Optimal Diversity-Multiplexing Tradeoff in Multiple Antenna Channels *

Lizhong Zheng David N.C. Tse
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720, USA
{lzheng,dtse}@eecs.berkeley.edu

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. We give a simple characterization of this tradeoff and use it to evaluate the performance of existing multiple antenna schemes.

1 Introduction

Multiple antennas are an important means to improve the performance of wireless systems. Traditionally, multiple antennas have been used to increase *diversity* to combat channel fading. By having multiple independently faded replicas at the receiver for each data symbol, more reliable reception is achieved. For example, to get a diversity gain (advantage) of d in a slow Rayleigh fading environment, one can use d *receive* antennas: it is well known that the average error probability can be made to decay like $1/\mathsf{SNR}^d$ at high SNR. More recent work has concentrated on coding for *transmit* diversity (some examples are space-time codes [4, 10] and orthogonal designs [1, 11]).

Transmit or receive diversity is a means to *combat* fading. A different line of thought suggests that in a system with multiple transmit *and* receive antennas (MIMO channel), fading can in fact be *beneficial* through increasing the *degrees of freedom* available for communication [2, 12]. Essentially, if the path gains between individual transmit-receive antenna pair fade independently, the channel matrix is well-conditioned with high probability, in which case multiple parallel *spatial channels* are created. This effect is also called *spatial multiplexing* [6], and is particular important in the high signal-to-noise (SNR) regime where the system is degree-of-freedom-limited (as opposed to energy-limited). Foschini [2] has shown that in the high SNR regime, the capacity of a channel with M

*This research is supported by a National Science Founda ion Early Faculty CAREER Award, with matching grants from A.T.&T., Lucent Technologies and Qualcomm Inc., and by a DARPA grant F30602-97-2-0346

transmit, N receive antennas and i.i.d. Rayleigh faded gains between each pair is given by:

$$C(\mathsf{SNR}) = \min\{M, N\} \log \mathsf{SNR} + O(1).$$

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

Most of current research focus on designing schemes to extract maximal diversity gain *or* maximal spatial multiplexing gain. In this paper, we put forth a new 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. We give a simple characterization of the tradeoff curve and use it as a unified framework to evaluate the performance of many existing diversity-based and multiplexing-based schemes.

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 $r \log \mathsf{SNR}$ and the average error probability decays like $1/\mathsf{SNR}^d$. The tradeoff curve yields for each multiplexing gain r the optimal diversity advantage $d^*(r)$ achievable by *any* scheme. Clearly, r cannot exceed the number of degrees of freedom in the channel and $d(r)$ cannot exceed the maximal diversity gain. For several well-known schemes, we compute the curves $d(r)$ and compare it to the optimal tradeoff curve.

The rest of the paper is outlined as follows. Section 2 presents the system model and the precise problem formulation. Our main result on the optimal diversity-multiplexing tradeoff curve is presented in Section 3. We establish a connection between our formulation with the outage capacity formulation in Section 4 and with the theory of error exponents in Section 5. We compare the performance of several schemes with this optimal tradeoff curve in Section 6. Section 7 contains the conclusions. The proofs of all the results presented here can be found in the journal version of the paper.

2 System Model and Problem Formulation

We consider the wireless link with M transmit and N receive antennas. The fading coefficient $\mathbf{h}_{ij}$ is the 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 focus on a single block of T symbols, within which the channel is assumed to remain constant. Within this block, the channel can be written as:

$$\mathbf{Y} = \sqrt{\frac{\mathsf{SNR}}{M}} \mathbf{H}\mathbf{X} + \mathbf{W} \tag{1}$$

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

The transmitted signal $\mathbf{X}$ is normalized to have the average transmitted power at each antenna in each symbol period to be 1. The power constraint on a codebook $\mathcal{C}$ can be written as

$$\frac{1}{|\mathcal{C}|}\sum_{i}\sum_{m=1}^{M}\sum_{t=1}^{T}\|\mathbf{x}_{mt}(i)\|^2 \le MT,$$

where the first sum is over all the codewords in the codebook. We now consider a family of codes $\{\mathcal{C}(\mathsf{SNR})\}$ of block length T, one at each SNR level, and define its spatial multiplexing gain and diversity advantage in the high SNR regime. Let $R(\mathsf{SNR})$ bits/symbol and $p_e(\mathsf{SNR})$ be the rate and average error probability of the code $\mathcal{C}(\mathsf{SNR})$ respectively.

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

$$R(\mathsf{SNR})/\log \mathsf{SNR} \to r$$

and average error probability

$$\limsup_{\mathsf{SNR}\to\infty} \frac{\log p_e(\mathsf{SNR})}{\log \mathsf{SNR}} = -d \tag{2}$$

For each r*, define* $d^*(r)$ *to be the supremum of the diversity advantage achieved by any scheme.*

For brevity, we will use the notation

$$p_e(\mathsf{SNR}) \doteq 1/\mathsf{SNR}^d$$

to denote

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_e(\mathsf{SNR})}{\log \mathsf{SNR}} = -d$$

The error probability $p_e(\mathsf{SNR})$ is averaged over the additive noise $\mathbf{W}$, the channel matrix $\mathbf{H}$ and the transmitted codewords (assumed equally likely). The definition of diversity advantage here differs from the standard definition in the space-time code literature (see for example [10]) 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, the code rate is *fixed* and the diversity gain describes how fast the error probability decays with increasing SNR. Here, the code rate also *increases* with SNR simultaneously. This is natural if one thinks of spatial multiplexing as achieving a *nonvanishing fraction* of the degrees of freedom in the channel. According to this definition, any fixed-rate scheme has a zero multiplexing gain.

3 Main Result

The main result of this paper is an exact characterization of the optimal tradeoff $d^*(r)$ for the case when the block length $T \geq M + N - 1$.

Theorem 2 *Assume $T \geq M+N-1$. Let $K = \min\{M, N\}$. The optimal tradeoff $d^*(r)$ is given by the piecewise linear function connecting the points $(K-k, d^*(K-k)), k = 0, \ldots, K$, where*

$$d^*(K-k) = k^2 + k|M-N| \tag{3}$$

Moreover, for each r, there exists a scheme $\{\mathcal{C}(\mathsf{SNR})\}$ with spatial multiplexing gain r and

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_e(\mathsf{SNR})}{\log \mathsf{SNR}} = -d^*(r),$$

i.e. the limit defined in (2) actually exists.

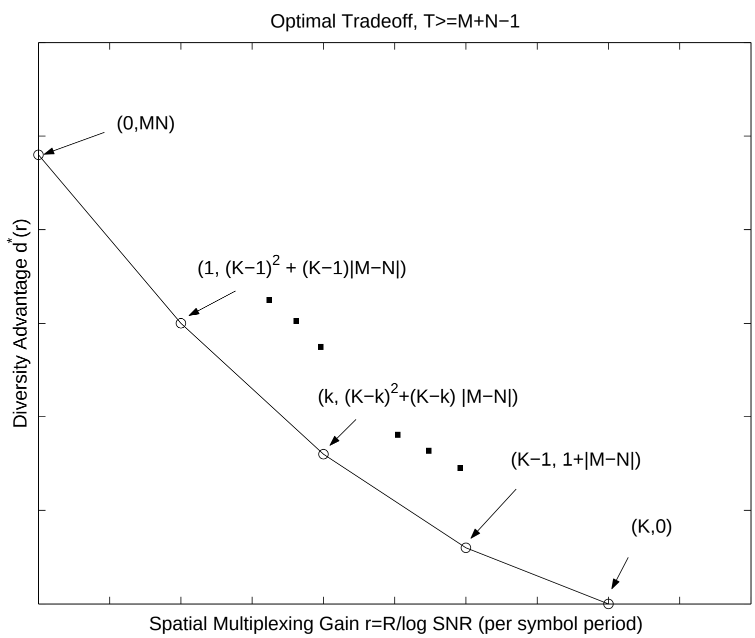

Figure 1: Diversity-multiplexing tradeoff, $d^*(r)$ for general M, N and $T \geq M+N-1$.

The function $d^*(r)$ is plotted in Figure 1. The curve intersects the r axis at $K = \min\{M, N\}$, which is the number of degrees of freedom in the channel. This is the maximal spatial multiplexing gain that can be obtained, but at this point there is no diversity advantage. Increasing the diversity advantage comes at a price of decrease in spatial multiplexing gain. The maximum diversity gain is MN, achievable by a fixed-rate scheme (i.e. zero multiplexing gain). This is also the number of random channel gains in the channel, and can be interpreted as the maximum number of independent random variables that a diversity scheme can average over.

The tradeoff curve for the special case $M = N$ is shown in Figure 2. In this case, the maximum diversity gain for a given multiplexing gain k (k integer) is simply given by $(N-k)^2$. Put it another way, to achieve a given diversity advantage of d (d perfect

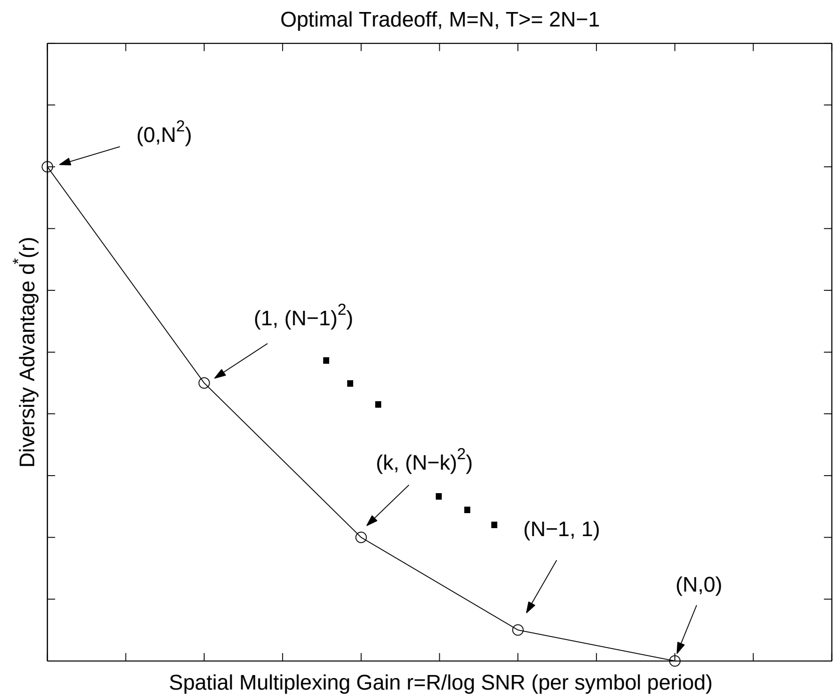

Figure 2: Diversity-multiplexing tradeoff for $M = N$ and $T \geq 2N - 1$.

square), the maximum spatial multiplexing gain achievable is $N - \sqrt{d}$. Since N is the total number of available degrees of freedom in the channel, this reveals exactly the price one must have to pay to achieve the desired level of diversity.

This result concerns the performance on transmission of a single block of length T. One can prove an analogous result for the case when one codes over L such blocks, each of which fades independently. This would model the case when antenna diversity is combined with other forms of diversity, such as over time or frequency. Consider now such schemes with rate scaling as $r \log \mathsf{SNR}$ bits/symbol and error probability decaying as $1/\mathsf{SNR}^{d_L(r)}$.

Theorem 3 *For schemes which code over L i.i.d. faded blocks, each of length $T \geq M + N - 1$, the optimal diversity advantage $d_L^*(r)$ for a given multiplexing gain of r is given by $d_L^*(r) = Ld^*(r)$ where $d^*(r)$ is given by Theorem 2.*

This means that the diversity gain simply adds across the L blocks.

4 Connection to Outage Capacity Formulation

There is an interesting and insightful connection between Theorem 2 and the outage capacity formulation, proposed in [7] for fading channels and applied to multi-antenna channels in [12]. This formulation is used to analyze non-ergodic situations where the fading channel, although random, remains fixed for all time (modeling the slow fading scenario). Following [12], for a given target data rate R, the outage probability

$$p_{\text{out}}(R) = \inf_{Q \geq 0, tr(Q) \leq M} P\left[\text{logdet}\left(I + \frac{\mathsf{SNR}}{M}\mathbf{H}Q\mathbf{H}^\dagger\right) < R\right]$$

is the probability that the mutual information per symbol falls below the target rate R, under the minimizing input distribution. Here, the probability is taken over the random channel $\mathbf{H}$. Note that without loss of generality the input distribution can be taken to be Gaussian, and the minimization is over the covariance matrix of the Gaussian distribution.

We have the following asymptotic result on the outage probability at high SNR.

Theorem 4 *If the target rate is scaled to be $R = r \log \mathsf{SNR}$, then*

$$p_{\text{out}}(r \log \mathsf{SNR}) \doteq \mathsf{SNR}^{-d^*(r)},$$

where $d^*(r)$ is precisely the same function as defined in Theorem 2. Moreover $Q^* = I$ is an asymptotically optimal input distribution at high SNR.

The proof of this result involves an analysis based on the distribution of the random eigenvalues of $\mathbf{HH}^\dagger$.

Theorem 4 provides a context for interpreting Theorem 2. If one believes that the outage formulation captures the performance under infinite coding block length, what Theorem 2 says is that as long as the block length $T \geq M+N-1$, the infinite block length performance is already achieved. This is because the tradeoff curve given in Theorem 2 actually does *not* depend on T as long as $T \geq M+N-1$. Thus, one cannot get more diversity gain by coding over a longer block length than $M+N-1$.

To prove Theorem 2, we show that the channel outage event is the dominant error event. We show that there exists codes which yield good performance (i.e. error probability much smaller than the outage probability) whenever the channel is not in outage. To prove this, we use a random coding argument.

The condition that the block length $T \geq M+N-1$ is not a technical one;if $T < M+N-1$, there are examples in which the outage performance cannot be achieved. The results for this case are more involved and can be found in the journal version of the paper.

5 Connection to Error Exponents

This is also an intimate connection of our results to the theory of error exponents (see for example [3]). Consider now codes over a single block of length T symbols and have rate R bits/symbol. We can think of these T symbols as one super-symbol. Applying the theory of error exponents to a block of length one such supersymbol, it says that the average error probability of the optimal code is upper bounded by $2^{-E(R)}$, where the random coding exponent $E(R)$) is given by:

$$E(R) = \max_{0 \leq \rho \leq 1} [E_0(\rho) - \rho RT],$$

and $E_0(\rho)$ is given by

$$E_0(\rho) = \sup_{q_X} -\log \mathcal{E}_H \left[\int \left[\int q_X(X) p(Y|X, \mathbf{H})^{1/(1+\rho)} dX \right]^{1+\rho} dY \right],$$

and the supremum is over all input distributions q_X on the N by T matrix $\mathbf{X}$. (The calculation of this follows almost directly from [3]; see also [12] which did a similar calculation for the case when $T = 1$.)

The following theorem connects the optimal tradeoff curve calculated in Theorem 2 and the random coding error exponent.

Theorem 5 *Suppose $T \geq M+N-1$. If the rate R scales as $r \log \mathsf{SNR}$ bits/symbol, then for any $r \geq 0$,*

$$\lim_{\mathsf{SNR}\to\infty} E(r \log \mathsf{SNR}) / \log \mathsf{SNR} = d^*(r),$$

where $d^*(r)$ is given in Theorem 2. Moreover the i.i.d. Gaussian input distrbution q_X is asymptotically optimal in achieving the error exponent at high SNR.

This theorem says that the tradeoff curve of r versus $d^*(r)$ is actually a scaled version of the plot R versus the error exponent $E(R)$, with both axes scaled by $1/\log \mathsf{SNR}$. The optimality of $d^*(r)$ implies that the random coding exponent not only yields an upper bound to the error probability, but is actually *tight* under the high SNR scaling considered.

Several comments are in order:

- For finite SNR level, the optimal input distribution is *not* Gaussian, but should be concentrated on a spherical shell [3]. Our result however says that at the high SNR limit, Gaussian distribution does asymptotically as well.
- The tightness of the random coding error exponent is usually discussed in the regime of long *block lengths*. Here, we are actually fixing the block length to be one super-symbol, but taking the high SNR limit. The tightness is in that regime.
- Even in the regime of long block lengths, the random coding exponent is in general not tight for all rates, but only rates above a critical rate R_{crit}. Our result says however that for $T \geq M+N-1$, the random coding error exponent is tight for *all* multiplexing gain r (albeit in a differnt asymptotic regime.) It is interesting that , for $T \geq M+N-1$, it can in fact be shown that $R_{\text{crit}}/\log \mathsf{SNR}$ approaches 0.

Another common upper bound for the average error probability of a random ensemble of codes is the *union bound*. The union bound is simply the number of codewords times the average error probability between a pair of random codewords. Under exactly the same scaling considered for the random coding exponent in Theorem 5, the exponent associated with the union bound also approaches a limit, as plotted in Figure 5. Under the assumption that $T > M+N-1$, we observe that compared to the optimal bound $d^*(r)$, the union bound is quite loose except for $r = 0$. This strongly suggests that to get significant multiplexing gain, a code design criterion based on pairwise error probability is not adequate. In fact, it can be shown that for the codes achieving the optimal curve $d^*(r)$, the typical error event is due to confusion with one of many codewords far away from the transmitted codeword, rather than confusion with one of the few nearest neighbors.

6 Evaluation of Existing Schemes

The diversity-multiplexing tradeoff can be used to evaluate and compare performance of many existing schemes. We only discuss a few representative examples here.

We first consider the Alamouti scheme [1] (also called orthogonal design [11]) for $M = 2$ transmit antennas. In this scheme, two symbols $\mathbf{x}_1, \mathbf{x}_2$ are transmitted over two symbol periods through the channel as follows:

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

In our framework, we view 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

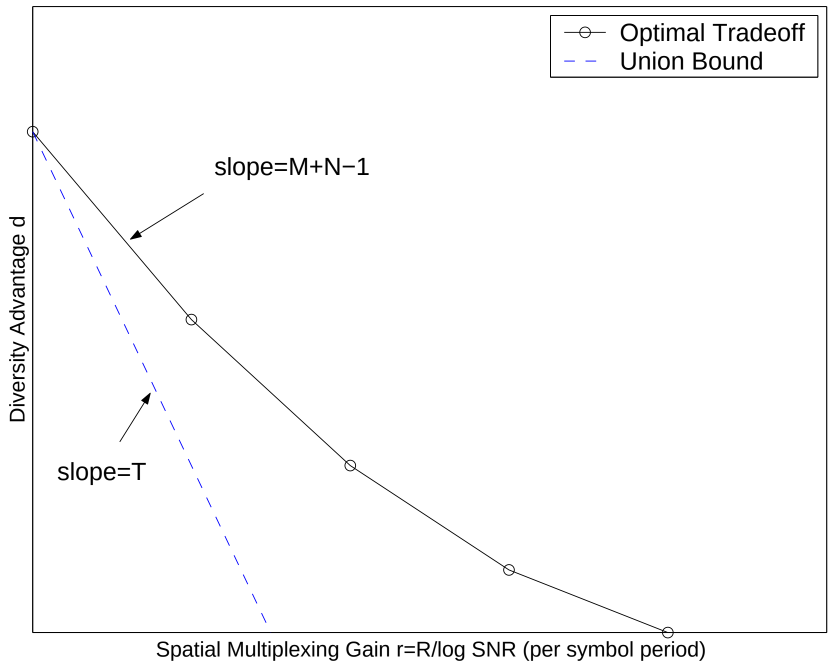

Figure 3: Comparison to Union Bound

as $R = r \log \mathsf{SNR}$. The tradeoff curve for Alamouti scheme can be computed for the best outer code (or, more precisely, best family of outer codes). This is shown in Fig. 4 for two cases: $N = 1$ receive antenna and $N = 2$ receive antennas. For the case of $N = 1$ and $T \geq 2$ (T even), Alamouti scheme is optimal, in the sense that it achieves the tradeoff curve $d^(r)$ for all r. For the case of $N = 2$ and $T \geq 4$, Alamouti scheme is sub-optimal: it achieves the maximum diversity gain of 2 but at all other points its tradeoff curve $d(r)$ is strictly below the optimal tradeoff curve $d^*(r)$. The fact that Alamouti scheme does not achieve the full degrees of freedom have already been pointed out [5]; this corresponds in our framework to the fact that $d(1) = 0$. Our results give a stronger conclusion here: the diversity-multiplexing tradeoff is strictly sub-optimal for all r other than $r = 0$.

Next we look at the BLAST schemes. We focus on one variant, D-BLAST [2]. We plot the performance for two versions of this scheme in Fig. 5, for the case when the block length $T = \infty$. and $M = N$. The version originally proposed in [2] is based on *nulling out* the uncancelled layers and the tradeoff curve is the lower curve, which is strictly sub-optimal. If we replace the nulling step by a linear MMSE receiver, on the other hand, we get the upper curve, which is precisely the optimal tradeoff curve $d^*(r)$. Thus, *for infinite block length, MMSE D-BLAST is optimal , but Nulling D-BLAST is not.* This conclusion is quite surprising, since in the high SNR regime, one would expect the MMSE receiver and the decorrelator (nuller) to have similar performance. The reason is somewhat subtle and is explained in the full paper.

For the case when the block length $T < \infty$, MMSE D-BLAST is strictly sub-optimal due to the loss in degrees of freedom in the first few symbols for the initialization of the procedure. As the block length T becomes large, this loss is amortized and becomes negligible for long block lengths.

Another variant of BLAST, V-BLAST, turns out to have a totally different tradeoff curve. Its analysis is more intricate and will be presented in the full paper.

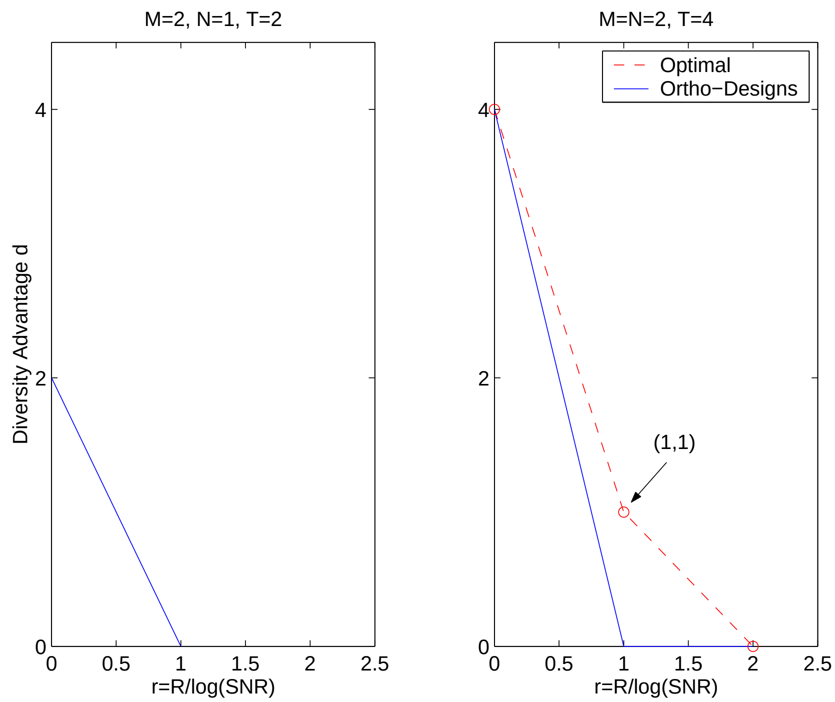

Figure 4: Diversity-multiplexing tradeoff for Alamouti scheme, $N = 1$ and $N = 2$.

7 Conclusions

Previous research on multi-antenna coding schemes has focused either on extracting maximum diversity gain or maximum spatial multiplexing gain from the channel. In this paper, we present a framework in which both diversity and spatial multiplexing gain have equal footing in the picture. We characterize the fundamental tradeoff between the two types of gains that can be extracted from a given MIMO channel. This framework is useful for evaluating and comparing existing schemes as well as providing insights for designing new schemes.

References

[1] S. Alamouti," A simple transmitter diversity scheme for wireless communications", *IEEE JSAC*, Oct 1998, pp. 1451-1458.

[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] R.G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

[4] J-C Guey et al., "Signal Designs for transmitter diversity wireless communication system over Rayleigh fading channels", Proc. VTC'96, pp. 136-140.

[5] B. Hassibi and H. Hochwald, "High-Rate Codes that are Linear in Space and Time", submitted to IEEE Trans. on Information Theory, August 2000.

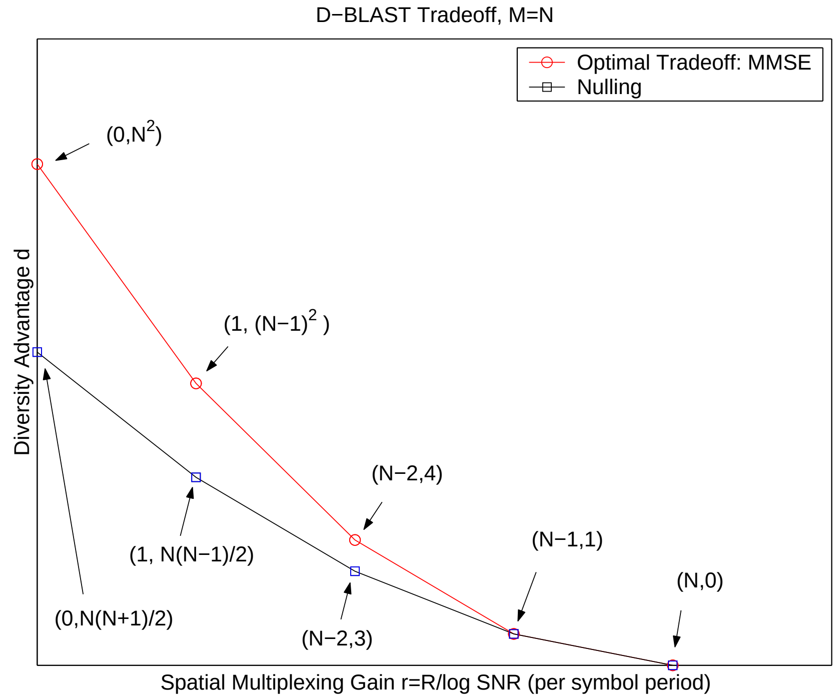

Figure 5: Diversity-multiplexing tradeoff for D-BLAST (Nulling) and D-BLAST (MMSE), block length $T = \infty$.

[6] R.W. Heath Jr and A.J. Paulraj, "Switching between multiplexing and diversity based on constellation distance", *Proc. of Allerton Conf. on Comm. Control and Comp.*, October 2000.

[7] Ozarow L. H., Shamai S., Wyner A. D. Information-theoretic considerations for cellular mobile radio. *IEEE Trans. on Vehicular Technology*, 43(2):359-378, May, 1994.

[8] A.J. Paulraj and T. Kalaith, "Increasing capacity in wireless broadcast systems using distributed transmission/directed reception, ", U.S. Patent, 1994.

[9] J. G. Proakis, *Digital Communications*, 2nd Edition, McGraw Hill, 1989.

[10] V. Tarokh, N. Seshadri and A.R. Calderbank,"Space-time coding for high data rate wireless communication: Performance analysis and code construction", *IEEE Trans. on Info. Theory*, v. 44, Mar 1998, pp. 744-765.

[11] V. Tarokh, H. Jafarkhani and A.R. Calderbank, "Space-time block codes from orthogonal designs", *IEEE Trans. on Info. Theory*, v.45, July 1999, pp. 1456-1467.

[12] I. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T Bell Laboratories internal Technical Memorandum*, 1995.