Ann. Telecommun. (2010) 65:87–99
DOI 10.1007/s12243-009-0139-5

General random coding bounds: AWGN channels to MIMO fading channels

Heung-No Lee · Jingqiao Zhang · Cheon Won Choi

Received: 22 June 2008 / Accepted: 16 November 2009 / Published online: 23 December 2009

Abstract Random coding bounds are obtained for multiple-input multiple-output (MIMO) fading channels. To derive the result in a compact and easy-to-evaluate form, a series of combinatorial codeword enumeration problems are solved for input-constrained MIMO fading channels. The bounds obtained in this paper are shown useful as performance prediction measures for MIMO systems which employ turbo-like block codes as the outer code to derive the space-time inner code. The error exponents for MIMO channels are also derived from the bounds, and then compared with the classical Gallager error exponents as well as the channel capacities. The random coding bounds associated with the maximum likelihood receiver exhibit good match with the extensive system simulation results obtained with a turbo-iterative receiver.

Heung-No Lee's work was supported in part by DASAN fund, GIST, Korea, and by the University of Pittsburgh CRDF award.

Jingqiao Zhang's work was supported by the University of Pittsburgh CRDF award.

Cheon Won Choi's work was supported by Dankook University project for funding RICT.

This work was presented in part at the IEEE International Conference on Communications 2006, June 2006, Istanbul, Turkey.

H.-N. Lee (✉)
Gwangju Institute of Science and Technology, GIST,
Gwangju, South Korea
e-mail: heungno@gist.ac.kr

J. Zhang
Amazon.com,
Seattle, USA

C. W. Choi
Dankook University,
Yongin, South Korea

Keywords Union bound · Random coding exponent · MIMO capacity

1 Introduction

Multiple-input multiple-output (MIMO) systems have drawn enormous attention ever since the potential for multifold increase in spectral efficiency over rich scattering fading channels has been identified [1, 2]. The use of turbo codes [3] and low-density parity-check (LDPC) codes [4] has shown extreme success, thanks to their capacity achieving/approaching performance in many communication channels [5–7].

It is therefore of our interest to investigate the application of turbo-like codes, LDPC codes in particular, over MIMO systems and to derive analytical bounds on their potential performance. In this regard, we have proposed a couple of performance-bounding techniques in the past; one for quasi-static fading channels in [8] and the other for fast fading channels in [9, 10]. They are canonical union bounds which require the knowledge of *distance spectrum* of *outer* binary block code (compared to *inner* space-time block code). These bounds were shown to be very useful as a performance measure in comparison with simulation results for a practical *sum-product* iterative decoding receiver.

In this paper, we focus on developing random coding bounds which do not require the knowledge of distance spectrum of outer code. In particular, the random block code takes the role as the outer binary block code. The distance spectra of random codes are straightforward and can be compactly described with a few parameters for any transmitter configuration. Random selection makes it possible. Each code is generated randomly; within each code, codewords are selected randomly as well. In a stark

contrast, the distance spectra of the turbo-like codes are not easy. Because of linear constraints, derivation of distance spectrum is quite involved and only numerical approaches are possible even for ensemble averages; furthermore, even the numerical approaches are available only for certain classes of linear codes [11–13]. Finding a non-ensemble average distance spectrum of a linear code is an NP-hard problem. Yet, searching for low weight codewords near the all-zero codeword and their multiplicity have shown some success using the so-called "error impulse" methods [14, 15].

From the perspective of the "spectral thinning" argument [16], the performance of random codes shall outperform that of the linear turbo-like codes. Namely, the distance spectrum of random block codes is *thinner* than that of turbo-like linear codes. Consider an ensemble of (L,K) random block codes. Let A_h denote the average number of codewords whose Hamming weight is h. Then, the average distance spectrum is simply given by $A_h = 2^{-L}2^K\binom{L}{h}$ for $h\in\{0,\cdots,L\}$. It is clear that A_h approaches to zero quickly for small h. For turbo-like codes, however, the distance spectrum is typically *thicker* due to the inherent linearity constraint, i.e., the linear code has non-negligible number of codewords with small Hamming weights. In fact, it has been one of code designer's wishes to eliminate the small weight codewords from turbo and low-density parity-check codes [16–18].

These observations have motivated our research in the following direction: random coding bounds can be derived in a concise form and the results can be utilized to serve as tight lower bounds to the canonical union bounds. The preliminary version of this approach has appeared in our short paper [9], yet given without detailed and rigorous proofs. We offer them here.

In this paper, our random coding bounds are compared with the classical results such as Gallager's random coding exponents [19] and the input constraint mutual information theoretic results [7, 20–23]. While Gallager's exponents and information theoretic capacities for MIMO channels are useful, they are available to date only in basic integral forms which require time-consuming numerical evaluation and nested loop integration. We obtain closed form error exponent expressions and show that the results are consistent with canonical measures. Furthermore, we show that our results are consistent with classical results obtained for additive white Gaussian noise (AWGN) channels.

Making connections to classical measures and comparison of our results to well-known results are aimed at (1) corroborating the accuracy of our results and at (2) showing the generality of new results obtained in this paper, which are extensive enough to subsuming previously known classical results.

The rest of this paper is organized as follows. Section 2 describes the MIMO system of interest. In Section 3, we provide the main result—the random coding bound—which is proved in Appendix. In Section 4 through 7, we discuss the results and make comparison with Gallager's random coding bound/exponent and MIMO capacity. Finally, we make a summary in Section 8.

2 System of interest

Consider an ensemble of (L,K) random codes with a code rate $R_c = K/L$. The ensemble is constructed by randomly selecting 2^K codewords out of a total number of 2^L distinct binary strings of length L without replacement.[1] The ensemble is thus composed of all $\binom{2^L}{2^K}$ distinct codes generated in this manner. Hereafter, it is assumed that each code in the ensemble is equi-probably selectable.

As illustrated in Fig. 1, a random binary code C is used to operate on the M-transmit N-receive MIMO system which employs 2^{K_b}-ary modulation (for example, K_b=2 for 4-QAM). In the system, a codeword $\mathbf{c}$ is equally likely chosen from a code $\boldsymbol{C}$ for transmission. A codeword of length L can be partitioned into a sequence of binary strings of size MK_b. Each string consists of MK_b bits. We assume the length of codeword L to be a T multiple of MK_b for convenience. We also assume that K_b stays the same over time and antenna index throughout the paper except in Section 7. There, we discuss the cases when K_b varies over space and time.

We use a modulation table in this paper which maps each string of MK_b bits one-to-one correspondingly to an $M\times 1$ vector s of channel symbols. Each entry of vector s takes a point from the channel-symbol constellation of size 2^{K_b}. We will call this *base constellation*. Note that a channel-symbol vector s is an element of the set of $J = 2^{MK_b}$ vectors. The set of the J vectors is referred to as *vector constellation* as compared to the base constellation. We will use $S := \{\mathbf{s}_0,\cdots,\mathbf{s}_{J-1}\}$ to denote the vector constellation. We assume the symbol vector s obeys the average energy constraint, i.e., $\mathrm{E}\left\{\|\mathbf{s}\|^2\right\} = E_s$.

A map from a code $\boldsymbol{C}$ to S^T is in general injective. Some $M\times T$ space-time matrices are not selected as codewords as some binary words of length L are not selected to be codeword. But given a code, the map we have defined in the previous paragraph will give us a set of space-time(ST) sequences of channel symbol vectors. This set of ST sequences, calling them ST codewords, is one-to-one correspondent with the binary code selected. That is, there

[1] Note that this is a little different from the classical definition of random codes in which repetition is allowed.

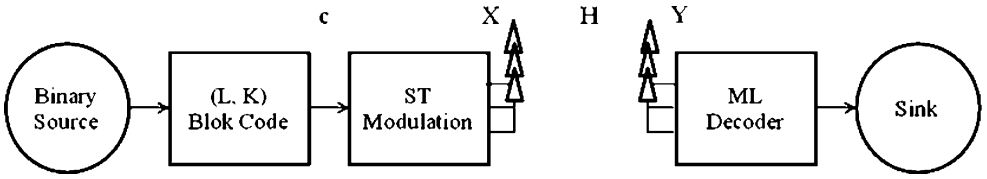

Fig. 1 System block diagram of interest. Notation: M number of transmit antennas; N number of receive antennas; $\boldsymbol{c}$ binary codeword of length L; $\boldsymbol{X}$ $[M\times T]$ space-time codeword one-to-one correspondingly mapped from $\mathbf{c}$. Its T columns are selected the J–ary vector constellation S; $\boldsymbol{H}$ is an $[N\times M]$ channel matrix. Its time index given in the input output relation in Eq. 1 is omitted in the figure. $\boldsymbol{Y}$ $[N\times T]$ matrix whose T columns are determined by the input-output relationship given in Eq. 1

are 2^K distinct binary codewords one-to-one correspondingly mapped to 2^K distinct space-time codewords. That is, each codeword $\mathbf{c}$ in a code is mapped to a distinct $M\times T$ space-time codeword matrix $\mathbf{X}=(\mathrm{x}_1,\cdots,\mathrm{x}_T)$. Note that a space-time codeword $\mathbf{X}$ consists of T columns and each column x_t takes a channel-symbol vector from the vector constellation for each $\mathrm{t}\in\{1,\cdots,T\}$.

The input output relation for a receive signal (an $N\times 1$ vector) $\mathbf{y}_t$ at the time epoch t is obtained as

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{n}_t \tag{1}$$

for $t\in\{1,\cdots,T\}$, where $\mathbf{H}_t$ is the $N\times M$ channel matrix whose entries are independent complex Gaussian distributed random variables with zero mean and variance of 1/2 in each dimension, and n_t is the $N\times 1$ complex spatially and temporally independent Gaussian noise with zero mean and variance $\frac{N_o}{2}$ in each dimension. The channel matrix $\mathbf{H}_\mathrm{t}$ is assumed to be known at the receiver and to be selected independently for each transmission of a channel-symbol vector. Note that the channel is thus assumed to be ergodic. This channel was considered in many previous studies [7, 9, 10, 22, 24].

We will omit the subscript t whenever there is no confusion not distinguishing the time epoch.

We finish this section with the following remark.

Remark 1 Please note that input symbol vectors are mutually independent and uniformly distributed (i.u.d.) on the set S. Accordingly the Gallager error exponent and the channel capacity will be obtained with the same condition in Section 6. Such capacity is called the i.u.d. capacity in the literature, see [23].

3 Random coding bounds

The main result of the paper is summarized in the following theorem. The derivation of the union upper bound leading to the Theorem is given in Appendix.

Theorem *Consider the random-coded modulation MIMO system described by Eq. 1. Let $\overline{P}_e$ denote the probability of maximum likelihood (ML) decoding error which averaged over the ensemble of random codes. Then, the error probability is upper-bounded by*

$$\overline{P}_e \le 2^{-T\cdot E(R)} \tag{2}$$

where, the exponent is defined as

$$E(R) := R_o - R, \tag{3}$$

$$R_o := -\log_2\left\{\frac{1}{2^{2MK_b}}\sum_{j=0}^{2^{MK_b}-1}\sum_{k=0}^{2^{MK_b}-1}\beta_{j,k}\right\}, \tag{4}$$

and

$$R = R_c M K_b \tag{5}$$

In the theorem, we note that R represents the transmission rate (bits per channel use) of the system. Also, the variable $\beta_{j,k}$ is defined as

$$\beta_{j,k} := \left(1+\frac{1}{4N_0}\left|s_j - s_k\right|^2\right)^{-N}, \tag{6}$$

where, s_k, $\mathrm{s}_j\in S$. It is the pairwise vector symbol error averaged over fading. The proof is given in Appendix.

4 Discussions

A closed-form union upper bound on the MIMO transmission system has been derived in the theorem. As expected, the bound decays exponentially fast to zero with increasing block length N as long as a transmission rate R leading to a positive error exponent $E(R)$ is selected.

We will compare the result with the Gallager's random coding exponent in Section 5 and with the MIMO channel capacity in Section 6. In Section 6, we will also show that the rate R_o given in Eq. 4 can collapse down to a value called the *cut-off* rate in the AWGN channel context. Hereafter, we thus refer to the rate R_o in Eq. 4 as the cut-off rate for MIMO channels. Later, we will also show that the cut-off rate can serve as a lower bound to the MIMO channel capacity C. It is worthwhile to note that this parameter is independent of the constellation map used for transition from a binary codeword to a space-time codeword. It rather depends on the choice of signal constellation. This indicates that it can be useful for code search problems.

4.1 An illustrative example

Consider a 2-transmit N-receive MIMO system with quadrature phase-shift keying (QPSK) signaling, (i.e., M=2 and K_h=2). We assume that signal points are chosen from the QPSK constellation $\{+1,-1,+\mathbf{i},-\mathbf{i}\}$, where $\mathbf{i}=\sqrt{-1}$.

In the system, there are $J=2^{MK_b}=16$ distinct values that a symbol vector s can have as follows:

$$\sqrt{\frac{2}{E_s}}s_j \in \left\{\begin{pmatrix}+1\\+1\end{pmatrix},\begin{pmatrix}+1\\+\mathbf{i}\end{pmatrix},\begin{pmatrix}+1\\-1\end{pmatrix},\begin{pmatrix}+1\\-\mathbf{i}\end{pmatrix},\begin{pmatrix}+\mathbf{i}\\+1\end{pmatrix},\begin{pmatrix}+\mathbf{i}\\+\mathbf{i}\end{pmatrix},\begin{pmatrix}+\mathbf{i}\\-1\end{pmatrix},\begin{pmatrix}+\mathbf{i}\\-\mathbf{i}\end{pmatrix},\begin{pmatrix}-1\\+1\end{pmatrix},\begin{pmatrix}-1\\+\mathbf{i}\end{pmatrix},\begin{pmatrix}-1\\-1\end{pmatrix},\begin{pmatrix}-1\\-\mathbf{i}\end{pmatrix},\begin{pmatrix}-\mathbf{i}\\+1\end{pmatrix},\begin{pmatrix}-\mathbf{i}\\+\mathbf{i}\end{pmatrix},\begin{pmatrix}-\mathbf{i}\\-1\end{pmatrix},\begin{pmatrix}-\mathbf{i}\\-\mathbf{i}\end{pmatrix}\right\}$$

for $j\in\{0,\cdots,15\}$. As a function of s_j and s_k, $\beta_{j,k}$ can be calculated from Eq. A24. By the use of $\beta_{j,k}$, the error exponent in Eq. 3 is finally obtained as:

$$E(R)=4-R-\log_2\left\{1+4\left(1+\frac{E_s}{4N_o}\right)^{-N}+6\left(1+\frac{E_s}{2N_o}\right)^{-N}+4\left(1+\frac{3E_s}{4N_o}\right)^{-N}+\left(1+\frac{E_s}{N_o}\right)^{-N}\right\}. \tag{7}$$

In Fig. 2, the random coding union bounds are compared with the union bounds as well as the system simulation results. At the transmitter, a Gallager's (3,6) LDPC code [4] generated randomly at the block length of 3,000 is used. The receiver employs the usual turbo-iterative detection and decoding which exchange the extrinsic log likelihood ratios on the coded bit sequence. The QPSK modulation is used to carry the coded bits over the 2×2 and 4×4 MIMO channels. The union bounds for the linear LDPC codes are derived using the technique reported in [9, 10], which is based on the distance spectrum of the (3,6) LDPC codes. The random coding bounds show a good prediction on the waterfall position (and at least provide coercive upper bounds to the simulation results). The comparison of the random coding bounds with the union bounds for the linear LDPC codes show that random coding bounds indeed are suitable for lower bounding the union bounds for the linear codes. Since the distance spectra of random codes possess vanishing spectral components for small Hamming weights, the random coding bounds show no sign of error floor (at least in the region of interests) unlike the union bounds for the linear code.

5 Comparison with Gallager's random coding exponent

In this section, we compare the exponent derived in Eq. 3 with the classical *random coding exponent* developed by Gallager ([19], see Section 5). Using the base-2 logarithm in the system of concern, the average probability of ML

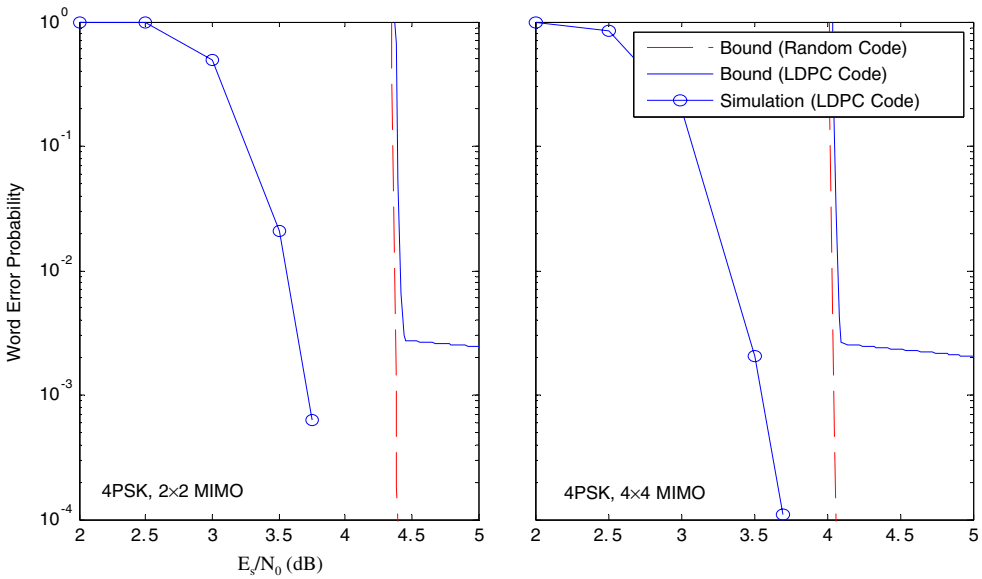

Fig. 2 The union bounds on random code (L=3,000, R_c=0.5) vs. on the LDPC (3,000, 3, 6) code: they are compared with the system simulation of iterative detection/decoding of the LDPC code transmitted over the MIMO channel

decoding error was shown to be upper bounded in [25] as follows:

$$\overline{P}_e \le 2^{-T\cdot E_r(R)} \tag{8}$$

where, T is the number of channel uses,

$$E_r(R) := \max_{0\le\rho\le 1}\{E_o(\rho)-\rho R\}, \text{ and} \tag{9}$$

$$E_o(\rho) := -\log_2\left(\int_{\mathbf{y},\mathbf{H}}\left[\sum_{j=0}^{2^{MK_b}-1}\frac{1}{2^{MK_b}}f\left(\mathbf{y},\mathbf{H}|\mathbf{s}_j\right)^{\frac{1}{1+\rho}}\right]^{1+\rho}d\mathbf{y}d\mathbf{H}\right). \tag{10}$$

Note that $f\left(\mathbf{y},\mathbf{H}\middle|\mathbf{s}_j\right)$ in Eq. 9 is the conditional probability density function for $\mathbf{y}$ and $\mathbf{H}$ given $s=s_j$. According to [19], the function E_o in Eq. 9, which is to be optimized, is concave for $\rho\ge 0$. Thus, it can be evaluated by an optimization method such as the golden section algorithm [26]. Nevertheless, the expression in Eq. 9 is still not easy to calculate due to the integration over $N(M+1)$ complex dimensions. To evaluate it by the Monte Carlo method, the following is more tractable alternative, which can be obtained through a simple manipulation [25]:

$$E_o(\rho) = -\log_2 \mathrm{E}\left(\frac{\left[\sum_{j=0}^{2^{MK_b}-1}\frac{1}{2^{MK_b}}f\left(\mathbf{y}\middle|\mathbf{H},\mathbf{s}_j\right)^{\frac{1}{1+\rho}}\right]^{1+\rho}}{\sum_{j=0}^{2^{MK_b}-1}\frac{1}{2^{MK_b}}f\left(\mathbf{y}\middle|\mathbf{H},\mathbf{s}_j\right)}\right) \tag{11}$$

where $f\left(\mathbf{y}\middle|\mathbf{H},\mathbf{s}_j\right)$ in Eq. 10 is the conditional probability density function for y given $\mathbf{H}$ and $s=s_j$.

As shown in Fig. 2, the exponent derived in Eq. 3 shows a good match with the Gallager's random-coding exponent. Over a wide range of region, covering from low rates below the cut-off rate, the optimal ρ in Eq. 9 is found to be equal to 1 (by a numerical method). Thus, we can practically use the exponent in Eq. 3 in place of the exponent in Eq. 9 while $R<R_o$. The two different measures coincide in this region. Further investigation in this direction seems interesting. As expected, the two exponents diverge in the high-rate region where the transmission rate passes beyond the cut-off rate and approaches the capacity. The final gap between the cut-off rate and the capacity is about 1 bit/channel use as observed in Figs. 3 and 4.

6 Comparison with MIMO channel capacity

In this section, we aim to compute the channel capacity and then compare it with the cut-off rate in Eq. 4. The channel capacity C is the mutual information between the input and output symbol vectors under the assumption that input symbol vectors are mutually i.u.d. on the set $S=\{s_0,\cdots,s_{J-1}\}$ (in this sense, such a channel capacity is also called i.u.d. capacity) [20, 23]. The capacity for the channel in Eq. 1 is thus the conditional mutual information between the i.u.d. input s and the output $\mathbf{y}$. Let h be the conditional entropy function. Note that $h(\mathbf{y},\mathbf{s}|\mathbf{H}) = N\log(\pi N_o e)$ for white Gaussian noise. Then, the channel capacity is expressed by

$$\begin{aligned} C &= h(\mathbf{y}|\mathbf{H}) - h(\mathbf{y}|\mathbf{H},\mathrm{s}) \\ &= MK_b - N\log_2(e) - \mathrm{E}\left\{\log_2\left[\sum_{j=0}^{2^{MK_b}-1}\exp\left(-\frac{|\mathbf{y}-\mathbf{H}\mathbf{s}|^2}{N_o}\right)\right]\right\} \end{aligned} \tag{12}$$

where the final expression of the channel capacity is yielded by formulating $h(\mathbf{y}|\mathbf{H})$ from the marginal probability density for y. Note that the marginal probability density for y is obtained from the conditional probability density for y given s as follows:

$$\begin{aligned} f(\mathbf{y}|\mathbf{H}) &= \frac{1}{2^{MK_b}}\sum_{j=0}^{2^{MK_b}-1} f\left(\mathbf{y}\middle|\mathbf{H},\mathbf{s}_j\right) \\ &= \frac{1}{2^{MK_b}(\pi N_o)^N}\sum_{j=0}^{2^{MK_b}-1}\exp\left(-\frac{\left|\mathbf{y}-\mathbf{H}\mathbf{s}_j\right|^2}{N_o}\right) \end{aligned} \tag{13}$$

where $f(\mathbf{y}|\mathbf{H})$ and $f(\mathbf{y}|\mathbf{H},\mathrm{s})$ are marginal and conditional probability densities for y, respectively.

In Figs. 3 and 4, the cut-off rates in Eq. 4 are plotted in comparison with the numerically obtained channel capacities for various modulation and channel scenarios. The cut-off rates are shown to be about 3.5 dB at maximum off from their respective channel capacities. It is observed that the signal-to-noise ratio (SNR) and the rate gaps do not significantly vary according to modulation sizes and the number of transmit and receive antennas used. In fact, the gap to the capacity shows in Fig. 5 that the union bound are not tight enough when the rate approaches the capacity.

Tight union bounds that continue to approach the capacity have been the subject of many researches. Two different tight bounding methods have been suggested by Gallager: they are so called (1) Gallager's first bounding method, discussed in Chapter 3 of Gallager's thesis [4] and (2) Gallager's second bounding method given in [19]. The second method is applied to our MIMO setting and the results obtained are given in Eqs. 8, 9, and 10. In the 2000s, the interest on tight bounds has been revived, and many papers have been published. Details can be found in [8] and the reference therein.

There exist several interesting relationships between the Gallager's random coding error exponent (Eq. 11) and the capacity (Eq. 12). One is that the derivative of the random coding exponent $E_o(\rho)$ with respect to ρ and evaluation of it

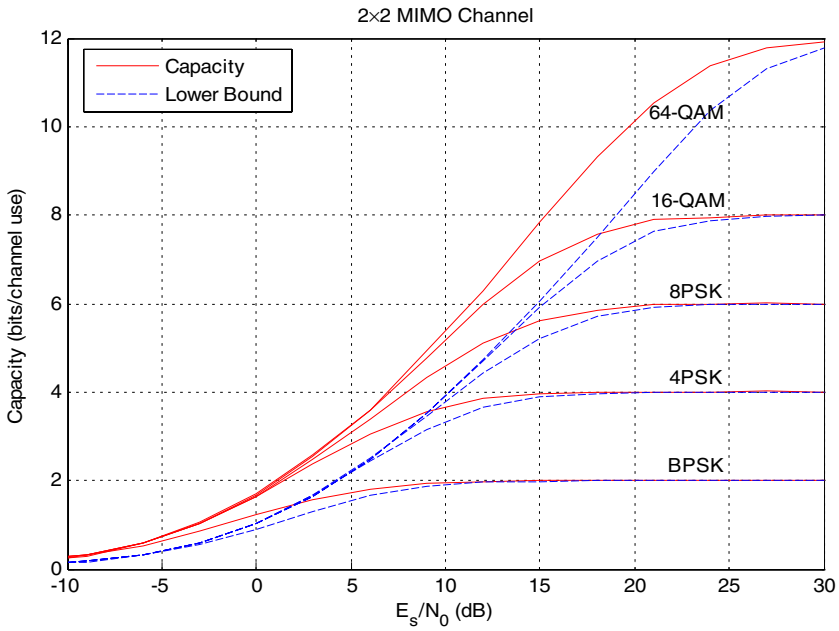

Fig. 3 The cutoff rate (lower bound) vs. capacity (2×2 MIMO channel)

at $\rho = 0$ is equal to the capacity, see Theorem 5.6.3 proved in [19]. Along with other relations, this shows that the maximum rate attainable with the random coding bound is the capacity.

Note, however, that the Gallager's results applied to MIMO channels—Eqs. 8, 9, and 10—require multiple nested numerical integration and numerical optimization at each rate with respect to the utility variable ρ.

Nevertheless, Gallager's exponent does show that it indeed stays useful up to the channel capacity. For example, let's take a look at 15dB E_s/N_o point with 16-QAM for 2×2 MIMO system in Fig. 3. Reading off a value from Fig. 3, we note that the capacity is around 7 bits/sec. The same value is obtained with the Gallager's random coding exponent curve shown in Fig. 5. The Gallager's exponent $E_r(R)$ approaches zero arbitrarily closely when R approaches 7 bits/sec.

6.1 Extension to ricean and AWGN channels

In this subsection, we discuss extensions of our current results to Ricean and AWGN channels. For this purpose, we start with the pairwise error probability (PEP) discussed in Appendix A. In Appendix A, the PEP is given for Rayleigh fading. Here, we define it for Ricean fading

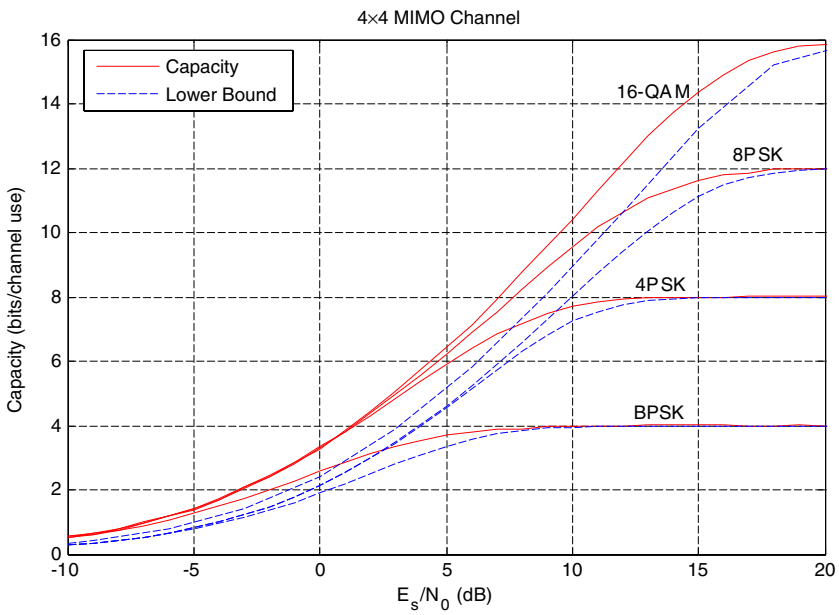

Fig. 4 The cutoff rate (lower bound) vs. capacity (4×4 MIMO channel)

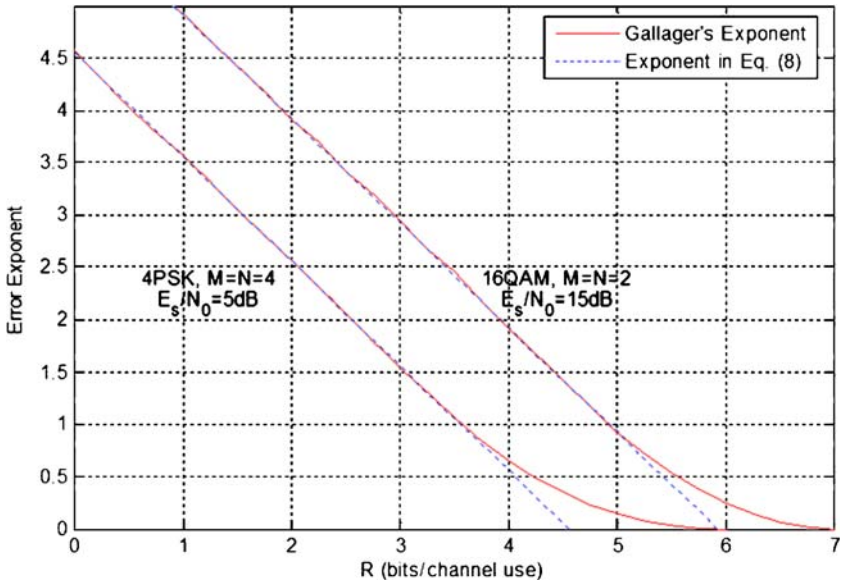

Fig. 5 Comparison of the two error exponents

channel, and will later on let the Ricean factor to be zero and obtain the result for the AWGN channel. Regardless if the PEP is for Ricean or Rayleigh, the derivation in Appendix holds.

For Ricean fading channels, for example, the PEP is given as the following,

$$P(\mathbf{c}\to\mathbf{c}') \le \prod_{t=1}^{T}\left(1+\frac{1}{4N_o}\left|\mathbf{x}_t-\mathbf{x}_t'\right|^2\right)^{-N}\exp\left(-N\frac{K_R\frac{1}{4N_o}\left|\mathbf{x}_t-\mathbf{x}_t'\right|^2}{1+\frac{1}{4N_o}\left|\mathbf{x}_t-\mathbf{x}_t'\right|^2}\right) \tag{14}$$

where K_R is the Ricean factor. It is clear that by setting K_R= 0, the right hand side becomes the PEP for Rayleigh fading which is the starting point of Appendix A. The right hand side of Eq. 13 can be rewritten in a form similar to Eq. A4 so that

$$P(\mathbf{c}\to\mathbf{c}') \le \prod_{j=0}^{J-1}\prod_{k=0}^{J-1}\left[\left(1+\frac{1}{4N_o}\left|\mathbf{s}_j-\mathbf{s}_k\right|^2\right)^{-N}\exp\left(-N\frac{K_R\frac{1}{4N_o}\left|\mathbf{s}_j-\mathbf{s}_k\right|^2}{1+\frac{1}{4N_o}\left|\mathbf{s}_j-\mathbf{s}_k\right|^2}\right)\right]^{\delta_{j,k}} \tag{15}$$

where, we can define $\beta_{j,k}$ as the term inside the square brackets for $j,k\in\{0,\cdots,J-1\}$. The PEP for Ricean is now in exactly the same form as that for Rayleigh. Thus, it is trivial to show that the theorem holds for the Ricean channels as well, but with a new definition $\beta_{j,k} := \left(1+\frac{1}{4N_o}\left|\mathbf{s}_j-\mathbf{s}_k\right|^2\right)^{-N}\exp\left(-N\frac{K_R\frac{1}{4N_o}\left|\mathbf{s}_j-\mathbf{s}_k\right|^2}{1+\frac{1}{4N_o}\left|\mathbf{s}_j-\mathbf{s}_k\right|^2}\right)$.

Another example of interest is the AWGN channel. For this, we let the number of antennas to be 1, i.e., M= N=1. In addition, we consider the most simple case in which a binary phase shift-keying modulation is assumed, i.e., K_b=1 and $\mathrm{x}=\sqrt{E_s}(1-2\mathbf{c})$. Then, the PEP can be written as

$$\begin{aligned} P(\mathbf{c}\to\mathbf{c}') &\le \exp\left(-\frac{1}{4N_o}\sum_{t=1}^{T}\left|\mathbf{x}_t-\mathbf{x}_t'\right|^2\right) \\ &= \prod_{t=1}^{T}\exp\left(-\frac{1}{4N_o}\left|\mathbf{x}_t-\mathbf{x}_t'\right|^2\right) \\ &= \prod_{j=0}^{J-1}\prod_{k=0}^{J-1}\beta_{j,k}{}^{\delta_{j,k}} \end{aligned} \tag{16}$$

where, $\boldsymbol{x}_t$ and $\boldsymbol{x}_t'$ are the tth components of $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively. It is trivial to show $\beta_{0,0}=\beta_{1,1}=1$ and $\beta_{0,1}=\beta_{1,0}=\exp(-E_s/N_o)$. Accordingly, the error exponent is obtained as

$$E(R) = 1-R-\log_2\left[1+\exp\left(-\frac{E_s}{N_o}\right)\right]. \tag{17}$$

We note that the rate $R_o = 1-\log_2\left[1+\exp\left(-\frac{E_s}{N_o}\right)\right]$ is the *cut-off* rate for the AWGN channel.

This extension to AWGN channel shows that our result is general and consistent so with the previous random-coding error exponent obtained in AWGN channel context.

7 Modification for different K_b in space and time

Throughout this paper so far, the size of base constellation 2^{K_b} has remained constant for different time epochs and

different antennas. During the review process, a very interesting question came up from an anonymous reviewer as to inquire if our framework is flexible enough to handle the cases when K_b varies over time and space. We consider both spatial and temporal variation cases and discuss each.

The first is the case when K_b varies over antenna. Different antennas can use different signal constellations. This is an easy case and can be handled with almost no additional effort. Let's use an example for easy exposition. Suppose M=2 and we want the first antenna to use $K_{b,1}$=2 bit-base constellation while the second antenna to use $K_{b,2}$=1 bit-base constellation. Then, the vector constellation is a set of 2×1 vectors which can be formed by taking its first element from the 2-bit constellation while taking the second element from the 1-bit constellation. One can construct the vector constellation this way whose size is $J = 2^{K_{b,1}+K_{b,2}} = 2^{2+1}$. The rest of the procedure stays exactly the same. A slightly general vector constellation is the only required modification.

The second is the case when K_b varies over time. This is more involved. The key idea of the proof given in Appendix was provided by the multinomial expansion lemma. A modified expansion is needed for the second case. Again let us take an example for easy exposition. Suppose T=10 and variation occurs at t=5 and on. Let $K_{b,1}$=1 for the first four time epochs. Then, T_1=4. Let $K_{b,2}$=2 for the last six time epochs; T_2=6. Then, consider two sets of utility variables. One is $\{z_{1,j}\}_{j=0,1,\ldots,J_1-1}$, and the expansion of T_1-th power of the sum is $\left(\sum_{j=0}^{J_1-1} z_{1,j}\right)^{T_1} = \sum_{\underline{v}_1\in\Omega_1} \binom{T_1}{v_{1,0}, v_{1,1}, \cdots v_{1,J-1}} \prod_{j=0}^{J_1-1} z_{1,j}^{v_{1,j}}$. The other is $\{z_{2,j}\}_{j=0,1,\ldots,J_2-1}$ and the expansion of T_2-th power of the sum is $\left(\sum_{j=0}^{J_2-1} z_{2,j}\right)^{T_2} = \sum_{\underline{v}_1\in\Omega_1} \binom{T_2}{v_{2,0}, v_{2,1}, \cdots v_{2,J-1}} \prod_{j=0}^{J_2-1} z_{2,j}^{v_{2,j}}$. The product of the two has the following expansion,

$$\left(\sum_{j=0}^{J_1-1} z_{1,j}\right)^{T_1}\left(\sum_{j=0}^{J_2-1} z_{2,j}\right)^{T_2} = \sum_{\underline{v}_1\in\Omega_1}\sum_{\underline{v}_2\in\Omega_2} \binom{T_1}{v_{1,0}, v_{1,1}, \cdots v_{1,J-1}}\binom{T_2}{v_{2,0}, v_{2,1}, \cdots v_{2,J-1}} \prod_{k=0}^{J_2-1} z_{2,k}^{v_{2,k}} \prod_{k=0}^{J_1-1} z_{1,k}^{v_{1,k}} \tag{18}$$

where, Ω_1 and Ω_2 are the collections of arrays defined as

$$\Omega_1 := \left\{ \underline{v}_1 \middle| v_{1,j} \in \{0,1,\ldots,T\}, \sum_{j=0}^{J_1-1} v_{1,j} = T_1 \right\} \text{ and}$$

$$\Omega_1 := \left\{ \underline{v}_1 \middle| v_{1,j} \in \{0,1,\ldots,T\}, \sum_{j=0}^{J_1-1} v_{1,j} = T_1 \right\}.$$

Corollary *Using Eq. 18 and following the steps similar to Appendix A and B, we obtain the following result:*

$$\overline{P}_e \le \frac{1}{2^L}\frac{2^K}{2^L}\left(\sum_{j,k=0}^{J_1-1}\beta_{1,j,k}\right)^{T_1}\left(\sum_{j,k=0}^{J_2-1}\beta_{2,j,k}\right)^{T_2} = 2^{-T(R_o-R)} \tag{19}$$

where, $R_o := -\log_2 2^{-2\frac{L}{T}} - \log_2\left(\sum_{j,k=0}^{J_1-1}\beta_{1,j,k}\right)^{\frac{T_1}{T}} - \log_2\left(\sum_{j,k=0}^{J_2-1}\beta_{2,j,k}\right)^{\frac{T_2}{T}}$.

Note that $L = T_1 M K_{b,1} + T_2 M K_{b,2}$, $\beta_{1,j,k} := \left(1+\frac{1}{4N_0}|s_{1,j}-s_{1,k}|^2\right)^{-N}$ is for the first vector constellation of size $MK_{b,1}$, and $\beta_{2,j,k} := \left(1+\frac{1}{4N_0}|s_{2,j}-s_{2,k}|^2\right)^{-N}$ is for the second vector constellation of size MK_{b2}. The proof is omitted.

8 Conclusions

In this paper, we have obtained random-coding bounds for MIMO systems. We have shown that these random-coding bounds are general and consistent with the classical measures such as Gallager's error exponents and MIMO channel capacities. We have shown that the obtained bounds are useful to benchmark the system simulation results of a practical coding scheme such as LDPC and turbo codes. In this paper, we use an LDPC code and a turbo-iterative message passing algorithm receiver. The results indicate the usefulness of the derived bounds.

For the design of space-time block codes, the random coding bound obtained in this paper can be used as a metric to search for good space-time block codes. Note that a better space-time block code consistently indicates a superior error exponent behavior throughout the whole rate region [25]. The evaluation of the Gallager's error exponent, however, involves both the statistical averaging—over the noise, the fading channel and the channel symbols—as well as the optimization over the parameter given in Eq. 8, hence highly intensive computation is required. Since the random coding

exponent can be more quickly evaluable than the classical measures, they may serve as a useful tool for code searches.

Appendix A. Pairwise error probability

In Appendix A and B, we will derive the main theorem. In this section, we start with discussion of pairwise error probability which will be used as the first building block to prove the Theorem.

The PEP from codeword $\mathbf{c}$ to codeword $\mathbf{c}'$ is defined as the probability that the receiver, when making an ML decision between a pair of codewords, erroneously decides in preference of $\mathbf{c}'$ when $\mathbf{c}$ was actually transmitted. Suppose $\mathbf{X}$ and $\mathbf{X}'$ are the two space-time words one-to-one correspondingly mapped from $\mathbf{c}$ and $\mathbf{c}'$, respectively.

In case of a Rayleigh MIMO channel, the PEP averaged over the fading channel distribution for the system described by Eq. 1 can be formulated as (see [24] for details)

$$P(\mathbf{c} \to \mathbf{c}') \le \prod_{t=1}^{T} \left(1 + \frac{1}{4N_o}\left|\mathbf{x}_t - \mathbf{x}_t'\right|^2\right)^{-N} \tag{A20}$$

where, T is the block length of the space-time word and $|\cdot|$ denotes the L_2 norm of the complex vector. Also, x_t and x_t' are the tth columns of space-time words $\mathbf{X}$ and $\mathbf{X}'$, respectively. Further note that $x_t, x_t' \in \{s_0, \cdots, s_{J-1}\}$.

One of the key steps involved in the derivation of the union bound is to determine the partition of a codebook into a number of smaller sets so that the PEP in Eq. A20 is to render an identical result within a set. The determination of this set and the calculation of its cardinality are thus the critical steps for deriving our result. To proceed, we introduce two metrics given in the form of definition for easy reference.

Recall that a codeword of length L is segmented into T binary strings of length MK_b and there are $J = 2^{MK_b}$ distinct strings. For each string, we keep track of the number of occurrence of the string within a codeword. Under a particular constellation map, each string is mapped to one of the J channel-symbol vectors in the vector constellation. From a straightforward tracking of the one-to-one correspondence in this manner, we will be able to resolve all the codeword enumeration problems.

Definition 1 Binary string weight profile. There are J binary strings which can be sequenced from 0 to J-1. Likewise, there are J channel symbol vectors which can be indexed from 0 to J-1. Let $\mathbf{b}_j$ denote the jth binary string of length MK_b that is modulated onto the jth symbol vector s_j. We will use δ_j to denote the number of occurrences of the jth string $\mathbf{b}_j$ in a codeword. They can be stored in an array, referred to here as the *binary string weight profile* (BSWP). We use $\underline{\hat{\delta}}(\mathbf{c}) = (\delta_0(\mathbf{c}), \cdots, \delta_{J-1}(\mathbf{c}))$ to denote a BSWP. Each BSWP must satisfy the following four constraints from its definition:

1. $\delta_j(\mathbf{c}) \in \{0, \cdots, T\}$
2. $$\sum_{j=0}^{J-1} \delta_j(\mathbf{c}) = T$$
3. $\delta_j(\mathbf{c}) \in \{0,1,2,...,T\}$ and
4. $$\sum_{j=0}^{J-1} \delta_j(\mathbf{c}) = T$$

When there is no ambiguity we will use $\hat{\underline{\delta}} = \hat{\underline{\delta}}(\mathbf{c})$.

Under a specific constellation map, each binary string b_j maps to a corresponding channel symbol vector $\mathbf{s}_j$; likewise each codeword, a sequence of T binary strings, maps to a space-time word, a sequence of channel symbols $\mathbf{s}_j$. Making use of definition 1, we note that there are $\delta_j(\mathbf{c})$ number of channel symbol vectors $\mathbf{s}_j$ in $\mathbf{X}$. Likewise, we can find the numbers of other channel-symbol vectors in $\mathbf{X}$.

Now the following definition will help us identify those pairwise error events which lead to an identical PEP under the input/output relationship given in Eq. 1. For this, we momentarily assume that a codeword $\mathbf{c}$ is selected and have it held fixed. Relating to definition 1, its BSWP $\hat{\underline{\delta}}$ is fixed as well.

Definition 2 Pairwise distance profile. In a pair of codewords $\mathbf{c}$ and $\mathbf{c}'$, there are a total of T binary string pairs. Likewise, in the corresponding pair of space-time words $\mathbf{X}$ and $\mathbf{X}'$, there are a total of T channel symbol pairs (x_t, x_t') for $t \in \{1, \cdots, T\}$. We use $\delta_{j,k}$ to denote the number of time indices that a particular channel symbol pair $(x_t = s_j, x_t' = s_k)$ appears in a pair of ST words. Note that all different combinations of $j,k \in \{0, \cdots, J\text{-}1\}$ are possible. The collection of all $\delta_{j,k}$ can be stored into an array of size T^2. The array is referred to as the *pairwise distance profile* (PDP) between $\mathbf{X}$ and $\mathbf{X}'$ (or between $\mathbf{c}$ and $\mathbf{c}'$). Let's use $\underline{\delta} := (\underline{\delta}_0, \underline{\delta}_1, \ldots, \underline{\delta}_{J-1})$ to denote the collection and each $\underline{\delta}_j$ is further defined as $\underline{\delta}_j := (\delta_{j,0}, \delta_{j,1}, \ldots, \delta_{j,J-1})$ for each $j \in \{0, \cdots, J\text{-}1\}$.

Using the definition of *pairwise distance profile* $\underline{\delta}$, one can succinctly represent a set of erroneous words $\mathbf{c}'$ each of which leads to an identical PEP. Namely, for a fixed $\mathbf{c}$, the group of words $\mathbf{c}'$ satisfy the following two constraints

$$0 \le \delta_{j,k} \le \delta_j(\mathbf{c}), \text{ and}$$
$$\sum_{k=0}^{J-1} \delta_{j,k} = \delta_j(\mathbf{c}) \tag{A21}$$

for each $j \in \{0,\cdots, J\text{-}1\}$. We note then that the sum of all elements in the profile should be equal to T because it is the total number of symbol vector pairs in any pair of space-time words.

In summary, we have the following defined:

1. $\delta_{j,k}$ is the count of occurrences of a channel-symbol pair in a pair of sequence for j, k=0, 1, …, J-1.
2. A *pairwise distance profile* $\underline{\delta}(\mathbf{c})=(\delta_0, \delta_1, \ldots, \delta_{J\text{-}1})$ is an array of collection of all $\delta_{j,k}$. This distance profile is between the two words in a pair. One is the test word **c** (or its corresponding space-time word x(**c**)). The other is an erroneous word c′ (or its corresponding x′(c′)). A single profile is sufficient to represent a group of erroneous binary words (c′) (or x′(c′)) whose PEPs (Eq. A20) are the same.
3. The collection of all words with a same PDP $\underline{\delta}(\mathbf{c})$ and its size: for any fixed word **c**, we may want to count all the words which share the same PDP $\underline{\delta}(\mathbf{c})$. Any erroneous error pattern belonging to this collection will generate the same PEP. The cardinality of this group of words is of interest and it can be written as the following,

$$\prod_{j=0}^{J-1} \binom{\delta_j(\mathbf{c})}{\delta_{j,0}, \ldots, \delta_{j,J-1}}, \tag{A22}$$

where $\binom{\delta_j(c)}{\delta_{j,0}, \ldots \delta_{j,J-1}}$ is the multinomial coefficient.

The following remarks show the usefulness of PDP.

Remark 2 The same PDP leads to the same PEP; but not vice versa.

(a) The set of candidate codewords c′ can be partitioned with respect to distinct PDP such that each partition contains codewords with an identical PDP.

(b) Each codewords c′ in a partition has the same PEP.

A codeword pair can be described by a PDP $\underline{\delta}$. A PDP is for each and every possible channel-symbol pair, and for each it specifies the total number of times a particular channel-symbol pair appears in a pair of ST words. Once a PDP is given, the PEP in Eq. A20 can be rewritten as,

$$\begin{aligned} P(\mathbf{c} \to \mathbf{c}') &\le \prod_{j=0}^{J-1}\prod_{k=0}^{J-1}\left[\left(1+\frac{1}{4N_o}\left|s_j - s_k\right|^2\right)^{-N}\right]^{-\delta_{j,k}} \\ &= \prod_{j=0}^{J-1}\prod_{k=0}^{J-1}\beta_{j,k}{}^{\delta_{j,k}} \end{aligned} \tag{A23}$$

by grouping the like terms under each power exponent $\delta_{j,k}$. Here, we made use of the memory-less property of the ergodic channel (Eq. 1). Note that the critical information needed to write (Eq. A23) is stored in the PDP. Note that terms $\beta_{j,k}$ are defined as

$$\beta_{j,k} := \left(1+\frac{1}{4N_0}\left|s_j - s_k\right|^2\right)^{-N} \tag{A24}$$

which are completely determined and held fixed once a vector constellation and power spectral density of the noise are given.

The rationale behind the definition of the PDP $\underline{\delta}$ as a distance metric should be clear now: For a given SNR, the pairwise distance profile $\underline{\delta}$ completely determines the upper bound formulation of the PEP. As will be noted in the subsequent sections, use of the two profiles greatly simplifies the union bound evaluation.

In the union bounds for binary transmission over AWGN channels, for example, the use of distance profiles based on Hamming weights and Hamming distances greatly simplifies the calculation of union bound. The calculation of union bounds becomes quite complex for a J-ary vector constellation. The two profiles play the roles similar to the Hamming weight and Hamming distance in AWGN channels, and simply the union bound evaluation for MIMO channels.

The following lemma is the multinomial expansion. It will prove useful to write it here; while we omit the proof.

Lemma *Consider a set of J utility variables* $\{z_j\}_{j=0,1,\ldots,J\text{-}1}$ *and the T-th power of the sum of the J utility variables* $\left(\sum_{j=0}^{J-1} z_j\right)^T$. *Then, for an array of integers* $\underline{v}=(v_0, v_1, \ldots, v_{J\text{-}1})$, *the T-th power of the sum can be expanded as,*

$$\left(\sum_{j=0}^{J-1} z_j\right)^T = \sum_{\underline{v}\in\Omega}\binom{T}{v_0, v_1, \cdots v_{J-1}}\prod_{j=0}^{J-1} z_j^{v_j} \tag{A25}$$

where, Ω is the collection of arrays $\underline{v}$, i.e.,

$$\Omega := \left\{\underline{v}\middle| v_j \in \{0,1,\ldots,T\}, \sum_{j=0}^{J-1} v_j = T\right\},$$

and

$$\binom{\sum v_i}{v_0, v_1, \cdots v_{n-1}} := \frac{(\sum v_i)!}{\prod v_i!}$$

is the multinomial coefficient.

Remark 3 Setting all utility variables to be equal to 1, we note that

$$\sum_{\underline{v}\in\Omega}\binom{T}{v_0, v_1, \cdots v_{J-1}} = J^T \tag{A26}$$

Appendix B. Proof of theorem

We now discuss the proof of Theorem. The sketch of proof goes as follows:

1. The random block code is not linear. Unlike linear codes, the all-zero codeword alone is not enough to be selected as the test codeword **c**. For a given codebook, one must take the average over all randomly selectable test codeword **c**. This is a difficult task.
2. The obstacle is circumvented by taking the ensemble average over all equally probable selection of random codebooks. The two profiles defined in Appendix A are useful to simplify the union bound.

Proof Consider the ensemble of randomly selectable (L,K) block codes. First, let us consider a code C in the ensemble and the calculation of probability of maximum-likelihood decoding error. A union bound to this error probability is given as follows

$$\begin{aligned} P_e(C) &= E_{c\in\mathcal{C}}\left[P_{e|c}\right] \\ &\le E_{c\in C}\left[\sum_{\mathbf{c}'\in C,\mathbf{c}'\neq c}\Pr(\mathbf{c}\to\mathbf{c}')\right] \\ &= \frac{1}{2^K}\sum_{\substack{\mathbf{c}\in C \\ \mathbf{c}'\in C,\mathbf{c}'\neq\mathbf{c}}}\Pr(\mathbf{c}\to\mathbf{c}'), \end{aligned} \tag{A27}$$

where, $E_c[\bullet]$ is the expectation over the choice of a test codeword out of 2^K equi-probably selectable codewords in the code C; $P_{e|c}$ denotes the error probability conditioned on the transmission of a test codeword **c**; and the inequality is due to the usual union bound argument.

Then, the average probability of decoding error over the ensemble of codes can be formulated according to Eq. A27 as follows,

$$\begin{aligned} \overline{P_e} &= E_{C\in\mathbb{C}}[P_e(C)]\frac{1}{|\mathbb{C}|}\left[\sum_{C\in\mathbb{C}}P_e(C)\right] \\ &\le \frac{1}{2^K|\mathbb{C}|}\sum_{C\in\mathbb{C}}\sum_{\substack{C\in\mathbb{C} \\ c'\in C,c'\neq c}}\Pr(\mathbf{c}\to\mathbf{c}') \end{aligned} \tag{A28}$$

where we make use of the assumption that each code in the ensemble is selected with equal probability.

In Eq. A28, note that (1) the inner summation shall be conducted over all codeword pairs **c** and **c**′ (**c**≠**c**′) where both should be element codewords in the same codebook C; that (2) the outer summation is implied for each and every code in the ensemble $\mathbb{C}$.

It should be noted that the pilot codeword **c** should be selected out of 2^K valid codewords within each codebook (the inner summation); but looking at it from the perspective of considering all codebooks in the ensemble, each and every possible 2^L distinct binary string of length L should be considered as the test codeword at least once.

Making use of this observation and changing the order of the summations, Eq. A28 can be rewritten as,

$$\overline{P}_e \le \frac{1}{2^K|\mathbb{C}|}\sum_{\mathbf{c}}\sum_{\substack{\mathbf{c}':\mathbf{c}'\neq\mathbf{c}, \\ \mathbf{c},\mathbf{c}'\in C,C\in\mathbb{C}}}\Pr(\mathbf{c}\to\mathbf{c}') \tag{A29}$$

where the outer sum is now over all 2^L distinct binary string **c** of length L. The inner sum is to count in all codeword **c**′ which are different from **c**, but must coexist with **c** in the same codebook. Of course, there are only a finite number of codebooks that possess both as its element codewords. Shortly later, this quantity will be obtained explicitly under the assumption of random coding argument (see Eq. A35).

Let us now consider the inner sum over all binary string **c**′ with respect to a binary string **c** that have a *binary string weigh profile* $(\delta_0(\mathbf{c}), \delta_1(\mathbf{c}), \ldots, \delta_{J\text{-}1}(\mathbf{c}))$. Recall that the pairwise error probability from **c** to **c**′ in Eq. A23 is completely determined by their pairwise distance profile $\underline{\delta}$. The summation over **c**′ thus can be re-arranged with respect to the PDP. We collect a single representative string per each group of strings **c**′ which possess the same PDP, and call it **c**″. That is, a string c″ represents all binary strings **c**′ each of which has the same *pairwise distance profile*, $\underline{\delta}(\mathbf{c})$, from the test word **c**.

Now Eq. A29 can be rewritten as

$$\begin{aligned} \overline{P}_e &\le \frac{1}{2^K|\mathbb{C}|}\sum_{\mathbf{c}}\sum_{\substack{\mathbf{c}'':\underline{\delta}\in\Omega(\mathbf{c}) \\ \underline{\delta}\neq\underline{\delta}^*}} S_{\underline{\delta}}(c)\Pr(\mathbf{c}\to\mathbf{c}'') \\ &= \frac{1}{2^K|\mathbb{C}|}\sum_{\mathbf{c}}\sum_{\substack{\mathbf{c}'':\underline{\delta}\in\Omega(\mathbf{c}) \\ \underline{\delta}\neq\underline{\delta}^*}} S_{\underline{\delta}}(\mathbf{c})\prod_{j,k=0}^{J-1}\beta_{j,k}^{\delta_{j,k}}. \end{aligned} \tag{A30}$$

A few explanations are in order per Eq. A30. First, $\Omega(\mathbf{c})$ denotes the set of all possible pairwise distance profile $\underline{\delta}(\mathbf{c}) := (\underline{\delta}_0, \underline{\delta}_1, \ldots, \underline{\delta}_{J\text{-}1})$ anchored at the test word **c**. Making use of our definition in Eq. A21, we have,

$$\Omega(\mathbf{c}) := \left\{\underline{\delta}\middle|\underline{\delta}_j \in \Omega_j(\mathbf{c}),\ \text{for } j = 0,1,2,\ldots,J-1\right\}, \tag{A31}$$

where,

$$\Omega_j(\mathbf{c}) := \left\{\underline{\delta}_j\middle|\delta_{j,k}\in\{0,1,\ldots,\delta_j(\mathbf{c})\}, \sum_{k=0}^{J-1}\delta_{j,k} = \delta_j(\mathbf{c})\right\}. \tag{A32}$$

Since the inner summation is taken over all the distinct strings $\mathbf{c}''$, each representing a group of equivalent strings with the same PDP, the size of the group should be calculable and it is the multinomial coefficient $\prod_{j=0}^{J-1}\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}$. It is the number of ways to come up with the equivalent strings which possess the given PDP. We will incorporate this factor into the parameter $S_{\underline{\delta}}(\mathbf{c})$; see Eq. A35.

Second, $\underline{\delta}^*$ denotes the *unique* PDP of a word anchored at itself, and thus $\underline{\delta}\neq\underline{\delta}^*$ is equivalent to $\mathbf{c}'\neq\mathbf{c}$. Notice that for $\underline{\delta}^*$ the entries are given as $\delta_{j,k}=\delta_j(\mathbf{c})$ for $j=k$ and $\delta_{jk}=0$ otherwise.

Just for a check, we take the sum of all coefficients and find:

$$\sum_{\mathbf{c}\in GF(2)^L}\sum_{\mathbf{c}'':\,\underline{\delta}\in\Omega(\mathbf{c}),\underline{\delta}\neq\underline{\delta}^*}\prod_{j=0}^{J-1}\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}=2^L\left(2^L-1\right). \tag{A33}$$

Third, we use $S_{\underline{\delta}}(c)$ to subsume the rest of the factors. It should point to the number of all erroneous codewords $\mathbf{c}'$ which have the pairwise distance profile $\underline{\delta}$ from $\mathbf{c}$, counted for all valid codebooks. Note that a test word $\mathbf{c}$ exists only in a certain number of codebooks. Such occasions should be counted properly in the parameter $S_{\underline{\delta}}(\mathbf{c})$. Thus, taking the summation of $S_{\underline{\delta}}(\mathbf{c})$ over all PDP and all codeword pairs shall give a number equal to the product of $2^K(2^K\text{-}1)$ and the cardinality of the (L,K) code ensemble, i.e.,

$$\sum_{\mathbf{c}\in GF(2)^L}\sum_{\mathbf{c}'':\,\underline{\delta}\in\Omega(\mathbf{c}),\underline{\delta}\neq\underline{\delta}^*}S_{\underline{\delta}}(\mathbf{c})=|\mathbb{C}|2^K\left(2^K-1\right). \tag{A34}$$

The value of $S_{\underline{\delta}}(\mathbf{c})$ can be calculated using the usual combinatorial methods:

$$S_{\underline{\delta}}(\mathbf{c})=\left[\frac{2^K}{2^L}|\mathbb{C}|\right]\cdot\left[\frac{2^K-1}{2^L-1}\right]\cdot\left[\prod_{j=0}^{J-1}\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}\right], \tag{A35}$$

for $\underline{\delta}\neq\underline{\delta}^*$. The first term is the number of codes in the ensemble that include a word $\mathbf{c}$ as a codeword. Only a certain fraction of these codes also include $\mathbf{c}''$ as its element codeword, which is the second term. There are 2^K–1 binary strings out of 2^L– 1 available to be selected as the erroneous codeword. Therefore, the number of codebooks which contain both $\mathbf{c}$ and $\mathbf{c}''$ simultaneously is the product of the first two terms in Eq. A35. The third term is the tally of all possible ways of having the binary strings for an erroneous codeword which possess a PDP $\underline{\delta}$ anchored at the test word $\mathbf{c}$ and thus satisfying all the constraints due in Eq. A31 and Eq. A32.

Substituting Eq. A35 into Eq. A30, we have

$$\overline{P}_e\le\frac{1}{2^L}\frac{2^K-1}{2^L-1}\sum_{\mathbf{c}\in GF(2)^L}\sum_{\mathbf{c}'':\,\underline{\delta}\in\Omega(\mathbf{c})\atop\underline{\delta}\neq\underline{\delta}^*}\prod_{j=0}^{J-1}\left[\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}\prod_{k=0}^{J-1}\beta_{j,k}^{\delta_{j,k}}\right]. \tag{A36}$$

Notice that, for $\underline{\delta}=\underline{\delta}^*$ (i.e., $\mathbf{c}'=\mathbf{c}$), the multinomial coefficient in Eq. A36 equals 1, and also $\beta_{j,k}=1$ according to Eq. A24. Thus, Eq. A36 can be rewritten by considering $\underline{\delta}\neq\underline{\delta}^*$ separately,

$$\overline{P}_e\le\frac{1}{2^L}\frac{2^K-1}{2^L-1}\sum_{\mathbf{c}}\sum_{\underline{\delta}\in\Omega(\mathbf{c})}\prod_{j=0}^{J-1}\left[\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}\prod_{k=0}^{J-1}\beta_{j,k}^{\delta_{j,k}}\right]-\frac{2^K-1}{2^L-1}. \tag{A37}$$

Recalling the definition in Eq. A31 that the J constraints for $\Omega(\mathbf{c})$ are not coupled with each other, the sum over $\underline{\delta}=(\delta_0,\delta_1,\ldots,\delta_{J-1})\in\Omega(\mathbf{c})$ can be simplified as follows,

$$\begin{aligned}&\sum_{\underline{\delta}\in\Omega(c)}\prod_{j=0}^{J-1}\left[\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}\prod_{k=0}^{J-1}\beta_{j,k}^{\delta_{j,k}}\right]\\&=\sum_{\underline{\delta}_0\in\Omega_0(\mathbf{c})}\cdots\sum_{\underline{\delta}_{J-1}\in\Omega_{J-1}(\mathbf{c})}\prod_{j=0}^{J-1}\left[\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}\prod_{k=0}^{J-1}\beta_{j,k}^{\delta_{j,k}}\right]\\&=\prod_{j=0}^{J-1}\sum_{\underline{\delta}_j\in\Omega_j(\mathbf{c})}\left[\binom{\delta_j(\mathbf{c})}{\delta_{j,0},\ldots,\delta_{j,J-1}}\prod_{k=0}^{J-1}\beta_{j,k}^{\delta_{j,k}}\right]\\&=\prod_{j=0}^{J-1}\left(\sum_{k=0}^{J-1}\beta_{j,k}\right)^{\delta_j(\mathbf{c})},\end{aligned} \tag{A38}$$

where, the last equality follows from the Lemma.

Substituting Eq. A38 into Eq. A36, we have,

$$\overline{P}_e\le\frac{1}{2^L}\cdot\frac{2^K-1}{2^L-1}\sum_{\mathbf{c}\in\mathrm{GF}(2)^L}\prod_{j=0}^{J-1}\left(\sum_{k=0}^{J-1}\beta_{j,k}\right)^{\delta_j(\mathbf{c})}-\frac{2^K-1}{2^L-1}. \tag{A39}$$

Now, let us move on to the summation over $\mathbf{c}$: it is over all 2^L distinct binary strings of length L, as mentioned before. Similar to the case for $\mathbf{c}'$, this summation can be reorganized with respect to the binary string weigh profile $\hat{\underline{\delta}}(\mathbf{c})=(\delta_0,\delta_1,\ldots,\delta_{J-1})$ associated with each $\mathbf{c}$. That is,

$$\overline{P}_e\le\frac{1}{2^L}\frac{2^K-1}{2^L-1}\sum_{\hat{\underline{\delta}}\in\widehat{\Omega}}\widehat{A}_{\hat{\underline{\delta}}}\prod_{j=0}^{J-1}\left(\sum_{k=0}^{J-1}\beta_{j,k}\right)^{\delta_j}-\frac{2^K-1}{2^L-1}. \tag{A40}$$

where, $\widehat{A}_{\underline{\hat{\delta}}}$ is the number of binary strings of length L that have a metric $\underline{\hat{\delta}}$; i.e., each of these strings can be regarded as a concatenation of a number δ_j of binary sub-string b_j (j=1, 2, …, J-1). Let $\widehat{\Omega}$ denote the set of all possible metric $\underline{\hat{\delta}}$ and according to definition 1, we have,

$$\widehat{\Omega} := \left\{ \underline{\hat{\delta}} \middle| \delta_j \in \{0, 1, ..., T\}, \sum_{j=0}^{J-1} \delta_j = T \right\}. \tag{A41}$$

Similar to $S_{\underline{\delta}}(\mathbf{c})$, we use the combinatorial analysis to calculate $\widehat{A}_{\underline{\hat{\delta}}}$,

$$\widehat{A}_{\underline{\hat{\delta}}} = \begin{pmatrix} T \\ \delta_0, ..., \delta_{J-1} \end{pmatrix}, \tag{A42}$$

which is the number of ways to arrange a number δ_j of binary sub-strings b_j (for j=0, 1, 2, … J-1). As expected, we can verify that $\sum_{\underline{\hat{\delta}} \in \widehat{\Omega}} \widehat{A}_{\underline{\hat{\delta}}} = J^T = 2^{MK_bT} = 2^L$.

Substituting Eq. A42 into Eq. A40, we have

$$\begin{aligned} \overline{P}_e &\le \frac{1}{2^L} \frac{2^K - 1}{2^L - 1} \sum_{\underline{\hat{\delta}} \in \widehat{\Omega}} \begin{pmatrix} T \\ \delta_0, ..., \delta_{J-1} \end{pmatrix} \prod_{j=0}^{J-1} \left[\sum_{k=0}^{J-1} \beta_{j,k} \right]^{\delta_j} - \frac{2^K - 1}{2^L - 1} \\ &= \frac{1}{2^L} \frac{2^K - 1}{2^L - 1} \left(\sum_{j,k=0}^{J-1} \beta_{j,k} \right)^T - \frac{2^K - 1}{2^L - 1}, \end{aligned} \tag{A43}$$

where the equality is obtained by applying the Lemma.

The bound can be further upper-bounded by

$$\overline{P}_e \le \frac{1}{2^L} \frac{2^K}{2^L} \left(\sum_{j,k=0}^{J-1} \beta_{j,k} \right)^T. \tag{A44}$$

Using L=TMK_b and R_c=K/L and rewriting Eq. A44 in an exponential form, we reach the result of the Theorem 2. □

References

1. Telatar IE (1999) Capacity of multi-antenna Gaussian channels. Eur Trans Telecommun 10:585–595
2. Foschini GJ, Gans MJ (1998) On limits of wireless communications in a fading environment when using multiple antennas. Wirel Pers Commun 6:311–335
3. Berrou C, Glavieux A, Thitimajshima P (1993) Near Shannon limit error-correcting coding and decoding: turbo-codes. In Proc IEEE Int Conf Commun, Geneva, Switzerland, pp. 1064–1070
4. Gallager RG (1963) Low density parity check codes, monograph, M.I.T. Press
5. Luby MG, Mitzenmacher M, Shokrollahi MA, Spielman DA (2001) Efficient erasure correcting codes. IEEE Trans Inf Theory 47:569–584
6. Richardson TJ, Amin Shokrollahi M, Urbanke RL (2001) Design of capacity-approaching irregular low-density parity-check codes. IEEE Trans Inf Theory 47(2):619–37
7. Hochwald BM, ten Brink S (2003) Achieving near-capacity on a multiple-antenna channel. IEEE Trans Commun 51(3):389–399
8. Zhang J, Lee H-N (2008) Performance analyses on LDPC coded system over quasi-static (MIMO) fading system. IEEE Trans Commun 56(12):2080–2093
9. Zhang J, Lee H-N (2006) A performance bound on random-coded MIMO systems. IEEE Commun Lett 10(3):168–170
10. Zhang J, Lee H-N (2005) Union bounds on LDPC coded modulation systems over fast fading MIMO channels. IEEE Commun Lett 9(9):796–798
11. Litsyn S, Shevelev V (2003) Distance distributions in ensembles of irregular low-density parity-check codes. IEEE Trans Inf Theory 49(12):3140–3159
12. Sason I, Shamai S (2001) On improved bounds on the decoding error probability of block codes over interleaved fading channels, with applications to turbo-like codes. IEEE Trans Inf Theory 47 (6):2275–2299
13. Burshtein D, Miller G (2004) Asymptotic enumeration methods for analyzing LDPC codes. IEEE Trans Inf Theory 50(6):1115–1131
14. Berrou C, Vaton S (2002) Computing the minimum distance of linear codes by the error impulse method. Proc IEEE Intl Symp Inform Theory, Lausanne, Switzerland
15. Xiao-Yu H, Fossorier MPC, Eleftheriou E (2004) On the computation of the minimum distance of low-density parity-check codes. Proc IEEE Int Conf Comm 2:767–771
16. Perez LC, Seghers J, Costello DJ Jr (1996) A distance spectrum interpretation of turbo codes. IEEE Trans Inf Theory 42(6):1698–1709
17. Kou Y, Lin S, Fossorier MPC (2001) Low-density parity-check codes based on finite geometries: a rediscovery and new results. IEEE Trans Inf Theory 47(7):2711–36
18. Vasic B, Milenkovic O (2004) Combinatorial constructions of low-density parity-check codes for iterative decoding. IEEE Trans Inf Theory 50(6):1156–76
19. Gallager RG (1968) Information theory and reliable communications, Wiley
20. Baccarelli E (2001) Evaluation of the reliable data rates supported by multiple-antenna coded wireless links for QAM transmissions. IEEE J Sel Areas Commun 19(2):295–304
21. He W, Georghiades CN (2005) Computing the capacity of a MIMO fading channel under PSK signaling. IEEE Trans Inform Theory 51(5):1794–1803
22. Huang J, Meyn SP (2005) Characterization and computation of optimal distributions for channel coding. IEEE Trans Inf Theory 51(7):2336–2351
23. Bellorado J, Kavcic A (2003) Approaching the capacity of the MIMO Rayleigh flat-fading channel with QAM constellations, independent across antennas and dimensions, in Proc. IEEE Intl. Symp. Inform. Theory, Yokohama, Japan, pp. 270
24. Tarokh V, Seshadri N, Calderbank AR (1998) Space–time codes for high data rate wireless communication: performance criterion and code construction. IEEE Trans Inf Theory 44(2):744–765
25. Jalden J, Skoglund M, Ottersten B (2004) On the random coding exponent of multiple antenna systems using space-time block codes, in Proc IEEE Intl Symposium Inform. Theory, Chicago, IL, pp.188
26. Venkataraman P (2009) Applied optimization with matlab programming. Wiley