda_i \leq \text{SNR}^{-\alpha_i}, \forall i)$ has an SNR exponent, $\sum (2i-1 + m-n)\alpha_i$. Among all the choices of α that lead to outage, one particular choice α^* , which minimizes the SNR exponent $\sum (2i-1 + m-n)\alpha_i$, has the dominating probability; this corresponds to the *typical* outage event. This is a manifestation of *Laplace's principle* [15].

The minimizing α^* can be explicitly computed. In the case that r takes an integer value k , we have $\alpha_i^* = 1$, for $i = 1, \dots, n-k$; and $\alpha_i^* = 0$ for $i = n-k+1, \dots, n$. Intuitively, since the smaller singular values have a much higher probability to be close to zero than the larger ones, the typical outage event has $n-k$ smallest singular values $\lambda_i \doteq \text{SNR}^{-1}$, k largest singular values are of order 1. This means that the typical outage event occurs when the channel matrix \mathbf{H} lies in a neighborhood of the sub-manifold \mathcal{R}_k , with the component in $mn - \dim(\mathcal{R}_k) = (m-k)(n-k)$ dimensions being of order SNR^{-1} , which has a probability $\text{SNR}^{-(m-k)(n-k)}$. For the case that r is not an integer, say, $r \in (k, k+1)$, we have $\alpha_i^* = 1$ for $i = 1, \dots, n-k-1$, $\alpha_i^* = 0$ for $i = n-k+1, \dots, n$, and $\alpha_{n-k}^* = k+1-r$. That is, by changing the multiplexing gain r between integers, only one singular value of \mathbf{H} corresponding to the typical outage event, is adjusted to be barely large enough to support the data rate; therefore, the SNR exponent of the outage probability, $d_{out}(r)$, is linear between integer points.

C. Proof of Theorem 2

Let us now return to our original diversity-multiplexing tradeoff formulation and prove Theorem 2. First, we show that the outage probability provides a lower bound on the error probability for channel (1).

Lemma 5: Outage Bound

For the channel in (1), let the data rate scale as $R = r \log \text{SNR}(\text{bps}/Hz)$. For any coding scheme, the probability of a detection error is lower bounded by

$$P_e(\text{SNR}) \geq \text{SNR}^{-d_{out}(r)} \quad (16)$$

where $d_{out}(r)$ is defined in (14).

Proof: Fix a codebook \mathcal{C} of size 2^{Rl} , and let $\mathbf{X} \in \mathcal{C}^{m \times l}$ be the input of the channel, which is uniformly drawn from the codebook \mathcal{C} . Since the channel fading coefficients in \mathbf{H} are not known at the transmitter, we can assume that \mathbf{X} is independent of \mathbf{H} .

Conditioned on a specific channel realization $\mathbf{H} = H$, write the mutual information of the channel as $I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = H)$, and the probability of detection error as $P(\text{error} | \mathbf{H} = H)$. By Fano's inequality, we have

$$Rl \leq 1 + P(\text{error} | \mathbf{H} = H)Rl + I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = H)$$

hence

$$P(\text{error} | \mathbf{H} = H) \geq 1 - \frac{I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = H)}{Rl} - \frac{1}{Rl}$$

Let the data rate be $R = r \log \text{SNR}$,

$$\begin{aligned} & P(\text{error} | \mathbf{H} = H) \\ & \geq 1 - \frac{I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = H)}{lr \log \text{SNR}} - \frac{1}{lr \log \text{SNR}} \end{aligned}$$

The last term goes to 0 as $\text{SNR} \rightarrow \infty$. Now average over \mathbf{H} to get the average error probability

$$P_e(\text{SNR}) = \mathcal{E}_{\mathbf{H}}[P(\text{error} | \mathbf{H} = H)]$$

Now for any $\delta > 0$, for any H in the set

$$\mathcal{D}_\delta \triangleq \{H : I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = H) < (r - \delta)l \log \text{SNR}\}$$

the probability of error is lower bounded by $1 - \frac{r - \delta}{r} + o(1)$; hence

$$P_e(\text{SNR}) \geq \left(1 - \frac{r - \delta}{r} + o(1)\right) P(\mathcal{D}_\delta)$$

Now choose the input \mathbf{X} to minimize $P(\mathcal{D}_\delta)$ and apply Theorem 4, we have

$$\begin{aligned} P_e(\text{SNR}) & \geq \left(1 - \frac{r - \delta}{r} + o(1)\right) \text{SNR}^{-d_{out}(r - \delta)} \\ & \doteq \text{SNR}^{-d_{out}(r - \delta)} \end{aligned}$$

Take $\delta \rightarrow 0$, by the continuity of $d_{out}(r)$, we have

$$P_e(\text{SNR}) \geq \text{SNR}^{-d_{out}(r)}$$

■

This result says that conditioned on the channel outage event, it is very likely that a detection error occurs; therefore, the outage probability is a lower bound on the error probability.

The outage formulation captures the performance under infinite coding block length, since by coding over an infinitely long block, the input can be reliably detected as long as the data rate is below the mutual information provided by the random realization of the channel. Intuitively, the performance improves as the block length increases; therefore, it is not too surprising that the outage probability is a lower bound on the error probability with any finite block length l . Since $d_{out}(\cdot) = d^*(\cdot)$, Theorem 2 however contains a stronger result: with a finite block length $l \geq m + n - 1$, this bound is tight. That is, no more diversity gain can be obtained by coding over

a block longer than $m + n - 1$, since the infinite block length performance is already achieved.

Consider now the use of a random code for the multi-antenna fading channel over a finite block length l . A detection error can occur as a result of the combination of the following three events: the channel matrix \mathbf{H} is atypically ill-conditioned, the additive noise is atypically large, or some codewords are atypically close together. By going to the outage formulation (effectively taking l to infinity), the problem is simplified by allowing us to focus only on the bad channel event, since for large l the randomness in the last two events is averaged out. Consequently, when there is no outage, the error probability is very small; the detection error is mainly caused by the bad channel event.

With a finite block length l , all three effects come into play, and the error probability given that there is no outage may not be negligible. In the following proof of Theorem 2, we will however show that under the assumption $l \geq m + n - 1$, given that there is no channel outage, the error probability (for an i.i.d. Gaussian input) has an SNR exponent that is not smaller than that of the outage probability; hence outage is still the dominating error event, as in the $l \rightarrow \infty$ case.

Proof: of Theorem 2

With Lemma 5 providing a lower bound on the error probability, to complete the proof we only need to derive an upper bound on the error probability (a lower bound on the optimal diversity gain). To do that, we choose the input to be the random code from the i.i.d. Gaussian ensemble.

Consider at data rate $R = r \log \text{SNR}$ (bits/symbol)

$$\begin{aligned} & P_e(\text{SNR}) \\ & = P_{out}(R)P(\text{error} | \text{outage}) + P(\text{error, no outage}) \\ & \leq P_{out}(R) + P(\text{error, no outage}) \end{aligned}$$

The second term can be upper bounded via a union bound. Assume $X(0), X(1)$ are two possible transmitted codewords, and $\Delta X = X(1) - X(0)$. Suppose $X(0)$ is transmitted, the probability that a ML receiver will make a detection error in favor of $X(1)$, conditioned on a certain realization of the channel, is

$$\begin{aligned} & P(X(0) \rightarrow X(1) | \mathbf{H} = H) \\ & = P\left(\frac{\text{SNR}}{m} \left\| \frac{1}{2} H(\Delta X) \right\|_F^2 \leq \|\mathbf{w}\|^2\right) \end{aligned} \quad (17)$$

where \mathbf{w} is the additive noise on the direction of $H(\Delta X)$, with variance $1/2$. With the standard approximation of the Gaussian tail function: $Q(t) \leq 1/2 \exp(-t^2/2)$, we have

$$P(X(0) \rightarrow X(1) | \mathbf{H} = H) \leq \exp\left[-\frac{\text{SNR}}{4m} \|H(\Delta X)\|^2\right]$$

Averaging over the ensemble of random codes, we have the average pairwise error probability given the channel realization [7]:

$$P(\mathbf{X}(0) \rightarrow \mathbf{X}(1) | \mathbf{H} = H) \leq \det\left(I + \frac{\text{SNR}}{2m} H H^\dagger\right)^{-l} \quad (18)$$

Now at a data rate $R = r \log \text{SNR}$ (bits/symbol), we have in total SNR^{lr} codewords. Apply the union bound, we have

$$\begin{aligned} & P(\text{ error } | \mathbf{H} = H) \\ & \leq \text{SNR}^{lr} \det \left(I + \frac{\text{SNR}}{2m} H H^\dagger \right)^{-l} \\ & = \text{SNR}^{lr} \prod_{i=1}^{\min\{m,n\}} \left(1 + \frac{\text{SNR}}{2m} \lambda_i \right)^{-l} \end{aligned}$$

This bound depends on H only through the singular values. Let $\lambda_i = \text{SNR}^{-\alpha_i}$ for $i = 1, \dots, \min\{m, n\}$, we have

$$P(\text{ error } | \alpha) \leq \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]} \quad (19)$$

Averaging with respect to the distribution of α given in Lemma 3, we have

$$\begin{aligned} P(\text{ error, no outage}) & = \int_{(\mathcal{A}')^c} p(\alpha) P(\text{ error } | \alpha) d\alpha \\ & \leq \int_{(\mathcal{A}')^c} p(\alpha) \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]} d\alpha \end{aligned}$$

where the $(\mathcal{A}')^c$ is the complement of the outage event \mathcal{A}' defined in (14). With a similar argument as in Theorem 4, we can approximate this as

$$\begin{aligned} & P(\text{ error, no outage}) \\ & \leq \int_{(\mathcal{A}')^c} \text{SNR}^{-\sum_i (|m-n|+2i-1)\alpha_i} \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]} d\alpha \\ & \doteq \int_{(\mathcal{A}')^c} \text{SNR}^{-d_G(r,\alpha)} d\alpha \end{aligned}$$

with

$$\begin{aligned} d_G(r,\alpha) & := \sum_{i=1}^{\min\{m,n\}} (2i-1 + |m-n|)\alpha_i \\ & \quad + l \left(\sum_{i=1}^{\min\{m,n\}} (1-\alpha_i)^+ - r \right) \quad (20) \end{aligned}$$

The probability is dominated by the term corresponding to α^* that minimizes $d_G(r,\alpha)$:

$$P(\text{ error, no outage}) \leq \text{SNR}^{-d_G(r)}$$

with

$$d_G(r) := d_G(r, \alpha^*) = \min_{\alpha \notin \mathcal{A}'} d_G(r, \alpha)$$

For $l \geq 2(\min\{m,n\}) - 1 + |m-n| = m+n-1$, the minimum always occurs with $\sum(1-\alpha_i^*) = r$; hence

$$d_G(r) = \min_{\sum \alpha_i = \min\{m,n\} - r} \sum_{i=1}^{\min\{m,n\}} (2i-1 + |m-n|)\alpha_i$$

Compare with (16), we have $d_G(r) = d_{out}(r), \forall r$. The overall error probability can be written as

$$\begin{aligned} P_e(\text{SNR}) & = P_{outage}(R) + P(\text{ error, no outage}) \\ & \doteq \text{SNR}^{-d_{out}(r)} + P(\text{ error, no outage}) \\ & \leq \text{SNR}^{-d_{out}(r)} + \text{SNR}^{-d_G(r)} \\ & \doteq \text{SNR}^{-d_{out}(r)} \end{aligned}$$

Notice the typical error is caused by the outage event, and the SNR exponent matches with that of the lower bound (16), which completes the proof. ■

An alternative derivation of the bound (18) on the pairwise error probability gives some insight to the typical way in which pairwise error occurs. Let $\lambda_i, i = 1, \dots, \min\{m, n\}$ be the non-zero eigenvalues of $H H^\dagger$, and $\Delta \mathbf{x}_i \in \mathcal{C}^l$ be the row vectors of $\Delta \mathbf{X}$. Since $\Delta \mathbf{X}$ is isotropic (i.e. its distribution is invariant to unitary transformations), we have

$$\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \stackrel{d}{=} \sum_{i=1}^{\min\{m,n\}} \lambda_i \|\Delta \mathbf{x}_i\|^2$$

where $\stackrel{d}{=}$ denotes equality in distribution. Consider

$$\begin{aligned} & P \left(\frac{\text{SNR}}{m} \left\| \frac{1}{2} H(\Delta \mathbf{X}) \right\|_F^2 \leq 1 \right) \\ & = P \left(\sum_{i=1}^{\min\{m,n\}} \lambda_i \|\Delta \mathbf{x}_i\|^2 \leq 4m \text{SNR}^{-1} \right) \end{aligned}$$

This probability is bounded by

$$\begin{aligned} & P \left(\lambda_i \|\Delta \mathbf{x}_i\|^2 \leq \frac{4m \text{SNR}^{-1}}{\min\{m,n\}}, i = 1, \dots, \min\{m,n\} \right) \\ & \leq P \left(\sum_{i=1}^{\min\{m,n\}} \lambda_i \|\Delta \mathbf{x}_i\|^2 \leq 4m \text{SNR}^{-1} \right) \\ & \leq P(\lambda_i \|\Delta \mathbf{x}_i\|^2 \leq 4m \text{SNR}^{-1}, i = 1, \dots, \min\{m,n\}) \end{aligned}$$

The upper and lower bounds have the same SNR exponent; hence

$$\begin{aligned} & P \left(\frac{\text{SNR}}{m} \left\| \frac{1}{2} H(\Delta \mathbf{X}) \right\|_F^2 \leq 1 \right) \\ & \doteq P(\lambda_i \|\Delta \mathbf{x}_i\|^2 \leq 4m \text{SNR}^{-1}, i = 1, \dots, \min\{m,n\}) \\ & = P(\|\Delta \mathbf{x}_i\|^2 \leq 4m(\text{SNR}\lambda_i)^{-1}, i = 1, \dots, \min\{m,n\}) \end{aligned}$$

Provided that $\lambda_i \geq \text{SNR}^{-1}$, from (15),

$$P(\|\Delta \mathbf{x}_i\|^2 \leq (\text{SNR}\lambda_i)^{-1}, \forall i) \doteq \prod_{i=1}^{\min\{m,n\}} (\text{SNR}\lambda_i)^{-l},$$

When $\lambda_i < \text{SNR}^{-1}$,

$$P(\|\Delta \mathbf{x}_i\|^2 \leq (\text{SNR}\lambda_i)^{-1}) \doteq \text{SNR}^0$$

Combining these, we have

$$\begin{aligned} & P \left(\frac{\text{SNR}}{m} \left\| \frac{1}{2} H(\Delta \mathbf{X}) \right\|_F^2 \leq 1 \right) \\ & \doteq \prod_{i=1}^{\min\{m,n\}} (\min\{1, \text{SNR}\lambda_i\})^{-l} \end{aligned}$$

which has the same SNR exponent as the right hand side of (18).

On the other hand, given that $\frac{\text{SNR}}{m} \left\| \frac{1}{2} \mathbf{H}(\Delta \mathbf{X}) \right\|_F^2 \leq 1$, there is a positive probability, $P(\|\mathbf{w}\|^2 > 1) > 0$, that an error occurs. Therefore,

$$\begin{aligned} & P(\mathbf{X}(0) \rightarrow \mathbf{X}(1) \mid \mathbf{H} = H) \\ & \doteq P\left(\frac{\text{SNR}}{m} \left\| \frac{1}{2} \mathbf{H}(\Delta \mathbf{X}) \right\|_F^2 \leq 1\right) \end{aligned} \quad (21)$$

This suggests that at high SNR, the pairwise error occurs typically when the difference between codewords (at the receiver end) is of order 1, i.e., has the same order of magnitude as the additive noise.

D. Relationship to the Naive Union Bound

The key idea of the proof of Theorem 2 is to find the right way to apply union bound to obtain a tight upper bound on the error probability. A more naive approach is to directly apply the union bound based on the pairwise error probability (PEP). However, the following argument shows that this union bound is not tight.

Consider the case when the i.i.d. Gaussian random code is used. It follows from (21) that the average pairwise error probability can be approximated as

$$P(\text{ pairwise error}) \doteq P\left(\frac{\text{SNR}}{4m} \|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq 1\right),$$

where $\Delta \mathbf{X}$ is the difference between codewords. Denote \mathcal{F} as the event that every entry of \mathbf{H} has norm $\|\mathbf{H}_{ij}\|^2 \leq \text{SNR}^{-1}$. Given that \mathcal{F} occurs, $\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq \text{SNR}^{-1}(mn)^2 \|\Delta \mathbf{X}\|_F^2$; hence

$$\begin{aligned} & P(\text{ pairwise error} \mid \mathcal{F}) \\ & \geq P(\text{SNR} \|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq 4m) \geq \text{SNR}^0 \end{aligned}$$

and

$$P(\text{ pairwise error}) \geq P(\mathcal{F}) \text{SNR}^0 \doteq \text{SNR}^{-mn}$$

Intuitively, when \mathcal{F} occurs, the channel is in deep fade and it is very likely that a detection error occurs. The average pairwise error probability is therefore lower bounded by SNR^{-mn} .

Now let the data rate $R = r \log \text{SNR}$, the union bound yields

$$\begin{aligned} P_{\text{union}}(R) & = \text{SNR}^{lr} P(\text{ pairwise error}) \\ & \geq \text{SNR}^{-(mn-lr)} \end{aligned} \quad (22)$$

The resulting SNR exponent as a function of r : $d_{\text{union}}(r) = mn - lr$, is plotted in Figure 4, in comparison to the optimal tradeoff curve $d^*(r)$. As spatial multiplexing gain r increases, the number of codewords increases as SNR^{lr} , hence the SNR exponent of the union bound $d_{\text{union}}(r)$ drops with a slope $-l$. Under the assumption $l \geq m + n - 1$, even when we applied (22) to have an ‘‘optimistic’’ bound, it is still below the optimal tradeoff curve. Therefore we conclude that the union bound on the average pairwise error probability is a loose bound on the actual error probability. This strongly suggests that to get significant multiplexing gain, a code design criterion based on pairwise error probability is not adequate.

The reason that this union bound is not tight is as follows. Suppose $\mathbf{X}(0)$ is the transmitted codeword. When the channel

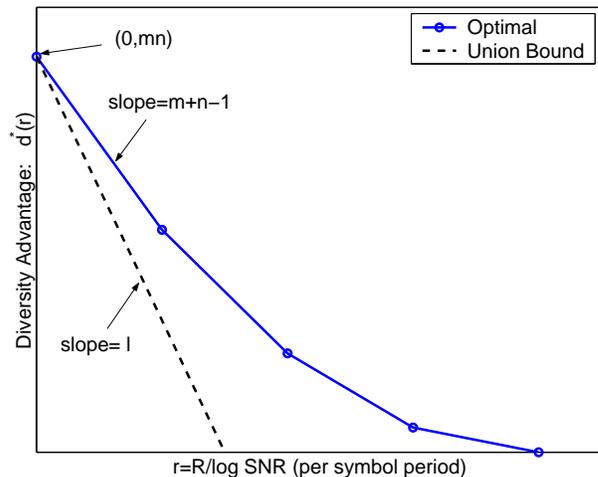


Fig. 4. Union Bound of PEP

matrix \mathbf{H} is ill-conditioned, $\mathbf{H}\mathbf{X}(i)$ is close to $\mathbf{H}\mathbf{X}(0)$ for many i 's. Now it is easy to get confused with many codewords, i.e., the overlap between many pairwise error events is significant. The union bound approach, by taking the sum of the pairwise error probability, over-counts this ‘‘bad channel’’ event, and is therefore not accurate.

To derive a tight bound, in the proof of Theorem 2, we first isolate the outage event,

$$P_e \leq P(\text{outage}) \times 1 + P(\text{error with no outage}),$$

and then bound the error probability conditioned on the channel having no outage with the union bound based on the conditional pairwise error probability. By doing this, we avoid the over-counting in the union bound, and get a tight upper bound of the error probability. It turns out that when $l \geq m + n - 1$, the second term above has the same SNR exponent as the outage probability, which leads to the matching upper and lower bounds on the diversity gain $d^*(r)$. The intuition of this will be further discussed in section IV.

IV. OPTIMAL TRADEOFF: $l < m + n - 1$ CASE

In the case $l < m + n - 1$, the techniques developed in the previous section no longer gives matching upper and lower bounds on the error probability. Intuitively, when the block length l is small, with a random code from the i.i.d. Gaussian ensemble, the probability that some codewords are atypically close to each other become significant, and the outage event is no longer the dominating error event. In this section, we will develop different techniques to obtain tighter bounds, which also provide more insights to the error mechanism of the multiple antenna channel.

A. Gaussian coding bound

In the proof of Theorem 2, we have developed an upper bound on the error probability, which in fact applies for systems with any values of m, n , and l . For convenience, we summarize this result in the following lemma.

Lemma 6: Gaussian coding bound

For the multiple antenna channel (1), let the data rate be $R =$

$r \log \text{SNR}(\text{bps}/\text{Hz})$. The optimal error probability is upper bounded by

$$P_e(\text{SNR}) \leq \text{SNR}^{-d_G(r)} \quad (23)$$

where

$$d_G(r) = \min_{\alpha} \sum_{i=1}^{\min\{m,n\}} (2i-1 + |m-n|)\alpha_i + l \left(\sum_{i=1}^{\min\{m,n\}} (1-\alpha_i) - r \right)$$

where the minimization is taken over the set

$$\mathcal{G} \triangleq \left\{ \alpha \in [0, 1]^{\min\{m,n\}} : \alpha_1 \geq \dots \geq \alpha_{\min\{m,n\}}, \sum_i (1-\alpha_i) > r \right\} \quad (24)$$

The function $d_G(r)$ can be computed explicitly. For convenience, we call a system with m transmit, n receive antennas and a block length l an (m, n, l) system, and define the function

$$G_{m,n,l}(x) = \min_{\alpha \in \mathcal{G}} \sum_{i=1}^{\min\{m,n\}} (2i-1 + |m-n|)\alpha_i + \left(l \sum_{i=1}^{\min\{m,n\}} (1-\alpha_i) - x \right) \quad (25)$$

Lemma 6 says that the optimal error probability is upper bounded by $\text{SNR}^{-d_G(r)}$ with $d_G(r) = G_{m,n,l}(lr)$. $G_{m,n,l}(x)$, also written as $G(x)$, is a piecewise linear function with $G(x) \geq 0$ for x in the range of $[0, l \min\{m, n\}]$. Let

$$k_1 = \left\lfloor \frac{l - |m-n| - 1}{2} \right\rfloor$$

For $i = 1, \dots, k_1$ and $x \in [l(\min\{m, n\} - i), l(\min\{m, n\} - i + 1)]$, $G(x)$ is a linear function with the slope $-(2i-1 + |m-n|)/l$, and $d_G(r) = G(lr)$ agrees with the upper bound on the SNR exponent, $d_{out}(r)$. For $x \leq l(\min\{m, n\} - k_1)$, $G(x)$ is linear with slope -1 , hence $d_G(r)$ has slope $-l$, which is strictly below $d_{out}(r)$.

In summary, for a system with $l < m + n - 1$, the optimal tradeoff curve $d^*(r)$ can be exactly characterized for the range $r \geq \min\{m, n\} - k_1$; in the range that $r < \min\{m, n\} - k_1$, however, the bounds $d_G(r)$ and $d_{out}(r)$ do not match. Examples for systems with $m = n = l = 2$ and $m = n = 4, l = 5$ are plotted in Figure 5, with $k_1 = 1$ and 2 , respectively.

In the next sub-section, we will explore how the Gaussian coding bound can be improved.

B. Typical Error Event

The key idea in the proof of Theorem 2 is to isolate a ‘‘bad channel’’ event $\mathbf{H} \in \mathcal{B}$,

$$P_e(\text{SNR}) \leq P(\mathbf{H} \in \mathcal{B}) \times 1 + P(\text{error}, \mathbf{H} \notin \mathcal{B}), \quad (26)$$

and compute the error probability in the second term with the union bound values of r . While this bound is tight for $l \geq m + n - 1$, it is loose for $l < m + n - 1$. A natural attempt to improve this bound is to optimize over the choices of \mathcal{B} to get the tightest bound. Does this work?

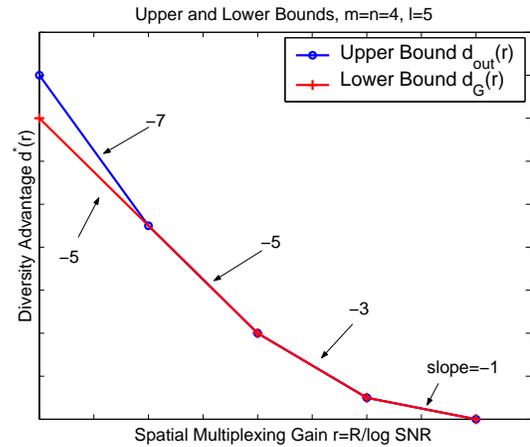
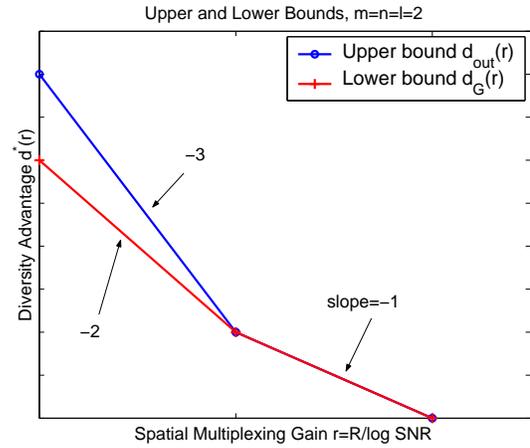


Fig. 5. Upper and Lower Bounds for the Optimal Tradeoff Curve

Let $\lambda_i, i = 1, \dots, \min\{m, n\}$ be the non-zero eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$, and define the random variables

$$\alpha_i = -\frac{\log \lambda_i}{\log \text{SNR}}, i = 1, \dots, \min\{m, n\}$$

Since the error probability depends on the channel matrix only through λ_i 's, we can rewrite the bad channel event in the space of α as $\alpha \in \mathcal{B}'$. (26) thus becomes

$$P_e(\text{SNR}) \leq P(\alpha \in \mathcal{B}') \times 1 + P(\text{error}, \alpha \notin \mathcal{B}') \quad (27)$$

To find the optimal choice of \mathcal{B}' , we first consider the error probability conditioned on a particular realization of α . From (19) we have

$$P(\text{error} | \alpha) \leq \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]}. \quad (28)$$

(27) essentially bounds this conditional error probability by 1 for all $\alpha \in \mathcal{B}'$. In order to obtain the tightest bound from (27), the optimal choice of \mathcal{B}' is exactly given by $\mathcal{B}^* \triangleq \{\alpha : \sum(1-\alpha_i)^+ < r\}$. To see this, we observe that for any $\alpha \in \mathcal{B}^*$, the right hand side of (28) is larger than 1; hence further bounding $P(\text{error} | \alpha)$ by 1 gives a tighter bound, which means the point α should be isolated. On the other hand, for any $\alpha \notin \mathcal{B}^*$, the right hand side of (28) is less than 1; hence it is loose to isolate this α and bound $P(\text{error} | \alpha)$ by 1.

The condition $\sum(1 - \alpha)^+ < r$ in fact describes the outage event at data rate $R = r \log \text{SNR}$, since

$$\begin{aligned} & \log \det \left(I + \frac{\text{SNR}}{m} \mathbf{H} \mathbf{H}^\dagger \right) < R \\ \Leftrightarrow & \prod_i \left(1 + \frac{\text{SNR}}{m} \lambda_i \right) < \text{SNR}^r \\ \stackrel{\text{SNR} \rightarrow \infty}{\Leftrightarrow} & \prod_i \text{SNR}^{(1-\alpha_i)^+} < \text{SNR}^r \end{aligned}$$

Consequently, we conclude that the optimal choice of \mathcal{B} to obtain the tightest upper bound from (26) is simply the outage event.

Discussion: Typical Error Event

Isolating the outage event essentially bounds the conditional error probability by

$$\begin{aligned} P(\text{error} \mid \alpha) & \leq \min\{1, \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]}\} \\ & \doteq \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]^+} \end{aligned}$$

Now the overall error probability can be bounded by

$$P_e(\text{SNR}) \leq \int_{\alpha} p(\alpha) \text{SNR}^{-l[\sum(1-\alpha_i)^+ - r]^+} d\alpha$$

where the integral is over the entire space of α . For convenience, we assume $\alpha_i \in [0, 1], \forall i$, which does not change the SNR exponent of the above bound. Under this assumption, $p(\alpha) \doteq \text{SNR}^{-\sum(2i-1+|m-n|\alpha_i)}$, hence

$$P_e(\text{SNR}) \leq \int_{\alpha} \text{SNR}^{-d'_G(r, \alpha)} d\alpha$$

for

$$\begin{aligned} & d'_G(r, \alpha) \\ = & \sum_{i=1}^{\min\{m, n\}} \left[\begin{aligned} & (2i - 1 + |m - n|)\alpha_i \\ & + l \left(\sum_{i=1}^{\min\{m, n\}} (1 - \alpha_i) - r \right)^+ \end{aligned} \right] \end{aligned}$$

This integral is dominated by the term corresponds to α^* that minimizes $d'_G(r, \alpha)$, i.e.,

$$P_e(\text{SNR}) \leq \text{SNR}^{-d_G(r)}$$

with

$$d_G(r) = \min_{\alpha} d'_G(r, \alpha)$$

$d_G(r)$ is in fact the same as defined in (20), since the optimizing α^* always satisfies $\sum(1 - \alpha_i^*) - r \geq 0$.

Here, the optimization over all α 's provides a closer view of the typical error event. From the above derivation, since $P_e(\text{SNR})$ is dominated by $\text{SNR}^{-d'_G(r, \alpha^*)}$, detection error typically occurs when α falls in a neighborhood of α^* ; in other words, when the channel has a singularity level of α^* .

In the case $l \geq m+n-1$, we have $\sum_i \alpha_i^* = \min\{m, n\} - r$, which is the same singularity level of the typical outage event; therefore detection error is typically caused by channel outage.

On the other hand, when $l < m+n-1$, we have for some r , $\sum \alpha_i^* < \min\{m, n\} - r$, corresponding to

$$\begin{aligned} I(\mathbf{H}) & = \log \prod_{i=1}^{\min\{m, n\}} \left(1 + \frac{\text{SNR}}{m} \lambda_i \right) \\ & > \log \prod_{i=1}^{\min\{m, n\}} \frac{\text{SNR}^{1-\alpha_i}}{m} \\ & > r \log \text{SNR} \end{aligned}$$

at high SNR. That is, the typical error event occurs when the channel \mathbf{H} is not in outage.

Discussion: Distance Between Codewords

Consider a random codebook \mathcal{C} of size SNR^{lr} generated from the i.i.d. Gaussian ensemble. Fix a channel realization $\mathbf{H} = H$. Assume that $\mathbf{X}(0)$ is the transmitted codeword. For any other codeword $\mathbf{X}(k), k \neq 0$ in the codebook, the pairwise error probability between $\mathbf{X}(0)$ and $\mathbf{X}(k)$, from (17), is

$$\begin{aligned} & P(\mathbf{X}(0) \rightarrow \mathbf{X}(k)) \\ = & P \left(\frac{\text{SNR}}{4m} \|\mathbf{H}(\mathbf{X}(k) - \mathbf{X}(0))\|_F^2 \leq \|\mathbf{w}\|^2 \right) \end{aligned}$$

Let $\lambda_i, i = 1, \dots, \min\{m, n\}$ be the ordered non-zero eigenvalues of $\mathbf{H} \mathbf{H}^\dagger$. Write $\Lambda = \text{diag}(\lambda_i, i = 1, \dots, \min\{m, n\})$, and $H = U \sqrt{\Lambda} V^\dagger$ for some unitary matrices U, V . Write $\Delta \mathbf{X} \triangleq \mathbf{X}(0) - \mathbf{X}(k)$. Since $\Delta \mathbf{X}$ is isotropic, it has the same distribution as $\Delta \mathbf{X}' \triangleq V^\dagger \Delta \mathbf{X}$. Following (21), the pairwise error probability can be approximated as

$$\begin{aligned} & P(\mathbf{X}(0) \rightarrow \mathbf{X}(k)) \\ = & P \left(\frac{\text{SNR}}{4m} \|\Lambda(\Delta \mathbf{X}')\|_F^2 \leq \|\mathbf{w}\|^2 \right) \\ \doteq & P(\|\Lambda(\Delta \mathbf{X}')\|_F^2 \leq 4m \text{SNR}^{-1}) \\ \doteq & P(\lambda_i \|\Delta \mathbf{x}_i\|^2 \leq \text{SNR}^{-1}, i = 1, \dots, \min\{m, n\}) \end{aligned}$$

where $\Delta \mathbf{x}_i$'s are the row vectors of $\Delta \mathbf{X}'$. Since $\Delta \mathbf{x}_i$'s have i.i.d. Gaussian distributed entries, we have for any $\beta \in [0, 1]^{\min\{m, n\}}$,

$$\begin{aligned} & P(\|\Delta \mathbf{x}_i\|^2 \leq \text{SNR}^{-\beta_i}, i = 1, \dots, \min\{m, n\}) \\ \doteq & \text{SNR}^{-l \sum \beta_i} \end{aligned} \quad (29)$$

Given a realization of the channel H with $\lambda_i = \text{SNR}^{-\alpha_i}$, an error occurs when $\|\Delta \mathbf{x}_i\|^2 \leq \text{SNR}^{-(1-\alpha_i)}$, for $i = 1, \dots, \min\{m, n\}$, with a probability of order $\text{SNR}^{-l \sum(1-\alpha_i)}$. In the following, we focus on a particular channel realization H^* that causes the typical error event. As shown in section IV-A, the typical error occurs when the channel has a singularity level of α^* , i.e., $\lambda_i = \text{SNR}^{-\alpha_i^*}$, for all i .

In the case that $\sum(1 - \alpha_i^*) - r = 0$, we have

$$\begin{aligned} & P(\mathbf{X}(0) \rightarrow \mathbf{X}(k) \mid \mathbf{H} = H^*) \\ \doteq & \text{SNR}^{-l \sum(1-\alpha_i^*)} = \text{SNR}^{-lr} \end{aligned}$$

Within the codebook \mathcal{C} containing SNR^{lr} codewords, with probability SNR^0 there will be some other codeword so close

to $\mathbf{X}(0)$ that causes confusion. In other words, when the channel $\mathbf{H} = H^*$, a random codebook drawn from the i.i.d. Gaussian ensemble with size SNR^{lr} will have error with high probability. This is natural since the capacity of the channel H^* can barely support this data rate.

On the other hand, if the typical error event occurs when $\sum(1 - \alpha_i^*) - r > 0$,

$$P\left(\exists \mathbf{X}_k \text{ in } \mathcal{C} : \begin{array}{l} \|\Delta \mathbf{x}_i\|^2 \leq \text{SNR}^{-(1-\alpha_i^*)}, \\ \text{for } i = 1, \dots, \min\{m, n\} \end{array}\right) \\ \doteq \text{SNR}^{lr-l\sum(1-\alpha_i^*)} = \text{SNR}^{-p}$$

for some $p > 0$.²

This means the typical error is caused by some codewords in \mathcal{C} that are atypically close to $\mathbf{X}(0)$. Such a ‘‘bad codeword’’ occurs rarely (with probability SNR^{-p}), but has a large probability to be confused with the transmitted codeword; hence this event dominates the overall error probability of the code. This result suggests that performance can be improved by *expurgating* these bad codewords.

C. Expurgated Bound

The expurgation of the bad codewords can be explicitly carried out by the following procedure³:

- **Step 1** Generate a random codebook of size SNR^{lr} , with each codeword $\mathbf{X}(k) \in \mathcal{C}^{m \times l}$.
- **Step 2** Define a set \mathcal{B}' . For the first codeword $\mathbf{X}(0)$, expurgate all the codewords $\mathbf{X}(k)$'s with $\mathbf{X}(k) - \mathbf{X}(0) \in \mathcal{B}'$.
- **Step 3** Repeat this procedure for each of the remaining codewords, until for every pair of codewords, the difference $\mathbf{X}(k) - \mathbf{X}(j) \notin \mathcal{B}'$.

By choosing \mathcal{B}' to obtain the tightest upper bound of the error probability, we get the following result.

Theorem 7: Expurgated Bound

For the channel (1) with data rate $R = r \log \text{SNR}$ (bps/Hz), the optimal error probability is upper bounded by

$$P_e(\text{SNR}) \leq \text{SNR}^{-d_{exp}(r)}$$

where

$$d_{exp}(r) = G_{m,l,n}^{-1}(lr)$$

for $G(x)$ defined in (25).

²This result depends on the i.i.d. Gaussian input distribution only through the fact that the difference between codewords $\Delta \mathbf{X}$ has row vectors $\Delta \mathbf{x}_i$'s satisfying

$$P(\|\Delta \mathbf{x}_i\|^2 \leq \epsilon) \geq \epsilon^l \quad (30)$$

for small ϵ . In fact, this property holds for any other distributions, from which the codewords are independently generated. To see this, assume $\mathbf{r}_1, \mathbf{r}_2 \in \mathcal{C}^l$ be two i.i.d. random vector with pdf. $f(r)$. Now $\Delta \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ has a density at 0 give by $f_{\Delta \mathbf{r}}(0) = \int f^2(r) dr > 0$. Hence $P(\|\Delta \mathbf{r}\|^2 \leq \epsilon) = f_{\Delta \mathbf{r}}(0)\epsilon^l$. Also, (30) is certainly true for the distributions with probability masses or $\int f^2(r) dr = \infty$. This implies that (29) holds for any random code; hence changing the ensemble of the random code cannot improve the bound in Lemma 6.

³This technique is borrowed from the theory of error exponents. The connection is explored in Section VI.

This theorem gives the following interesting dual relation between the Gaussian coding bound derived in Lemma 6 and the expurgated bound: for an (m, n, l) system with m transmit, n receive antennas and a block length of l , using the i.i.d. Gaussian input, if at a spatial multiplexing gain r , one can get, from isolating the outage event, a diversity gain of d , then with an (m, l, n) system of m transmit, l receive antennas and a block length of n , at a spatial multiplexing gain of d/n , one can get a diversity gain of lr , from the expurgated bound. Besides the complete proof of the theorem, we will discuss in the following the intuition behind this result.

In the proof of Lemma 6, we isolate a ‘‘bad channel’’ event $\mathbf{H} \in \mathcal{B}$, and compute the upper bound on the error probability:

$$P_e \leq P(\mathbf{H} \in \mathcal{B}) \times 1 + P(\text{error}, \mathbf{H} \notin \mathcal{B}) \quad (31)$$

The second term, following (21), can be approximated as

$$P(\text{error}, \mathbf{H} \notin \mathcal{B}) \leq \text{SNR}^{lr} P(\text{pairwise error}, \mathbf{H} \notin \mathcal{B}) \\ \doteq \text{SNR}^{lr} P(\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq 1)$$

where the last probability is taken over $\mathbf{H} \notin \mathcal{B}$, and $\Delta \mathbf{X}$. Since $P(\mathbf{H} \in \mathcal{B})$ approaches 0, this can also be written as

$$P(\text{error}, \mathbf{H} \notin \mathcal{B}) \\ \leq \text{SNR}^{lr} P(\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq \text{SNR}^{-1} \mid \mathbf{H} \notin \mathcal{B})$$

At spatial multiplexing gain r , a diversity d gain can be obtained only if there exists a choice of \mathcal{B} such that both terms in (31) are upper bounded by SNR^{-d} , i.e.,

$$d_G(r) = \max \left\{ d \left| \begin{array}{l} \exists \mathcal{B} \subset \mathcal{C}^{n \times m} \text{ s.t. :} \\ P(\mathbf{H} \in \mathcal{B}) \leq \text{SNR}^{-d} \\ P(\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq \text{SNR}^{-1} \mid \mathbf{H} \notin \mathcal{B}) \\ \leq \text{SNR}^{-(lr+d)} \end{array} \right. \right\} \quad (32)$$

Now consider a system with m' transmit, n' receive antennas, and a block length of l' . Let the input data rate be $R' = r' \log \text{SNR}$. We need to find a ‘‘bad codewords’’ set \mathcal{B}' to be expurgated to improve the error probability of the remaining codebook. Clearly, the more we expurgate, the better error probability we can get. However, to make sure that there are enough codewords left to carry the desired data rate, we need

$$P(\Delta \mathbf{X} \in \mathcal{B}') \leq \text{SNR}^{-l'r'}$$

That is, for one particular codeword, the average number of other codewords that need to be expurgated is of order $\text{SNR}^{l'r'} P(\Delta \mathbf{X} \in \mathcal{B}') \leq \text{SNR}^0$. Hence the total number of codewords that need to be expurgated is much less than $\text{SNR}^{l'r'}$, which does not affect the spatial multiplexing gain. Now with the expurgated codebook, the error probability (from union bound and (21)) is:

$$P_e \leq \text{SNR}^{l'r'} P(\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq \text{SNR}^{-1} \mid \Delta \mathbf{X} \notin \mathcal{B}')$$

With the expurgation, for a diversity requirement d' , the highest spatial multiplexing gain that can be supported is

$d_{exp}^{-1}(d')$ with

$$= l' \max \left\{ r' \left| \begin{array}{l} \exists \mathcal{B}' \subset \mathcal{C}^{m' \times l'} \text{ s.t. :} \\ P(\Delta \mathbf{X} \in \mathcal{B}') \leq \text{SNR}^{-l'r'} \\ P(\|\mathbf{H}(\Delta \mathbf{X})\|_F^2 \leq \text{SNR}^{-1} \mid \Delta \mathbf{X} \notin \mathcal{B}') \\ \leq \text{SNR}^{-(l'r'+d')} \end{array} \right. \right\} \quad (33)$$

Now compare (32) and (33), notice that both \mathbf{H} and $\Delta \mathbf{X}$ are i.i.d. Gaussian distributed. If we exchange \mathbf{H} and $\Delta \mathbf{X}$, and equate the parameters $m' = m, n' = l, l' = n, d' = lr = n'r, l'r' = d$, the above two problems become the same. Now in an (m, n, l) system, $ld_{exp}^{-1}(nx)$ is the same function of x as $d_G(x)$ in an (m, l, n) system; hence $ld_{exp}^{-1}(d) = G_{m,l,n}(d)$, and $d_{exp}(r) = G_{m,l,n}^{-1}(lr)$. Theorem 7 follows.

Combining the bounds from Lemma 6 and Theorem 7, yield,

$$P_e(\text{SNR}) \leq \text{SNR}^{-d_l(r)}$$

where

$$d_l(r) \triangleq \max\{d_G(r), d_{exp}(r)\} \quad (34)$$

The example of a system with $m = n = l = 2$ is plotted in Figure 6. In general, for an (m, n, l) system with $l \geq m$, the SNR exponent of this upper bound (to the error probability) is a piecewise linear function described as follows: let $k_1 = \lceil (l - |m - n| - 1)/2 \rceil, k_2 = \lceil (n - |l - m| - 1)/2 \rceil, k_3 = \min\{m, n\}, k_4 = \min\{l, m\}$, the lower bound $d_l(r) = \max\{d_G(r), d_{exp}(r)\}$ connects points

$$\begin{aligned} & (i, (m-i)(n-i)) \quad \text{for } i = k_3, k_3 - 1, \dots, k_3 - k_1 \\ & \left(\frac{(m-j)(l-j)}{l}, nj \right) \quad \text{for } j = k_4 - k_2, \dots, k_4 - 1, k_4 \end{aligned}$$

The connecting points for $i = k_3 - k_1$ and $j = k_4 - k_2$ are respectively $(k_3 - k_1, k_1^2 + k_1|m - n|)$ and $((k_2^2 + k_2|l - m|)/l, n(k_4 - k_2))$. One can check that these two points always lie on the same line with slope $-l$.

$d_l(r)$ matches with $d_{out}(r)$ for all $r \geq k_3 - k_1$, and yields a gap for $r < k_3 - k_1$. At multiplexing gain $r = 0$, corresponding to the points with $j = 0$, $d_l(0) = nk_4$. For any block length $l \geq m$, this gives a diversity gain mn , which again matches with the upper bound $d_{out}(r)$.

When $l < m$, the upper and lower bounds do not match even at $r = 0$. In this case, the maximal diversity gain mn in the channel is not achievable, and $d_l(0)$ gives the optimal diversity gain at $r = 0$. To see this, consider binary detection with $\mathbf{X}(0)$ and $\mathbf{X}(1)$ being the two possible codewords. Let $\Delta \mathbf{X} = \mathbf{X}(1) - \mathbf{X}(0)$, and define $\Omega_{\Delta \mathbf{X}}$ as the l -dimensional subspace of \mathcal{C}^m , spanned by the column vectors of $\Delta \mathbf{X}$. We can decompose the row vectors of \mathbf{H} into the components in $\Omega_{\Delta \mathbf{X}}$ and perpendicular to it, i.e., $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2$, with $\mathbf{H}_2 v = 0$ for any $v \in \Omega_{\Delta \mathbf{X}}$. Since \mathbf{H} is isotropic, it follows that \mathbf{H}_1 contains the component of \mathbf{H} in nl dimensions, and \mathbf{H}_2 contains the rest in $n(m-l)$ dimensions. Now a detection

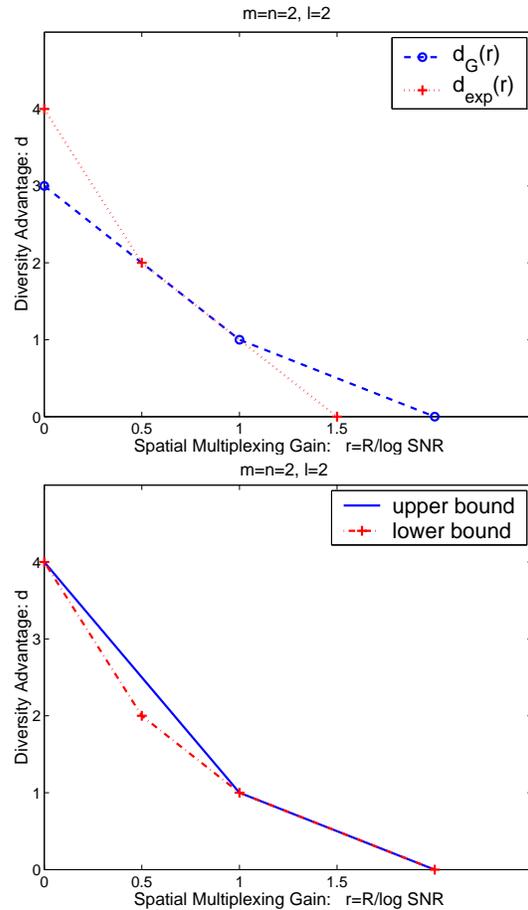


Fig. 6. Upper and Lower Bounds for System with $m = n = l = 2$.

error occurs with a probability

$$\begin{aligned} P(\mathbf{X}(0) \rightarrow \mathbf{X}(1)) & \doteq P\left(\|\mathbf{H}\Delta \mathbf{X}\|_F^2 \leq \text{SNR}^{-1}\right) \\ & = P\left(\|\mathbf{H}_1\Delta \mathbf{X}\|_F^2 \leq \text{SNR}^{-1}\right) \end{aligned}$$

since $\mathbf{H}_2\Delta \mathbf{X} = 0$. If $\|\mathbf{H}_1\|_F^2 \leq \text{SNR}^{-1}$, which has a probability of order SNR^{-nl} , the transmitted signal is lost. Intuitively, when $l < m$, the code is too short to average over all the fading coefficients; thus the diversity is decreased.

D. Space-Only Codes

In general, our upper and lower bounds on the diversity-multiplexing tradeoff curve do not match for the entire range of rates whenever the block length $l < m + n - 1$. However, for the special case of $l = 1$ and $m \leq n$, the lower bound $d_l(r)$ is in fact tight and the optimal tradeoff curve is again completely characterized. This characterizes the performance achievable by coding only over space and not over time.

In this case, it can be calculated from the above formulas that $d_G(r) = m - r$ and $d_{exp}(r) = n(1 - r/m)$. The expurgated bound dominates the Gaussian bound for all $r \in [0, m]$, and hence $d_l(r) = n(1 - r/m)$, a straight line connecting the points $(0, n)$ and $(\min\{m, n\}, 0)$. This provides a lower bound to the optimal tradeoff curve, i.e. an upper bound to the error probability.

We now show that this bound is tight, i.e. the optimal error probability is also lower bounded by:

$$P_e(\text{SNR}) \geq \text{SNR}^{-n(1-\frac{r}{\min\{m,n\}})}. \quad (35)$$

To prove this, suppose there exists a scheme \mathcal{C} can achieve a diversity gain d and multiplexing gain r such that $r/m+d/n > 1$. First we can construct another scheme \mathcal{C}' with the same multiplexing gain, such that the minimum distance between any pair of the codewords in \mathcal{C}' is bounded by

$$\|\Delta X\|^2 \geq \text{SNR}^{1-d/n}$$

To see this, fix any codeword $X(0)$ to be the transmitted codeword, let $X(1)$ be it's nearest neighbor and ΔX be the minimum distance. We have

$$\begin{aligned} & P(\text{error} \mid X(0) \text{ transmitted}) \\ & \geq P(X(0) \rightarrow X(1)) \\ & \doteq P(\text{SNR} \|\mathbf{H}(\Delta X)\|^2 \leq 1) \\ & = P(\text{SNR} \|\mathbf{h}\|^2 \|\Delta X\|^2 \leq 1) \end{aligned}$$

where \mathbf{h} is the n -dimensional component of \mathbf{H} in the direction of ΔX . Now if the minimum distance $\|\Delta X\|$ is shorter than $\text{SNR}^{-(1-d/n)/2}$, the error probability given that $X(0)$ is transmitted is strictly larger than SNR^{-d} . Since the average error probability is SNR^{-d} , there must be a majority of codewords, say, half of them, for which the nearest neighbor is at least $\text{SNR}^{-(1-d/n)/2}$ away. Now take these one half of the codewords to form a new scheme \mathcal{C}' (or to be more precise take half of the codeword for each code in the family), it has the desired minimum distance, and the multiplexing gain r is not changed.

This scheme \mathcal{C}' can be viewed as SNR^r spheres, each of radius at least $\text{SNR}^{-(1-d/n)/2}$, packed in the space \mathcal{C}^m . Notice that each sphere has a volume of $\text{SNR}^{-m(1-d/n)}$. Now since $r/m + d/n > 1$, for small enough $\delta > 0$, we have

$$\text{SNR}^{r-\delta-m(1-d/n)} \geq \text{SNR}^{m\epsilon}$$

for some $\epsilon > 0$. That is, with in a sphere of radius $\text{SNR}^{\epsilon/2}$, there are at most $\text{SNR}^{r-\delta}$ codewords. Consequently, all the other $\text{SNR}^r(1 - \text{SNR}^{-\delta})$ codewords are strictly outside the sphere with radius $\text{SNR}^{\epsilon/2}$. This violates the power constraint (2), since

$$\frac{1}{\text{SNR}^r} \sum_{\mathcal{C}'} \|X_i\|_F^2 \geq \text{SNR}^\epsilon (1 - \text{SNR}^{-\delta}),$$

Thus, we prove that for $l = 1$, $m \leq n$, the optimal diversity-multiplexing curve $d(r) = n(1 - r/m)$.

V. CODING OVER MULTIPLE BLOCKS

So far we have considered the multiple antenna channel (1) in a single block of length l . In this section we consider the case when one can code over k such blocks, each of which fades independently. This is the block fading model. Having multiple independently faded blocks allows us to combine the antenna diversity with other forms of diversity, such as time and frequency diversity.

Corollary 8: Coding over k blocks

For the block fading channel, with a scheme that codes over k blocks, each of which are independently faded, let the input data rate be $R = r \log \text{SNR}$ (bps/Hz) ($lr \log \text{SNR}$ bits/block), the optimal error probability is upper bounded by

$$P_e^{(k)}(\text{SNR}) \leq \text{SNR}^{-k \times d_l(r)}$$

for $d_l(r)$ defined in (34); and lower bounded by

$$P_e^{(k)}(\text{SNR}) \geq \text{SNR}^{-k \times d_{out}(r)}$$

for $d_{out}(r)$ defined in (14).

This means that the diversity gain simply adds across the k blocks. Hence if we can afford to increase the code length k , we can reduce our requirement for the antenna diversity $d^*(r)$ in each channel use, and trade that for a higher data rate.

Compare to the case when coding over single block, since both the upper and lower bounds on the SNR exponent are multiplied by the same factor k , the bounds match for all r when $l \geq m + n - 1$; and for $r \geq \min\{m, n\} - k_1$, with $k_1 = \lceil (l - |m - n| - 1)/2 \rceil$ for $l < m + n - 1$.

This corollary can be proved by directly applying the techniques we developed in the previous sections. Intuitively, with a code of length k blocks, an error occurs only when the transmitted codeword is confused with another codeword in all the blocks; thus the SNR exponent of the error probability is multiplied by k . As an example, we consider the pairwise error probability. In contrast to (21), with coding over k blocks, error occurs between two codewords when

$$\frac{\text{SNR}}{4m} \sum_{t=1}^k \|\mathbf{H}_t(\Delta X_t)\|_F^2 \leq \|\mathbf{w}\|^2$$

where \mathbf{H}_t and ΔX_t are the channel matrix and the difference between the codewords, respectively, in block t . This requires $\text{SNR}/(4m) \|\mathbf{H}_t(\Delta X_t)\|^2 \leq \|\mathbf{w}\|^2$, for $t = 1, \dots, k$. The probability of this event has an SNR exponent of kmn , which is also the total number of random fading coefficients in the channel during k blocks.

VI. CONNECTION TO ERROR EXPONENTS

The diversity-multiplexing tradeoff is essentially the tradeoff between the error probability and the data rate of a communication system. A commonly used approach to study this tradeoff for memoryless channels is through the theory of error exponents [10]. In this section, we will discuss the relation between our results and the theory of error exponents.

For convenience, we quote some results from [10].

Lemma 9: Error Exponents

For a memoryless channel with transition probability density $p(y|x)$, consider block codes with length k and rate \bar{R} (bit per channel use). The minimum achievable error probability has the following bounds:

random coding bound:

$$P_e \leq \exp[-kE_{ran}(\bar{R})] \quad (36)$$

sphere-packing bound:

$$P_e \geq \exp[-k(E_{sp}[\bar{R} - o_1(1)] + o_2(1))] \quad (37)$$

where $o_1(1), o_2(1)$ are terms that go to 0 as $k \rightarrow \infty$, and

$$\begin{aligned} E_{ran}(\bar{R}) &= \max_{0 \leq \rho \leq 1} [E_0(\rho) - \rho \bar{R}] \\ E_{sp}(\bar{R}) &= \sup_{\rho > 0} [E_0(\rho) - \rho \bar{R}] \end{aligned}$$

$E_0(\rho) = \max_q E_0(\rho, q)$, where the maximization is taken over all input distributions q satisfying the input constraint, and

$$E_0(\rho, q) = -\log \int \left[\int q(x)p(y|x)^{1/(1+\rho)} dx \right]^{1+\rho} dy$$

In the block fading model considered in Section V, one can think of l symbol times as one channel use, with the input super symbol of dimension $m \times l$. In this way, the channel is memoryless, since for each use of the channel an independent realization of \mathbf{H} is drawn. One approach to analyze the diversity-multiplexing tradeoff is to calculate the upper and lower bounds on the optimal error probability as given in Lemma 9. There are two difficulties with this approach:

- The computation of the error exponents involves optimization over all input distributions; a difficult task in general.
- Even if the sphere-packing exponent $E_{sp}(\bar{R})$ can be computed, it does not give us directly an upper bound on the diversity-multiplexing curve. Since we are interested in analyzing the error probability for a fixed k (actually, we considered $k = 1$ for most of the paper), the $o(1)$ terms have to be computed as well. Thus, while the theory of error exponents is catered for characterizing the error probability for large block length k , we are more interested in what happens for fixed k but at the high SNR regime.

Because of these difficulties, we took an alternative approach to study the diversity-multiplexing tradeoff curve, exploiting the special properties of the multiple antenna fading channel. We however conjecture that there is a one-to-one correspondence between our results and the theory of error exponents.

Conjecture 10: For the multiple antenna fading channel, the error exponents $E_{ran}(\bar{R}), E_{sp}(\bar{R})$ satisfy

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \frac{E_{ran}(lr \log \text{SNR})}{\log \text{SNR}} &= d_G(r) \\ \lim_{\text{SNR} \rightarrow \infty} \frac{E_{sp}(lr \log \text{SNR})}{\log \text{SNR}} &= d_{out}(r) \end{aligned}$$

i.e. the diversity-multiplexing tradeoff bounds are scaled versions of the error exponent bounds, with both the rate and the exponent scaled by a factor of $1/\log \text{SNR}$.

While we have not been able to verify this conjecture, we have shown that if we fix the input distribution q to be i.i.d. Gaussian, the result is true.

There is also a similar correspondence between the expurgated bound derived in Section IV-C and the expurgated exponent, the definition of which is quoted in the following lemma.

Lemma 11: Expurgated Bound For a memoryless channel characterized with the transition probability density $p(y|x)$,

consider block codes with a given length k and rate \bar{R} (bits per channel use). The achievable error probability is upper bounded by

$$P_e \leq \exp[-kE_{ex}(\bar{R})]$$

where

$$E_{ex}(\bar{R}) = \sup_{\rho \geq 1} [E_x(\rho) - \rho \bar{R}]$$

$E_x(\rho) = \max_q E_x(\rho, q)$, where the maximization is taken over all input distributions q satisfying the input constraint, and

$$\begin{aligned} E_x(\rho, q) &= -\rho \log \int_x \int_{x'} q(x)q(x') \left[\int \sqrt{p(y|x)p(y|x')} dy \right]^{1/\rho} dx dx' \end{aligned}$$

We conjecture that

$$\lim_{\text{SNR} \rightarrow \infty} \frac{E_{ex}(lr \log \text{SNR})}{\log \text{SNR}} = d_{exp}(r) \quad (38)$$

for $d_{exp}(r)$ given in Theorem 7. Again, we have only been able to verify this conjecture for the i.i.d. Gaussian input distribution.

VII. EVALUATION OF EXISTING SCHEMES

The diversity-multiplexing tradeoff can be used as a new performance metric to compare different schemes. As shown in the example of a 2-by-2 system in section III, the tradeoff curve provides a more complete view of the problem than just looking at the maximal diversity gain or the maximal spatial multiplexing gain.

In this section, we will use the tradeoff curve to evaluate the performance of several well-known space-time coding schemes. For each scheme, we will compute the achievable diversity-multiplexing tradeoff curve $d(r)$, and compare it against the optimal tradeoff curve $d^*(r)$. By doing this, we take into consideration both the capability of a scheme to provide diversity and to exploit the spatial degrees of freedom available. Especially for schemes that were originally designed according to different design goals (e.g. to maximize the data rate or minimize the error probability), the tradeoff curve provides a unified framework to make fair comparisons and helps us understand the characteristic of a particular scheme more completely.

A. Orthogonal Designs

Orthogonal designs, first used to design space-time codes in [3], provide an efficient means to generate codes that achieve the full diversity gain. In this section, we will first consider a special case with $m = 2$ transmit antennas, in which case the orthogonal design is also known as the Alamouti scheme [8]. In this scheme, two symbols $\mathbf{x}_1, \mathbf{x}_2$ are transmitted over two symbol periods through the channel as

$$\mathbf{Y} = \sqrt{\frac{\text{SNR}}{2}} \mathbf{H} \begin{bmatrix} \mathbf{x}_1 & -\mathbf{x}_2^\dagger \\ \mathbf{x}_2 & \mathbf{x}_1^\dagger \end{bmatrix} + \mathbf{W}$$

The ML receiver performs linear combinations on the received signals, yielding the equivalent scalar fading channel:

$$\mathbf{y}_i = \sqrt{\frac{\text{SNR} \|\mathbf{H}\|_F^2}{2}} \mathbf{x}_i + \mathbf{w}_i \quad \text{for } i = 1, 2 \quad (39)$$

where $\|\mathbf{H}\|_F$ is the Frobenius norm of \mathbf{H} . $\|\mathbf{H}\|_F^2$ is chi-square distributed with dimension $2mn$: $\|\mathbf{H}\|_F^2 \sim \chi_{2mn}^2$. It is easy to check that for small ϵ ,

$$P(\|\mathbf{H}\|_F^2 \leq \epsilon) \approx \epsilon^{mn}$$

In our framework, we view the Alamouti scheme as an inner code to be used in conjunction with an outer code which generates the symbols \mathbf{x}_i 's. The rate of the overall code scales as $R = r \log \text{SNR}$ (bits/symbol). Now for the scalar channel (39), using similar approach discussed in section III-C, we can compute the tradeoff curve for the Alamouti scheme with the best outer code (or, more precisely, the best family of outer codes). To be specific, we compute the SNR exponent of the minimum achievable error probability for channel (39) at rate $R = r \log \text{SNR}$. To do that, we lower bound this error probability by the outage probability, and upper bound it by choosing a specific outer code.

Conditioned on any realization of the channel matrix $\mathbf{H} = H$, channel (39) has capacity $\log(1 + \text{SNR} \|\mathbf{H}\|_F^2 / 2)$. The outage event for this channel at a target data rate R is thus defined as

$$\left\{ H : \log \left[1 + \frac{\text{SNR} \|\mathbf{H}\|_F^2}{2} \right] < R \right\}$$

It follows from Lemma 5 that when outage occurs, there is a significant probability that a detection error occurs; hence the outage probability is a lower bound to the error probability with any input, up to the SNR exponent.

Let $R = r \log \text{SNR}$, the outage probability

$$\begin{aligned} P_{out}(R) &= P \left(\log \left[1 + \frac{\text{SNR} \|\mathbf{H}\|_F^2}{2} \right] < R \right) \\ &= P \left(1 + \frac{\text{SNR} \|\mathbf{H}\|_F^2}{2} < \text{SNR}^r \right) \\ &\doteq P(\|\mathbf{H}\|_F^2 \leq \text{SNR}^{-(1-r)+}) \\ &\doteq \text{SNR}^{-mn(1-r)+} \end{aligned}$$

That is, for the Alamouti scheme, the tradeoff curve $d(r)$ is upper bounded (lower bound on the error probability) by $d_{out}(r) = mn(1-r)^+$.

To find an upper bound on the error probability, we can use a QAM constellation for the symbol \mathbf{x}_i 's. For each symbol, to have a constellation of size SNR^r , with $r < 1$, we choose the distance between grid points to be $\text{SNR}^{-r/2}$. Assume that one point \mathbf{c}_0 in the constellation is transmitted. Let \mathbf{c}_1 be one of the nearest neighbors. From (21), the pairwise error probability

$$\begin{aligned} P(\mathbf{c}_0 \rightarrow \mathbf{c}_1) &\doteq P \left(\sqrt{\frac{\text{SNR} \|\mathbf{H}\|_F^2}{2}} \text{SNR}^{-r/2} < 1 \right) \\ &\doteq P(\|\mathbf{H}\|_F^2 \leq \text{SNR}^{-(1-r)}) \\ &\doteq \text{SNR}^{-mn(1-r)} \end{aligned}$$

Now since for the QAM constellation there are at most 4 nearest neighbors to \mathbf{c}_0 , the overall error event is simply the union of these 4 pairwise error event. Therefore, the error probability is upper bounded by 4 times the pairwise error probability, and has the same SNR exponent $mn(1-r)$.

Another approach to obtain this upper bound is by using the duality argument developed in section IV-C. Channel (39) is essentially a channel with 1 transmit, mn receive antennas and a block length $l = 1$. Consider the dual system with 1 transmit, 1 receive antenna and a block length $l = mn$. It is easy to verify that the random coding bound for the dual system is $G_{1,1,mn}(mnr) = 1 - r/(mn)$. Therefore for the original system, the expurgated exponent is $G_{1,1,mn}^{-1}(r) = mn(1-r)$.

Combining the upper and lower bounds, we conclude that for the Alamouti scheme, the optimal tradeoff curve is $d_{alamouti}(r) = mn(1-r)^+$. This curve is shown in Figure 7 for cases with $n = 1$ receive antenna and $n = 2$ receive antennas, with different block length l .

For the case $n = 1$ and $l \geq 2$, the Alamouti scheme is optimal, in the sense that it achieves the optimal tradeoff curve $d^*(r)$ for all r . Therefore the structure introduced by the Alamouti scheme, while greatly simplifies the transmitter and receiver designs, does not lose optimality in terms of the tradeoff.

In the case $n = 2$, however, the Alamouti scheme is in general not optimal: it achieves the maximal diversity gain of 4 at $r = 0$, but falls below the optimal for positive values of r . In the case $l = 2$, it achieves the first line segment of the lower bound $g_2(r)$; for the case that $l \geq 3$, its tradeoff curve is strictly below optimal for any positive value of r .

The fact that the Alamouti scheme does not achieve the full degrees of freedom has already been pointed out in [16]; this corresponds in Figure 7 to $d(1) = 0$. Our results give a stronger conclusion: the achieved diversity-multiplexing tradeoff curve is sub-optimal for all $r > 0$.

It is shown in [3] [17] that a "full rate" orthogonal design does not exist for systems with $m > 2$ transmit antennas. A full rate design corresponds to the equivalent channel (39), with a larger matrix \mathbf{H} . Even if such a full rate design exists, the maximal spatial multiplexing gain achieved is just $r = 1$, since "full rate" essentially means that only one symbol is transmitted per symbol time. Therefore, the potential of a multiple antenna channel to support higher degrees of freedom is not fully exploited by the orthogonal designs.

B. V-BLAST

Orthogonal designs can be viewed as an effort primarily to maximize the diversity gain. Another well-known scheme that mainly focuses on maximizing the spatial multiplexing gain is V-BLAST [4].

We consider V-BLAST for a square system with n transmit and n receive antennas. With V-BLAST, the input data is divided into independent sub-streams which are transmitted on different antennas. The receiver first demodulates one of the sub-streams by nulling out the others with a decorrelator. After this sub-stream is decoded, its contribution is subtracted from the received signal and the second sub-stream is in

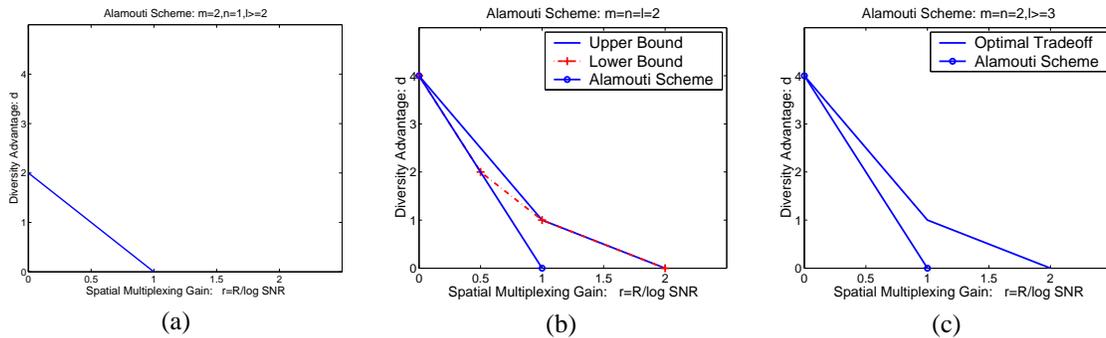


Fig. 7. Tradeoff Curves for Alamouti Scheme

turn demodulated by nulling out the remaining interference. Suppose for each sub-stream, a block code of length l symbols is used. With this successive nulling and canceling process, the channel is equivalent to:

$$\mathbf{y}_i = \sqrt{\frac{\text{SNR}}{n}} \mathbf{g}_i \mathbf{x}_i + \mathbf{w}_i \quad \text{for } i = 1, \dots, n \quad (40)$$

where $\mathbf{x}_i, \mathbf{y}_i, \mathbf{w}_i \in \mathcal{C}^l$ are the transmitted, received signals and the noise for the i^{th} sub-stream; \mathbf{g}_i^2 is the signal to noise ratio at the output of the i^{th} decorrelator. Again, we apply an outer code in the input \mathbf{x}_i 's so that the overall input data rate $R = r \log \text{SNR}$ (bits/symbol), and compute the tradeoff curve achieved by the best outer code.

This equivalent channel model is not precise since error propagation is ignored. In the V-BLAST system, an erroneous decision made in an intermediate stage affects the reliability for the successive decisions. However, in the following, we will focus only on the frame error probability. That is, a frame of length l symbols is said to be successfully decoded only if all the sub-streams are correctly demodulated; whenever there is error in any of the stages, the entire frame is said to be in error. To this end, (40) suffices to indicate the frame error performance of V-BLAST.

The performance of V-BLAST depends on the order in which the sub-streams are detected and the data rates assigned to the sub-streams. We will start with the simplest case: the same data rate is assigned to all sub-streams; and the receiver detects the sub-streams in a prescribed order regardless of the realization of the channel matrix \mathbf{H} . In this case, the equivalent channel gains are chi-square distributed: $\mathbf{g}_i^2 \sim \chi_{2i}^2$, with $P(\mathbf{g}_i^2 \leq \epsilon) \approx \epsilon^{2i}$. The data rates in all sub-streams are $R_i = r/n \log \text{SNR}$ (bits/symbol), for $i = 1, \dots, n$. Now each sub-stream passes through a scalar channel with gain \mathbf{g}_i . Using the same argument for the orthogonal designs, it can be seen that an error occurs at the i^{th} sub-stream with probability

$$P_e^{(i)}(\text{SNR}) \doteq \text{SNR}^{-i(1-r/n)^+}$$

with the first sub-stream having the worst error probability. The frame error probability $P_e(\text{SNR})$ is bounded by $P_e^{(1)}(\text{SNR}) \leq P_e(\text{SNR}) \leq \sum_{i=1}^n P_e^{(i)}(\text{SNR})$. Since the upper and lower bounds have the same SNR exponent, we have

$$P_e(\text{SNR}) \doteq \text{SNR}^{-(1-r/n)^+}$$

The tradeoff curve achieved by this scheme is thus $d(r) = (1-r/n)^+$. The maximal achievable spatial multiplexing gain

is n , which is the total number of degrees of freedom provided by the channel. However, the maximal diversity gain is 1, which is far below the maximal diversity gain n^2 provided by the channel. This tradeoff curve is plotted in Figure 8 under the name ‘‘V-BLAST(1)’’.

We observe that in the above version of V-BLAST, the first stage (detecting the first sub-stream) is the bottleneck stage. There are various ways to improve the performance of V-BLAST, by improving the reliability at the early stages. Clearly, the order in which the sub-streams are demodulated affects the performance. In [4], it is shown that fixing the same data rate for each sub-stream, the optimal ordering is to choose the sub-stream in each stage such that the SNR at the output of the corresponding decorrelator is maximized. Simulation results in [4] show that a significant gain can be obtained by applying this ordering. Essentially, choosing the order of detection based on the realization of \mathbf{H} changes the distribution of the effective channel gains \mathbf{g}_i in (40). For example for the first detected sub-stream, the channel gain \mathbf{g}_1 is the maximum gain of n possible decorrelators, the reliability of detecting this sub-stream and hence the entire frame is therefore improved.

Since the \mathbf{g}_i 's are not independent of each other, it is complicated to characterize the tradeoff curve exactly. However, a simple lower bound to the error probability can be derived as follows. Assume that for all sub-streams $3, \dots, n$, the correct transmitted symbols are given to the receiver by a genie, and hence are canceled. With the remaining two sub-streams, let the corresponding column vectors in \mathbf{H} be \mathbf{h}_1 and \mathbf{h}_2 . Write $\mathbf{h}_i = \|\mathbf{h}_i\| \theta_i$, where $\theta_i \in \mathcal{C}^n$ has unit length, for $i = 1, 2$. θ_i 's are independent of the norms $\|\mathbf{h}_i\|$, and are independently isotropically distributed on the surface of the unit sphere in \mathcal{C}^n . It is easy to check that

$$P(1 - \langle \theta_1, \theta_2 \rangle^2 < \epsilon) \approx \epsilon^{n-1}$$

for small ϵ . The gains of the two possible decorrelators are $\|\mathbf{h}_i\|^2 (1 - \langle \theta_1, \theta_2 \rangle^2)$, $i = 1, 2$. Given that $1 - \langle \theta_1, \theta_2 \rangle^2 < \epsilon$, with high probability (of order 1) both gains are small. In other words,

$$P(\mathbf{g}_{n-1}^2 \leq \text{SNR}^{-\alpha}) \doteq \text{SNR}^{-(n-1)\alpha}$$

Consequently, the error probability of this scheme is lower bounded by

$$P_e(\text{SNR}) \geq P_e^{(n-1)}(\text{SNR}) \doteq \text{SNR}^{-(n-1)(1-r/n)}$$

The upper bound of the tradeoff curve $d(r) \leq (n-1)(1-r/n)$ is plotted in Figure 8 under the name “V-BLAST(2)”.

Another way to improve the performance of V-BLAST is to fix the detection order but assign different data rates to different sub-streams. As proposed in [18], since the first sub-stream passes through the most unreliable channel (g_1 being small with the largest probability), it is desirable to have a lower data rate transmitted by that sub-stream. In our framework, we can in fact optimize the data rate allocation between sub-streams to get the best performance. Let the data rate transmitted in the i^{th} sub-stream be $r_i \log \text{SNR}$, for $i = 1, \dots, n$. The probability of error for the i^{th} sub-stream is

$$P_e^{(i)}(\text{SNR}) \doteq \text{SNR}^{-i(1-r_i)}$$

Now the overall error probability has a SNR exponent $\min_{i:r_i>0} i(1-r_i)$. The minimization is taken over all the sub-streams that are actually used, i.e., $r_i > 0$. We can choose the values of r_i 's to maximize this exponent:

$$d(r) = \max_{r_1, \dots, r_n} \left[\min_{i:r_i>0} i(1-r_i) \right]$$

subject to: $\sum_i r_i = r; r_i \in [0, 1] \forall i$

The optimal rate allocation is described as follows:

- For $r \leq 1/n$, only one sub-stream is used. i.e., $r_n = r$, and $r_i = 0$ for $i = 1, \dots, n-1$. The tradeoff curve is thus $n(1-r)$.
- For $r \in [1/n, 2/n + 1/(n-1)]$, two sub-streams are used with $n(1-r_n) = (n-1)(1-r_{n-1})$ and $r = r_n + r_{n-1}$. Hence $r_n = 1/n + (n-1)/(2n-1)(r-1/n)$ and $r_{n-1} = n/(2n-1)(r-1/n)$. The tradeoff curve is $n-1 - \frac{n(n-1)}{(2n-1)}(\frac{1}{n} - r)$.
- Define

$$r_k = \sum_{i=0}^{k-1} \frac{k-i}{n-i}$$

Let $r_0 = 0$. The tradeoff curve $d(r)$ connects points $(r_k, n-k)$ for $k = 0, \dots, n$. For $r \in [r_{k-1}, r_k]$, k sub-streams are used, with rate r_i satisfying $\sum_{i=n-k+1}^n r_i = r$ and $d(r) = i(1-r_i)$ for $i = n-k+1, \dots, n$.

The resulting tradeoff curve is plotted in Figure 8 as “V-BLAST(3)”.

We observe that for all versions of V-BLAST the achieved tradeoff curve is lower than the optimal, especially when the multiplexing gain r is small. This is because independent sub-streams are transmitted over different antennas; hence each sub-stream sees only n random fading coefficients. Even if we assume no interference between sub-streams, the tradeoff curve is just $\bar{d}(r) = n - r$. None of the above approaches can reach above this line. The reliability of V-BLAST is thus limited by the lack of coding between sub-streams.

Now to compare V-BLAST with the orthogonal designs, we observe from the tradeoff curves in Figure 7 and 8 that for low multiplexing gain, the orthogonal designs yield higher diversity gain, while for high multiplexing gain, V-BLAST is better. Similar comparison is also made in [16], where it is

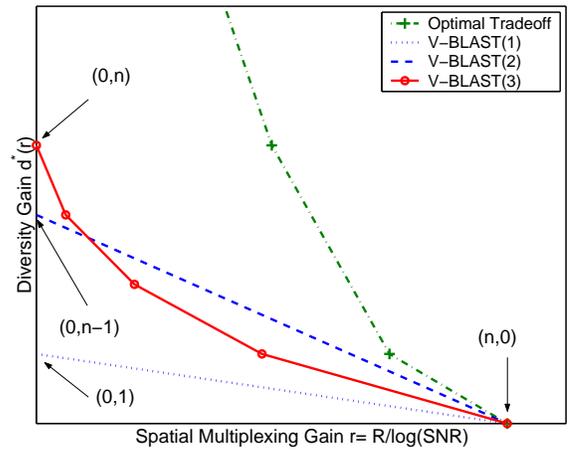
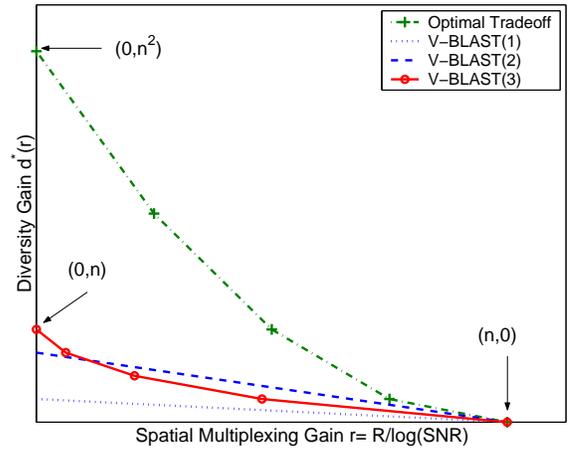


Fig. 8. Tradeoff Performance of V-BLAST

pointed out that comparing to V-BLAST “orthogonal designs are not suitable for very high-rate communications”. Note that we used the notion of high/low multiplexing gain instead of high/low data rate. As discussed in Section II, the multiplexing gain indicates the data rate normalized by the channel capacity as a function of SNR. This notion is more appropriate since otherwise comparing schemes at a certain data rate may yield different results at different SNR levels.

C. D-BLAST

While the tradeoff performance of V-BLAST is limited due to the independence over space, D-BLAST [2], with coding over the signals transmitted on different antennas, promises a higher diversity gain.

In D-BLAST, the input data stream is divided into sub-streams, each of which is transmitted on different antennas time slots in a diagonal fashion. For example, in a 2×2 system, the transmitted signal in matrix form is

$$\begin{bmatrix} 0 & \mathbf{x}_1^{(1)} & \mathbf{x}_1^{(2)} & \dots \\ \mathbf{x}_2^{(1)} & \mathbf{x}_2^{(2)} & \mathbf{x}_2^{(3)} & \dots \end{bmatrix} \in \mathcal{C}^{2 \times l} \quad (41)$$

where $\mathbf{x}_i^{(k)}$ denotes the symbols transmitted on the i^{th} antenna for sub-stream k . The receiver also uses a successive nulling

and canceling process. In the above example, the receiver first estimates $\mathbf{x}_2^{(1)}$ and then estimates $\mathbf{x}_1^{(1)}$ by treating $\mathbf{x}_2^{(2)}$ as interference and nulling it out using a decorrelator. The estimates of $\mathbf{x}_1^{(1)}$ and $\mathbf{x}_2^{(1)}$ are then fed to a joint decoder to decode the first sub-stream. After decoding the first sub-stream, the receiver cancels the contribution of this sub-stream from the received signals, and start to decode the next sub-stream, etc. Here, an overhead is required to start the detection process; corresponding to the 0 symbol in the above example.

In this section, we assume that each $\mathbf{x}_i^{(k)}$ is a block of symbols of length l' . In a frame of length l , there are $l/l' - m(m-1)$ sub-streams. A frame error occurs when any of these sub-streams is incorrectly decoded; therefore the probability of a frame error is, at high SNR, $l/l' - m(m-1)$ times the error probability in each sub-stream; and thus has the same SNR exponent as the sub-stream error probability.⁴ In the following, we will focus on the detection of only one sub-stream, denoted as $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{C}^{l'}$. In a square $n \times n$ system, each sub-stream passes through an equivalent channel as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \sqrt{\frac{\text{SNR}}{n}} \begin{bmatrix} \mathbf{g}_1 & 0 & \dots & 0 \\ 0 & \mathbf{g}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{g}_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} + \mathbf{w} \quad (42)$$

where $\mathbf{x}_i, \mathbf{y}_i \in \mathcal{C}^{l'}$, and \mathbf{g}_i^2 is the gain of the nulling decorrelator used for \mathbf{x}_i . We first ignore the overhead in D-BLAST and write the data rate R as the number of bits transmitted in each use of the equivalent channel (42).

The important difference between (42) and (40), the equivalent channel of V-BLAST, is that here the transmitted symbols \mathbf{x}_i 's belong to the same sub-stream and one can apply an outer code to code over these symbols; in contrast, the \mathbf{x}_i 's in V-BLAST correspond to independent data streams. The advantage of this is that each individual sub-stream passes through all the sub-channels; hence an error in one of the sub-channels, protected by the code, does not necessarily cause the loss of the stream.

It can be shown that \mathbf{g}_i 's are independent with distribution $\mathbf{g}_i^2 \sim \chi_{2i}^2$, with $P(\mathbf{g}_i^2 \leq \epsilon) \approx \epsilon^{2i}$. The tradeoff curve achieved by the optimal outer code can be derived using a similar approach as the proof of Theorem 2: by deriving a lower bound on the optimal error probability from the outage analysis; and an upper bound by picking the i.i.d. Gaussian random code as the input.

Given a channel realization $\mathbf{H} = H$ (unknown to the transmitter), the capacity of (42) is

$$\sum_{i=1}^n \log \left(1 + \frac{\text{SNR}}{n} \mathbf{g}_i^2 \right)$$

At data rate $R = r \log \text{SNR}$ (bps/Hz), the outage probability

⁴In practice, the error probability does depend on l' ; choosing a small value of l' makes the error propagation more severe; on the other hand, increasing l' requires a larger overhead to start the detection process.

for this channel is defined as

$$\begin{aligned} P_{out}^{BLAST}(R) & \triangleq P \left(\mathbf{H} : \sum \log \left(1 + \frac{\text{SNR}}{n} \mathbf{g}_i^2 \right) \leq r \log \text{SNR} \right) \\ & = P \left[\prod_{i=1}^n \left(1 + \frac{\text{SNR}}{n} \mathbf{g}_i^2 \right) \leq \text{SNR}^r \right] \end{aligned}$$

Write $\mathbf{g}_i^2 = \text{SNR}^{-\alpha_i}$, with $\alpha_i \geq 0$, we have

$$P_{out}^{BLAST}(R) \doteq P \left(\sum (1 - \alpha_i)^+ \leq r \right)$$

\mathbf{g}_i^2 's are independent and chi-square distributed, with density

$$f_{g_i^2}(x) = K_i x^{i-1} e^{-x/2}$$

where K_i is a normalizing constant. Now the random variables $\alpha_i = -\log g_i^2 / \log \text{SNR}$, $i = 1, \dots, n$ are also independent; hence the pdf. of $\alpha = [\alpha_1, \dots, \alpha_n]$ is

$$\begin{aligned} f_{\alpha}(x) & = \prod_{i=1}^n f_{\alpha_i}(x_i) \\ & = \prod_{i=1}^n K_i (\log \text{SNR}) \text{SNR}^{-ix_i} \exp \left(-\frac{\text{SNR}^{-x_i}}{2} \right) \end{aligned}$$

The outage probability is thus

$$P_{out}^{BLAST}(R) \doteq \int_{\mathcal{A}} f_{\alpha}(x) dx$$

where $\mathcal{A} = \{\alpha : \sum (1 - \alpha_i)^+ \leq r\}$ corresponds to the outage event. At high SNR, we have:

$$P_{out}^{BLAST}(r \log \text{SNR}) \doteq \int_{\mathcal{A}} \prod_{i=1}^n \text{SNR}^{-ix_i} dx$$

As $\text{SNR} \rightarrow \infty$, the integral is dominated by the term with the largest SNR exponent, and

$$P_{out}^{BLAST}(r \log \text{SNR}) \doteq \text{SNR}^{-d_{out}^{BLAST}(r)}$$

with

$$d_{out}^{BLAST}(r) = \inf_{\alpha: \sum (1 - \alpha_i)^+ \leq r} \sum_{i=1}^N i \alpha_i \quad (43)$$

With the same argument as Lemma 5, it can be shown that the curve $d_{out}^{BLAST}(r)$ is an upper bound on the optimal achievable tradeoff curve in D-BLAST. On the other hand, by picking the input to be the i.i.d. Gaussian random code and using a similar approach as in section III-C, one can show that for $l' \geq n$, the diversity achieved at any spatial multiplexing gain r is given by $d_{out}^{BLAST}(r)$. Therefore, the optimal tradeoff curve achieved by D-BLAST is $d^{BLAST}(r) = d_{out}^{BLAST}(r)$ as given in (43).

The optimization (43) can be explicitly solved. For $k = 0, \dots, n$, the minimizing α^* for $r \in [n - k - 1, n - k]$ is: $\alpha_i^* = 1$ for $i = 1, \dots, n - k - 1$, $\alpha_{n-k}^* = r - (n - k - 1)$, and $\alpha_i^* = 0$ for $i \geq n - k$. The resulting tradeoff curve $d^{BLAST}(r)$ connects points $(n - k, k(k+1)/2)$ for $k = 0, \dots, n$. $d^{BLAST}(r)$ is plotted in Figure 9, and compared to

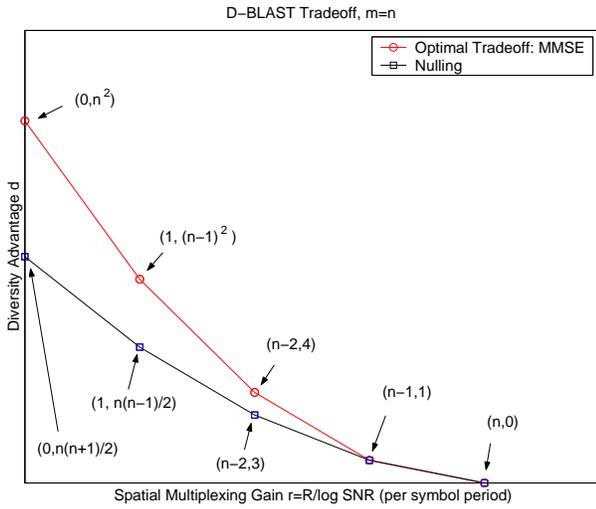


Fig. 9. Tradeoff Curve for D-BLAST

the optimal tradeoff curve. We observe that the tradeoff of D-BLAST is strictly sub-optimal for all r . In particular, while it can reach the maximal spatial multiplexing gain at $r = n$, the maximal diversity gain one can get at $r = 0$ is just $n(n+1)/2$, out of n^2 provided by the channel.

The reason for this loss of diversity is illustrated in the following example.

Example: D-BLAST for 2×2 system

Consider a 2×2 system, for which the channel matrix is denoted as $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]$, where $\mathbf{h}_i \in \mathcal{C}^2$ has i.i.d. $\mathcal{CN}(0, 1)$ entries. Decompose \mathbf{h}_1 into the component in the direction of \mathbf{h}_2 ,

$$\mathbf{h}_{1\parallel 2} = \frac{\langle \mathbf{h}_1, \mathbf{h}_2 \rangle}{\|\mathbf{h}_2\|^2} \mathbf{h}_2$$

and the component that is perpendicular to \mathbf{h}_2 , $\mathbf{h}_{1\perp 2} = \mathbf{h}_1 - \mathbf{h}_{1\parallel 2}$. The equivalent channel (42) has channel gains given by

$$\begin{aligned} \mathbf{g}_1^2 &= \|\mathbf{h}_{1\perp 2}\|^2 \sim \chi_2^2 \\ \mathbf{g}_2^2 &= \|\mathbf{h}_2\|^2 \sim \chi_2^4 \end{aligned}$$

Note that $\|\mathbf{h}_{1\parallel 2}\|^2$ and $\|\mathbf{h}_{1\perp 2}\|^2$ are exponentially distributed. However, in D-BLAST, the term of $\mathbf{h}_{1\parallel 2}\mathbf{x}_1$ is discarded by the nulling process, and the symbol \mathbf{x}_1 passes through an equivalent channel with gain $\mathbf{g}_1 = \|\mathbf{h}_{1\perp 2}\|$. Consequently, the received signals from each sub-stream depend only on 3 independent fading coefficients, and the maximal diversity for the nulling D-BLAST in this case is just 3. In a general $n \times n$ system, the component of \mathbf{H} in $n(n-1)/2$ dimensions is nulled out, and the maximal diversity is only $n(n+1)/2$.

Since nulling causes the degradation of diversity, it is natural to replace the nulling step with a linear MMSE receiver. The equivalent channel for MMSE D-BLAST is of the same form as in (42), except that the channel gains \mathbf{g}_i 's are changed. Denote the channel gains for the original nulling D-BLAST as \mathbf{g}_i^{dec} , and for MMSE D-BLAST as \mathbf{g}_i^{mmse} . It turns out the MMSE D-BLAST achieves the entire optimal tradeoff curve $d^*(r)$. This is a direct consequence of the fact that given any

realization of \mathbf{H} , the mutual information of the channel is achieved by successive cancellation and MMSE receivers [19]:

$$I(\mathbf{X}; \mathbf{Y}) = \sum_{k=1}^n \log(1 + \mathbf{g}_k^{mmse});$$

therefore, the optimal outage performance can be achieved, i.e., the SNR exponent of the outage probability for the MMSE D-BLAST matches with $d_{out}(r)$ defined in Theorem 4. Now using the same argument which proved Theorem 2, it can be shown that for $l' \geq 2n - 1$, with the i.i.d. Gaussian random code, the optimal tradeoff curve $d^*(r) = d_{out}(r)$ is achieved.

The difference in performance between the decorrelator and the MMSE receiver is quite surprising, since in the high SNR regime, one would expect that they have similar performance. What causes the difference in performance? For the 2×2 example, the equivalent channel gains for the MMSE D-BLAST are

$$\begin{aligned} (\mathbf{g}_1^{mmse})^2 &= \|\mathbf{h}_{1\perp 2}\|^2 + \frac{\|\mathbf{h}_{1\parallel 2}\|^2}{\text{SNR} \|\mathbf{h}_2\|^2 + 1} \\ (\mathbf{g}_2^{mmse})^2 &= \|\mathbf{h}_2\|^2 = (\mathbf{g}_2^{dec})^2 \end{aligned}$$

An explicit calculation shows that at high SNR, the *marginal distributions* of the gains \mathbf{g}_i^{dec} and \mathbf{g}_i^{mmse} are asymptotically the same, for each stage i . The difference is in their *statistical dependency*: under the MMSE receiver, the gains \mathbf{g}_1^{mmse} and \mathbf{g}_2^{mmse} from the two stages are *negatively correlated*; under the decorrelator, the gains \mathbf{g}_1^{dec} and \mathbf{g}_2^{dec} are *independent*. The gain \mathbf{g}_2^{mmse} is small when the channel gain \mathbf{h}_2 from the second transmit antenna is small, but in that case the interference caused by transmit antenna 2 during the demodulation of \mathbf{x}_1 is also weak. The MMSE receiver takes advantage of that weak interference in demodulating \mathbf{x}_1 in stage 1. The decorrelator, on the other hand, is insensitive to the *strength* of the interference as it simply nulls out the *direction* occupied by the signal. Combined with coding across the two transmit antennas, the negative correlation of the channel gains in the two stages provides more diversity in MMSE D-BLAST than in decorrelating D-BLAST.

The above results are derived by ignoring the overhead that is required to start the D-BLAST processing. With the overhead, the actual achieved data rate is decreased, therefore both the nulling and MMSE D-BLAST do not achieve the optimal tradeoff curve.

In summary, for the examples shown in this section, we observe that the tradeoff curve is powerful enough to distinguish the performance of schemes, even with quite subtle variations. Therefore, it can serve as a good performance metric in comparing existing schemes and designing new schemes for multiple antenna channels. It should be noted that other than for the 2×1 channel (for which Alamouti scheme is optimal), there is no explicitly constructed coding scheme that achieves the optimal tradeoff curve for *any* $r > 0$. This remains an open problem.

VIII. CONCLUSIONS

Previous research on multi-antenna coding schemes has focused either on extracting the maximal diversity gain or the

maximal spatial multiplexing gain of a channel. In this paper, we present a new point of view that both types of gain can in fact be simultaneously achievable in a given channel, but there is a tradeoff between them. The diversity-multiplexing tradeoff achievable by a scheme is a more fundamental measure of its performance than just its maximal diversity gain or its maximal multiplexing gain alone. We give a simple characterization of the optimal diversity-multiplexing tradeoff achievable by any scheme and use it to evaluate the performance of many existing schemes. Our framework is useful for evaluating and comparing existing schemes as well as providing insights for designing new schemes.

APPENDIX

Proof of Theorem 4

From (11), we only need to prove

$$\begin{aligned} F(\text{SNR}) & \triangleq \int_{\mathcal{A}'} \prod_{i=1}^{\min\{m,n\}} \text{SNR}^{-(|m-n|+1)\alpha_i} \\ & \quad \prod_{i<j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 d\alpha \\ & \doteq \text{SNR}^{-d_{out}(r)} \end{aligned}$$

where

$$d_{out}(r) = \inf_{\alpha \in \mathcal{A}'} \sum_{i=1}^{\min\{m,n\}} (2i-1 + |m-n|)\alpha_i$$

and

$$\mathcal{A}' = \{\alpha : \alpha_1 \geq \dots \geq \alpha_{\min\{m,n\}} \geq 0, \sum_i (1 - \alpha_i)^+ < r\}$$

As before we can assume without loss of generality $m \geq n$.

We first derive an upper bound on $F(\text{SNR})$. Consider

$$\begin{aligned} F(\text{SNR}) & \leq \bar{F}(\text{SNR}) \\ & \triangleq \int_{\mathcal{A}'} \prod_{i=1}^n \text{SNR}^{-(|m-n|+1)\alpha_i} \prod_{i<j} (\text{SNR}^{-\alpha_j} - 0)^2 d\alpha \\ & = \int_{\mathcal{A}'} \prod_i \text{SNR}^{-(|m-n|+2i-1)\alpha_i} d\alpha \\ & = \int_{\mathcal{A}'} \text{SNR}^{-f(\alpha)} d\alpha \end{aligned}$$

where

$$f(\alpha) = \sum_i f_i(\alpha_i) = \sum_i (|m-n| + 2i-1)\alpha_i$$

Denote $\alpha^* = \arg \inf_{\mathcal{A}'} f(\alpha)$, we claim $\bar{F}(\text{SNR}) \doteq \text{SNR}^{-f(\alpha^*)}$, that is

$$\lim_{\text{SNR} \rightarrow \infty} \frac{\log \bar{F}(\text{SNR})}{\log \text{SNR}} = -f(\alpha^*) \quad (44)$$

To see that, first let $I = [0, mn]^n$ and consider

$$\begin{aligned} \bar{F}(\text{SNR}) & \leq \int_{\mathcal{A}' \cap I} \text{SNR}^{-f(\alpha)} d\alpha + \int_{I^c} \text{SNR}^{-f(\alpha)} d\alpha \\ & \leq \text{vol}[\mathcal{A}' \cap I] \text{SNR}^{-f(\alpha^*)} + \int_{I^c} \prod_{i=1}^n \text{SNR}^{-f_i(\alpha)} d\alpha \quad (45) \end{aligned}$$

For the second term, since $\alpha \notin [0, mn]^n$, we must have $\alpha_i > mn$ for some i . Without loss of generality, we assume $\alpha_1 > mn$. Now

$$\begin{aligned} & \int_{mn}^{\infty} \text{SNR}^{-f_1(\alpha_1)} d\alpha_1 \\ & \leq \int_0^{\infty} \text{SNR}^{-(|m-n|+1)mn} \text{SNR}^{-f_1(t)} dt \\ & \leq k \text{SNR}^{-mn} \end{aligned}$$

where $t = \alpha_1 - mn$ and k is a finite constant. Also notice $\int_0^{\infty} \text{SNR}^{-f_i(t)} dt < \infty$, we have

$$\begin{aligned} \bar{F}(\text{SNR}) & \leq \text{vol}[I] \text{SNR}^{-f(\alpha^*)} \\ & \quad + k \text{SNR}^{-mn} \prod_{i=2}^n \int_0^{\infty} \text{SNR}^{-f_i(\alpha_i)} d\alpha_i \\ & \leq \text{SNR}^{-f(\alpha^*)} \end{aligned}$$

To find a lower bound on $\bar{F}(\text{SNR})$, note that $f(\alpha)$ is continuous, therefore for any $\delta > 0$, there exists a neighborhood I of α^* , within which $f(\alpha) \geq f(\alpha^*) + \delta$. Now

$$\begin{aligned} \bar{F}(\text{SNR}) & \geq \int_{I \cap \mathcal{A}'} \text{SNR}^{-(f(\alpha^*)+\delta)} d\alpha \\ & = \text{vol}[I \cap \mathcal{A}'] \text{SNR}^{-(f(\alpha^*)+\delta)} \end{aligned}$$

Since $\bar{F}(\text{SNR}) \geq \text{SNR}^{-f(\alpha^*)+\delta}$ for any $\delta > 0$, we have $\bar{F}(\text{SNR}) \doteq \text{SNR}^{-f(\alpha^*)}$, which proves (44)

(44) says as $\text{SNR} \rightarrow \infty$, the integral $\int_{\mathcal{A}} \text{SNR}^{-f(\alpha)} d\alpha$ is dominated by the term corresponding to the minimum SNR exponent, $\text{SNR}^{-f(\alpha^*)}$. The above proof of (44) is in fact a special case of the *Laplace's method*, which is widely used in obtaining asymptotic expansions of special functions such as Bessel functions [15].

Now to derive a lower bound on $F(\text{SNR})$, for any $\delta > 0$, define the set

$$S_\delta = \{\alpha : |\alpha_i - \alpha_j| > \delta, \forall i \neq j\}$$

Now

$$\begin{aligned} F(\text{SNR}) & = \int_{\mathcal{A}'} \prod_{i=1}^n \text{SNR}^{-(|m-n|+1)\alpha_i} \\ & \quad \prod_{i<j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 d\alpha \\ & \geq \int_{\mathcal{A}' \cap S_\delta} \prod_i \text{SNR}^{-(|m-n|+1)\alpha_i} \\ & \quad \prod_{i<j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 d\alpha \\ & \geq \int_{\mathcal{A}' \cap S_\delta} \prod_i \text{SNR}^{-(|m-n|+1)\alpha_i} \\ & \quad \prod_{i<j} ((1 - \text{SNR}^{-\delta}) \text{SNR}^{-\alpha_j})^2 d\alpha \\ & = (1 - \text{SNR}^{-\delta})^{n^2} \int_{\mathcal{A}' \cap S_\delta} \text{SNR}^{-f(\alpha)} d\alpha \end{aligned}$$

Following the above argument $\int_{\mathcal{A}' \cap S_\delta} \text{SNR}^{-f(\alpha)} d\alpha$ has SNR exponent

$$\inf_{\mathcal{A}' \cap S_\delta} f(\alpha)$$

which, by the continuity of f , approaches $f(\alpha^*)$ as $\delta \rightarrow 0$.

Combining the upper and lower bound, we have the desired result.

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. On Telecommunications*, vol. 10, pp. 585–95, Nov/Dec 1999.
- [2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block code from orthogonal designs," *IEEE Trans. Info. Theory*, vol. 45, pp. 1456–67, July 1999.
- [4] G. Foschini, G. Golden, R. Valenzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE JSAC*, vol. 17, Nov. 1999.
- [5] R. Heath Jr. and A. Paulraj, "Switching between multiplexing and diversity based on constellation distance," in *Proc. of Allerton Conf. on Comm. Control and Comp.*, October 2000.
- [6] J.-C. Guey et al., "Signal designs for transmitter diversity wireless communication system over rayleigh fading channels," in *Proc. VTC'96*, pp. 136–40.
- [7] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Info. Theory*, vol. 44, pp. 744–65, 1998.
- [8] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE JSAC*, pp. 1451–8, October 1998.
- [9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *Submitted to IEEE Trans. Info. Theory*.
- [10] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.
- [11] L. Ozarow, S. Shamai, and A. Wyner, "Information-theoretic considerations in cellular mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 43, pp. 359–78, May 1994.
- [12] J. Proakis, *Digital Communications*. McGraw Hill, 4th ed.
- [13] J. G. D. Forney and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *IEEE Trans. Info. Theory*, 1998.
- [14] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, 1982.
- [15] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*. Jones and Bartlett Publishers, 1992.
- [16] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *Submitted to IEEE Trans. Info. Theory*, 2000.
- [17] H. Wang and X. Xia, "Upper bounds of rates of complex orthogonal space-time block codes," in *Proc. International Symposium on Information Theory*, June 2002.
- [18] N. Prasad and M. Varanasi, "Optimum efficiently decodable layered space-time block code," *Proc. Asilomar Conf. Signals, Systems, and Computers*, Nov. 2001.
- [19] M. K. Varanasi and T. Guess, "Optimal decision feedback multiuser equalization with successive decoding achieves the total capacity of the gaussian multiple-access channel," in *Proc. 31st Asilomar Conf. Singals, Systems, and Computers*, pp. 1405–9, Nov. 1997.

David N.C. Tse David Tse received the B.A.Sc. degree in systems design engineering from University of Waterloo, Canada in 1989, and the M.S. and Ph.D. degrees in electrical engineering from Massachusetts Institute of Technology in 1991 and 1994 respectively. From 1994 to 1995, he was a postdoctoral member of technical staff at A.T. & T. Bell Laboratories. Since 1995, he has been at the Department of Electrical Engineering and Computer Sciences in the University of California at Berkeley, where he is currently a Professor. He received a 1967 NSERC 4-year graduate fellowship from the government of Canada in 1989, a NSF CAREER award in 1998, the Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, and the IEEE Communications and Information Theory Society Joint Paper Award in 2001. He is currently an Associate Editor for the IEEE Transactions on Information Theory. His research interests are in information theory, wireless communications and networking.

Lizhong Zheng Lizhong Zheng received the B.S. and M.S. degrees in 1994 and 1997 respectively, both from the Department of Electronic Engineering, Tsinghua University, Beijing, China; and the Ph.D. degree in 2002, from the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA. He is currently an assistant professor in the Department of Electrical Engineering and Computer Sciences at Massachusetts Institute of Technology. His research interests includes wireless communications, networks, and information theory.