Approximately Universal Codes Over Slow-Fading Channels

Saurabha Tavildar and Pramod Viswanath, *Member, IEEE*

***Abstract*—Performance of reliable communication over a coherent slow-fading multiple-input multiple-output (MIMO) channel at high signal-to-noise ratio (SNR) is succinctly captured as a fundamental tradeoff between diversity and multiplexing gains. This paper studies the problem of designing codes that optimally tradeoff the diversity and multiplexing gains. The main contribution is a precise characterization of codes that are *universally* tradeoff-optimal, i.e., they optimally tradeoff the diversity and multiplexing gains for *every* statistical characterization of the fading channel. This characterization is referred to as *approximate universality*; the approximation is in the connection between error probability and outage capacity with diversity and multiplexing gains, respectively. The characterization of approximate universality is then used to construct new coding schemes as well as to show optimality of several schemes proposed in the space–time coding literature.**

***Index Terms*—Compound channel, diversity-multiplexing tradeoff, fading channel, multiple-input multiple-output (MIMO), space–time codes, universal codes.**

I. Introduction

RELIABLE communication over *slow* fading point-to-point channels, where the (random) channel realization is fixed over the time scale of communication, is characterized by the tradeoff between data rate and error probability: typical fading distributions have a nonzero probability of being very small and thus arbitrarily reliable communication is not possible at any nonzero rate. The tradeoff between the data rate and the error probability is captured by the outage capacity, the largest rate of reliable communication for a fixed error probability. The information theoretic view is that of a *compound* channel: the slow fading channel is composed of a class of channels parameterized by the different channel realizations that are not in outage. The outage capacity is achieved by *universal* codes, those that work reliably over *every* one of the channel realizations not in outage.

At high signal-to-noise ratio (SNR), the precise (but too involved to derive code design principles) tradeoff between error probability and data rate is coarsely captured in terms of a tradeoff between diversity and multiplexing gains [1]: these are the rate of decay of error probability and the increase of data rate with increasing SNR. Since the tradeoff is captured at a coarser scale, we shall denote codes that optimally tradeoff diversity and multiplexing gains for *every* slow fading channel as *approximately universal*; the approximation here refers to the coarseness in the definition of diversity and multiplexing gains as opposed to studying error probability and data rate directly. Our main result is a *precise* characterization of approximately universal codes. We use this characterization to show the approximate universality of some codes proposed in the literature and to also construct new space–time codes that are approximately universal. These codes are robust to statistical channel modeling errors, hence their engineering appeal is clear. This approach of using compound channel viewpoint to construct robust codes for multiple-input multiple-output MIMO channels has also been taken in a series of works in [2]–[4].

We are interested in codes that achieve reliable communication over all channel realizations not in outage: this suggests, as done in [3], asking for the performance of the code for the *worst* channel not in outage. This is in contrast to the traditional performance analysis where the error probability is *averaged* over the statistics of the fading channel. In particular, if the worst-case pairwise error probability decays *exponentially* with increasing SNR then such a code is approximately universal. For a parallel channel, the worst channel for a given pair of codewords is “inverse waterfilling” over the pairwise squared codeword differences. For a MIMO channel, the worst channel (derived in [3]) aligns its singular vectors in the same directions as those of the pairwise codeword difference matrix and then the singular values inverse waterfill the singular values of the pairwise codeword difference matrix. While the exact expression of the worst-case pairwise error is somewhat involved, a simple worst-case code design criterion emerges at high SNR for both the parallel channel and the MIMO channel.

For a parallel channel, somewhat surprisingly, the worst-case code design criterion at high SNR simplifies to the product distance criterion which was derived initially for the independent and identically distributed (i.i.d.) Ricean-fading channel [5], though is better known for the i.i.d. Rayleigh-fading channel (see [6, Ch.3]). In a compound channel setting the criterion was heuristically derived in [2], here we give a more precise statement for the criterion. In particular, we show that if the products of all normalized squared codeword differences is larger than 2^{-R} where R is the communication rate, then the code is approximately universal. This design criterion suggests a class of codes based on permutations of the quadrature amplitude modulation (QAM) constellation that we call *permutation codes*. Even random permutation codes are approximately

Manuscript received February 24, 2005; revised December 6, 2005. This work was supported in part by the National Science Foundation under Grants NSF CAREER 0237549 and NSF ITR 0325924, by a Vodafone graduate fellowship, and by Motorola Inc. The material in this paper was presented in part at the Conference on Information Systems and Sciences, Princeton, NJ, March 2004.

The authors are with the Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: tavildar@uiuc.edu; pramodv@uiuc.edu).

Communicated by M. Médard, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2006.876226

universal and we provide examples of simple and explicit permutation codes that are approximately universal. We show that a code based on a rotated QAM constellation proposed in the literature [7] also satisfies the desired product distance property and is hence approximately universal.

For a MIMO channel, the worst-case code design criterion is in general not simply to maximize the determinant of the codeword difference matrix, the criterion derived for the i.i.d. Rayleigh-fading channel [8]. This can be explicitly seen in the case of the multiple transmit but single receive antenna (MISO) channel: the worst channel chooses the most susceptible direction to confuse between a pair of codeword matrices–this is the direction of the smallest singular value of the codeword difference matrix. Thus the worst-case code design criterion for the MISO channel is to maximize the *smallest* singular value of the codeword difference matrix; different from the determinant criterion derived for the i.i.d. Rayleigh-fading channel. More generally, the worst-case code design criterion at high SNR for a MIMO channel (with n_t transmit and n_r receive antennas) is to maximize the product of the smallest $\min(n_t, n_r)$ singular values of the codeword difference matrix. With more receive than transmit antennas, the worst-case code design criterion reduces to the determinant criterion derived for the i.i.d. Rayleigh-fading channel.

An important implication of our worst-case code design criterion is the following: if a code is approximately universal on an $n_t \times n_t$ MIMO channel, then it is also approximately universal for $n_t \times n_r$ MIMO channel for *every* n_r. Several space–time codes proposed in the literature satisfy the worst-case code design criterion and are hence approximately universal. In particular, the QAM rotation codes in [7], [9] are approximately universal for every MIMO channel with two transmit antennas. The recently proposed codes in [10]–[12] that are derived from cyclic division algebra are also approximately universal. In fact, it follows from the results in [10] that explicit approximately universal codes exist for the *shortest* theoretically possible block-length $(= n_t)$ for *every* $n_t \times n_r$ MIMO channel.

V-BLAST [13] and D-BLAST [14] are classical architectures for communication over a MIMO channel. While they are not approximately universal, we show that they are tradeoff optimal in some rate regime universally over a (restricted) class of channels which are rotationally invariant. In particular, this class of channels includes the i.i.d. Rayleigh-fading channel: we show that V-BLAST with simple QAM constellations as the independent data streams achieves the last segment of the tradeoff curve for the $n \times n$ i.i.d. Rayleigh-fading MIMO channel and D-BLAST achieves the first segment of every $n_t \times 2$ i.i.d. Rayleigh fading MIMO channel. These results are illustrated in the context of a 2×2 i.i.d. Rayleigh fading MIMO channel in Fig. 1.

We have organized this paper into two distinct parts: first, we present a precise characterization of approximate universal codes for the general MIMO channel. In the second part, we discuss explicit approximately universal codes, starting with simpler channel models and moving on to the more involved ones. In particular, we start with the scalar channel and show that a simple QAM is approximately universal (this is done in Section IV). Next, we study the parallel channel and the MISO channel in Sections V and VI, respectively. Finally we consider the general MIMO channel in Section VII by demonstrating the approximately universality of some codes proposed in the literature, and then analyzing the approximately universal performance of V-BLAST and D-BLAST in Sections VIII and IX, respectively.

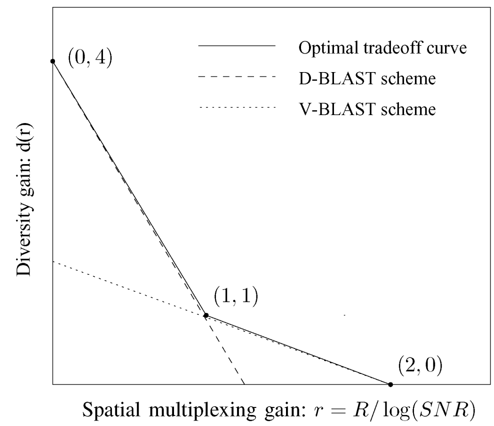

Fig. 1. Tradeoff curves: $n_t = n_r = 2$.

II. Channel Model and the Outage Formulation

The main focus of this paper is on the slow-fading (point-to-point) MIMO channel

$$\mathbf{y}[m] = \mathbf{H}\mathbf{x}[m] + \mathbf{w}[m] \tag{1}$$

where m is the time index and $\mathbf{y}$ and $\mathbf{x}$ denote the output and the input vectors respectively. The complex $n_r \times n_t$ matrix $\mathbf{H}$ of fading gains is randomly picked, but stays constant over the time-scale of communication; we suppose that the exact realization of $\mathbf{H}$ is known at the receiver. The additive noise $\mathbf{w}$ has i.i.d. complex Gaussian $(\mathcal{CN}(0,1))$ entries. We are interested in one-shot communication over this channel over a (small) length of time T. There is a transmit power constraint of $Tn_t\mathsf{SNR}$ for any transmit codeword of length T.

In this paper, we focus on the scaling at high SNR introduced in [1]: the data rate is measured on a scale of $\log \mathsf{SNR}$ and the decay rate of error probability is measured on a scale which is a negative exponent of SNR. All logarithms in this paper are to the base 2. More precisely, the *multiplexing* and *diversity* gains are defined as follows. A sequence of coding schemes (sequence in SNR) achieves a multiplexing rate of r and diversity gain of d if

$$\lim_{\mathsf{SNR}\to\infty} \frac{R(\mathsf{SNR})}{\log \mathsf{SNR}} = r, \quad \text{and}$$
$$\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}_e(\mathsf{SNR})}{\log \mathsf{SNR}} = -d$$

where $R(\mathsf{SNR})$ is the rate of the scheme and $\mathbb{P}_e(\mathsf{SNR})$ is the probability of error with maximum-likelihood (ML) decoding for the scheme. For a given multiplexing gain r, the largest diversity gain supported by any coding scheme is denoted by $d^*(r)$. The goal is to find a characterization of this optimal diversity-multiplexing tradeoff, $d^*(r)$, for any correlated channel and then to find (simple) coding schemes with as small a block length (T) as possible that achieve this optimal tradeoff curve.

The outage event turns out to be closely related to the problem of characterizing $d^*(r)$. It is defined as the set of channel realizations for which the mutual information is below the data rate

$$\{\mathbf{H} : I(\mathbf{x}; \mathbf{H}\mathbf{x} + \mathbf{w} \,|\, \mathbf{H}) < R\} \tag{2}$$

where the input distribution is independent of the realization of $\mathbf{H}$. It is shown in [1] that, in the scale of interest, the input distribution $\mathbb{P}_{\mathbf{x}}$ can be taken to be i.i.d. complex Gaussian for the Rayleigh fading channel; a similar argument for any fading distribution shows that the input distribution can be taken to be i.i.d. complex Gaussian. This means that the outage curve can be defined as shown in (3) at the bottom of the page.

The outage curve $d_{\text{out}}(r)$ is an upper bound to $d^*(r)$ [1]. On the other hand, the set of channel realizations that are not in outage constitute a compound channel, the capacity of which is $r \log \mathsf{SNR}$. The compound channel coding theorem guarantees the existence of universal codes: codes that achieve reliable communication over *every* MIMO channel realization that is not in outage. This means, that by coding over possibly long block lengths, one can actually achieve the outer bound of $d_{\text{out}}(r)$. Therefore for the rest of this paper, *we identify the outage curve with the optimal diversity-multiplexing tradeoff curve*. Note that, we are mainly interested in fading distributions such that the eigen-values are not bounded away from zero (e.g., AWGN channel can be considered as a fading channel). Otherwise, the outage curve will be infinite, and an approximately universal code will achieve it. But, the diversity-multiplexing tradeoff is not the right setup to study this problem.

We are interested in universal codes that achieve the upper bound of $d_{\text{out}}(r)$ only to the extent that they are tradeoff-optimal; we call such codes approximately universal. Our main focus is on a characterization of approximately codes with *small block-length*.

III. Main Result

Our main result is a precise characterization of approximately universal codes. We define n_m to be $\min(n_r, n_t)$.

Theorem 3.1: A sequence of codes of rate $R(\mathsf{SNR})$ bits/symbol is approximately universal over the MIMO channel if and only if, for every pair of codewords

$$\lambda_1^2 \lambda_2^2 \cdots \lambda_{n_m}^2 \geq \frac{1}{2^{R(\mathsf{SNR}) + o(\log(\mathsf{SNR}))}} \tag{4}$$

where $\lambda_1, \ldots, \lambda_{n_m}$ are the smallest n_m singular values of the normalized (by $\sqrt{\mathsf{SNR}}$) codeword difference matrix for a pair of codewords in $R(\mathsf{SNR})$.

For $n_r \geq n_t$, (4) turns out to be the same as the "nonvanishing determinant" criterion introduced in the context of i.i.d. Rayleigh-fading channels in [15]. This criterion was also studied in [7][10], also in the context of i.i.d. Rayleigh-fading channels. In [7], it was shown that for two transmit antennas, if a code satisfies this nonvanishing determinant criterion, then it is tradeoff-optimal for the i.i.d. Rayleigh-fading channel.

Our result is much stronger: for an $n_t \times n_r$ MIMO channel, if a code satisfies the nonvanishing determinant criterion, then it is tradeoff-optimal for *every* fading distribution. Thus, our result gives the well-known determinant criterion a precise *operational* interpretation in terms of approximate universality. Through this characterization, we will see that codes with small block lengths can be approximately universal. We start with a few implications of this criterion and then prove the sufficiency part of the criterion. The necessity part is proved in Appendix I.

A. Approximately Universal Codes in the Downlink

Some interesting observations follow from our characterization of approximately universal codes.

- If a code is approximately universal over an $n_t \times n_r$ MIMO channel with $n_r \geq n_t$, i.e., the number of receive antennas is equal to or larger than the number of transmit antennas, then it is also approximately universal for an $n_t \times l$ MIMO channel with $l \geq n_t$.
- The singular values of the normalized codeword difference matrices are upper bounded by a fixed number $(\sqrt{n_t T})$. Thus, a code that is approximately universal over an $n_t \times n_r$ MIMO channel is also approximately universal over an $n_t \times l$ MIMO channel with $l \leq n_r$.
- Consider the downlink of a cellular system where the base stations are equipped with multiple transmit antennas. Suppose we want to broadcast common information to all the users in the cell. We would like our transmission scheme to not depend on the number of receive antennas at the users: each user could have a different number of receive antennas, depending on the model, age, and type of the mobile device. Universal MIMO codes provide an attractive solution to this problem. Suppose we broadcast the common information at rate R using an approximately universal space time code over an $n_t \times n_t$ MIMO channel. Since this code is approximately universal for every $n_t \times n_r$ MIMO channel, the diversity seen by each user is *simultaneously* the best possible at rate R. To summarize: the diversity gain obtained by each user is the best possible with respect to both
  — the number of receive antennas the user has;
  — the statistics of the fading channel the user is currently experiencing.

B. Characterization of Approximately Universal Codes

Toward our goal of characterizing approximately universal codes, we first calculate the pairwise error probability for a pair of codewords based on the *worst* channel realization not in outage, i.e., we consider the realization (not in outage) as a *function* of the specific pair of codewords so as to yield the worst

$$d_{\text{out}}(r) := -\lim_{\mathsf{SNR}\to\infty} \frac{\log \mathbb{P}\{\mathbf{H} : \log\det(\mathbf{I} + \mathsf{SNR}\mathbf{H}\mathbf{H}^*) < r \log \mathsf{SNR}\}}{\log \mathsf{SNR}}. \tag{3}$$

pairwise error probability. If this worst-case pairwise error probability decays exponentially with SNR for every pair of codewords (we allow the worst channel to change as a function of the pair of codewords), then a simple union bound argument shows that the error probability conditioned on the channel realization not in outage decays exponentially with SNR: the total number of codewords is only polynomial in SNR; for example if the multiplexing rate is r, the rate is $R = r \log \mathsf{SNR}$ and the total number of codewords is SNR^r. Since the error probability is lower bounded by the outage probability, we arrive at a *sufficient* condition for approximate universality of a code:

> the worst-case (over channels not in outage) pairwise error probability for every pair of codewords should decay exponentially with SNR.

It turns out that this condition is *necessary* as well; thus we have an exact characterization of approximately universal codes.

In Section III-B1 we derive an expression for the worst-case pairwise error for a pair of codewords. This derivation allows us to explicitly characterize approximate universality of a code in terms of a condition on its pairwise difference codewords. It is fruitful to contrast our approach with the traditional "code design criterion" for space-time codes in the literature where the pairwise error probability is *averaged* over the channel statistics. This criterion indeed depends on the specific channel statistics being considered. This is in stark contrast to the worst-case analysis we have proposed; the corresponding "universal code design criterion" does not depend on the channel statistics and characterizes properties of a universal code: the engineering appeal of the universal code design criterion is natural; modeling channel statistics is a bit of an "art" in practice and it is useful to have a code that is robust to a variety of channel statistics.

The classical code design criterion for the i.i.d. Rayleigh-fading channel is the *determinant criterion*; as we will see in Section III-B1, the universal code design criterion at any specific SNR is quite different from the determinant criterion. However, it is also somewhat involved and is not directly suited to verify or to design approximately universal codes. In Section III-B3 we derive a simplified condition for approximate universality taking the high SNR scaling into consideration and this high SNR criterion is indeed very closely related to the determinant criterion.

1) Worst-Case Pairwise Error Probability: Our approach is to study the worst-case pairwise error probability of the code over MIMO channel realizations not in outage. The pairwise error probability between two codeword matrices $\mathbf{X}_A$ and $\mathbf{X}_B$ (of length $T \geq n_t$), conditioned on a specific realization of the MIMO channel $\mathbf{H}$, is

$$Q\left(\sqrt{\frac{\mathsf{SNR}}{2}\|\mathbf{H}\mathbf{D}\|^2}\right) \tag{5}$$

where $\mathbf{D}$ is the normalized codeword difference matrix

$$\mathbf{D} = \frac{1}{\sqrt{\mathsf{SNR}}}(\mathbf{X}_A - \mathbf{X}_B).$$

Expanding the channel and codeword difference matrices using the singular value decomposition (SVD)

$$\mathbf{H}:=\mathbf{U}_1\mathbf{\Psi}\mathbf{V}_1^* \quad \text{and} \quad \mathbf{D}:=\mathbf{U}_2\mathbf{\Lambda}\mathbf{V}_2^* \tag{6}$$

the pairwise error probability in (5) can be rewritten as

$$Q\left(\sqrt{\frac{\mathsf{SNR}}{2}\|\mathbf{\Psi}\mathbf{V}_1^*\mathbf{U}_2\mathbf{\Lambda}\|^2}\right). \tag{7}$$

Suppose the absolute values are increasingly ordered in $\mathbf{\Lambda}$ and decreasingly ordered in $\mathbf{\Psi}$:

$$\begin{aligned}\mathbf{\Psi} &:= \operatorname{diag}\{\psi_1, \ldots, \psi_{n_m}, 0, \ldots, 0\},\\ \mathbf{\Lambda} &:= \operatorname{diag}\{\lambda_1, \ldots, \lambda_{n_t}\}.\end{aligned}$$

Then the worst-case rotation can be determined and it turns out to be the one that aligns the weaker singular values of the channel matrix with the stronger singular values of the codeword difference matrix [3]. More precisely, the channel eigen-directions $\mathbf{V}_1$ that maximize the pairwise error probability in (7) is [3]

$$\mathbf{V}_1 = \mathbf{U}_2. \tag{8}$$

Now, the no-outage condition is only a condition on the nonzero n_m singular values of the fading matrix and is given by:

$$\sum_{\ell=1}^{n_m} \log(1 + \mathsf{SNR}\,|\,\psi_\ell|^2) \geq R. \tag{9}$$

Hence the worst-case pairwise error probability for the MIMO channel reduces to the optimization problem

$$\min_{\psi_1,\ldots,\psi_{n_m}} \frac{\mathsf{SNR}}{2}\sum_{\ell=1}^{n_m}|\psi_\ell|^2|\lambda_\ell|^2, \tag{10}$$

subject to the constraint in (9).

If we define $Q_\ell := \mathsf{SNR} \cdot |\psi_\ell|^2|\lambda_\ell|^2$, then the optimization problem can be rewritten as

$$\min_{Q_1\geq 0,\ldots,Q_{n_m}\geq 0} \frac{1}{2}\sum_{\ell=1}^{n_m} Q_\ell$$

subject to the constraint

$$\sum_{\ell=1}^{n_m}\log(1 + \frac{Q_\ell}{|d_\ell|^2}) \geq R.$$

This is the dual of the problem of minimizing the total power required to support a target rate R bits/symbol per sub-channel over a parallel Gaussian channel; the solution is just standard waterfilling, and is given by

$$\begin{aligned}Q_\ell &:= \mathsf{SNR}\cdot|\psi_\ell|^2|\lambda_\ell|^2\\ &= \left(\frac{1}{\lambda} - |\lambda_\ell|^2\right)^+, \quad \ell = 1,\ldots,n_m.\end{aligned} \tag{11}$$

Here λ is the Lagrange multiplier chosen such that the channel in (11) satisfies (9) with equality. The worst-case pairwise error probability is

$$Q\left(\sqrt{\frac{1}{2}\sum_{\ell=1}^{n_m}\left(\frac{1}{\lambda}-|\lambda_\ell|^2\right)^+}\right) \tag{12}$$

where λ satisfies

$$\sum_{\ell=1}^{n_m}\left[\log\left(\frac{1}{\lambda|\lambda_\ell|^2}\right)\right]^+ = R. \tag{13}$$

For convenience, we denote the argument of the $Q(\sqrt{\frac{(\cdot)}{2}})$ function at the worst-case channel realization as the universal code construction criterion for the given difference codeword pair. In general, the goal is to maximize this universal code construction criterion

$$\sum_{\ell=1}^{n_m}\left(\frac{1}{\lambda}-|\lambda_\ell|^2\right)^+. \tag{14}$$

2) A Closer Look at the Universal Criterion: To get a feel for the universal criterion in (14), consider the simple case when codeword difference eigenvalues have the same magnitude, i.e., $|\lambda_1| = \cdots = |\lambda_{n_t}|$. Then λ can be explicitly calculated

$$\frac{1}{\lambda} = 2^{R/n_m}|\lambda_1|^2.$$

Thus, the universal criterion is given by

$$n_m(2^{R/n_m}-1)|\lambda_1|^2$$

a simple function of the magnitude of the normalized codeword difference. To understand the situation in general, let us suppose without any loss of generality that $|\lambda_1| \le \cdots \le |\lambda_{n_m}|$. Now consider the largest k such that

$$|\lambda_k|^2 \le 2^{R/k}|\lambda_1\cdots\lambda_k|^{2/k} \le |\lambda_{k+1}|^2, \tag{15}$$

with $|\lambda_{n_m+1}|$ defined as $+\infty$. Then λ can be calculated explicitly

$$\frac{1}{\lambda} = 2^{R/k}|\lambda_1\cdots\lambda_k|^{2/k} \tag{16}$$

satisfies (13). Thus, the universal code design criterion turns out to be

$$\left(k(2^R|\lambda_1\lambda_2\cdots\lambda_k|^2)^{1/k} - \sum_{\ell=1}^{k}|\lambda_\ell|^2\right) \tag{17}$$

a combination of the geometric and arithmetic means of the magnitudes of the k smallest singular values of normalized codeword differences. While this calculation sheds some insight into the nature of the universal code design criterion, it still does not lend itself to designing or verifying approximately universal codes. Toward making this expression more amenable to code design, we would like to develop a high SNR approximation; this is done next.

3) Proof of Theorem 3.1: Our goal here is to show that for a sequence of codes satisfying (4), the probability of error has the same decay rate as that of the outage probability for all fading distributions. The probability of error can be upper bounded using a smart union bound (as in [1])

$$\mathbb{P}_e \le \mathbb{P}\{\mathcal{O}\} + \mathbb{P}(\text{error}, \mathcal{O}^c). \tag{18}$$

Here we have denoted the outage event by $\mathcal{O}$. Similar to the union bound, the second term can be upper bounded by a sum of pairwise errors averaged over all channel realizations not in $\mathcal{O}$. This sum can be further upper bounded by the sum of the *worst-case* (over all channel realizations not in $\mathcal{O}$) pairwise error probabilities. For the probability of error to behave like the probability of outage for every fading distribution, we *require* the second term in (18) to decay exponentially in SNR ($= e^{-\mathsf{SNR}^\delta}$ for some $\delta > 0$). One way to do this is to make *every* worst-case pairwise error decay exponentially in SNR.

Instead of considering a single outage event, we consider a sequence of outage events $\mathcal{O}_\epsilon$, parameterized by $\epsilon > 0$: the channel realizations not in $\mathcal{O}_\epsilon$ are those that are *strictly* inside the no-outage region

$$\sum_{\ell=1}^{n_m}\log(1+|\psi_\ell|^2\mathsf{SNR}) \ge R(1+\epsilon).$$

For a pair of codewords, the worst case pairwise error probability is (12)

$$Q\left(\sqrt{\frac{\sum_{\ell=1}^{n_m}\left(\frac{1}{\lambda}-|\lambda_\ell|^2\right)^+}{2}}\right)$$

where λ satisfies (see (13))

$$\sum_{\ell=1}^{n_m}\left[\log\left(\frac{1}{\lambda|\lambda_\ell|^2}\right)\right]^+ = R(1+\epsilon). \tag{19}$$

Since the codeword differences satisfy the condition in (4), λ can be explicitly calculated (see (15) and (16))

$$\frac{1}{\lambda} = 2^{R(1+\epsilon)}(|\lambda_1|\cdots|\lambda_{n_m}|)^{\frac{2}{n_m}}. \tag{20}$$

Thus the worst-case pairwise error probability can be upper bounded by (see (17))

$$Q\left(\frac{\sqrt{n_m 2^{R(1+\epsilon)}\left(|\lambda_1|\cdots|\lambda_{n_m}|\right)^{\frac{2}{n_m}} - \sum_{\ell=1}^{n_m}|\lambda_\ell|^2}}{\sqrt{2}}\right). \tag{21}$$

Again using the supposition in (4), the first term in (21) is growing unbounded with increasing SNR, while the second term in (21) is bounded above by $2n_tT$ (a constant) because of the power constraint. Thus, the second term can be ignored for increasing SNR and we can write the following upper bound to the worst-case pairwise probability of error (using (4))

$$Q\left(\frac{2^{R\epsilon}}{\sqrt{2}}\right) < \exp\left(\frac{-2^{R\epsilon}}{2}\right).$$

With $R = r \log \mathsf{SNR}$, we conclude that the pairwise error probability conditioned on the channel realization not in $\mathcal{O}_\epsilon$ decays exponentially with SNR. Since the number of codewords is polynomial in SNR, the overall error probability conditioned on the channel realization not in $\mathcal{O}_\epsilon$ decays exponentially with SNR. Thus the error probability decays at the same rate as $\mathbb{P}\{\mathcal{O}_\epsilon\}$. Letting ϵ become arbitrarily close to zero, this decay rate can be made arbitrarily close to that of the outage probability. Thus the sequence of codes achieves the optimal tradeoff curve, and further for every fading distribution, we conclude that the sequence of codes satisfying (4) is approximately universal. This completes the sufficiency part of Theorem 3.1; necessity is proved in Appendix I. Next, we discuss some explicit schemes that are approximately universal, starting with the simple scalar channel and then moving onto more complex channel models.

IV. QAM is Approximately Universal for the Scalar Channel

The single antenna (transmit and receive) channel model can be written as (dropping the time index)

$$y = hx + w.$$

The criterion for approximate universality (cf. Theorem 3.1) simply translates into a minimum distance one for the code

$$d_{\min}^2 > \frac{1}{2^{R(\mathsf{SNR})+o(\log \mathsf{SNR})}} \tag{22}$$

where $d_{\min}$ is the normalized minimum distance over all the codeword pairs for the coding scheme. Now, consider a simple coding scheme with unit block length: QAM of size 2^R. The normalized minimum distance of this QAM has the property

$$d_{\min}^2 \approx \frac{1}{2^R}$$

and is therefore approximately universal for the scalar fading channel.

V. The Parallel Channel

The parallel fading channel with L diversity branches at time m is

$$y_\ell[m] = h_\ell x_\ell[m] + w_\ell[m], \qquad \ell = 1, \dots, L. \tag{23}$$

Here $w_1[m], \dots, w_L[m]$ are i.i.d. $\mathcal{CN}(0,1)$. The approximate universality criterion for the parallel channel is stated in the following theorem. The proof is very much similar to the general approximate universality proof in Section III-B3, hence we omit it here.

Theorem 5.1: A sequence of codes with rate $R(\mathsf{SNR})$ bits/symbol is approximately universal if and only if, for every pair of codewords, the normalized codeword differences $\mathbf{d}_1, \dots, \mathbf{d}_L$ (the rows of the difference codeword matrix) satisfy

$$\|\mathbf{d}_1\|^2 \cdot \|\mathbf{d}_2\|^2 \cdots \|\mathbf{d}_L\|^2 > \frac{1}{2^{R(\mathsf{SNR})+o(\log \mathsf{SNR})}}. \tag{24}$$

In the rest of the section, we study a simple class of codes that are approximately universal. Our main focus is on unit block length codes based on permutations of a QAM constellation that we call *permutation* codes.[1] We show in Section V-B that even a random permutation code is approximately universal; thus space-only approximately universal codes exist. Finally, we demonstrate simple examples of approximately universal permutation codes: these codes are easy to represent (so the storage complexity is low) and very easy to encode and decode (so the run time complexity is small as well). The parallel channel with two subchannels is studied in Section V-C where a *bit-reversal permutation* is shown to be approximately universal; this scheme also provides an operational interpretation to the outage condition (defined based on an information theoretic underpinning) of the parallel channel. Simple permutation codes for the parallel channel with more than two sub-channels are the topic of Section V-D.

A. Approximate Universality of Codes Based on Rotation of PAM

The criterion of maximizing the product-distance has been known in the context of the i.i.d. Rayleigh-fading channel. A code construction based on rotations of PAM constellations is discussed in [16]: the transmit codeword vector $\mathbf{x} := [x_1, \dots, x_L]$ is defined as

$$\mathbf{x} = \mathbf{u}\mathbf{M} \tag{25}$$

where $u_1, \dots, u_L$ are independent PAM constellations and $\mathbf{M}$ is an orthonormal matrix. [16] shows existence of $\mathbf{M}$ such that the code has the maximum diversity possible, i.e., a nonzero product distance. The problem of explicitly maximizing the minimum product distance was later considered in [17]: it was treated as an optimization problem over $\mathbf{M}$ for fixed input constellations. For $L = 2$, the $\mathbf{M}$ that maximizes the product distance was explicitly found using computer simulations. Later, a similar idea of rotating QAM constellations was proposed in [7] as a part of the 2×2 code construction. It follows from Theorem 2 in [7] that these codes are also in fact approximately universal for the parallel channel.

Unfortunately, no generalizations of the rotation based codes exist when there are more than two sub-channels. Further, these codes are hard to decode for large constellation sizes. Therefore, we propose another approach: QAM constellations are the basis of the code design but we consider mappings that utilize the *algebraic* structure of the constellation; these mappings are *nonlinear* with respect to the Euclidean vector space in which the QAM constellations are embedded—this is in contrast to the rotation operation which is a linear mapping.

[1] These codes are intimately related to interleaver designs in turbo codes.

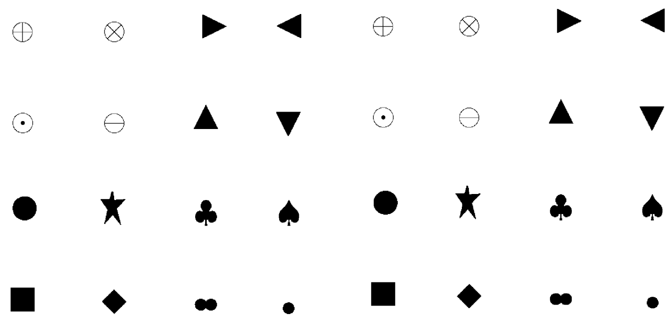

Fig. 2. Repetition coding: $L = 2, R = 4$.

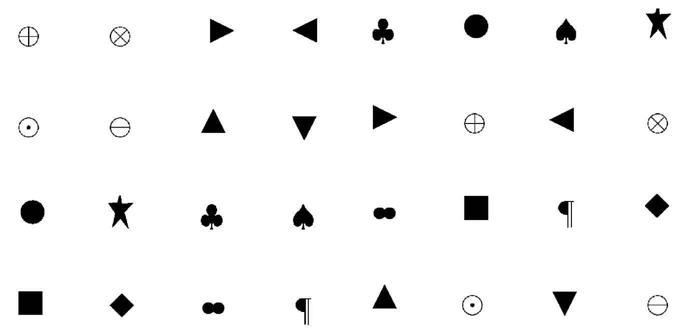

Fig. 3. Permutation code: $L = 2, R = 4$.

B. Permutation Codes

We would like to construct simple space-only (i.e., unit block length) approximately universal codes. As a step toward simple encoding and decoding, suppose the QAM constellation to be the alphabet for each subchannel. We need to protect every codeword by coding it across every subchannel: for the code to have any chance of being approximately universal, it should allow reliable communication for every channel realization not in outage and, in particular, over the parallel channel where all but one subchannel is zero. Two design implications are suggested.

1) With a rate of R bits/symbol, each of the QAM constellations on the subchannels has 2^R points.
2) With 2^R-point QAM as the alphabet for each subchannel, the points in the constellation over each subchannel can be identified one-one with points in the constellation of the other subchannels. In other words, the QAM constellation over one subchannel is a *permutation* of the points in the QAM constellation over any other subchannel.

Mathematically, the permutation code can be represented as

$$\mathbf{C} = \left\{ \sqrt{\frac{\mathsf{SNR}}{2^R}}(q, f_2(q), \ldots, f_L(q)) | q \in \mathbf{Q}_Z \right\}$$

where

$$\mathbf{Q}_Z = \left\{ (a+ib) : -\frac{2^{\frac{R}{2}}}{2} \leq a, b \leq \frac{2^{\frac{R}{2}}}{2} \right\} \tag{26}$$

is the integer-QAM with 2^R points, and $f_2, \ldots, f_L$ are *permutations* of $\mathbf{Q}_Z$.

1) Examples: Repetition coding is a simple example of a permutation code: the permutations are just the identity. Fig. 2 illustrates the permutation code with identity permutation for $L = 2$. Here $\mathbf{Q}_Z$ is the QAM with 16 points.

For $L = 2$, Fig. 3 shows a permutation code with 16 codewords that is designed to maximize the minimum product distance. Product distance of this code is an improvement over the repetition code in Fig. 2 by a factor of 4. The code in Fig. 3 and its generalization to larger L is discussed in [18] using the theory of spreading transforms. The focus in [18] is on finding codes that have a nonzero product distance and can be efficiently constructed from smaller constellations (QPSK) using spreading transforms.

2) A Random Permutation Code Ensemble: Our search for permutation codes that are approximately universal leads us to study permutations with large QAM alphabet sizes. To get a feel for whether there indeed exist permutation codes with large enough product distance, we can look at an appropriate random permutation ensemble and see if the product distance *averaged* over this ensemble of permutation codes has the desired property. If this is the case, then there must have been at least one permutation code in the ensemble that is approximately universal. Averaging the product-distance itself is not good enough; we look at the inverse of the product distance and average it over all possible permutation codes with the uniform measure. Our main result is the demonstration of existence of permutation codes that are approximately universal.

Theorem 5.2: There exists a sequence of permutation codes that is approximately universal over the parallel channel.

The details of the proof are relegated to Appendix II.

C. Two Subchannels: Bit-Reversal Permutation Code

While it is encouraging to know the existence of permutation codes that are approximately universal, it is of engineering interest to actually construct simple approximately universal codes from this ensemble. It turns out that an *operational interpretation* of the outage condition (which was defined based on an information theoretic understanding of the compound channel) suggests natural permutation codes that are approximately universal. In this section, we focus on the special case when the parallel channel has just two subchannels, i.e., $L = 2$.

1) Operational Interpretation to the Outage Condition: If we communicate at a rate of R bits/symbol over the parallel channel, the no-outage condition is

$$\log(1 + |h_1|^2\mathsf{SNR}) + \log(1 + |h_2|^2\mathsf{SNR}) > R. \tag{27}$$

One way of interpreting this condition is as though the first sub-channel provides $\log(1+|h_1|^2\mathsf{SNR})$ bits of information and the second sub-channel provides $\log(1+|h_2|^2\mathsf{SNR})$ bits of information, and as long as the total number of bits provided exceed the target rate, then reliable communication is possible. In the high SNR regime, we exhibit below a permutation code that makes the outage condition concrete.

Suppose we independently code over the I and Q channels of the two subchannels. So we can focus on only one of them, say, the I channel. We wish to communicate $R/2$ bits over two uses of the I-channel. Analogous to the typical event analysis for the

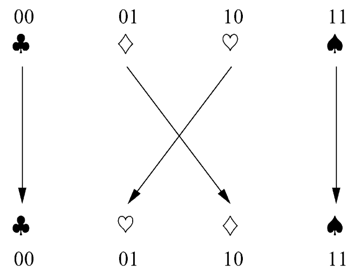

Fig. 4. The bit-reversal map for a 4-PAM.

scalar channel, we can exactly recover all the $R/2$ information bits from the first I sub-channel alone if:

$$\frac{1}{2}\log\left(1+|h_1|^2\mathsf{SNR}\right) > \frac{R}{2}.$$

However, we do not need to use just the first I subchannel to recover all the information bits: the second I sub-channel also contains the same information and can be used in the recovery process. Indeed, if we create x_1^{I} by treating the ordered $R/2$ bits as the binary representation of the points x_1^{I}, then one would intuitively expect that if

$$\frac{1}{2}\log(1+|h_1|^2\mathsf{SNR}) > k_1 \tag{28}$$

then one should be able to recover at least k_1 of the most significant bits of information. Now, if we create x_2^{I} by treating the *reversal* of the $R/2$ bits as its binary representation, then one should be able to recover at least k_2 of the most significant bits, if

$$\frac{1}{2}\log(1+|h_2|^2\mathsf{SNR}) > k_2. \tag{29}$$

But due to the reversal, the most significant bits in the representation in the second I sub-channel are the least significant bits in the representation in the first I sub-channel. Hence, as long as $k_1+k_2 \geq R/2$, then we can recover *all* $R/2$ bits. This translates to the condition

$$\log(1+|h_1|^2\mathsf{SNR}) + \log(1+|h_2|^2\mathsf{SNR}) > R \tag{30}$$

which is precisely the no-outage condition (27). Thus, the bit-reversal scheme gives an operational meaning to the outage condition.

2) Bit-Reversal Permutation Code: To make this idea concrete, first we need to define bit reversal. A QAM can be thought of as two independent PAM's, and using I and Q channels separately is equivalent to taking the QAM permutation as two independent PAM permutations. Therefore we concentrate on one of the PAM's and define the bit-reversal permutation for it. For a PAM with $2^{R/2}$ points, we number the points from left to right by 0 to $2^{R/2}-1$. Based on this numbering, a canonical bit sequence of length R represents each point in the PAM constellation. Bit reversals are defined based on this representation. The bit-reversal map for the 4-PAM is illustrated in Fig. 4.

3) Product Distance and Bit Reversals: To show that the bit reversal scheme is approximately universal, we have to show that it satisfies the criterion in (24). However, the plain bit-reversal is *not* approximately universal. The problem is the inherent assumption in the operational interpretation that if two points have different MSB, then they are far apart geometrically and hence cannot be confused with each other. This, however, is not true. Consider the points with the binary representations

$$011\cdots10 \quad \text{and} \quad 100\cdots01.$$

Even though their MSB is different, they are separated by a fixed distance of 3 independent of the length $R/2$ of the binary representation. The same is true for their bit-reversals. Thus, the product distance between this codeword pair is $\frac{9}{2^{2R}}$ and it does not satisfy (24) for large R.

Even though the simple bit-reversal is not optimal, it can be modified so that it essentially retains the operational interpretation (so it is still easy to decode) and is approximately universal. We discuss two such modifications here: irregularly spaced PAM and alternate-bit-flipping.

4) Irregularly Spaced Pam Permutation Code: We have seen that the problem with the bit-reversal scheme is the inherent assumption that the two points having different MSB are geometrically far apart. A simple way to get around this problem is to put gaps in the PAM constellation. That is, we introduce a gap of $g2^{R/2}$ between $011\cdots1$ and $100\cdots0$ so that any two points with different MSB are indeed far apart. More precisely, to retain the operational interpretation, one has to put a gap of $g2^m$ for every mth bit-change to ensure that the product distance condition is met. The PAM constellation is now *irregularly spaced*.[2]

Consider any two points in the irregularly spaced PAM constellation. Suppose the first MSB they differ in their bit representation is the mth one: then by construction the normalized distance between the two points is lower bounded by

$$g2^{-m}$$

. The bit-reversals of these two points must have the same $m-1$ LSBs but a different mth LSB; so the normalized distance between the bit-reversals of these two points is lower bounded by

$$2^{m-R/2}.$$

Putting these two together, we conclude that the normalized product distance between a pair of codewords in the bit-reversed irregularly spaced permutation code is lower bounded as

$$\begin{aligned}|d_1d_2| &\geq g2^{-m}2^{m-R/2}\\ &= \frac{g}{2^{R/2}}.\end{aligned}$$

Comparing this with (24), we conclude that the code is approximately universal.

A potential drawback of this approach is that the extra gaps translate into an increase in the amount of power used for the

[2]The same idea of introducing gaps is also present in the Cantor set based representation in [19].

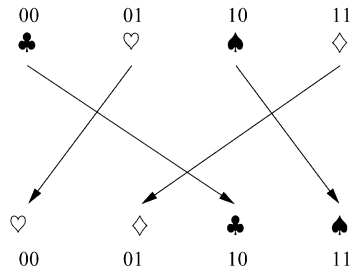

Fig. 5. Bit Reversals with alternate bits flipped.

same rate. Thus, for a PAM of size $2^{R/2}$, the normalized increase in size is given by

$$\begin{aligned}\sum_{m=1}^{R/2} & g2^{m-R/2} \text{ (number of } m\text{th bit} - \text{changes)} \\ &= \sum_{m=1}^{R/2} g2^{m-R/2}(2^{R/2-m}) \\ &= gR/2.\end{aligned}$$

With $R = r \log \mathsf{SNR}$, the SNR of this scheme is increased by a factor of $(1 + gr \log \mathsf{SNR}/2)$. In the diversity-multiplexing scaling of our interest, this is an insignificant increase and thus the code is still approximately universal.

5) Alternate-Bit-Flipping Permutation Code: Another modification of the plain bit-reversal scheme is to flip every alternate bit after reversing. For example, the point in the PAM constellation with bit representation 111111 is mapped to the point in the PAM constellation with bit representation 010101. The scheme is illustrated for the 4-PAM constellation in Fig. 5.

In general, consider the $R/2$-bit representation of integers a_1 and a_2 between 0 and $2^{R/2} - 1$

$$a_1 = b^1_{R/2} \cdots b^1_1,$$
$$a_2 = b^2_{R/2} \cdots b^2_1.$$

The the alternate-flip bit-reversal map B is defined as (assuming R is even)

$$B(a_1) = \overline{b^1_1} b^1_2 \cdots \overline{b^1_{R/2-1}} b^1_{R/2},$$
$$B(a_2) = \overline{b^2_1} b^2_2 \cdots \overline{b^2_{R/2-1}} b^2_{R/2}.$$

An easy observation is that this scheme maintains the integrity of the operational interpretation since the decoder can always flip the bits back after estimating the flipped bits. Further, this scheme turns out to be approximately universal.

Theorem 5.3: For every a_1 and a_2 between 0 and $2^{R/2} - 1$

$$\frac{|a_1 - a_2|}{2^{R/2}} \frac{|B(a_1) - B(a_2)|}{2^{R/2}} \geq \frac{1}{8 \cdot 2^{R/2}}. \tag{31}$$

The details of the proof are somewhat involved and are relegated to Appendix III.

D. Explicit Permutation Codes for General Parallel Channel

In an effort to generalize the bit-reversal scheme consider the following alternative, but equivalent, view of the same scheme (for $L = 2$).

1) Bit-Reversal as a Linear Operation: Each codeword in the bit-reversal permutation code is represented by a sequence of, say $2n$ bits. The first n bits correspond to a point in a 2^n-PAM constellation. The corresponding PAM constellation point is then transmitted over the I channel of the first sub-channel. The last n bits similarly correspond to a point in another 2^n-PAM constellation which is then transmitted over the Q channel of the first subchannel. The transmissions over the I and Q channels of the second subchannel are the points in the PAM constellation that correspond to bit-reversals of the first and last n bits, respectively, of the total $2n$ bits that define the codeword.

If we fix the mapping between the sequence of bits and points in a PAM constellation, the bit-reversal scheme can be viewed entirely as an operation on the $2n$ bits that represent the codeword. Further more, if we decide to do the *same operation* over both the I and Q channels (as in the bit-reversal scheme), then we only need to consider operations over the first n bits that represent the codeword. In the rest of this discussion, we consider only the operation on the first n bits representing the codeword. The operation involved in bit-reversal is particularly simple: it is a *linear* operation on the vector of bits (over the field $\mathbb{F}_2$). Linear operations can be represented by matrices and the bit reversal scheme corresponds to two matrices: the *identity* matrix $(\mathbf{I}_n)$ for the first subchannel and the *cross-diagonal* matrix with unit entries on the cross diagonal $(\mathbf{D}_n)$ for the second subchannel.

The outage interpretation implies that the decoder can deduce k_1 most significant bits from the first subchannel (see (28)) and k_2 most significant bits from the second subchannel (see (29)). Because of the simple mappings in this case, the k_1 bits from the first subchannel correspond to the first k_1 bits of the vector of n bits representing the codeword and k_2 bits from the second subchannel that correspond to the last k_2 bits of the vector of n bits representing the codeword. As long as $k_1 + k_2 \geq n$, the decoder can determine the codeword correctly.

2) Universally Decodable Matrices: This view of the bit-reversal scheme suggests a natural generalization to more than two sub-channels. We first generalize the bit representation of the integers points of the PAM constellation: we allow q-digit representation over a finite field $\mathbb{F}_q$. Next we consider a (sequence of) collection of L matrices $\{\mathbf{A}^{(n)}_1, \ldots, \mathbf{A}^{(n)}_L\}_n$ of size $n \times n$ with entries selected from the finite field $\mathbb{F}_q$. These matrices naturally generate a sequence of permutation codes: for a permutation code conveying $2n$ q-digits of information, we transmit over the I channel of the ℓth subchannel the point in the 2^n-PAM constellation that corresponds to the q-digit sequence that results from the linear operation of $\mathbf{A}^{(n)}_\ell$ over the first nq-digits of the $2n$ information q-digits. This is done for each of the $\ell = 1, \ldots, L$ subchannels. Further, the same linear operations are used on the last n information q-digits to transmit points from the PAM constellation on the Q channels of the L subchannels.

We say that this collection of matrices is *universally decodable* if for any $k_1, \ldots, k_L$ such that

$$k_\ell \geq 0, \quad \ell = 1, \ldots, L \quad \text{and} \quad \sum_{\ell=1}^{L} k_\ell \geq n \tag{32}$$

the collection of the first $k_1, \ldots, k_L$ rows of the matrices $\mathbf{A}_1^{(n)}, \ldots, \mathbf{A}_L^{(n)}$ respectively is *full rank*, i.e., spans the vector space $\mathbb{F}_q^n$.

Universally decodable matrices (UDMs) provide an operational interpretation to the information theoretically defined outage condition. The number k_ℓ can be interpreted as the amount of q-digits provided by the ℓth subchannel; this depends on the corresponding channel amplitude $|h_\ell|$. If the channel is not in outage, then (32) holds. The full rank condition implies that a unique codeword can be decoded whenever the channel is not in outage. We formally state the implication of this operational interpretation to outage below; the proof is relegated to Appendix IV.

Theorem 5.4: A sequence of UDMs leads to an approximately universal permutation code sequence.

Observe that the encoding and decoding complexity of the code based on UDMs is simply *linear* in the number of bits n and the number of subchannels L. The representation of the code involves storing the L matrices with a total of Ln^2 entries, again a very small number.

In the rest of this section, we focus on explicit construction of UDMs. First, we show how UDMs can be easily constructed from maximum-distance separable codes (MDS) (though these constructions require a field size that grows with n). In Section V-D4, we present fixed field size constructions for $L = 3$ and then discuss a recent construction [20], for arbitrary L.

3) Reed–Solomon Codes Are Approximately Universal: In general, some progress on the search for universally decodable matrices can be made by strengthening the requirement on the collection of matrices by requiring the collection of *any* n rows from the matrix

$$\mathbf{A}^{(n)} = \begin{bmatrix} \mathbf{A}_1^{(n)^t} & \mathbf{A}_2^{(n)^t} & \cdots & \mathbf{A}_L^{(n)^t} \end{bmatrix}$$

to be full rank. Note that such a collection of matrices is still universally decodable. This problem is same as designing a maximum distance separable (MDS) codes with $\mathbf{A}^{(n)}$ as its parity check matrix. The condition universal decodability condition is the same as requiring that the minimum distance of the code to be at least $n+1$. Since $\mathbf{A}^{(n)}$ is an $n \times Ln$ matrix, such a code has length Ln and rate $Ln-n$. A simple singleton bound shows that then the code must be $[Ln, Ln-n, n+1]$.[3]

Simple examples of such a code exist and this allows us to explicitly construct the parity check matrix $\mathbf{A}^{(n)}$. For a finite field $\mathbb{F}_q$, a $[q+1, k, q-k+2]$ extended *Reed–Solomon* code can be explicitly constructed for *every* $k \leq q+1$ (see Chapter 6.8 of [21] for the exact parity check matrix). For the extended Reed-Solomon codes, the field size grows with the block-length. In fact, the field size is at least $Ln-1$. In our setting, n grows as $\log \mathsf{SNR}$, thus the field size grows like $\log \mathsf{SNR}$. As noted in the proof of Theorem 5.4, this still gives an approximately universal code.

[3]An $[n,k,d]$ code over $\mathbb{F}_q$ is a linear, length-n code with q^k codewords and a minimum Hamming distance of d. Its parity check matrix is a $n-k \times n$ matrix over $\mathbb{F}_q$. Codes for which $d = n-k+1$ meet the singleton bound (see Chapter 3.2 in [21]) and are called MDS codes. These codes are well-studied in coding theory and explicit codes like the Reed-Solomon codes are MDS codes.

Next, we focus on the situation of practical and theoretical interest: constructing UDM's with a field size not growing with n. With $L = 2$, we have already seen an example: $\{\mathbf{I}_n, \mathbf{D}_n\}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{D}_n$ is the $n \times n$ cross-diagonal matrix with all unit entries on the cross diagonal; here the field size $q = 2$.

4) $L = 3$: Universally Decodable Matrices: Consider the following collection of binary matrices (i.e., the field size $q = 2$): $\{\mathbf{I}_n, \mathbf{D}_n, \mathbf{T}_n\}_n$, where $\mathbf{I}_n$ and $\mathbf{D}_n$ are, as before, the $n \times n$ identity and cross-diagonal matrix with unit cross diagonal entries, respectively. $\mathbf{T}_n$ is defined using the recursive definition

$$\mathbf{T}_{2n} := \begin{bmatrix} \mathbf{T}_n & \mathbf{T}_n \\ \mathbf{0} & \mathbf{T}_n \end{bmatrix} \tag{33}$$

with $\mathbf{T}_1 = [1]$. Equivalently, $\mathbf{T}_{2n} = \mathbf{T}_2 \otimes \mathbf{T}_n$, where $\otimes$ denotes the *tensor* or *Kronecker* product operation between two matrices (cf. [22, Ch 4.2]). For $2^{m-1} < n < 2^m$, we define $\mathbf{T}_n$ to be the principal submatrix of $\mathbf{T}_{2^m}$. We omit our original proof of this result (it is still available in an earlier version of this paper [23]), in light of a crisper proof that follows from a more general result in [20]; this generalization was motivated by the present construction for $L = 3$.

For $L = 4, q = 3$ computer simulations are used in [24] to justify the conjecture that the following collection of matrices is universally decodable: $\{\mathbf{I}_n, \mathbf{D}_n, \mathbf{T}_n, \mathbf{R}_n\}_n$ where the first two matrices are, as before, the $n \times n$ identity and cross-diagonal matrix with unit cross diagonal entries, respectively. With $n = 3$, define

$$\mathbf{T}_3 := \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{R}_3 := \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}. \tag{34}$$

For n a power of 3, we define, recursively, $\mathbf{T}_{3n} = \mathbf{T}_3 \otimes \mathbf{T}_n$ and $\mathbf{R}_{3n} = \mathbf{R}_3 \otimes \mathbf{R}_n$, with the multiplication operations in the context of the field $\mathbb{F}_3$. For $3^{m-1} < n < 3^m$, we define $\mathbf{T}_n$ and $\mathbf{R}_n$ to be the principal submatrices of $\mathbf{T}_{3^m}$ and $\mathbf{R}_{3^m}$, respectively. This conjecture has now been verified as a special case of the general result in [20].

5) A Complete Characterization of UDMs: Motivated by the results in the previous two subsections, the authors in [20], have recently completely solved the problem of constructing UDMs. They show for any n the condition $L \leq q+1$ is both necessary and sufficient. They construct UDMs based on *Pascal's triangle*. We state their construction (see Proposition 9, [20]), for completeness:

Theorem: Let q be a prime power and let $L \leq q+1$. Suppose α is a primitive element over $\mathbb{F}_q$. Then the following matrices are UDMs:

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{I}_n \\ \mathbf{A}_2 &= \mathbf{D}_n \\ [\mathbf{A}_\ell]_{(j,k)} &= \binom{k}{j} \alpha^{(\ell-2)(k-j)} \\ &\quad \text{for } 1 \leq j,k \leq n \quad \text{and} \quad 3 \leq \ell \leq L \end{aligned}$$

where $\binom{k}{j}$ is defined as the natural mapping to prime subfield of $\mathbb{F}_q$ of the natural number

$$\binom{k}{j} := \frac{k(k-1)\cdots(k-j+1)}{j(j-1)\cdots 1}.$$

VI. The MISO Channel

The parallel channel allowed us to study approximately universal codes on channels with solely multiplexing gain. We now turn to study channels that offer solely diversity gain: the MISO and SIMO channels, with multiple transmit (receive) and single receive (transmit) antennas, respectively. The SIMO channel can be reduced to a scalar channel by considering a scalar sufficient statistic: receive beamformed vector. Therefore, any approximately universal scheme for the scalar channel, such as the QAM scheme (see Section IV), will also be approximately universal for the SIMO channel. In this section, we focus on the MISO channel and understand properties of approximately universal codes over this channel.

The scalar output of a MISO channel with n_t transmit antennas at time m can be written as

$$y[m] = \mathbf{h}^t \mathbf{x}[m] + w[m]$$

where $\mathbf{x}[m]$ is an n_t dimensional vector input and $\mathbf{h}$ is the n_t-dimensional vector of fading gains h_is.

A. Characterization of Approximately Universal Codes

The approximate universality criterion for the MISO channel can be stated as (see Theorem 3.1), for every codeword difference matrix

$$\lambda_1^2 > \frac{1}{2^{R(\mathsf{SNR})+o(\log \mathsf{SNR})}} \tag{35}$$

where λ_1 is the minimum singular value of the normalized codeword difference matrix.

There is an intuitive explanation for this result: a universal code has to protect itself against the worst channel that is not in outage. The condition of no-outage only puts a constraint on the *norm* of the channel vector $\mathbf{h}$ but not on its direction. So, the worst channel aligns itself to the "weakest direction" of the codeword difference matrix. The corresponding worst-case pairwise error probability is governed by the smallest singular value of the codeword difference matrix.

On the other hand, the i.i.d. Rayleigh channel does not prefer any specific direction: thus the design criterion tailored to its statistics requires that the *average* direction be well protected and this translates to the determinant criterion. While the two criteria are different, codes with large determinant tend to also have a large value for the smallest singular value; the two criteria (based on worst-case and average-case) are related in this aspect.

For the case when $n_t = 2$, the Alamouti scheme [25] converts the MISO channel to a scalar channel with gain $||\mathbf{h}||$ and the *total* SNR reduced by a factor of 2. Hence, the outage behavior is exactly the same as in the original MISO channel, and the Alamouti scheme provides a *universal* conversion of the 2×1 MISO channel to a scalar channel. Any approximately universal scheme for the scalar channel, such as a QAM, when used in conjunction with the Alamouti scheme will be approximately universal for the MISO channel.

In the general case when the number of transmit antennas is greater than 2, there is no equivalent to the Alamouti scheme. Here we explore one approach to construct approximately universal schemes for the general MISO channel: we consider a simple scheme that converts the MISO channel into a parallel channel and show that the scheme is approximately universal over a restricted class of MISO channel statistics.

B. MISO Channel Viewed as a Parallel Channel

Consider the simple scheme of using one antenna at a time to communicate at a rate of R bits/symbol on the MISO channel. By using one transmit antenna at a time, we arrive at a parallel channel with n_t subchannels and the data rate of communication is $n_t R$ bits/symbol per subchannel. We code over the antennas using a parallel channel code, e.g., a permutation code. Our first result is that this simple scheme is tradeoff optimal for the i.i.d. Rayleigh-fading MISO channel.

Can this conversion be approximately universal? To see that this could not be the case, consider the following (worst-case) MISO channel model: the channels from all but the first transmit antenna are very poor. To make this example concrete, set $h_\ell = 0, \ell = 2, \ldots, n_t$. The tradeoff curve depends on the outage probability (which depends only on the statistics of the first channel). Using one transmit antenna at a time is a waste of degrees of freedom: since the channels from the all but the first antenna are zero, there is no point in transmitting any signal on them. Thus the scheme could not have been tradeoff optimal over a MISO channel with such statistics.

Essentially, using one antenna at a time equates temporal degrees of freedom with spatial ones. All temporal degrees of freedom are the same, but the spatial ones need not be the same: in the extreme example above, the spatial channels from all but the first transmit antenna are zero. Thus, it seems reasonable that when all the spatial channels are *symmetric* then the parallel channel conversion of the MISO channel is tradeoff-optimal. This intuitive argument is formalized in the proposition below; the proof is provided in Appendix V.

Proposition 6.1: An approximately universal parallel channel code sequence used over the antennas of a MISO channel, one antenna at a time, is tradeoff-optimal for the class of MISO channels with i.i.d. fading coefficients. Further, the optimal tradeoff curve of the MISO channel is given by

$$d^*(r) = a n_t (1-r), \quad 0 \le r \le 1 \tag{36}$$

where

$$a := \lim_{x \to 0} \frac{\log \mathbb{P}(|h_\ell|^2 \le x)}{\log x}, \quad \forall \ell = 1, \ldots, n_t. \tag{37}$$

We have seen that the conversion of the MISO channel into a parallel channel is tradeoff-optimal for the i.i.d. Rayleigh fading channel. To get a practical feel for how much loss the conversion of the MISO channel into a parallel channel entails with respect to the optimal outage performance, we plot the error probabilities of two schemes with the same rate ($R = 2$ bits/symbol): uncoded QAMs over the Alamouti scheme and the permutation code in Fig. 3. This performance is plotted in Fig. 6 where we see that the conversion of the MISO channel into a parallel channel entails a loss of about 1.5 dB in SNR for the same error

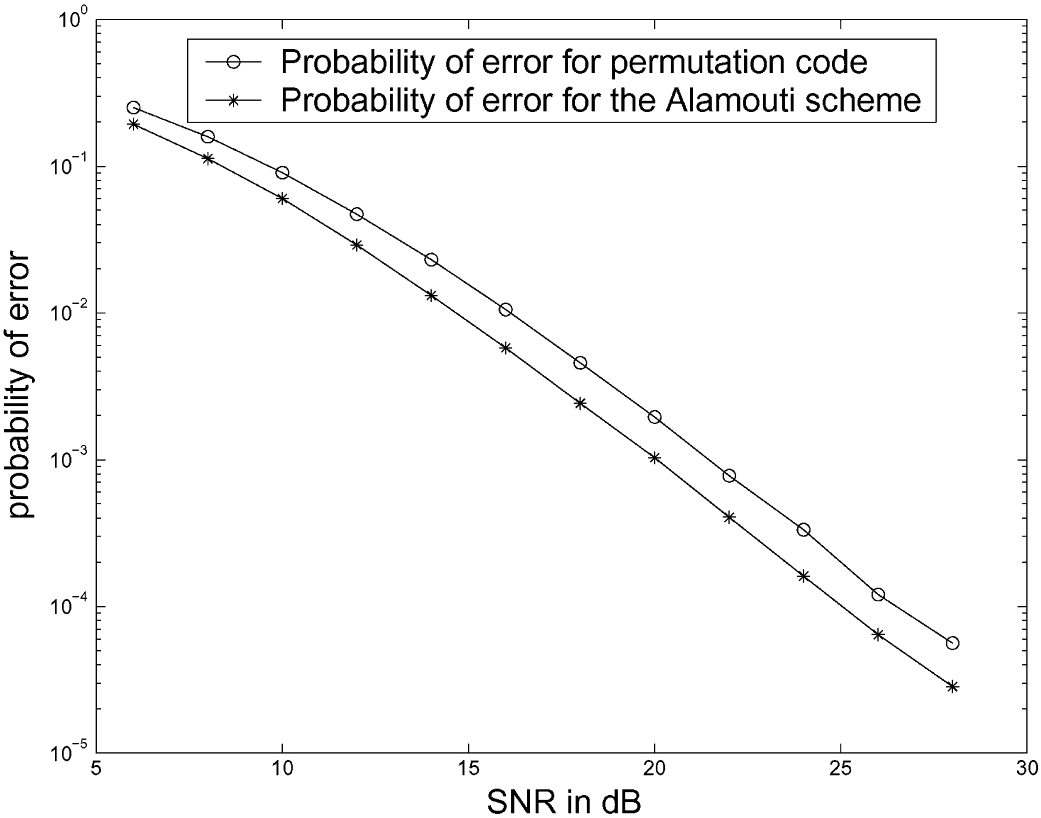

Fig. 6. The error probability of uncoded QAM with the Alamouti scheme and that of a permutation code over one antenna at a time for the Rayleigh fading MISO channel with two transmit antennas: the permutation code is only about 1.5 dB worse than the Alamouti scheme over the plotted error probability range.

probability performance. This is a fairly small loss and suggests the practical utility of the conversion of the MISO channel with larger number of receive antennas to a parallel channel.

VII. The MIMO Channel

Having studied the construction of approximately universal codes over the parallel and the MISO channel, we are now ready to move over the general MIMO channel: we first conclude the approximate universality of some recently proposed codes and then explore the approximate universality properties of two classical space time coding architectures: D-BLAST and V-BLAST.

A. Approximate Universality of Number-Theoretic Codes

Some of the recent space time code constructions in the literature have a number-theoretic flavor. In particular, a rotated QAM constellation was used to construct a two transmit antenna space time code in [7], [9], [26]. For arbitrary n_t, [10] proposes codes derived from cyclic division algebras that have the shortest block-length possible $(T = n_t)$.[4] Some constructions based on cyclic division algebras are also presented in [11][12]. All these two codes satisfy the nonvanishing determinant criterion. The authors in [7][9] used this property to conclude the tradeoff optimality over the i.i.d. Rayleigh-fading channel. In the light of our characterization of approximate universality (cf. Theorem 3.1), we can conclude that all these codes are approximately universal; further more, in the light of the discussion in Section III-A, we can conclude that these codes are approximately universal simultaneously for every MIMO channel with n_t transmit antennas ($n_t = 2$ for the code in [7][9]) and arbitrary n_r. To see this formally, we discuss the two transmit antenna code in [7] in some detail.

The rotated code QAM code in [7] spans two symbols and is designed to work over the two transmit MIMO channel. The entries of the 2×2 transmit codeword matrix $\mathbf{X}{:=}[x_{ij}]$ are

$$\begin{bmatrix} x_{11} \\ x_{22} \end{bmatrix} := \mathbf{Q}(\theta_1) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$
$$\begin{bmatrix} x_{21} \\ x_{12} \end{bmatrix} := \mathbf{Q}(\theta_2) \begin{bmatrix} u_3 \\ u_4 \end{bmatrix}. \quad (38)$$

Here u_1, u_2, u_3, u_4 are independent QAMs of size $2^{R/2}$ each (so the data rate of this scheme is R bits/symbol). The rotation matrix $\mathbf{Q}(\theta)$ is

$$\mathbf{Q}(\theta){:=} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

With the choice of the angles θ_1, θ_2 equal to $1/2\tan^{-1} 2$ and $1/2\tan^{-1}(1/2)$ radians respectively, Theorem 2 of [7] shows that the determinant of every normalized codeword difference matrix $\mathbf{D}$ satisfies

$$|\det \mathbf{D}|^2 \geq \frac{1}{10 \cdot 2^R}.$$

[4]Thus, it follows from the results in [10] that the optimal diversity-multiplexing tradeoff for arbitrary fading MIMO channels can be achieved for any block-length $T \geq n_t$.

Our discussion so far is summarized in the following formal statement characterizing of the performance of this code.

Proposition 7.1: The code described in (38), with $\theta_1 = 1/2\tan^{-1}2$ and $\theta_2 = 1/2\tan^{-1}(1/2)$, is approximately universal for every MIMO channel with two transmit antennas.

A) Discussion: While the two codes discussed above are explicit and easy to encode, they lack a computationally simple decoding algorithm. In general, it appears hard to design explicit approximately universal codes for the MIMO channel with a computationally simple decoding algorithm; it still remains an open problem. For the parallel channel we have been able to answer this question to a reasonable extent. The difference in the two models arises due to the *rotation matrix* in the SVD decomposition (6): a parallel channel code has to be optimal for a fixed rotation matrix (the identity matrix) while a MIMO channel code has to be optimal for *every* rotation matrix. This difference seems to naturally lead to codes with a number-theoretic flavor: they are delicately designed so as to cope with every possible rotation. Such a code with a computationally simple decoding algorithm has not yet been found.

An alternate view point is proposed in [27] where a lattice based space–time code is constructed. The authors show that the structure of these codes resembles random Gaussian codes and then conclude the tradeoff optimality of an ensemble of lattice codes for a decoder based on a generalized MMSE estimator for the i.i.d. Rayleigh-fading channel. A typical code in this ensemble is very unlikely to be approximately universal. In fact, one of the important conclusions of the the authors of [27] is that their construction shows that maximizing the determinant criterion is not a necessary requirement for achieving the tradeoff for *specific* fading distributions. However, as we see here, maximizing the determinant criterion is a *necessary and sufficient* condition to design robust codes that are tradeoff-optimal for *every* fading distribution.

B. The V-BLAST Architecture

The V-BLAST architecture was proposed for high rate communication over the MIMO channel [13]. It splits the data stream into independent streams that are sent over the different transmit antennas. It is very clear that V-BLAST is not tradeoff optimal at low rates: the largest diversity of any data stream is limited by the number of receive antennas. However, it is also clear that the V-BLAST scheme cannot be approximately universal even at high rates: over the 2×1 MIMO channel suppose the channel from one of the transmit antennas is zero and the other channel is $\mathcal{CN}(0,1)$. Then the diversity obtained by the data stream sent over the first transmit antenna for any multiplexing gain is zero whereas the overall channel has a nonzero diversity-multiplexing tradeoff. Since the V-BLAST scheme does not code across the transmit antennas it takes a hit when the transmit antennas have asymmetric fading statistics. When all transmit antennas are statistically similar to one another, V-BLAST indeed turns out to be tradeoff optimal at high rates; we explore this aspect in detail in Section VIII.

C. The D-BLAST Architecture

The D-BLAST architecture has been proposed to attain high diversity gains over the MIMO channel [14]. The data is split into independent streams that are sent over the MIMO channel in a diagonal fashion. The coding scheme can be written as

$$\begin{bmatrix} 0 & \cdots & 0 & p_1^{(1)} & \cdots & p_1^{(T-n_t+1)} \\ \vdots & \iddots & \cdots & \iddots & \vdots & \\ 0 & p_{n_t-1}^{(1)} & \iddots & & \iddots & 0 \\ p_{n_t}^{(1)} & \cdots & \cdots & 0 & 0 & \end{bmatrix} \tag{39}$$

where $\mathbf{p}^{(k)} = [p_1^{(k)}, \dots, p_{n_t}^{(k)}]$ are the independent data streams.

It is well known that the D-BLAST architecture with MMSE-SIC receiver preserves mutual information over any deterministic MIMO channel with Gaussian inputs; thus it converts a MIMO channel into an equivalent parallel channel (a tutorial description of this conversion is described in Ch. 8.5 of [6]). Therefore, an approximately universal code over the parallel channel, such as the permutation code, when used as the streams of the D-BLAST architecture for the MIMO channel will be approximately universal for the MIMO channel. This approach of converting the MIMO channel into a parallel channel has also been used by Matache and Wesel in [4].

Alternatively, one can see its approximate universality by explicitly verifying that it satisfies the condition in (4) for $n_t = n_r$. The product of singular values of the codeword difference matrix for (39) turns out to be lower bounded by the product distance of the permutation code. Thus, if $\mathbf{p}^{(k)}$ is a permutation code that is approximately universal for the parallel channel, then the D-BLAST scheme (39) is approximately universal for the MIMO channel (see and compare (24) and (4)).

A potential drawback is the initialization loss due to the zero padding in (39) which reduces the effective rate. For a 2×2 channel with block-length three, a rate of R bits/stream corresponds to a rate of $2R/3$ bits/symbol on the MIMO channel. In general, the actual tradeoff curve achieved by this scheme is

$$d_{\text{out}}\left(\frac{T}{T-n_t+1}r\right) \tag{40}$$

where r is the multiplexing gain per symbol. For the block length T large, D-BLAST approaches approximate universality. For finite block-length, this scheme is strictly suboptimal. The precise characterization for approximate universality also implies that this performance cannot be universally improved upon using a better decoding strategy (than MMSE and successive interference cancelation). In Section IX, we see that the performance can indeed be improved upon for a certain *restricted* class of fading distributions using a better decoding strategy.

VIII. The V-BLAST Architecture

The V-BLAST architecture transmits independent data streams over the transmit antennas. This is closely related to how a multiple access channel is operated, the tradeoff performance of which under i.i.d., Rayleigh fading is studied (using random Gaussian codes) in [28][29]. In this section, we study the performance of simple modulation schemes over the V-BLAST architecture: in particular, QAM constellations.

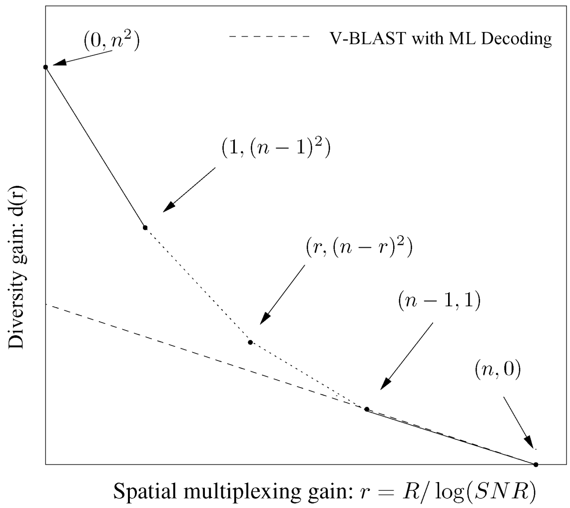

Fig. 7. The i.i.d. Rayleigh-fading channel with $n_t = n_r = n$.

While we have seen that the V-BLAST architecture can never be approximately universal, it still performs very well for an interesting *restricted* class of channels.

A. Tradeoff Optimality Over Rayleigh Fading Channels

Consider operating the V-BLAST architecture over an $n_t \times n_r$ i.i.d. Rayleigh-fading channel: we transmit independent data streams over each of the n_t antennas; each data stream is transmitted *uncoded* using a QAM constellation (with SNR^{r/n_t} points at each time symbol). This scheme corresponds to a total data rate of $r \log \mathsf{SNR}$ bits/symbol over the MIMO channel. Our main result is the precise characterization of the tradeoff performance; the proof is available in Appendix VII.

Proposition 8.1: Uncoded independent QAMs of size SNR^{r/n_t} points over the antennas of an $n_t \times n_r$ i.i.d. Rayleigh fading MIMO channel are protected by a diversity gain, $d(r)$, where

$$d(r) = n_r - \frac{n_r r}{n_t} \quad \text{if } n_r \geq n_t \tag{41}$$

$$\geq n_r - r \quad \text{if } n_r < n_t. \tag{42}$$

Several interesting observations follow from this result.

1) Apart from the fact that the channel can be in outage, there is an additional error event in the V-BLAST architecture: the presence of the other simultaneously transmitted streams impacts the reliable reception of any particular data stream. However, the reliability performance represented in (41) is as if the other streams did not exist at all. This suggests that the typical way error occurs is not due to the interstream interference but because of the channel being in outage.
2) With $n_t = n_r = n$, the diversity gain of uncoded QAMs is equal to $n - r$; this matches the optimal diversity gain characterized in [1] for large enough r ($\geq n-1$). This observation is graphically illustrated in Fig. 7.
3) In a multiple access setting with
   - n_t users with one transmit antenna each;
   - a symmetric multiplexing gain of r/n_t per user;
   - $n_r \geq n_t$ receive antennas;

   the diversity-multiplexing tradeoff is given by [28]:

$$n_r - \frac{n_r r}{n_t}. \tag{43}$$

   Therefore this simple scheme is tradeoff-optimal.
4) With $n_t \neq n_r$, the performance of uncoded QAM's is never equal to the optimal diversity gain of the channel.

Rayleigh fading is a physically relevant fading model and we have seen the tradeoff optimality at high rates of plain uncoded QAMs using the V-BLAST architecture. We can conclude the robustness of this performance if it continues to hold for a wider class of fading distributions; this is the focus of the next section.

B. Tradeoff Optimality Over Isotropic Fading Channels

The key property of a fading distribution determining the diversity performance is the *near zero* behavior of its singular values. In particular, denoting $\phi_1, \ldots, \phi_{n_m}$ to be the increasingly ordered *squared* singular values of $\mathbf{H}$, suppose

$$\mathbb{P}\left\{\phi_1 \leq \epsilon_1, \ldots, \phi_{n_m} \leq \epsilon_{n_m}\right\} \doteq \epsilon_1^{k_1+1} \cdots \epsilon_{n_m}^{k_{n_m}+1} \tag{44}$$

for $\epsilon_1 < \cdots < \epsilon_{n_m}$. Here our notation $f(\epsilon_1, \ldots, \epsilon_{n_m}) \doteq g(\epsilon_1, \ldots, \epsilon_{n_m})$ is in the sense of

$$\lim_{\epsilon_1 \to 0} \lim_{\epsilon_2 \to 0} \cdots \lim_{\epsilon_{n_m} \to 0} \frac{\log f(\epsilon_1, \ldots, \epsilon_{n_m})}{\log g(\epsilon_1, \ldots, \epsilon_{n_m})} = 1. \tag{45}$$

We also assume that all the singular values have an exponential tail, i.e., for there exists an ϵ such that for large enough x,

$$\mathbb{P}\{\phi_\ell \geq x\} \leq e^{-\epsilon x}, \quad \forall\, \ell. \tag{46}$$

For a given near zero behavior of singular values, the tradeoff curve can be explicitly determined. We compute it for the case when k_is are increasingly ordered (as is the case for i.i.d. Rayleigh fading).

Theorem 8.1: If $k_1 < k_2 < \cdots < k_{n_m}$, then the tradeoff curve is piecewise linear with n_m segments and the sth segment (i.e., $s \leq r < s+1$) is given by:

$$(k_{n_m - s} + 1)(s + 1 - r) + \sum_{\ell = n_m - s + 1}^{n_m} (k_\ell + 1).$$

Furthermore, random Gaussian codes with block-length $T \geq k_{n_m - s} + 1$ will achieve this performance.

Proof: See Appendix VIII-A for the outage curve calculation. The proof of achievability for random Gaussian codes is a simple generalization of achievability proof in [1] and we omit it here. □

The key property of the i.i.d. Rayleigh-fading channel used in the calculation of the performance of uncoded V-BLAST transmission is the rotational symmetry of its statistics. We can thus generalize this calculation and characterize the performance of uncoded V-BLAST transmission over *isotropic* distributions on the $n \times n$ MIMO channel $\mathbf{H}$:

$$\mathbf{HQ} \text{ has the same distribution as } \mathbf{H} \quad \text{for every unitary matrix } \mathbf{Q}. \tag{47}$$

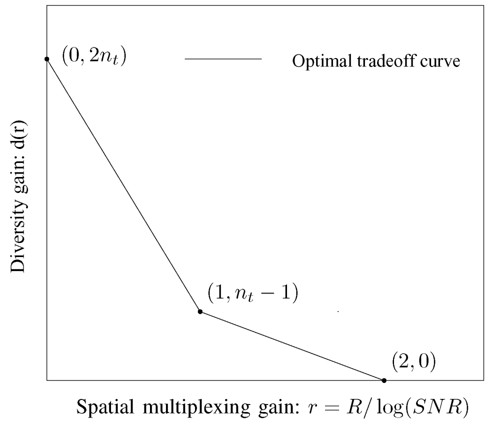

Fig. 8. The tradeoff behavior for the $n_t \times 2$ i.i.d. Rayleigh-fading channel.

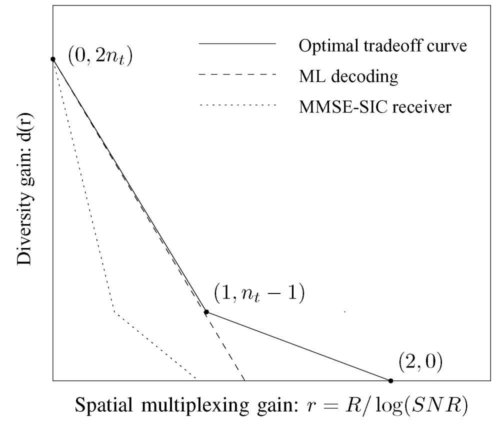

Fig. 9. Diversity performance of the D-BLAST architecture.

If the ordered singular values of the $n \times n$ MIMO channel $\mathbf{H}$ decay *slower* than the corresponding decay rate of ordered singular values of $\mathbf{H}$ with i.i.d. Rayleigh fading, then we can extend our earlier observation of tradeoff optimality of the transmission of uncoded QAM's over the V-BLAST architecture at multiplexing gains $r \geq n - 1$ on the i.i.d. Rayleigh-fading channel. We make this precise in the following proposition, delegating the proof to Appendix VIII.

Proposition 8.2: Consider $n \times n$ isotropic MIMO channels with the polynomial decay rates of its squared singular values as defined in (44). The uncoded QAM transmission over the V-BLAST architecture at multiplexing rates $r \geq n - 1$ is tradeoff optimal for every isotropic MIMO channel satisfying

$$k_i > (2i - 2) + k_1 \quad i = 2, \ldots, n,$$
$$k_1 \leq 0.$$

IX. The D-BLAST Architecture

We have seen (cf. Section VII.C) that the D-BLAST architecture with approximately universal parallel channel codes over its independent constituent data streams approaches approximately universality for large block length (cf. (40)). For any finite block length, the architecture is strictly *not* approximately universal. However, we will see in this section that by restricting the class of MIMO channels over which we demand universality, the performance of the D-BLAST architecture can be significantly improved. In particular, our focus throughout this section is with isotropic MIMO channels. We characterize the diversity performance of the D-BLAST architecture with exactly two data streams; our main result is the observation of a restricted universality result for channels with two receive antennas.

The i.i.d. Rayleigh-fading MIMO channel is also isotropic and we state our results first in this context; the calculations are relatively simple and shed insight as to why we can expect robustness when generalized to arbitrary isotropic channel distributions.

A. Tradeoff Optimality Over Rayleigh Fading Channels

Consider the $n_t \times 2$ i.i.d. Rayleigh-fading MIMO channel: the tradeoff curve is composed of two linear segments, as illustrated in Fig. 8.

A) D-BLAST and the First Segment: Consider the D-BLAST architecture with only two independent data streams

$$\begin{bmatrix} 0 & \cdots & 0 & p_{n_t} & q_{n_t} \\ \vdots & \iddots & \iddots & \iddots & 0 \\ 0 & p_2 & \iddots & \iddots & \vdots \\ p_1 & q_1 & 0 & \cdots & 0 \end{bmatrix} \tag{48}$$

here $[p_1, \ldots, p_{n_t}]$ and $[q_1, \ldots, q_{n_t}]$ are unit block-length approximately universal codes for a parallel channel with n_t sub-channels. Suppose both these codes have a data rate of

$$\frac{(n_t + 1)r}{2n_t} \log \mathsf{SNR} \quad \text{bits per symbol.} \tag{49}$$

Since the overall architecture is composed of two data streams and the transmission lasts $n_t + 1$ time symbols long, the overall data rate of the architecture is $r \log \mathsf{SNR}$ bits/symbol. Our main result is a precise characterization of the diversity performance under joint ML decoding of the streams; the proof is available in Appendix VI.

Proposition 8.1: The D-BLAST architecture in (48) with approximately universal parallel channel codes as its two data streams operated at a total multiplexing gain of r over the i.i.d. Rayleigh-fading $n_t \times n_r$ MIMO channel with $n_r \geq 2$ sees a diversity gain equal to

$$n_r \left(n_t - \frac{n_t + 1}{2} r \right). \tag{50}$$

A couple of observations follow.

1) If we set $n_r = 2$, the diversity performance in (50) is equal to $2n_t - (n_t + 1)r$; this overlaps with the optimal tradeoff curve of the channel for small enough multiplexing gains, i.e., $r \leq 1$, thus achieving the first segment for the $n_t \times 2$ i.i.d. Rayleigh fading channel (see Fig. 9).
2) From the perspective of one of the streams in the D-BLAST architecture, the best diversity performance is obtained if the other stream did not exist at all (or was decoded correctly and thus canceled exactly). Suppose this is the case: then each data stream sees a parallel channel with n_t scalar sub-channels, each of whose squared amplitudes are i.i.d.

with distribution $\chi^2_{2n_r}$. The optimal tradeoff curve for this parallel channel with a data rate of $(n_t+1)/2$ bits/symbol (cf. (49)) is

$$n_r\left(n_t - \frac{n_t+1}{2}r\right). \tag{51}$$

The diversity performance of any data stream with the other stream being perfectly canceled cannot be any more than the gain in (51). However, from the claim in Proposition 9.1 (cf. (50)), we observe that this upper bound is *exactly* equal to the diversity gain achieved even when there is inter-stream interference. There we conclude the following:

> Under the joint ML decoder, interstream interference is not the typical error event.

We study the joint ML decoder in some detail in the next section.

3) Finally, we observe that we crucially used the symmetry between the two streams in the above argument. With more than two streams, the middle streams see more interference than the outer two streams and an extension to this situation is not natural.

B) D-BLAST and ML Decoding: In this section, we discuss the ML decoding of the two data streams in the D-BLAST architecture in some detail. To make our discussions simple and concrete we focus on the simple case of $n_t = 2$; the received signal spans three time symbols and can be written as

$$[\mathbf{y}_1\mathbf{y}_2\mathbf{y}_3] = [\mathbf{h}_1\mathbf{h}_2]\begin{bmatrix} 0 & p_2 & q_2 \\ p_1 & q_1 & 0 \end{bmatrix} + [\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3].$$

The two data streams $[p_1, p_2]$ and $[q_1, q_2]$ are unit block-length approximately universal codes for a parallel channel with 2 subchannels and independent of each other. For concreteness, suppose p_1 (q_1) and p_2 (q_2) are points from a QAM constellation and correspond to bit reversal with alternative bits flipped of each other (cf. Section V.C.5). The ML decoder makes a joint decision on both these codes using the three received vectors $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$. However, due to the specific structure of the zeros in the D-BLAST architecture, the joint ML decoder can be broken down algorithmically into three separate steps.

1) We observe that the received vector at the first time symbol $\mathbf{y}_1$ gives information only about the the QAM symbol p_1:

$$\mathbf{y}_1 = p_1\mathbf{h}_2 + \mathbf{w}_1. \tag{52}$$

In particular, $\mathbf{y}_1$ specifies exactly the most significant bits of the bit representation of the QAM point p_1 (cf. Section V-C). More specifically, the number of MSB's of p_1 that can be deduced from $\mathbf{y}_1$ is with high probability equal to $\lfloor \log(|\mathbf{h}_2|^2\mathsf{SNR})\rfloor$; further more, the information about the remaining bits of p_1 depends on the noise $\mathbf{w}_1$ that is independent of the received signals at the other two time symbols. Since the QAM points p_1 and p_2 correspond to bit reversals (with alternate bits flipped) of each other, we have also deduced the *least* significant bits of $\lfloor \log(|\mathbf{h}_2|^2\mathsf{SNR})\rfloor$ of p_2.

2) The scenario at the third time symbol is identical to that at the first time symbol except that p_1 is replaced by q_2 and p_2 by q_1. In particular, we can deduce $\lfloor \log(|\mathbf{h}_1|^2\mathsf{SNR})\rfloor$ MSB's of q_2 (and the $\lfloor \log(|\mathbf{h}_1|^2\mathsf{SNR})\rfloor$ LSBs of q_1) from $\mathbf{y}_3$; further more, the information about the remaining bits of q_2 (and hence q_1) depends on the noise vector $\mathbf{w}_3$ that is independent of the received vector at the first two time symbols.

3) We are now ready to focus on the received vector at the second time symbol:

$$\mathbf{y}_2 = p_2\mathbf{h}_1 + q_1\mathbf{h}_2 + \mathbf{w}_2. \tag{53}$$

Here we know some of the LSB's of both p_2 and q_1 (due to processing of the received vector at the first and third time symbols, respectively); this reduces the randomnes s in p_2 and q_1 to another sparser QAM which is a subset of the original QAM from which they were drawn. We see from (53) is exactly the output of a 2×2 MIMO channel with uncoded QAMs transmitted over the two transmit antennas, i.e., uncoded QAM transmission over the V-BLAST architecture. Thus, the ML decoding of the two streams of the D-BLAST architecture reduces to that of a decoding uncoded QAM transmission over the V-BLAST architecture.

C) A Time–Space Code and the Second Segment: While we have seen the tradeoff optimality of the D-BLAST architecture in achieving the first segment of the $n_t \times 2$ i.i.d. Rayleigh-fading channel, there is a simple transformation of this architecture that achieves the *second segment* of the same channel. The key is to consider a *time–space* version of the space–time D-BLAST architecture: replace the transmit symbol at time symbol m over the transmit antenna k by the transmit symbol at time symbol k and transmit antenna m. In particular, the time–space version of the space–time code in (48) is

$$\begin{bmatrix} 0 & \cdots & 0 & p_1 \\ \vdots & \iddots & p_2 & q_1 \\ 0 & \iddots & q_2 & 0 \\ p_{n_t} & \iddots & \iddots & \vdots \\ q_{n_t} & 0 & \cdots & 0 \end{bmatrix}. \tag{54}$$

It is meant to be used over a channel with $n_t + 1$ transmit antennas and spans n_t time symbols long; observe that the original code in (48) is meant to be used over a channel with n_t transmit antennas and spans $n_t + 1$ time symbols long. Suppose that $[p_1, \ldots, p_{n_t}]$ and $[q_1, \ldots, q_{n_t}]$ independent unit block-length approximately universal codes for the parallel channel at rate $0.5rn_t \log \mathsf{SNR}$ bits per symbol; this corresponds to the overall code in (54) to have a total multiplexing gain on r. Our main result is a precise characterization of the diversity performance of this space–time code over the i.i.d. Rayleigh-fading channel; the proof is available in Appendix VI-A.

Proposition 9.2: The diversity gain of joint ML decoding the data streams of the time-space code in (54) at a total multiplexing rate of r bits/symbol over the $(n_t+1)\times n_r$ i.i.d. Rayleigh fading MIMO channel with $n_r \geq 2$ is equal to

$$\frac{n_t n_r}{2}(2-r). \tag{55}$$

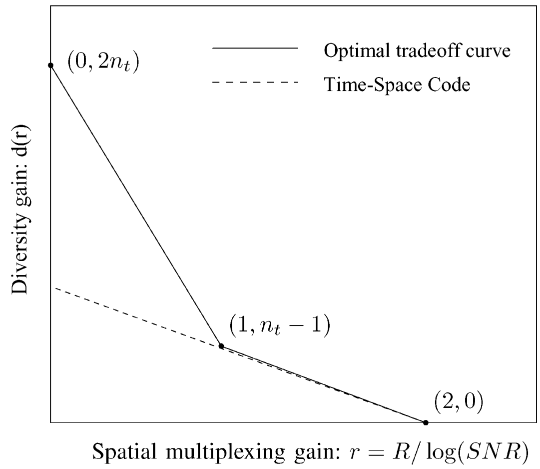

Fig. 10. D-BLAST curves versus the optimal tradeoff curve

Setting $n_r = 2$, we see that the diversity gain in (55) is equal to $2n_t - n_t r$ which overlaps with the optimal tradeoff curve for that channel for large enough multiplexing gains, i.e., $r \geq 1$; in particular, this achieves the second segment of the tradeoff curve (see Fig. 10).

D) Tradeoff Optimality Over Isotropic Channels: We demonstrate the robustness of the performance results a time–space code for the i.i.d. Rayleigh-fading channel by generalizing them to the class of isotropic fading distributions: in particular, we are interested in MIMO channel distributions which satisfy the property in (47). Further recall the definition of the polynomial decay rates of the squared singular values of the $n \times n$ MIMO channel in (44). The proofs of the results in this section are available in Appendix VIII.

Our result is the restricted approximate universality of the time–space version of the D-BLAST architecture with two data streams in achieving the second segment of the tradeoff curve; this generalizes the result in Proposition 9.2. The proof of this result is available in Appendix VIII.

Theorem 9.1: The diversity gain of joint ML decoding the data streams of the time-space code in (54) at a total multiplexing rate of r bits/symbol over any isotropic $n_{t+1} \times 2$ MIMO channel achieves the second segment of its tradeoff curve, provided

$$\begin{aligned} k_2 - k_1 &\geq 2, \\ k_1 &\leq 0. \end{aligned}$$

X. Conclusion

We have presented a precise characterization of universally tradeoff optimal codes for the MIMO channel. We also presented explicit codes for the parallel channel that are simple to encode and decode. These codes, along with the general construction in [20], completely solves the code design problem for the parallel channel. For the MIMO channel, we suggest using the D-BLAST architecture to reduce it to a parallel channel and using codes designed for the parallel channel. This approach is reasonable when the block-length is large, since in this case the initialization overhead in D-BLAST is insignificant. While, finite block length approximately universal codes for the MIMO channel have been constructed, they are not known to be simple to decode; construction of simple codes for the MIMO channel remains an open problem.

Alternative to approximately universal codes for MIMO channel, we have seen the existence of simple codes for the MIMO channel that are approximately universal for a restricted class of fading distributions. Our construction has been restricted for specific number of antenna elements; a generalization of this construction is also an interesting future research direction.

Appendix I
Converse for Approximate Universality

We want to show that if a coding scheme does not satisfy the universal code design criterion, then there exists a fading distribution such that the coding scheme is not tradeoff optimal. In the high SNR scaling of [1], a coding scheme is defined by a *discrete sequence* of codes $C(\mathsf{SNR})$ with rate $r \log \mathsf{SNR}$. If this sequence does not satisfy the approximate universality criterion, then there exists a subsequence of $C(\mathsf{SNR})$ such that for every code in the sub-sequence there exists a codeword pair such that it does not satisfy the universal criterion. For proving the existence of a fading distribution such that the original sequence is not tradeoff optimal, it is enough to find a fading distribution for which this subsequence of codes is not tradeoff-optimal. Therefore we assume that for every code in the sequence we can find a codeword pair that does not satisfy the universal criterion.

A brief note regarding our notation: we use the symbols $\doteq$ $(\dot{\geq}, \dot{\leq})$ to denote exponential equality (inequality), i.e.,

$$f(\mathsf{SNR}) \doteq \mathsf{SNR}^b \Rightarrow \lim_{\mathsf{SNR}\to\infty} \frac{\log f(\mathsf{SNR})}{\log \mathsf{SNR}} = b.$$

A) Proof of Theorem 5.1: Here we focus on the necessity of the condition for approximate universality for the MIMO channel. If a sequence of codes is not approximately universal, we show that there exists an i.i.d. distribution on ψ_ℓs such that this sequence of codes is not tradeoff optimal.

For codewords $\mathbf{X}_A$ and $\mathbf{X}_B$, the pairwise error conditioned on a channel realization, $\mathbf{H}$, can be written as (cf. (5))

$$\mathbb{P}_e(\mathbf{X}_A \to \mathbf{X}_B | \mathbf{H}) = Q\left(\sqrt{\frac{\mathsf{SNR}\sum_{\ell=1}^{n_m} |\lambda_\ell|^2 |\psi_\ell|^2}{2}}\right).$$

The approximate universality condition can then be written as

$$\min_{\sum_{\ell=1}^{n_m} \log(1+|\psi_\ell|^2\mathsf{SNR}) > r\log\mathsf{SNR}} \mathsf{SNR}\sum_{\ell=1}^{n_m} |\lambda_\ell|^2 |\psi_\ell|^2 \dot{\geq} 1.$$

Thus, if a sequence of codes does not satisfy the universal criterion then there exists a sequence of codeword pair differences, $\mathbf{D}(\mathsf{SNR})$, and a corresponding realization $\mathbf{H}^a(\mathsf{SNR})$ such that

$$\mathsf{SNR}\sum_{\ell=1}^{n_m} |\lambda_\ell(\mathsf{SNR})|^2 |\psi_\ell^a(\mathsf{SNR})|^2 < 2^{-r\epsilon \log \mathsf{SNR}} \tag{56}$$

for some positive ϵ, where $\mathbf{H}^a(\mathsf{SNR})$ satisfies

$$\sum_{\ell=1}^{n_m} \log(1+|\psi_\ell^a(\mathsf{SNR})|^2 \mathsf{SNR}) = r \log \mathsf{SNR}. \quad (57)$$

Now define $\mathbf{H}^b(\mathsf{SNR})$ as

$$|\psi_\ell^b(\mathsf{SNR})|^2 = |\psi_\ell^a(\mathsf{SNR})|^2 \cdot 2^{r\epsilon \log \mathsf{SNR}}, \quad l = 1, \ldots, n.$$

Then using (56) and (57), $\mathbf{H}^b(\mathsf{SNR})$ satisfies

$$\mathsf{SNR} \sum_{\ell=1}^{n} |\lambda_\ell(\mathsf{SNR})|^2 |\psi_\ell^b(\mathsf{SNR})|^2 < 1 \quad \text{and} \quad (58)$$

$$\sum_{\ell=1}^{n} \log\left(1+|\psi_\ell^b(\mathsf{SNR})|^2 \mathsf{SNR}\right) \geq r(1+\epsilon) \log \mathsf{SNR}. \quad (59)$$

Now, consider the i.i.d. fading distribution on $\mathbf{H}$ such that:

$$\mathbb{P}\left\{|\psi_\ell|^2 \leq \frac{1}{x}\right\} \doteq \frac{1}{x^{\frac{2}{\epsilon}}}, \quad \forall\, \ell = 1, \ldots, n. \quad (60)$$

The diversity for the code-sequence can then be upper bounded using the following sequence of steps:

i) The pairwise error for the codeword difference $D(\mathsf{SNR})$ can be lower bounded by a constant, $Q(\sqrt{0.5})$, for a range of channels such that (see (58))

$$\left\{\mathbf{H} : |\psi_\ell|^2 < |\psi_\ell^b(\mathsf{SNR})|^2, \quad \ell = 1, \ldots, n\right\}.$$

Furthermore, because of the power constraint on the input we can assume that

$$|\psi_\ell^b(\mathsf{SNR})|^2 \dot{\geq} \frac{1}{\mathsf{SNR}}, \quad \forall\, \ell = 1, \ldots, n_m \quad (61)$$

If this is not true, we can increase $\psi_\ell^b(\mathsf{SNR})$ to $\frac{1}{n_m \mathsf{SNR}}$ such that (59) and (58) still hold.

ii) Hence the probability of error can be lower bounded by

$$\mathbb{P}_e(C(\mathsf{SNR})) \geq \frac{Q(\sqrt{0.5}) \prod_{\ell=1}^{n_m} \mathbb{P}\left\{|\psi_\ell|^2 \leq |\psi_\ell^b(\mathsf{SNR})|^2\right\}}{\mathsf{SNR}^r}. \quad (62)$$

Writing

$$|\psi_\ell^b(\mathsf{SNR})|^2 = \mathsf{SNR}^{-\alpha_\ell}$$

the probability of error expression (62) can be written as (also see (60)):

$$\mathbb{P}_e(C(\mathsf{SNR})) \dot{\geq} \mathsf{SNR}^{-\left(\frac{2}{\epsilon}\sum_\ell \alpha_\ell + r\right)} \quad (63)$$

where α_ℓs satisfy (see (59) and (61)):

$$\sum_\ell (1-\alpha_\ell)^+ \geq r(1+\epsilon) \quad \text{and} \quad \alpha_\ell \leq 1.$$

Therefore

$$\sum_\ell \alpha_\ell \leq n_m - r(1+\epsilon).$$

Then the probability of error for $C(\mathsf{SNR})$ is lower bounded by (see (60) and (63)):

$$\mathbb{P}_e(C(\mathsf{SNR})) \dot{\geq} \mathsf{SNR}^{-\left(\frac{2}{\epsilon}(n_m - r(1+\epsilon)) + r\right)} = \mathsf{SNR}^{-\left(\frac{2}{\epsilon}(n_m - r) - r\right)}.$$

Thus, the diversity of the sequence of codes is upper bounded by

$$\frac{2}{\epsilon}(n_m - r) - r. \quad (64)$$

The outage curve on the other hand is given by[5]

$$\frac{2}{\epsilon}(n_m - r).$$

Thus, comparing with (64), this sequence of codes is not tradeoff optimal and hence not approximately universal.

APPENDIX II
PROOF OF THEOREM 5.2

Consider a parallel slow fading channel with L sub-channels. A permutation code over this channel can be rewritten as

$$\mathbf{C} = \left\{\sqrt{\frac{\mathsf{SNR}}{\mathsf{SNR}^r}}(q, f_2(q), \ldots, f_L(q)) | q \in \mathbf{Q}_Z\right\}$$

where

$$\mathbf{Q}_Z = \left\{(a+ib) : -\frac{\mathsf{SNR}^{r/2}}{2} \leq a, b \leq \frac{\mathsf{SNR}^{r/2}}{2}\right\} \quad (65)$$

is the integer-QAM with SNR^r points, and $f_2, \ldots, f_L$ are permutations of $\mathbf{Q}_Z$. We define the normalized product distance between two codewords as

$$\pi_d(q_1, q_2) = \frac{|q_1 - q_2|^2}{\mathsf{SNR}^r} \prod_{k=2}^{L} \frac{|f_k(q_1) - f_k(q_2)|^2}{\mathsf{SNR}^r}. \quad (66)$$

The condition for approximate universality, (24), on the other hand, can be written as

$$\pi_d(q_1, q_2) \dot{\geq} \frac{1}{\mathsf{SNR}^r}, \quad \forall\, q_1 \neq q_2. \quad (67)$$

[5] A proof for this result can be seen from Appendix V, (80), with n_t replaced by n_m and $n_t r$ replaced by r and a replaced by $\frac{2}{\epsilon}$.

The number of permutation codes with SNR^r points is given by

$$((\mathsf{SNR}^r)!)^{L-1}.$$

We now prove existence of a permutation code in this ensemble such that (67) is satisfied. We average of the *inverse* of product distance over all such codes under the uniform measure (all codes have the same probability). The intuition behind averaging the inverse of product distance is to capture the codeword differences that have small product distance, which is the event of interest.

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\pi_d}\right] &= \frac{1}{(\mathsf{SNR}^r!)^{L-1}\mathsf{SNR}^{2r}} \sum_{\substack{f_2,\ldots,f_L,\\ q_1,q_2\neq q_1}} \frac{\mathsf{SNR}^r}{|q_1-q_2|^2} \\ &\quad\times \prod_{k=2}^{L} \frac{\mathsf{SNR}^r}{|f_k(q_1)-f_k(q_2)|^2} &(68)\\ &= \frac{((\mathsf{SNR}^r-2)!)^{L-1}\mathsf{SNR}^{rL}}{(\mathsf{SNR}^r!)^{L-1}\mathsf{SNR}^{2r}} \sum_{\substack{q_1,q_2\neq q_1,\\ q_1^k,q_2^k\neq q_1^k\\ k=2,\ldots,L}} \frac{1}{|q_1-q_2|^2} \\ &\quad\times \prod_{k=2}^{L} \frac{1}{|q_1^k-q_2^k|^2}. &(69)\end{aligned}$$

The second equality is obtained by considering all permutations f_k's that map q_1 to q_1^k and q_2 to q_2^k; the number of such permutations is $((\mathsf{SNR}^r-2)!)^{L-1}$. Therefore,

$$E\left[\frac{1}{\pi_d}\right] \leq \frac{1}{\mathsf{SNR}^{Lr}}\left[\sum_{q_1,q_2\neq q_1} \frac{1}{|q_1-q_2|^2}\right]^L.$$

Because of the symmetry of the QAM, the average inverse product distance can be further upper bounded as

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\pi_d}\right] &\leq \frac{1}{\mathsf{SNR}^{Lr}}\left[\mathsf{SNR}^r\sum_{q_1\neq 0}\frac{1}{|q_1|^2}\right]^L \\ &= \left[\sum_{q_1\neq 0}\frac{1}{|q_1|^2}\right]^L. &(70)\end{aligned}$$

The summation inside the parantheses in (70) can be upper bounded by $(\log \mathsf{SNR})^2$. This implies that the expectation can be upper bounded by

$$\mathbb{E}\left[\frac{1}{\pi_d}\right] \dot{\leq} 1.$$

We conclude that there exists at least one permutation code $\mathbf{C}^a$ with the average inverse product distance less than 11. We now use this code $\mathbf{C}^a$ with good average behavior to construct a code that has a good worst-case behavior. For $\mathbf{C}^a$

$$\frac{1}{\mathsf{SNR}^{2r}}\sum_{q_1\neq q_2\in\mathbf{Q}_Z}\frac{1}{\pi_d(q_1,q_2)}\dot{\leq}1.$$

Therefore,

$$\frac{1}{\mathsf{SNR}^r}\sum_{q_1\in\mathbf{Q}_Z} g(q_1)\dot{\leq}\mathsf{SNR}^r$$

where

$$g(q_1) = \sum_{q_2\in\mathbf{Q}_Z, q_2\neq q_1}\frac{1}{\pi_d(q_1,q_2)}.$$

Thus, at least half of the q_1's have $g(q_1)\dot{\leq}\mathsf{SNR}^r$. By expurgating at most half the codewords, we can construct a code $\mathbf{C}^b$ such that

$$g(q_1)\dot{\leq}\mathsf{SNR}^r, \quad \forall\, q_1. \tag{71}$$

This implies that for every $q_2 \neq q_1$,

$$\pi_d(q_1,q_2)\dot{\geq}\frac{1}{\mathsf{SNR}^r};$$

this is precisely the criterion for approximate universality (67). Finally, expurgating at most of half of the codeword reduces the rate of the code by at most one and hence does no change the multiplexing gain. Thus, there exist approximately universal permutation codes.

A. Product Distance Distribution

A statement much more stronger that that made about the code $\mathbf{C}^b$ constructed in Section II. The result below characterizes the behavior of the product distance π_d, cf. (66), (rather than just a lower bound, which is what was required for approximate universality), and hence can be thought of as a weight distribution result for the product distance.

Theorem 2.1: Consider a parallel slow fading channel with L subchannels. There exists a permutation code with SNR^r points over this channel such that the number of codeword pairs that have a product distance less than SNR^{k-r} is $\Theta(\mathsf{SNR}^{r+k})$, for k in $[0,r]$.

Proof: We start with the code $\mathbf{C}^b$ constructed above that satisfies (71): then for each q_1, the number of codewords which are at a product distance less than SNR^{k-r} is $\Theta(\mathsf{SNR}^k)$, for k in $[0,r]$ (otherwise such a code will not satisfy (71)). Considering all possible values of q_1, the number of codeword difference that have product distance less than SNR^{k-r} is $\Theta(\mathsf{SNR}^{r+k})$, for k in $[0,r]$. □

Appendix III
Proof of Theorem 5.3

Let the binary representation of integers a_1 and a_2 be

$$\begin{aligned}a_1 &= b_{R/2}^1\cdots b_1^1\\ a_2 &= b_{R/2}^2\cdots b_1^2.\end{aligned}$$

Let k be the largest integer such that $b_k^1 \neq b_k^2$. Then without any loss of generality we can assume that $b_k^1 = 1$ and $b_k^2 = 0$. We also write

$$b_i = b_i^1 = b_i^2, \quad \forall\, k+1 \leq i \leq R/2$$

for notational convenience as well as to emphasize that the largest $R/2 - k$ bits are the same. Now, let l be the smallest integer such that $b_{k-l}^1 \geq b_{k-l}^2$. Note that this implies that

$$b_{k-i}^1 = 0 \quad b_{k-i}^2 = 1, \quad \forall\, 1 \leq i \leq l-1$$

which is similar to the codeword pair that was the counter example given for the fact that simple bit-reversal is not universal (see Section V-C3). Here we essentially prove that such pairs are the only reason that the simple bit reversal is not approximately universal and bit reversal with alternated bit flipping can tackle this problem. We consider the following subcases.

- If no such l exists, Then a_is can be written as

$$a_1 = b_{R/2} \cdots b_{k+1} 10 \cdots 0$$
$$a_2 = b_{R/2} \cdots b_{k+1} 01 \cdots 1$$

  and $B(a_i)$s can be written as

$$B(a_1) = 101 \cdots$$
$$B(a_2) = 010 \cdots$$

  Thus $B(a_1) - B(a_2)$ is lower bounded by $2^{R/2-2}$, hence (31) is satisfied.
- $l \geq 2$: Then a_is can be written as

$$a_1 = b_{R/2} \cdots b_{k+1} 10 \cdots 0 b_{k-l}^1 b_{k-l-1}^1 \cdots b_1^1$$
$$a_2 = b_{R/2} \cdots b_{k+1} 01 \cdots 1 b_{k-l}^2 b_{k-l-1}^2 \cdots b_1^2$$

  then, the difference $a_1 - a_2$ can be lower bounded by 2^{k-l-1} and $B(a_i)$s can be written as

$$B(a_1) = \overline{b_1^1} b_2^1 \cdots \overline{b_{k-l-1}^1} b_{k-l}^1 1101 \cdots 0 b_{k+1} \cdots b_{R/2}$$
$$B(a_2) = \overline{b_1^1} b_2^2 \cdots \overline{b_{k-l-1}^2} b_{k-l}^2 0010 \cdots 1 b_{k+1} \cdots b_{R/2}.$$

  Then the difference $|B(a_1) - B(a_2)|$ is lower bounded by $2^{R/2-(k-l)-2}$ (here we have assumed that b_{k-l} is not flipped, i.e., $k-l$ is even; if $k-l$ is odd, then same argument hold with a_2 and a_1 reversed). Thus, the product distance is lower bounded by $\frac{1}{8 \cdot 2^{R/2}}$ (which is the one in (31)).
- If $l = 1$: then a_is can be written as

$$a_1 = b_{R/2} \cdots b_{k+1} 1 b_{k-1}^1 \cdots b_1^1$$
$$a_2 = b_{R/2} \cdots b_{k+1} 0 b_{k-1}^2 \cdots b_1^2$$

  then the difference $a_1 - a_2$ can be lower bounded by 2^{k-2} (since $b_{k-1}^1 \geq b_{k-1}^2$). The $B(a_i)$s can be written as

$$B(a_1) = \overline{b_1^1} b_2^1 \cdots \overline{b_{k-1}^1} 1 \overline{b_{k+1}} b_{k+2} \cdots b_{R/2}$$
$$B(a_2) = \overline{b_1^2} b_2^1 \cdots \overline{b_{k-1}^2} 0 \overline{b_{k+1}} b_{k+2} \cdots b_{R/2}$$

  and the difference $|B(a_1) - B(a_2)|$ is lower bounded by $2^{R/2-k}$ (here we have assumed that b_{k-1} is flipped, i.e., k is even; same is true if k is odd). Thus, the product distance is lower bounded by $\frac{1}{4 \cdot 2^{R/2}}$.

APPENDIX IV
PROOF OF THEOREM 5.4

We again consider the I and Q channels separately. Then we want to define $L - 1$ permutations of the PAM such that the corresponding permutation code is approximately universal. We consider the q-digit representation of the PAM. For a PAM with q^n points and number it from left to right by 0 to $q^n - 1$ (in term of the rate , behaves like $\frac{\log_2 R/2}{log_2 q}$). For showing that a universally decodable system satisfies the product distance criterion, we have to resort to irregularly spaced PAM's. For every mth least significant q-bit change, we put a gap of gq^{m-1}. Similar to the two subchannel case, using this construction we prove that any universally decodable scheme satisfies the condition for approximate universality: consider any two codewords; suppose for the ℓth subchannel their k_ℓ MSB's are the same and $(k_\ell + 1)$th MSB is different. By construction of the irregularly spaced QAM, the normalized (by q^n) separation in the ℓth coordinate is lower bounded by

$$\frac{gq^{n-k_\ell-1}}{q^n} = gq^{-(k_\ell+1)}.$$

The universal decodability condition implies that if $\sum_\ell k_\ell \geq n$, then there exists a unique codeword corresponding to the MSBs. Therefore, the k_ℓs must satisfy

$$\sum_\ell k_\ell < n.$$

Thus, the product distance can be lower bounded by

$$\begin{aligned} |d_1 d_2 \cdots d_L|^{2/L} &\geq \left(\prod_\ell g^2 q^{-2k_\ell - 2}\right)^{1/L} \\ &= \left(q^{-2\sum_\ell k_\ell}\right)^{1/L} \\ &\geq \frac{g^2}{q^2} q^{-2n/L} \\ &= \frac{g^2}{q^2 2^R} \\ &\doteq \frac{1}{2^R} \end{aligned} \tag{72}$$

implying that the code satisfies the approximate universality condition (24). For a PAM of size q^n, the (normalized) increase in size is given by

$$\begin{aligned} &\sum_{m=1}^{n} gq^{m-1-n} (\text{number of } m\text{th LS } q\text{-bit changes}) \\ &= \sum_{m=1}^{n} gq^{m-n} (q^{n-m}) \\ &= gn. \end{aligned} \tag{73}$$

In the high SNR scaling

$$q^n = \mathsf{SNR}^r \Longrightarrow gn \doteq \log \mathsf{SNR}.$$

Thus the extra spacing does not affect the multiplexing gain.

We also note the Theorem 5.4 is true even if the field size q is growing like $\log \mathsf{SNR}$. Note that if q grew like a polynomial in

SNR, i.e., like SNR^{ϵ}, then we can no longer ignore q in (72) and such a code then will not be approximately universal. We also have to show that the power gain because of the gaps still increases slowly enough so as to not affect the multiplexing gain. For a PAM of size q^n, the increase in size is, cf. (73)

$$gn \dot{\leq} \mathsf{SNR}^{\epsilon}, \quad \forall\, \epsilon > 0.$$

Therefore, the extra spacing does not affect the diversity-multiplex tradeoff.

Appendix V
Proof of Propositions 6.1

We use an approximately universal parallel channel code, (e.g., a permutation code $[p_1, \ldots, p_{n_t}]$ with total rate Rn_t) over the MISO channel in a diagonal fashion

$$\begin{bmatrix} p_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_{n_t} \end{bmatrix}. \tag{74}$$

We prove that scheme (74) is tradeoff optimal for MISO channel with i.i.d. fading coefficients. Since it operationally converts the MISO channel into a parallel channel, we only need to match the outage probabilities of the MISO channel and the corresponding parallel channel. The outage probability of the MISO channel is given by

$$\mathbb{P}\left\{\log(1 + \sum_{i=1}^{n_t} |h_i|^2 \mathsf{SNR}) \leq r \log \mathsf{SNR}\right\} \tag{75}$$

For the equivalent parallel channel, the outage probability is given by

$$\mathbb{P}\left\{\sum_{i=1}^{n_t} \log\left(1 + |h_i|^2 \mathsf{SNR}\right) \leq n_t r \log \mathsf{SNR}\right\} \tag{76}$$

The near zero behavior of sum of $|h_i|^2$s can be upper and lower bounded as

$$\begin{aligned}\left(\mathbb{P}\left\{|h_1|^2 < \frac{x}{n_t}\right\}\right)^{n_t} &\leq \mathbb{P}\left\{\sum_{i=1}^{n_t} |h_i|^2 \leq x\right\} \\ &\leq (\mathbb{P}\{|h_i|^2 \leq x\})^{n_t}.\end{aligned}$$

Since the upper and lower bound have the same decay rate, the probability of outage of the MISO channel, (75), has a decay rate of

$$\left(\frac{\mathsf{SNR}^r}{\mathsf{SNR}}\right)^{an_t}. \tag{77}$$

Thus, the outage curve of the MISO channel with i.i.d. fading coefficients with the a denoting the decay rate of $|h_1|^2$ near zero is

$$d_{\text{out}}(r) = an_t(1 - r).$$

The second outage probability, (76), is somewhat more involved. Define α_i by

$$|h_i|^2 = \frac{\mathsf{SNR}^{\alpha_i}}{\mathsf{SNR}}.$$

In this notation, the outage condition for the parallel channel can be written as

$$\sum_i \alpha_i \leq n_t r. \tag{78}$$

Since the subchannels are independent, the outage probability (cf. (76)) has the decay rate

$$\max_{\alpha_1, \ldots, \alpha_{n_t}} \prod_{i=1}^{n_t} \left(\frac{\mathsf{SNR}^{\alpha_i}}{\mathsf{SNR}}\right)^a \tag{79}$$

where the maximization is under the constraint in (78). Thus, the decay rate of the outage probability expression in (79) is

$$\left(\frac{\mathsf{SNR}^r}{\mathsf{SNR}}\right)^{n_t a} \tag{80}$$

the same as that in (77); this completes the proof.

Appendix VI
Proof of Proposition 9.1

We prove that the diversity obtained by the code in (48) is $n_r(n_t - \tilde{r})$, where $\tilde{r} \log \mathsf{SNR}$ is the rate of codes $[p_1, \ldots, p_{n_t}]$ and $[q_1, \ldots, q_{n_t}]$.

The pairwise probability of error, averaged over the Rayleigh fading channel with n_r receive antennas is given by [8]

$$\mathbb{P}(\mathbf{X}_0 \to \mathbf{X}_1) \leq \frac{1}{\det\left(\mathbf{I} + (\mathbf{X}_0 - \mathbf{X}_1)(\mathbf{X}_0 - \mathbf{X}_1)^\dagger\right)^{n_r}}.$$

The difference codeword pair can be written as

$$\mathbf{X}_0 - \mathbf{X}_1 = \begin{bmatrix} 0 & \cdots & 0 & d^p_{n_t} & d^q_{n_t} \\ \vdots & \iddots & \iddots & \iddots & 0 \\ 0 & d^p_2 & \iddots & \iddots & \vdots \\ d^p_1 & d^q_1 & 0 & \cdots & 0 \end{bmatrix} \tag{81}$$

where $\mathbf{d}^p = [d^p_1, \ldots, d^p_{n_t}]$ and $\mathbf{d}^q = [d^q_1, \ldots, d^q_{n_t}]$ are the codeword difference for a permutation code.

Expanding $(\mathbf{X}_0 - \mathbf{X}_1)(\mathbf{X}_0 - \mathbf{X}_1)^\dagger$ in terms of the streams, we get

$$\begin{aligned}&\det\left(\mathbf{I} + (\mathbf{X}_0 - \mathbf{X}_1)(\mathbf{X}_0 - \mathbf{X}_1)^\dagger\right) \\ &\qquad \geq \left|d^p_1 d^p_2 \cdots d^p_{n_t}\right|^2 + |d^q_1 d^q_2 \cdots d^q_{n_t}|^2.\end{aligned}$$

The probability of error can be upper bounded using the union bound:

$$\mathbb{P}_{\mathrm{e}} \leq \frac{1}{\mathsf{SNR}^{2\tilde{r}}} \sum_{\substack{\mathbf{x}_0, \\ \mathbf{X}_1 \neq \mathbf{X}_0}} \frac{1}{\left(\left|d^p_1 d^p_2 \cdots d^p_{n_t}\right|^2 + \left|d^q_1 d^q_2 \cdots d^q_{n_t}\right|^2\right)^{n_r}}.$$

This upper bound can be broken into two summations: one corresponding to where both the streams are different and the other summation where one of the streams is the same. Suppose the same code is used for both the streams; now the upper bound can be simplified

$$\mathbb{P}_{\mathrm{e}} \leq \frac{1}{\mathsf{SNR}^{2\tilde{r}}} \sum_{\substack{\mathbf{d}^p \neq \mathbf{0},\\ \mathbf{d}^q \neq \mathbf{0}}} \frac{1}{\left(|d_1^q d_2^q \cdots d_{n_t}^q|^2 + |d_1^q d_2^q \cdots d_{n_t}^q|\right)^{n_r}} + \frac{2}{\mathsf{SNR}^{\tilde{r}}} \sum_{\mathbf{d}^p \neq \mathbf{0}} \frac{1}{|d_1^p d_2^p \cdots d_{n_t}^p|^{2n_r}}.$$

The arithmetic mean-geometric mean inequality for the term inside the first summation yields

$$\begin{aligned}\mathbb{P}_{\mathrm{e}} \dot{\leq} & \frac{1}{\mathsf{SNR}^{2\tilde{r}}} \sum_{\mathbf{d}^p \neq \mathbf{0}, \mathbf{d}^q \neq \mathbf{0}} \frac{1}{|d_1^q d_2^q \cdots d_{n_t}^q|^{n_r} |d_1^q d_2^q \cdots d_{n_t}^q|^{n_r}} \\ & + \frac{2}{\mathsf{SNR}^{\tilde{r}}} \sum_{\mathbf{d}^p \neq \mathbf{0}} \frac{1}{|d_1^p d_2^p \cdots d_{n_t}^p|^{2n_r}}, \\ = & \left(\frac{1}{\mathsf{SNR}^{\tilde{r}}} \sum_{\mathbf{d}_1^p \neq 0} \frac{1}{|d_1^p d_2^p \cdots d_{n_t}^p|^{n_r}}\right)^2 \\ & + \frac{2}{\mathsf{SNR}^{\tilde{r}}} \sum_{\mathbf{d}_1^p} \frac{1}{|d_1^p d_2^p \cdots d_{n_t}^p|^{2n_r}}.\end{aligned}$$

Now, we use the product distance distribution result in Appendix II.A to separately bound the two summations on the RHS. The result says that the number of codeword differences pairs with $|d_1^p \cdots d_{n_t}^p|^2$ less than $\frac{\mathsf{SNR}^{n_t}}{\mathsf{SNR}^{\tilde{r}-k}}$ is

$$\mathsf{SNR}^{\tilde{r}+k}$$

for k in $[0, r]$. Using this result, the first term can be upper bounded as

$$\begin{aligned}&\left(\frac{1}{\mathsf{SNR}^{\tilde{r}}} \sum_{p_1 \neq 0} \frac{1}{|d_1^p d_2^p \cdots d_{n_t}^p|^{n_r}}\right)^2 \\ &\dot{\leq} \left(\max_{k \in [0,\tilde{r}]} \frac{1}{\mathsf{SNR}^{\tilde{r}}} \mathsf{SNR}^{k+\tilde{r}} \frac{\mathsf{SNR}^{n_r(\tilde{r}-k)/2}}{\mathsf{SNR}^{n_t n_r/2}}\right)^2 \\ &= \left(\max_{k \in [0,\tilde{r}]} \mathsf{SNR}^{(1-n_r/2)k} \frac{\mathsf{SNR}^{n_r \tilde{r}/2}}{\mathsf{SNR}^{n_t n_r/2}}\right)^2 \\ &= \mathsf{SNR}^{-(n_r(n_t-\tilde{r}))}\end{aligned}$$

for $n_r \geq 2$. The second term corresponds to the error when one of the streams is decoded correctly and can be directly verified to be of the correct order. Alternatively,

$$\begin{aligned}&\frac{2}{\mathsf{SNR}^{\tilde{r}}} \sum_{\mathbf{d}_1^p \neq \mathbf{0}} \frac{1}{|d_1^p d_2^p \cdots d_{n_t}^p|^{2n_r}} \\ &\dot{\leq} \max_{k \in [0,\tilde{r}]} \frac{1}{\mathsf{SNR}^{\tilde{r}}} \mathsf{SNR}^{k+\tilde{r}} \frac{\mathsf{SNR}^{n_r(\tilde{r}-k)}}{\mathsf{SNR}^{n_r n_t}} \\ &= \max_{k \in [0,\tilde{r}]} \mathsf{SNR}^{k(1-n_r)} \frac{\mathsf{SNR}^{n_r \tilde{r}}}{\mathsf{SNR}^{n_r n_t}} \\ &= \mathsf{SNR}^{-(n_r(n_t-\tilde{r}))}.\end{aligned}$$

Thus, combining the two upper bounds, for $n_r \geq 2$ there exists a code such that the diversity gain is

$$n_r(n_t - \tilde{r}).$$

Taking $\tilde{r} = \frac{n_t+1}{2} r$ proves Proposition 9.1.

A. *Proof of Proposition 9.2*

For the $(n_t + 1) \times 2$ channel, we transposed the code in (48) which was used for achieving the first segment $n_t \times 2$ channel. The probability of error can be calculated using a union bound calculation. The pairwise probability of error is given by

$$1/\det(\mathbf{I} + (\mathbf{X}_0 - \mathbf{X}_1)(\mathbf{X}_0 - \mathbf{X}_1)^\dagger)^{n_r}.$$

Since

$$\det(\mathbf{I} + (\mathbf{X}_0 - \mathbf{X}_1)(\mathbf{X}_0 - \mathbf{X}_1)^\dagger) = \det(\mathbf{I} + (\mathbf{X}_0 - \mathbf{X}_1)^\dagger(\mathbf{X}_0 - \mathbf{X}_1))$$

the union bound calculation for calculating the probability of error is exactly the same as same as (48) case. Therefore the diversity obtained by this scheme is given by $n_r n_t - n_r \tilde{r}$. But in this case we are coding over a block-length of n_t, thus the actual tradeoff curve is $n_r n_t (2 - r)/2$, where $r \log \mathsf{SNR}$ is the per symbol rate of the channel.

APPENDIX VII
PROOF OF PROPOSITION 8.1

The scheme of sending n_t QAM constellations can be written as

$$\left\{\mathbf{q} = \sqrt{\frac{\mathsf{SNR}}{\mathsf{SNR}^{\frac{r}{n_t}}}}(i_1, \ldots, i_{2n_t}) | i_j \in \mathbf{P}_Z, \forall\, j = 1, \ldots, 2n_t\right\}$$

where $\mathbf{P}_Z$ is the integer PAM constellation with $\mathsf{SNR}^{\frac{r}{2n_t}}$ points. For a Rayleigh fading channel, the probability of pairwise error averaged over the fading statistics is given by [8]

$$\mathbb{P}(\mathbf{q}_1 \to \mathbf{q}_2) \leq \left[\frac{1}{1 + ||\mathbf{q}_1 - \mathbf{q}_2||^2}\right]^{n_r}.$$

Using the union bound the probability of error is bounded by

$$\begin{aligned}\mathbb{P}_e &\leq \frac{1}{\mathsf{SNR}^r} \sum_{\mathbf{q}_1 \neq \mathbf{q}_2} \left[\frac{1}{1 + ||\mathbf{q}_1 - \mathbf{q}_2||^2}\right]^{n_r} \\ &\leq \sum_{\mathbf{q} \neq \mathbf{0}} \frac{1}{||\mathbf{q}||^{2n_r}}\end{aligned}$$

where $\mathbf{0}$ is the $2n_t$ dimensional vector of zeros. The second step follows from the symmetry of the QAM. To compute the summation in on RHS, we split into a summation over vectors such that all its components are nonzero and then use the arithmetic mean-geometric mean (am/gm) inequality. We denote a subset of the index set, $\{1, 2, \dots, 2n_t\}$, by S. Then the summation can be simplified as

$$\begin{aligned}
&\sum_{\mathbf{q}\neq\mathbf{0}} \frac{1}{||\mathbf{q}||^{2n_r}} \\
&= \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \sum_{(i_1,\dots,i_{2n_t})\neq\mathbf{0}} \frac{1}{(|i_1|^2+\cdots+|i_{n_t}|^2)^{n_r}} \\
&= \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \sum_{S} \sum_{i_j\neq 0: j\in S} \frac{1}{(|i_1|^2+\cdots+|i_{2n_t}|^2)^{n_r}} \\
&\le \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \sum_{S} \sum_{i_j\neq 0: j\in S} \prod_{j\in S} \frac{1}{|i_j|^{2n_r/|S|}} \\
&\quad \text{using A.M./G.M. inequality} \\
&\le \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \sum_{S} \left(\sum_{i_1\neq 0} \frac{1}{|i_1|^{2n_r/|S|}} \right)^{|S|}
\end{aligned}$$

Since the range of summation $|i_1|$ is growing with SNR, the inner summation has different behavior for depending on whether $|S|$ is larger/smaller than $2n_r$

$$\begin{aligned}
\sum_{i_1\neq 0} \frac{1}{|i_1|^{2n_r/|S|}} &\doteq (\mathsf{SNR}^{r/2n_t})^{1-2n_r/|S|}, \quad \text{if } 2n_r < |S| \\
&\doteq 1, \quad \text{otherwise.}
\end{aligned}$$

But because of the definition of S, $|S|$ is naturally upper bounded by $2n_t$. Thus, for $n_r \ge n_t$, the probability of error can be upper bounded by

$$\mathbb{P}_e \dot{\le} \mathsf{SNR}^{-(n_r - \frac{rn_r}{n_t})}. \tag{82}$$

On the other hand, if $n_r < n_t$, then the probability of error can be upper bounded as

$$\begin{aligned}
\mathbb{P}_e &\le \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \sum_{S} \left(\sum_{i_1\neq 0} \frac{1}{|i_1|^{2n_r/|S|}} \right)^{|S|} \\
&\dot{\le} \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \sum_{2n_r\le|S|\le 2n_t} (\mathsf{SNR}^{r/2n_t})^{|S|-2n_r} \\
&\dot{\le} \frac{\mathsf{SNR}^{\frac{rn_r}{n_t}}}{\mathsf{SNR}^{n_r}} \max_{2n_r\le|S|\le 2n_t} (\mathsf{SNR}^{r/2n_t})^{|S|-2n_r} \\
&= \mathsf{SNR}^{-(n_r-r)}.
\end{aligned}$$

APPENDIX VIII
ISOTROPIC MIMO CHANNELS

We concentrate on the rotationally invariant distributions. For this class, the singular value distribution determines the channel statistics completely. Let $f(\boldsymbol{\phi})$ be the density function of the ordered *squared* singular values, $\boldsymbol{\phi}$, of the channel gain matrix. In terms of notation of Section VII, we have

$$\phi_\ell = \psi_\ell^2 \quad \text{for } \ell = 1, \dots, n_m$$

where ψ_ℓs are the singular values of $\mathbf{H}$. In the high SNR regime, we are only interested in the near zero behavior of $\boldsymbol{\phi}$. Therefore, in the scaling of interest, f can be assumed to be of the form

$$f(\boldsymbol{\phi}) \doteq \phi_1^{k_1} \cdots \phi_{n_m}^{k_{n_m}} \mathbb{1}_{\phi_1 \le \phi_2 \le \cdots \le \phi_{n_m}} \tag{83}$$

This is same as the earlier definition of distribution of the squared singular values

$$\mathbb{P}\{\phi_1 \le \epsilon_1, \dots, \phi_{n_m} \le \epsilon_{n_m}\} \doteq \epsilon_1^{k_1+1} \cdots \epsilon_{n_m}^{k_{n_m}+1},$$

for $\epsilon_1 < \cdots < \epsilon_{n_m}$.

For Rayleigh fading distribution, $\boldsymbol{\phi}$ has the Wishart distribution which can be reduced to this polynomial form by ignoring the exponential terms in the Wishart distribution (for the exact expression, see [1])

$$r(\boldsymbol{\phi}) \doteq \phi_1^{r_1} \phi_2^{r_2} \cdots \phi_{n_m}^{r_{n_m}} \mathbb{1}_{\phi_1 \le \phi_2 \le \cdots \le \phi_{n_m}} \tag{84}$$

where $r_\ell = |n_t - n_r| + 2(\ell - 1)$.

In this appendix, first we characterize the outage curve in terms of k_is for general f. Then, we use this characterization to characterize restricted universality for codes based on the V-BLAST and D-BLAST architecture proposed in Section VIII and Section IX respectively.

A. *The Outage Curve for General Fading Distributions*

For a general fading distribution, F, we want to calculate the probability of outage. The outage event can be written as

$$\sum_{\ell=1}^{n_m} \log(1 + \phi_\ell \mathsf{SNR}) \le r \log \mathsf{SNR}.$$

If we write

$$\phi_\ell = \mathsf{SNR}^{-\alpha_\ell} \tag{85}$$

then the induced distribution (from (83)) on the ordered vector $\boldsymbol{\alpha}$ is

$$p(\boldsymbol{\alpha}) \doteq \mathsf{SNR}^{-\alpha_1(k_1+1)} \cdots \mathsf{SNR}^{-\alpha_{n_m}(k_{n_m}+1)} \tag{86}$$

which can obtained by change of variables (85). The outage probability will be dominated by the $\boldsymbol{\alpha}$ that is on the boundary of outage and has smallest SNR exponent. More precisely, using *Laplace's method* as in [1], the outage curve is the solution to the optimization problem

$$\inf_{\boldsymbol{\alpha}\in A'} \sum_{\ell} (k_\ell + 1)\alpha_\ell \tag{87}$$

where

$$A^{'} = \left\{ \boldsymbol{\alpha} : \alpha_1 \geq \cdots \geq \alpha_{n_m} \geq 0 \text{ and } \sum_{\ell} (1-\alpha_\ell)^+ \leq r \right\}.$$

The fact that α_ℓs are positive uses our assumption that the singular values have an exponential tail. Let's assume for some integer s, $s \leq r < s+1$. Then, if

$$\begin{aligned} k_\ell < k_{n_m - s} \quad & \text{for } \ell = 1, \ldots, n_m - s - 1 \\ k_\ell > k_{n_m - s} \quad & \text{for } \ell = n_m - s + 1, \ldots, n_m \end{aligned} \tag{88}$$

then the optimizing $\boldsymbol{\alpha}$ in (87) is given by

$$\begin{aligned} \alpha_\ell^* &= 1 \quad \text{for } \ell = 1, \ldots, n_m - s - 1 \\ \alpha_{n_m - s}^* &= s + 1 - r \\ \alpha_\ell^* &= 0 \quad \text{for } \ell = n_m - s + 1, \ldots, n_m. \end{aligned}$$

The corresponding outage curve is given by

$$\begin{aligned} d_{\text{out}}(r) = (k_{n_m - s} + 1)(s + 1 - r) \\ + \sum_{\ell = n_m - s + 1}^{n_m} (k_\ell + 1), \quad s \leq r < s + 1. \end{aligned} \tag{89}$$

In particular, we would like to stress that if all the k_is are increasingly ordered then the $\boldsymbol{\alpha}$ that dominates the outage probability for fading density f is the same one that dominates the outage probability for for i.i.d. Rayleigh fading.

B. Restricted Universality of V-BLAST and D-BLAST

We want to prove that the simple QAM code for the V-BLAST architecture and codes based on using permutation codes over the D-BLAST architecture are universal over a class of isotropic fading distributions. We know that all these codes are tradeoff optimal for the i.i.d. Rayleigh fading channel under the union bound calculation. We exploit this fact to prove optimality over isotropic distributions that fade *slower* than i.i.d. Rayleigh fading.

We denote the diagonal matrices with entries $\boldsymbol{\psi}$, the singular values of the the channel gain matrix, and $\boldsymbol{\lambda}$, the singular values of the codeword difference matrix, as $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$. Then the probability of pairwise error averaged over the channel statistics can be written as (see (5))

$$\begin{aligned} &\mathbb{P}_e(\mathbf{X}_A \to \mathbf{X}_B) \\ &\quad = \int_{\mathbf{H}} Q\left(\sqrt{\frac{\|\mathbf{HD}\|^2}{2}}\right) d\mathbf{H} \\ &\quad = \int_{\boldsymbol{\Psi}} \int_{\mathbf{V}_1} Q\left(\sqrt{\frac{\|\boldsymbol{\Psi}\mathbf{V}_1\mathbf{U}_2\boldsymbol{\Lambda}\|^2}{2}}\right) d\mathbf{V}_1 d\boldsymbol{\Psi} \\ &\quad = \int_{\mathbf{Phi}} \int_{\mathbf{V}_1} Q\left(\sqrt{\frac{\|\boldsymbol{\Phi}^{1/2}\mathbf{V}_1\mathbf{U}_2\boldsymbol{\Lambda}\|^2}{2}}\right) d\mathbf{V}_1 d\boldsymbol{\Phi} \\ &\quad = \int_{\boldsymbol{\Phi}} \int_{\mathbf{V}} Q\left(\sqrt{\frac{\|\boldsymbol{\Phi}^{1/2}\mathbf{V}\boldsymbol{\Lambda}\|^2}{2}}\right) d\mathbf{V} d\boldsymbol{\Phi} \end{aligned} \tag{90}$$

where the last two steps use the independence of $\boldsymbol{\Phi}$ and $\mathbf{V}_1$ and rotational invariance of $\mathbf{V}_1$, respectively. The integral with respect to $\mathbf{V}$ is taken with respect to the Haar measure and does not depend on the distribution of $\boldsymbol{\Phi}$ and is only a function of the realization $\boldsymbol{\Psi}$ and the code.

Now, the probability of error can be upper bounded using a union bound

$$\mathbb{P}_e \leq \frac{1}{\mathsf{SNR}^r} \sum_{\boldsymbol{\Lambda}} \int_{\boldsymbol{\Phi}} \int_{\mathbf{V}} Q\left(\sqrt{\frac{\|\boldsymbol{\Phi}^{1/2}\mathbf{V}\boldsymbol{\Lambda}\|^2}{2}}\right) d\mathbf{V} d\boldsymbol{\Psi}$$

where the summation is over all possible codeword difference pairs. Since all the terms are positive, interchanging the order of the summation and integration the union bound can be written as

$$\mathbb{P}_e \leq \int_{\boldsymbol{\Phi}} \left(\frac{1}{\mathsf{SNR}^r} \sum_{\boldsymbol{\Lambda}} \int_{\mathbf{V}} Q\left(\sqrt{\frac{\|\boldsymbol{\Phi}^{1/2}\mathbf{V}\boldsymbol{\Lambda}\|^2}{2}}\right) d\mathbf{V} \right) d\boldsymbol{\Phi}.$$

The term inside the outer integral only depends on the code and the channel realization $\boldsymbol{\Phi}$ and not on the fading distribution. We denote it by $g(\boldsymbol{\phi})$. Then the smart union bound can be written as

$$\mathbb{P}_e \leq \int_{\boldsymbol{\phi}} g(\boldsymbol{\phi}) f(\boldsymbol{\phi}) d\boldsymbol{\phi}$$

where f is the density function of $\boldsymbol{\phi}$. Similarly the upper bound corresponding to the smart union bound is given by

$$\mathbb{P}_e \leq \mathbb{P}(\mathcal{H}) + \int_{\boldsymbol{\Psi} \notin \mathcal{H}} g(\boldsymbol{\phi}) f(\boldsymbol{\phi}) d\boldsymbol{\phi}$$

where $\mathcal{H}$ is the set of all channel realizations in outage. If we assume that the union bound is tight for Rayleigh fading, then it implies

$$\int_{\boldsymbol{\phi} \notin \mathcal{H}} g(\boldsymbol{\phi}) r(\boldsymbol{\phi}) d\boldsymbol{\phi} \dot{\leq} \mathsf{SNR}^{-d_R^*(r)} \tag{91}$$

where $r(\boldsymbol{\phi})$ is the is density for the i.i.d. Rayleigh-fading channel and $d_R^*(r)$ is the corresponding outage curve. We use $d_F^*(r)$ to denote the optimal curve for a generic density f.

Then, for any f the second term in (91) can be upper bounded as

$$\begin{aligned} &\int_{\boldsymbol{\phi} \notin \mathcal{H}} g(\boldsymbol{\phi}) f(\boldsymbol{\phi}) d\boldsymbol{\phi} \\ &\quad = \int_{\boldsymbol{\phi} \notin \mathcal{H}} \frac{f(\boldsymbol{\phi})}{r(\boldsymbol{\phi})} g(\boldsymbol{\phi}) r(\boldsymbol{\phi}) d\boldsymbol{\phi} \\ &\quad \leq \left(\max_{\boldsymbol{\phi} \notin \mathcal{H}} \frac{f(\boldsymbol{\phi})}{r(\boldsymbol{\phi})} \right) \int_{\boldsymbol{\phi} \notin \mathcal{H}} g(\boldsymbol{\phi}) r(\boldsymbol{\phi}) d\boldsymbol{\phi} \\ &\quad \dot{\leq} \left(\max_{\boldsymbol{\phi} \notin \mathcal{H}} \frac{f(\boldsymbol{\phi})}{r(\boldsymbol{\phi})} \right) \mathsf{SNR}^{-d_R^*(r)}. \end{aligned} \tag{92}$$

The expression to be maximized can be written as (see (83) and (84))

$$\max_{\boldsymbol{\phi} \notin \mathcal{H}} \prod_{\ell=1}^{n_m} \phi_\ell^{k_\ell - r_\ell} = \min_{\boldsymbol{\phi} \notin \mathcal{H}} \prod_{\ell=1}^{n_m} \phi_\ell^{u_\ell}. \tag{93}$$

where $u_\ell = r_\ell - k_\ell$ Now, we consider the codes from Secstions VIII and IX and explicitly compute the maximization (93).

A) V-BLAST: For the last segment of an $n \times n$ channel, none of the singular values can die completely (i.e., become less than $\frac{1}{\mathsf{SNR}}$), therefore the no-outage condition can be written as

$$\prod_{\ell=1}^{n} \phi_\ell \geq \frac{\mathsf{SNR}^r}{\mathsf{SNR}^n}. \tag{94}$$

Therefore the minimization (93) can be written as

$$\min_{\prod_{\ell=1}^{n} \phi_\ell \geq \frac{\mathsf{SNR}^r}{\mathsf{SNR}^n}} \prod_{\ell=1}^{n} \phi_\ell^{u_\ell} \tag{95}$$

with an additional constraint that the ϕ_ℓs are bounded by one (using the exponential tail assumption). If we assume that u_1 is larger than u_ℓ for every $\ell \geq 2$, then

$$\min_{\prod_{\ell=1}^{n} \phi_\ell \geq \frac{\mathsf{SNR}^r}{\mathsf{SNR}^n}} \prod_{\ell=1}^{n} \phi_\ell^{u_\ell} \leq \frac{\min_{\prod_{\ell=1}^{n} \phi_\ell \geq \frac{\mathsf{SNR}^r}{\mathsf{SNR}^n}} \left(\prod_{\ell=1}^{n} \phi_\ell\right)^{u_1}}{\max_{\boldsymbol{\phi}} \prod_{\ell=2}^{n} \phi_\ell^{u_1 - u_\ell}}$$

If we assume that $u_1 \geq 0$ and $u_1 - u_\ell \geq 0$ for every $\ell \geq 2$, then the optimizing solution is given by

$$\begin{aligned} \phi_1^* &= \frac{\mathsf{SNR}^r}{\mathsf{SNR}^n} \\ \phi_\ell^* &= 1 \quad \text{for } \ell = 2, \ldots, L. \end{aligned}$$

This optimal point is same as the point (in terms of $\boldsymbol{\alpha}$), that optimized the outage probability calculation in (87). Then, at the optimal point we can write

$$\begin{aligned} \prod_{\ell=1}^{n} {\phi_\ell^*}^{k_\ell - r_\ell} &= \frac{\prod_{\ell=1}^{n} {\phi_\ell^*}^{k_\ell + 1}}{\prod_{\ell=1}^{n} {\phi_\ell^*}^{r_\ell + 1}} \\ &= \frac{\mathsf{SNR}^{-d_F^*(r)}}{\mathsf{SNR}^{d_R^*(r)}}. \end{aligned}$$

Therefore, using (92) and (91) the probability of error can be upper bounded by

$$\begin{aligned} \mathbb{P}_e &\dot{\leq} \mathsf{SNR}^{-d_F^*(r)} + \frac{\mathsf{SNR}^{-d_F^*(r)}}{\mathsf{SNR}^{d_R^*(r)}} \mathsf{SNR}^{-d_R^*(r)} \\ &\dot{\leq} \mathsf{SNR}^{-d_F^*(r)}. \end{aligned}$$

Thus, the code is also tradeoff optimal for the channel with fading density f, where f satisfies the following conditions

$$\begin{aligned} r_1 &\geq k_1 \\ r_1 - k_1 &\geq r_1 + 2(\ell - 1) - k_\ell, \quad \text{for } \ell = 2, \ldots, L. \end{aligned}$$

Combining these two conditions, we get

$$\begin{aligned} k_\ell - 2(\ell - 1) &\geq k_1, \quad \text{for } \ell = 2, \ldots, L \\ k_1 &\leq 0. \end{aligned}$$

B) D-BLAST: For the last segment of an $n_t \times 2$ channel, none of the singular values can fade completely (i.e., become less than $\frac{1}{\mathsf{SNR}}$), and hence the no-outage condition can be written as

$$\phi_1 \phi_2 \geq \frac{\mathsf{SNR}^r}{\mathsf{SNR}^2}. \tag{96}$$

This means that the minimization (93) can be written as

$$\min_{\phi_1 \phi_2 \geq \frac{\mathsf{SNR}^r}{\mathsf{SNR}^2}} \phi_1^{u_1} \phi_2^{u_2} \tag{97}$$

Now, this optimization problem is the same as the V-BLAST optimization problem in (95), with $n = 2$. Hence, the optimality condition on k_1 and k_2 turns out to be

$$\begin{aligned} k_2 - k_1 &\geq 2 \\ k_1 &\leq 0 \end{aligned}$$

REFERENCES

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

[2] R. D. Wesel, "Trellis code design for correlated fading and achievable rates for Tomlinson-Harashima precoding," Ph.D. dissertation, Stanford University, Stanford, CA, 1996.

[3] C. Köse and R. D. Wesel, "Universal space-time trellis codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2717–2727, Oct. 2003.

[4] A. Matache and R. D. Wesel, "Universal trellis codes for diagonally layered space-time systems," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 1073–1096, Nov. 2773-2783.

[5] D. Divsalar and M. Simon, "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Trans. Commun.*, vol. 36, no. 9, pp. 1004–1012, Sept. 1988.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press, 2005.

[7] H. Yao and G. Wornell, "Achieving the the full MIMO diversity-multiplexing frontier with rotation based space-time codes," in *Proc. Allerton Conf. Commun., Contr. Comput.*, Oct. 2003.

[8] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[9] P. Dayal and M. Varanasi, "An optimal two transmit antenna space-time code and its stacked extension," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003.

[10] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H. F. Lu, Explicit space-time codes achieving the diversity-multiplexing gain tradeoff [Online]. Available: http://arxiv.org/abs/cs.IT/0602054

[11] F. Oggier, G. Rekhaya, J.-C. Belfiore, and E. Viterbo, "Perfect space time block codes," *IEEE Trans. Inf. Theory*, available online at http://arxiv.org/abs/cs.IT/0604093, submitted for publication.

[12] T. Kiran and B. S. Rajan, "STBC-schemes with nonvanishing determinant for certain number of transmit antennas," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2984–2992, Aug. 2005.

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

[14] G. J. Foschini, "Layered space time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[15] J.-C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," in *Proc. Inf. Theory Workshop*, Paris, France, Mar. 2003.

[16] J. Boutros, E. Viterbo, C. Rastello, and J. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 502–518, Mar. 1996.

[17] J. Boutros and E. Viterbo, "Signal space diversity: A power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1453–1467, July 1998.

[18] J. Yedidia, K. Pedagani, and A. Molisch, "New spreading transforms for fading channels," in *Proc. Allerton Conf. Commun., Contr. Comput.*, Oct. 2004.

[19] A. Sahai, "Anytime information theory," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 2001.

[20] P. Vontobel and A. Ganesan, An Explicit Construction of Universally Decodable Matrices [Online]. Available: http://arxiv.org/abs/cs.IT/0508098

[21] R. E. Blahut, *Algebraic Codes for Data Transmission*. Cambridge, U.K.: Cambridge University Press, 2003.

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

[23] S. Tavildar and P. Viswanath, Approximately universal codes for over slow fading channels [Online]. Available: http://arxiv.org/abs/cs.IT/0512017v1

[24] V. Doshi, "Explicit permutation codes for the slow fading parallel channel," Bachelors thesis, University of Illinois at Urbana-Champaign, Urbana, IL, 2005.

[25] S. M. Alamouti, "A simple transmit divesity technique for wireless communication," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[26] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A 2×2 full rate space-time code with nonvanishing determinants," in *Proc. Int. IEEE Symp. Inf. Theory*, Jun. 2004, p. 308.

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

[28] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sept. 2004.

[29] N. Prasad and M. Varanasi, "Outage analysis and optimization for multiaccess/V-BLAST architecture over MIMO Rayleigh fading channels," in *Proc. 41st Annu. Allerton Conf. Commun. Contr. Comput.*, Oct. 2003.