The Throughput-Reliability Tradeoff in Block-Fading MIMO Channels

Kambiz Azarian and Hesham El Gamal*

Abstract

We build on Zheng and Tse's elegant formulation of diversity-multiplexing tradeoff to provide a better understanding of the asymptotic interplay between transmission rate, error probability and signal-to-noise ratio in block-fading MIMO channels. In particular, we identify the limitation imposed by the notion of multiplexing gain and develop a new formulation called the throughput-reliability tradeoff, that avoids this limitation. The new characterization is then used to elucidate the asymptotic trends exhibited by the outage probability curves of block-fading MIMO channels.

1 Problem Formulation

This paper revolves around the following question: What does a 3 dB gain in signal-to-noise ratio (SNR) provide in a block-fading multi-input multi-output (MIMO) channel? For a *non-fading* MIMO with m transmit and n receive antennas, it is well known that a 3 dB gain in SNR translates into $\min\{m,n\}$ extra bits in channel capacity in the high SNR regime. The scenario considered in this paper, however, is more involved. We study a *block-fading* case where the path gains remain constant over l consecutive symbol-intervals (*i.e.,* a block), but change independently from one block to another. We further assume a coherent communication model implying the availability of channel state information (CSI) only at the receiver. More specifically, we adopt the following input-output relation for the channel under study

$$\mathbf{y} = \sqrt{\frac{\rho}{m}}\mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^n$ has entries y_i representing the signal received at antenna $i \in \{1, \cdots, n\}$, $\mathbf{x} \in \mathbb{C}^m$ has entries x_j representing the signal transmitted by antenna $j \in \{1, \cdots, m\}$, $\mathbf{H} \in \mathbb{C}^{n\times m}$ has entries h_{ij} representing the path gain between receive antenna $i \in \{1, \cdots, n\}$ and transmit antenna $j \in \{1, \cdots, m\}$, and $\mathbf{w} \in \mathbb{C}^n$ represents the unit-variance additive white Gaussian noise. We model $\{h_{ij}\}$ as i.i.d complex Gaussian random variables with zero-mean and unit-variance. Finally, ρ corresponds to the SNR at each receive antenna. Under these assumptions, the randomness of the instantaneous mutual information results in a *non-zero* lower bound on the error probability for any

*The authors are with the ECE Department at the Ohio State University (email:azariany, helgamal@ece.osu.edu).

fixed transmission rate. In such a scenario, the SNR gain can either be used to increase the transmission rate or lower the error probability or a combination of the two. Hence, a fundamental tradeoff between throughput (i.e., transmission rate) and reliability (i.e., error probability) arises. In [1], Zheng and Tse provided an elegant formulation for this tradeoff in the high SNR regime called the diversity-multiplexing tradeoff (DMT). In this formulation, throughput and reliability are quantified through the notions of multiplexing and diversity gains, respectively. More specifically, the DMT assumes a family of codes $\{\mathcal{C}_\rho\}$, such that the code $\mathcal{C}_\rho$ (corresponding to SNR ρ) has rate $R(\rho)$, in bits per channel use (bpcu), and error probability $P_e(\rho)$. For this family, multiplexing gain r and diversity gain d are defined by[1]

$$r \triangleq \lim_{\rho\to\infty} \frac{R(\rho)}{\log\rho} \quad \text{and} \quad d \triangleq -\lim_{\rho\to\infty} \frac{\log P_e(\rho)}{\log\rho}. \tag{2}$$

Now, a protocol's DMT characterizes the diversity gain $d(r)$ that the protocol achieves for a multiplexing gain r. Also, the best DMT achieved by any protocol in a channel is called that channel's optimal DMT. For example, the following theorem gives the optimal DMT for block-fading MIMO channels [1].

Theorem 1 *[1] The optimal DMT for a coherent block-fading MIMO channel with m transmit and n receive antennas is given by $d(r) = f(r)$, where $f(\cdot)$ is the piecewise linear function joining the points $(k, (m-k)(n-k))$, for $k = 0, \ldots, \min\{m, n\}$. Moreover, there exists a code that achieves $d(r)$ for all block lengths $l \geq m + n - 1$.*

To motivate our work, we next use the DMT to make a first attempt towards answering our central question on the utility of a 3 dB SNR gain in block-fading MIMO channels. In particular, we borrow the following rule from [2].

- At high enough SNRs and operating point $(r, d(r))$, where $r \triangleq R/\log\rho$, a 3 dB SNR gain can be used to approximately increase the transmission rate by r bpcu *and* reduce the error probability by a factor of $2^{-d(r)}$.

This rule elucidates two very important aspects of the DMT. First, since the DMT is an *asymptotic* formulation, it only yields *approximate* predictions for any *finite* SNR. Second, since the DMT requires the transmission rate and error probability to change in a rather *restricted* way, it only offers a partial answer to our question. To further illustrate this point, we consider in some details the following two special cases of the aforesaid rule.

- Case One ($r = 0, d = d_{max}$): At high enough SNRs and operating point $(0, d_{max})$, one can fix the transmission rate and obtain approximately d_{max} orders of decay in outage probability (on a log-log scale) for every 10 dB gain in SNR.

- Case Two ($r = r_{max}, d = 0$): At high enough SNRs and operating point $(r_{max}, 0)$, one can fix the outage probability and obtain approximately r_{max} bpcu increase in rate for every 3 dB gain in SNR.

These two cases are illustrated, for a 2×2 block-fading MIMO channel, in Fig. 1 and Fig. 2, respectively.

[1]Unless otherwise stated, in this paper all logarithms are assumed to be in base 2.

- The slope of the outage probability curves in Fig. 1 is shown to approach the asymptotic value of $d_{max} = 4$, as predicted by our first special case. Here, however, we can not use this rule to predict the SNR gain required for increasing the transmission rate (at a fixed outage probability) from 4 to 8 bpcu, which the figure suggests to be 9 dB. Interestingly, this horizontal spacing of 9 dB seems to remain valid for operating points for which $R/\log\rho$ is as large as 0.8 (*e.g.*, $R = 4$ bpcu and $\rho = 15$ dB).

- The SNR gain required for increasing the transmission rate by 4 bpcu in Fig. 2 is seen to be 6 dB, as predicted by our second special case. Here, however, we fail to predict the slope of the outage probability curves for fixed transmission rates. Interestingly, this slope, which is seen from the figure to be 2, remains valid for operating points for which $R/\log\rho$ is as small as 1.2 (*e.g.*, $R = 28$ bpcu and $\rho = 70$ dB).

Inspired by these observations, this paper aims at developing a better understanding of the fundamental tradeoff between throughput and reliability in block-fading MIMO channels. It turns out that such an understanding requires a more general formulation which is not limited by the multiplexing gain notion as defined in (2). In particular, the multiplexing gain notion limits the scenarios of interest to those where R asymptotically scales *linearly* with $\log\rho$. Our formulation, on the other hand, allows for investigating more general scenarios by relaxing this constraint. Specifically, we shed more light on the relationship between the three quantities $(R, \log\rho, P_e(R,\rho))$, in the asymptotic limit of large ρ, when

$$\limsup_{\rho\to\infty} \frac{R}{\log\rho} \neq \liminf_{\rho\to\infty} \frac{R}{\log\rho}. \tag{3}$$

It is clear that (3) allows for investigating scenarios defined by arbitrary asymptotic trajectories in the $R - \log\rho$ plane where the multiplexing gain may not be defined. As argued in the sequel, this *freedom* of walking along arbitrary trajectories on the $R - \log\rho$ plane, is the *key* for obtaining accurate predictions for the outage probability slopes and horizontal spacings in different operating regions[2]. While our characterization, like the DMT, is only *rigorous* in the asymptotic scenario where SNR grows to infinity (i.e., $\rho \to \infty$), we will demonstrate, via numerical results, that it yields very accurate predictions for practically relevant values of SNR.

The rest of the paper is organized as follows. In Section 2, we state our main result formulating the throughput-reliability tradeoff (TRT) for block-fading MIMO channels, along with its proof. In this section, we also present intuitive arguments and numerical results that demonstrate the utility of our results in predicting the *behavior* of outage probability curves in the high SNR regime. Section 3 utilizes the TRT to shed more light on the performance of various space-time architectures and further extends our investigation to automatic repeat request (ARQ) channels. We offer few concluding remarks in Section 4. In order to enhance the flow of the paper, we collect the proofs of the theorems (except for Theorem 2) in the Appendix.

[2]A more formal definition of an operating region is presented in the sequel.

2 The Throughput-Reliability Tradeoff (TRT)

An outage is defined as the event that the instantaneous mutual information does not support the intended rate [8, 9], i.e.,

$$O_p \triangleq \{H \in \mathbb{C}^{n\times m} | I(\mathbf{x};\mathbf{y}|\mathbf{H}=H) < R\}. \tag{4}$$

Notice that the instantaneous mutual information $I(\mathbf{x};\mathbf{y}|\mathbf{H}=H)$ depends on both the channel realization H and the input distribution $p(\mathbf{x})$. The outage probability $P_o(R,\rho)$ is then defined as [8, 9]

$$P_o(R,\rho) = \inf_p \Pr\{O_p\}.$$

The following theorem characterizes the asymptotic relationship between R, ρ, and $P_o(R,\rho)$.

Theorem 2 *For the $m \times n$ coherent block-fading MIMO channel described by (1),*

$$\lim_{\substack{\rho\to\infty \\ (R,\rho)\in\mathcal{R}(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log\rho} = -g(k), \tag{5}$$

where $P_o(R,\rho)$ denotes the outage probability at rate R and SNR ρ, and $\mathcal{R}(k)$ denotes the kth operating region (for integer k) defined by

$$\mathcal{R}(k) \triangleq \{(R,\rho) | k+1 > \frac{R}{\log\rho} > k\} \quad for \quad k\in\mathbb{Z}, \min\{m,n\} > k \geq 0. \tag{6}$$

In (5), $c(k)$ and $g(k)$ are given by

$$c(k) \triangleq m+n-(2k+1), \tag{7}$$

and

$$g(k) \triangleq mn - k(k+1). \tag{8}$$

Moreover, in the degenerate region where $R > \min\{m,n\}\log\rho$, $\lim_{\rho\to\infty} \log P_o(R,\rho)/\log\rho = 0$. In this formulation, $g(k)$ is referred to as the reliability gain coefficient and $t(k) \triangleq g(k)/c(k)$ as the throughput gain coefficient.

Proof: The proof follows the same lines as that of Theorem 4 in [1], except for the fundamental challenge that the multiplexing gain is not defined here. This means that the notion of exponential equality (i.e., $\doteq$), which is instrumental in deriving the corresponding result in [1], is unapplicable. To handle this challenge, we judicially choose the operating region $\mathcal{R}(k)$ and find a lower bound on

$$\liminf_{\substack{\rho\to\infty \\ (R,\rho)\in\mathcal{R}(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log\rho},$$

and an upper bound on

$$\limsup_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho}$$

and show that the two bounds coincide for the choice of $c(k)$ and $g(k)$ given by (7) and (8), respectively. In particular, we start with the following two bounds on the outage probability [1], i.e.

$$P_o(R,\rho) \leq \Pr\{\log\det(I_n + \frac{\rho}{m} HH^H) < R\} \text{ and}$$
$$P_o(R,\rho) \geq \Pr\{\log\det(I_n + \rho HH^H) < R\},$$

to write

$$P_o(R,\rho) \leq \Pr\{\log(\prod_{i=1}^{q}(1+\frac{\rho}{m}\mu_i)) < R\} \text{ and} \tag{9}$$
$$P_o(R,\rho) \geq \Pr\{\log(\prod_{i=1}^{q}(1+\rho\mu_i)) < R\}, \tag{10}$$

where $q \triangleq \min\{m,n\}$ and $\mu_q \geq \cdots \geq \mu_1 \geq 0$ represent the ordered eigenvalues of HH^H. Notice that the joint probability density function (PDF) of $(\mu_1, \cdots, \mu_q)$ is given by the Wishart distribution [1]. Now, let us focus on (9) and introduce the change of variables

$$\alpha_i \triangleq \frac{\log(1+\frac{\rho}{m}\mu_i)}{R}. \tag{11}$$

This implies that $\alpha_q \geq \cdots \geq \alpha_1 \geq 0$. In terms of the new variables, (9) can be written as

$$P_o(R,\rho) \leq \Pr\{\mathcal{A}\}, \tag{12}$$

where

$$\mathcal{A} \triangleq \{\alpha | \alpha_q \geq \cdots \geq \alpha_1 \geq 0, 1 - \sum_{i}^{q} \alpha_i > 0\}, \tag{13}$$

with $\alpha \triangleq (\alpha_1, \cdots, \alpha_q)$ being distributed according to

$$p(\alpha) = KR^q \rho^{-mn} 2^{R\sum_i \alpha_i} \times$$
$$\prod_{i=1}^{q}(2^{\alpha_i R}-1)^{|m-n|} \prod_{i<j}(2^{\alpha_i R} - 2^{\alpha_j R})^2 e^{-\sum_i \frac{m(2^{\alpha_i R}-1)}{\rho}}.$$

In this expression, K is a normalizing factor. Next, we define $\mathcal{R}_\delta(k)$, for integer k's, as

$$\mathcal{R}_\delta(k) \triangleq \begin{cases} \{(R,\rho) | \frac{1}{\delta} > \frac{\log\rho}{R} > 1+\delta\} & \text{if} \quad k=0 \\ \{(R,\rho) | \frac{1}{k} - \delta > \frac{\log\rho}{R} > \frac{1}{k+1} + \delta\} & \text{if} \quad k \in \mathbb{Z}, q > k > 0 \end{cases}, \tag{14}$$

where δ denotes a small positive value. Notice that $\delta = 0$ reduces $\mathcal{R}_\delta(k)$ to $\mathcal{R}(k)$, as given by (6). Now, it follows from (12) that

$$P_o(R,\rho)2^{-c(k)R} \leq 2^{-c(k)R}\int_{\mathcal{A}} p(\alpha)\mathrm{d}\alpha, \quad (R,\rho)\in\mathcal{R}_\delta(k)$$

Note that this expression holds, regardless of the choice for $c(k)$, i.e., at this point we regard $c(k)$ as an *arbitrary* function of k. This inequality can be written as

$$P_o(R,\rho)2^{-c(k)R} \leq A_1(R,\rho,\epsilon) + A_2(R,\rho,\epsilon), \quad (R,\rho)\in\mathcal{R}_\delta(k)$$

where

$$A_1(R,\rho,\epsilon) \triangleq 2^{-c(k)R}\int_{\mathcal{A}_1} p(\alpha)\mathrm{d}\alpha, \qquad A_2(R,\rho,\epsilon) \triangleq 2^{-c(k)R}\int_{\mathcal{A}_2} p(\alpha)\mathrm{d}\alpha$$

and

$$\mathcal{A}_1 \triangleq \{\alpha\in\mathcal{A}|\alpha_q > \frac{\log\rho}{R}+\epsilon\}, \qquad \mathcal{A}_2 \triangleq \{\alpha\in\mathcal{A}|\frac{\log\rho}{R}+\epsilon \geq \alpha_q\}. \tag{15}$$

This means that

$$\begin{aligned}\limsup_{\substack{\rho\to\infty\\(R,\rho)\in\mathcal{R}_\delta(k)}} \frac{\log P_o(R,\rho)-c(k)R}{\log\rho} \leq & \limsup_{\substack{\rho\to\infty\\(R,\rho)\in\mathcal{R}_\delta(k)}} \frac{\log\left(1+A_1(R,\rho,\epsilon)/A_2(R,\rho,\epsilon)\right)}{\log\rho} + \\ & \limsup_{\substack{\rho\to\infty\\(R,\rho)\in\mathcal{R}_\delta(k)}} \frac{\log A_2(R,\rho,\epsilon)}{\log\rho},\end{aligned} \tag{16}$$

To characterize the first term in the right-hand side of (16), we notice that

$$A_1(R,\rho,\epsilon) \leq KR^q e^{\frac{m}{\rho}}\rho^{-mn}2^{-c(k)R}\int_{\mathcal{A}_1} 2^{f(\alpha)R}e^{-m2^{\epsilon R}}\mathrm{d}\alpha,$$

where

$$f(\alpha) \triangleq \sum_{i=1}^{q}(|m-n|+2i-1)\alpha_i. \tag{17}$$

Realizing that since $\mathcal{A}_1 \subseteq \mathcal{A} \subseteq \{\alpha|1\geq\alpha_i\geq 0,\forall i\}$, $\mathrm{Vol}\{\mathcal{A}_1\}\leq 1$, we conclude

$$A_1(R,\rho,\epsilon) \leq KR^q(2^{\epsilon R})^{\frac{f_1-c(k)}{\epsilon}}e^{-m2^{\epsilon R}}e^{\frac{m}{\rho}}\rho^{-mn}, \tag{18}$$

where

$$f_1 \triangleq \sup_{\mathcal{A}_1} f(\alpha).$$

On the other hand

$$\begin{aligned}A_2(R,\rho,\epsilon) \geq & KR^q e^{\frac{mq}{\rho}}\rho^{-mn}2^{-c(k)R}\int_{\mathcal{A}_{\epsilon_1}} e^{-mq2^{-\epsilon_1 R}}2^{R\sum_i\alpha_i}\times \\ & \prod_{i=1}^{q}(1-2^{-\epsilon_1 R})^{|m-n|}2^{|m-n|\alpha_i R}\prod_{i<j}(1-2^{-\epsilon_1 R})^2 2^{2\alpha_j R}\mathrm{d}\alpha,\end{aligned} \tag{19}$$

where

$$\mathcal{A}_{\epsilon_1} \triangleq \{\alpha \in \mathcal{A}_2 | \frac{\log \rho}{R} - \alpha_q > \epsilon_1, \alpha_1 > \epsilon_1, |\alpha_j - \alpha_i| > \epsilon_1 \ \forall i \neq j\}.$$

Realizing that $e^{-2^{-\epsilon_1 R}} \geq (1 - 2^{-\epsilon_1 R})$, (19) yields

$$A_2(R, \rho, \epsilon) \geq K R^q e^{\frac{mq}{\rho}} (1 - 2^{-\epsilon_1 R})^{m(n+q)} \rho^{-mn} 2^{-c(k)R} \int_{\mathcal{A}_{\epsilon_1}} 2^{f(\alpha)R} \mathrm{d}\alpha,$$

where, as before, $f(.)$ is given by (17). Let us define α^* as

$$\alpha^* \triangleq \arg \sup_{\alpha \in \mathcal{A}_{\epsilon_1}} f(\alpha).$$

Then it follows from the continuity of $f(.)$ that, for any $\epsilon_2 > 0$, there exists a neighborhood I_{ϵ_2} of α^*, within which

$$f(\alpha) \geq f(\alpha^*) - \epsilon_2.$$

This means that

$$A_2(R, \rho, \epsilon) \geq K R^q e^{\frac{mq}{\rho}} (1 - 2^{-\epsilon_1 R})^{m(n+q)} \rho^{-mn} 2^{(f(\alpha^*) - c(k) - \epsilon_2)R} \mathrm{Vol}\{\mathcal{A}_{\epsilon_1} \cap I_{\epsilon_2}\}. \tag{20}$$

Now, from (18), (20) and the fact that $e^{-\frac{m(q-1)}{\rho}} \leq 1$, we conclude that

$$\frac{A_1(R, \rho, \epsilon)}{A_2(R, \rho, \epsilon)} \leq \big(1 - (2^{\epsilon R})^{-\frac{\epsilon_1}{\epsilon}}\big)^{-m(n+q)} (2^{\epsilon R})^{\frac{f_1 - f(\alpha^*) + \epsilon_2}{\epsilon}} e^{-m(2^{\epsilon R})} \mathrm{Vol}^{-1}\{\mathcal{A}_{\epsilon_1} \cap I_{\epsilon_2}\}. \tag{21}$$

This means that

$$\limsup_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log\big(1 + A_1(R, \rho, \epsilon)/A_2(R, \rho, \epsilon)\big)}{\log \rho} = 0 \tag{22}$$

Note that (22) holds, whether ρ growing to infinity and $(R, \rho) \in \mathcal{R}_\delta(k)$ result in R growing to infinity or not. This is because the right hand side of (21) decays *exponentially* with $2^{\epsilon R}$, while it only grows *polynomially* with the same variable. To characterize the second term on the right-hand side of (16), we note that

$$A_2(R, \rho, \epsilon) \leq K R^q 2^{(f_2 - c(k))R} \rho^{-mn}.$$

This means that

$$\frac{\log A_2(R, \rho, \epsilon)}{\log \rho} \leq \frac{\log(K R^q)}{\log \rho} + (f_2 - c(k)) \frac{R}{\log \rho} - mn, \tag{23}$$

where

$$f_2 \triangleq \sup_{\mathcal{A}_2} f(\alpha),$$

with $f(.)$ given by (17). To derive f_2, one needs to consider two different cases. The first case is when $(R, \rho) \in \mathcal{R}_\delta(0)$, in which case

$$f_2 = m + n - 1, \quad (R, \rho) \in \mathcal{R}_\delta(0). \tag{24}$$

In this case, the supremum is achieved at $\alpha^* = (0, \cdots, 0, 1)$. The second case is when $(R, \rho) \in \mathcal{R}_\delta(k)$, for $k \in \mathbb{Z}$ and $q > k > 0$, where

$$f_2 = [m + n - (2k + 1)] + k(k + 1)(\frac{\log \rho}{R} + \epsilon), \qquad (R, \rho) \in \mathcal{R}_\delta(k), q > k > 0. \tag{25}$$

The supremum happens at

$$\alpha^* = (0, \cdots, 1 - k(\frac{\log \rho}{R} + \epsilon), \underbrace{\frac{\log \rho}{R} + \epsilon, \cdots, \frac{\log \rho}{R} + \epsilon}_{k \text{ times}}).$$

Notice that, assuming $\epsilon \leq \delta$, (14) guarantees that $1 - k(\frac{\log \rho}{R} + \epsilon) > 0$. Plugging in for f_2 in (23), we conclude

$$\begin{aligned} \frac{\log A_2(R, \rho, \epsilon)}{\log \rho} \leq & \frac{\log(KR^q)}{\log \rho} - g(k) + (mn - g(k))\frac{R}{\log \rho}\epsilon + \\ & (\tilde{c}(k) - c(k))\frac{R}{\log \rho}, \quad \text{for} \quad (R, \rho) \in \mathcal{R}_\delta(k). \end{aligned} \tag{26}$$

In this expression, $g(k)$ is given by (8) and $\tilde{c}(k)$ is defined as

$$\tilde{c}(k) \triangleq m + n - (2k + 1). \tag{27}$$

Now, from (16), (22) and (26), together with the fact that ϵ can be made arbitrarily small, one concludes

$$\begin{aligned} \limsup_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R, \rho) - c(k)R}{\log \rho} \leq & -g(k) + \\ (\tilde{c}(k) - c(k)) \times & \limsup_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{R}{\log \rho}. \end{aligned} \tag{28}$$

Next we turn our attention to (10) and introduce the following change of variables, i.e.,

$$\beta_i \triangleq \frac{\log(1 + \rho\mu_i)}{R}.$$

In terms of the new variables, (10) can be written as

$$P_o(R, \rho) \geq \Pr\{\mathcal{B}\}, \tag{29}$$

where

$$\mathcal{B} \triangleq \{\beta | \beta_q \geq \cdots \geq \beta_1 \geq 0, 1 - \sum_i \beta_i > 0\}.$$

Now, taking steps very similar to those outlined by (19) through (20) yields (for further details, please refer to [10])

$$\begin{aligned} \liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R, \rho) - c(k)R}{\log \rho} \geq & -g(k) + \\ (\tilde{c}(k) - c(k)) \times & \liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{R}{\log \rho}, \end{aligned} \tag{30}$$

Examining (28) and (30) reveals that the choice

$$c(k) = \tilde{c}(k), \quad \forall k \tag{31}$$

guarantees the existence of

$$\lim_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho} = -g(k), \tag{32}$$

regardless of whether $\lim_{\rho\to\infty} \frac{R}{\log \rho}$ exits or not. Now, since (32) holds for arbitrarily small values of δ, we get (5). Note that (31), together with (27), result in (7) and thus complete the proof.

Before using Theorem 2 to answer our central question on the utility of a 3 dB SNR gain in block-fading MIMO channels, let us investigate some of the connections between the two formulations (i.e., the TRT and DMT). It is immediate to check that if the multiplexing gain is well defined, i.e., if

$$r = \lim_{\rho\to\infty} \frac{R(\rho)}{\log \rho}$$

exists, then Theorem 2 reduces to Theorem 1. In the more general case, however, Theorem 2 replaces the restrictive notion of multiplexing gain with the new concept of operating region $\mathcal{R}(k)$. It is worth noting that every operating region in the TRT corresponds to a line segment in the DMT. In fact, this correspondence inspires the following observation

$$g(k) = d(k) - kd'(k^+), \tag{33}$$

$$c(k) = -d'(k^+), \tag{34}$$

where $d'(k^+)$ is the slope of the line segment connecting $d(k)$ and $d(k+1)$. The key observation here is that in order to fix the transmission rate (or outage probability), while staying in the same operating region, one *has to* deviate from the linear trajectory imposed by the multiplexing gain notion. The following *heuristic* derivation of (33) and (34) further illustrates this point. To derive (33), we start with two approximate relationships obtained from the DMT (please refer to Fig. 3), i.e.,

$$\log P_o\left(R, 2^{\log\rho}\right) \approx -d\left(\frac{R}{\log\rho}\right)\log\rho, \quad \text{and} \tag{35}$$

$$\log P_o\left(R, 2^{\log\rho+\Delta\log\rho}\right) \approx -d\left(\frac{R}{\log\rho+\Delta\log\rho}\right)(\log\rho+\Delta\log\rho). \tag{36}$$

We further approximate $d\left(\frac{R}{\log\rho+\Delta\log\rho}\right)$ with the first two terms of its Taylor series expansion, i.e.,

$$d\left(\frac{R}{\log\rho+\Delta\log\rho}\right) \approx d\left(\frac{R}{\log\rho}\right) - \frac{R\times\Delta\log\rho}{\log\rho(\log\rho+\Delta\log\rho)} d'\left(\frac{R}{\log\rho}\right). \tag{37}$$

Now (37), together with (35) and (36), gives

$$\frac{\log P_o(R, 2^{\log\rho+\Delta\log\rho}) - \log P_o(R, 2^{\log\rho})}{\Delta\log\rho} \approx -d\left(\frac{R}{\log\rho}\right) + \frac{R}{\log\rho} d'\left(\frac{R}{\log\rho}\right). \tag{38}$$

Realizing that the left-hand side of (38) gives the slope of $\log P_o(R,\rho)$ with respect to $\log\rho$, i.e., $-g(k)$, and that $d(r)-rd'(r)$ remains constant over the line segments of $d(r)$, we get (33). Deriving (34) follows the same lines (please refer to Fig. 4). In particular, we first compute the horizontal spacing, $\Delta\log\rho$, between the outage curves corresponding to rates R and $R+\Delta R$. For this purpose, we use (35) to write

$$\log P_o\left(R+\Delta R, 2^{\log\rho+\Delta\log\rho}\right) \approx -d\left(\frac{R+\Delta R}{\log\rho+\Delta\log\rho}\right)(\log\rho+\Delta\log\rho) \tag{39}$$

Then we expand $d(\frac{R+\Delta R}{\log\rho+\Delta\log\rho})$ in a way similar to (37) and equate (35) with (39) to get

$$\begin{aligned}\Delta\log\rho &\approx \frac{-d'(\frac{R}{\log\rho})}{d(\frac{R}{\log\rho})-\frac{R}{\log\rho}d'(\frac{R}{\log\rho})}\Delta R,\\ \Delta\log\rho &\approx \frac{-d'(\frac{R}{\log\rho})}{g(k)}\Delta R.\end{aligned} \tag{40}$$

Realizing that $\Delta\log\rho \approx \frac{c(k)}{g(k)}\Delta R$ (see (41) below), and that $d'(r)$ remains constant over the line segments of $d(r)$, we get (34). It is now clear from (36) and (39) that to maintain the transmission rate or outage probability constant, one has to *deviate* from multiplexing gain $R/\log\rho$. Finally, we emphasize that our derivations for (33) and (34) are based on *heuristics*, i.e., in general, a channel or protocol's TRT may not be related to its DMT through these two equations.

Next, we use Theorem 2 to investigate the asymptotic trends exhibited by the outage probability curves of block-fading MIMO channels. Our investigation hinges on the following *piecewise linear* approximation of outage probability suggested by (5) (please refer to Fig. 5 and Fig. 6), i.e.,

$$\log P_o(R,\rho) \approx c(k)R - g(k)\log\rho \quad \text{for} \quad (R,\rho)\in\mathcal{R}(k), \tag{41}$$

where the approximation becomes more accurate as ρ increases. In fact, one can use the following *rule of thumb* as a measure of accuracy for (41), i.e., for operating point $(R,\rho)\in\mathcal{R}(k)$, find the smallest value of ϵ such that

$$\frac{\rho^k}{2^R}\le\epsilon, \quad \text{and} \quad \frac{2^R}{\rho^{k+1}}\le\epsilon.$$

Now the smaller ϵ is, the better the approximation becomes. Back to (41), we notice that the slope of the outage curve, for a fixed rate, is given by $-g(k)$, while the horizontal spacing in dB between two outage curves with a ΔR rate difference is given by $3\Delta R/t(k)$. Thus, the larger $g(k)$ is, the faster the outage probability decays with SNR, hence $g(k)$ is called the reliability gain coefficient. On the other hand, the larger $t(k)$ is, the smaller the SNR gain needs to be, to support the increased rate. Hence $t(k)$ is called the throughput gain coefficient. Revisiting the two special cases given earlier in Section 1, we now see that the TRT predicts a reliability gain coefficient of $g(0)=mn=d_{max}$ for operating points for which $R/\log\rho<1$ (i.e., $(R,\rho)\in\mathcal{R}(0)$). This not only comes in full agreement with special case one, but also explains the observation made in Fig. 1 regarding the validity of this rule for operating points for which $R/\log\rho$ is as large as 0.8. The same thing holds regarding special case two. The TRT predicts a throughput gain coefficient of $t(\min\{m,n\}-1)=\min\{m,n\}=r_{max}$ for operating points for which $\min\{m,n\}-$

$1 < R/\log\rho < \min\{m,n\}$ (i.e., $(R,\rho) \in \mathcal{R}(\min\{m,n\}-1)$). This again comes in agreement with special case two and furthermore accounts for the observation made in Fig. 2 regarding the validity of this rule for operating points for which $R/\log\rho$ is as small as 1.2 (note that in this case $m = n = 2$). As we will show in the sequel, the TRT, in contrast to these two special cases, characterizes both, the slope and the spacing between outage curves and as so, gives the complete answer to our question. Furthermore, it is not restricted to certain operating points.

Finally, we present numerical results that validate the predictions of the TRT. In particular, it is now straightforward to see that the high SNR segment of Fig. 1 falls in the region $\mathcal{R}(0)$ and, indeed, the 4 levels of diversity and 4.5 dB spacing (for every 2 bpcu throughput increase) in this figure correspond precisely to $g(0) = 4$ and $t(0) = 4/3$. Similarly, the high SNR segment of Fig. 2 falls in $\mathcal{R}(1)$ and, again, the 2 levels of diversity and 3 dB spacing, for every 2 bpcu throughput increase, agree with $g(1) = 2$ and $t(1) = 2$. Fig. 8 and Fig. 9 are different from the previous two figures in that the high SNR segments of the outage curves fall within *both* of the two regions. As a result, the slope of the curves and the spacing between them change as the operating point leaves one operating region and enters another. Again, the values of the slopes, spacings, and operating points at which the change occurs (which can be read from Fig. 7) match nicely with the predictions of the TRT as formulated by Theorem 2. Fig. 10 through Fig. 13 correspond to a 3×3 MIMO system. As can be seen from Fig. 10, the solid segment of the curve corresponding to $R = 10$ bpcu, falls in $\mathcal{R}(1)$ and, as predicted, we observe a slope of $g(1) = 7$. It should be noted, however, that the tail of the curve corresponding to $R = 4$ bpcu is leaving $\mathcal{R}(1)$ and entering $\mathcal{R}(0)$ and thus the slope of this curve is larger than 7 (about 7.7). For the same reason, the horizontal spacing between the two curves (almost 10 dB) is larger than the value predicted for $k = 1$ (i.e., 7.7 dB). The solid segments of the outage curves corresponding to $R = 58$ and 64 bpcu in Fig. 11 fall in $\mathcal{R}(2)$, and therefore, we observe 3 levels of diversity and 3 dB of spacing, for every 3 bpcu throughput increase, which correspond precisely to $g(2) = 3$ and $t(2) = 3$. Fig. 12 and Fig. 13 depict the case where the high SNR segments fall within two different regions ($k = 2$ and $k = 1$). Again the slopes and spacings are in agreement with the predictions of the TRT.

3 Applications

The following result establishes the operational significance of Theorem 2 by showing that the error probability for the optimal space-time code exhibits the same asymptotic behavior as the outage probability.

Theorem 3 *The probability of error for the* **optimal** *coding/decoding scheme used in conjunction with channel (1) satisfies*

$$\lim_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log\rho} = -g(k), \tag{42}$$

where $\mathcal{R}(k)$, $c(k)$, and $g(k)$ are given by (6), (7), and (8), respectively. Moreover, there exists a coding scheme that achieves (42) for $l \geq m + n - 1$.

Proof: (Sketch) The proof follows the same lines as that of Theorem 2 in [1]. In particular, the converse is obtained via a careful application of Fano's inequality (compare

to Lemma 5 in [1]). The achievability is established using an ensemble of Gaussian codebooks along with the appropriate use of the union bound (compare to Theorem 2 in [1]). The detailed proof is reported in the Appendix.

One can also characterize the TRT achieved by certain suboptimal space-time architectures. In this paper, we restrict our study to square V-BLAST protocols, orthogonal space-time constellations and MIMO-ARQ channels. In the V-BLAST architecture, the input stream is split into m sub-streams. These sub-streams are then encoded independently and transmitted over the m transmit antennas [3]. The following theorem characterizes the achieved TRT for this protocol, when a *maximum likelihood* (ML) decoder is employed.

Theorem 4 *The ML error probability for a V-BLAST communication system with m transmit and m receive antennas satisfies*[3]

$$\lim_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_{vb}}} \frac{\log P_e(R,\rho) - R}{\log \rho} = -m, \tag{43}$$

where $\mathcal{R}_{vb}$ is given by

$$\mathcal{R}_{vb} \triangleq \{(R,\rho) | m > \frac{R}{\log \rho} > 0\}.$$

Moreover, there exists a coding scheme that achieves (43) *for $l \geq 2m-1$.*

Proof: Please refer to the appendix for a detailed proof.

Fig. 14 depicts the outage curves corresponding to $R = 8$ and 12 bpcu for a 2×2 V-BLAST scheme with ML decoding. As can be seen from this figure, the high SNR segments of the outage curves achieve 2 levels of diversity with a horizontal spacing of 6 dB. These values agree with $g_{vb}(0) = 2$ and $t_{vb}(0) = 2$. Fig. 15, on the other hand, compares the outage behavior of 2×2 MIMO and ML V-BLAST schemes for $R = 4, 16$ and 32 bpcu. As can be seen from this figure, the outage curves for the two schemes coincide while $R/\log \rho > 1$. In particular, the curves corresponding to $R = 32$ bpcu are almost identical. However, for $R/\log \rho < 1$, the sub-optimality of the V-BLAST becomes evident. In fact, the curve corresponding to the 2×2 MIMO with $R = 4$ bpcu approaches 4 levels of diversity very rapidly, while the curve corresponding to the V-BLAST only attains 2 levels.

Similarly, orthogonal space-time constellations allow for a simple TRT characterization. An orthogonal constellation of size m, length l, and rate k/l (in symbols per channel use (spcu)) is a space-time code $\mathbf{X} \in \mathbb{C}^{m \times l}$ such that

$$\mathbf{X}\mathbf{X}^H = \left(\sum_{i=1}^{k} |x_i|^2 \right) \times \mathbf{I}_m, \tag{44}$$

where $\{x_i\}_{i=1}^k$ denote the symbols to be sent, $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{X}^H$ denotes the hermitian of matrix $\mathbf{X}$ [5]. As an example, consider the orthogonal

[3]The subscript "vb" stands for V-BLAST.

constellation with $m = 2$, $l = 2$, and rate one, which is known as the Alamouti code [6]. In this case

$$\mathbf{X} = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix},$$

where x^* denotes the complex conjugate of x. Notice that

$$\mathbf{X}\mathbf{X}^H = (|x_1|^2 + |x_2|^2) \times \mathbf{I}_2,$$

as required by (44). The received signal matrix $\mathbf{Y} \in \mathbb{C}^{n \times l}$ at the destination can be written as

$$\mathbf{Y} = \sqrt{\frac{\rho}{m}}\mathbf{H}\mathbf{X} + \mathbf{W}. \tag{45}$$

The ML receiver performs linear processing on $\mathbf{Y}$ to yield the following equivalent parallel channel model

$$\tilde{y}_i = \sqrt{\frac{\rho}{m}\|\mathbf{H}\|^2}x_i + \tilde{w}_i, \quad \text{for} \quad i = 1, \cdots, k. \tag{46}$$

In (46), $\|\mathbf{H}\|^2$ denotes the Frobenius norm of $\mathbf{H}$ (i.e., $\|\mathbf{H}\|^2 \triangleq \sum |h_{ij}|^2$, where $\{h_{ij}\}$ are the entries of $\mathbf{H}$) and $\{\tilde{w}_i\}_{i=1}^k$ are i.i.d complex Gaussian random variables of zero mean and unit variance. The following theorem gives the throughput-reliability tradeoff for orthogonal constellations.

Theorem 5 *The optimal error probability for an orthogonal constellation of size m, length l, rate k/l spcu (R bpcu) and n receive antennas satisfies*[4]

$$\lim_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_{oc}}} \frac{\log P_e(R,\rho) - \frac{l}{k}mnR}{\log \rho} = -mn, \tag{47}$$

where $\mathcal{R}_{oc}$ is given by

$$\mathcal{R}_{oc} \triangleq \{(R,\rho) | \frac{k}{l} > \frac{R}{\log \rho} > 0\},$$

Moreover, there exists an outer coding scheme (one that maps the information bits into symbols $\{x_i\}_{i=1}^k$) that achieves (47) *for $k \geq mn$.*

Proof: The proof follows immediately from Theorem 3 and the fact that the orthogonal constellation of interest effectively converts the underlying $m \times n$ MIMO channel of rate R (as given by (45)) into a $1 \times mn$ channel of rate $\frac{l}{k}R$ (as given by (46)).

Fig. 16 depicts the outage curves corresponding to $R = 4$ and 8 bpcu for a 2×2 Alamouti scheme. As can be seen from this figure, the high SNR segments of the outage curves achieve 4 levels of diversity with a horizontal spacing of 12 dB. These values agree with $g_{oc}(0) = 4$ and $t_{oc}(0) = 1$. Fig. 17, on the other hand, compares the outage behavior of the 2×2 MIMO channel and the Alamouti constellation for $R = 4, 16$ and 32 bpcu. As can be seen from this figure, while the outage curves for the two schemes coincide for

[4]The subscript "oc" stands for an orthogonal constellation.

small rates, the sub-optimality of the Alamouti scheme becomes evident at higher values of R. In particular, the curves corresponding to $R = 4$ bpcu are almost identical and for $R = 32$ bpcu the curve corresponding to the Alamouti scheme lags that of MIMO by more than 40 dB.

Finally, we extend our analysis to MIMO-ARQ channels. In this setup, the transmitter starts by picking up a message from the transmission buffer. It then uses a space-time encoder to map the message to a sequence of blocks $\mathbf{X}_p \in \mathbb{C}^{m\times l}, L \geq p \geq 1$. During transmission-round p, the transmitter sends $\mathbf{X}_p$ one column at a time over its m antennas. The receiver then tries to decode the message. If successful, it sends back a positive acknowledgement signal (ACK), which causes the transmitter to start sending the next message. However, if the receiver detects an error, it requests another round of transmission by feeding back a negative acknowledgement signal (NACK). The only exception to this rule is when L rounds of transmission have already been sent, in which case the transmitter abandons sending the current message and goes to the next one. In this paper, we address the long-term static channel scenario, where all of the transmission-rounds corresponding to a message take place over the same channel realization. We also impose a short term power constraint on the transmitter, such that power-control is not possible [7]. At this point, we need to distinguish between two closely related parameters, namely, the first-round transmission rate and the long-term average throughput. Assume that each message consists of b information bits which means that the first-round transmission rate is $R_1 = b/l$ bpcu. Since some of the messages take more than one transmission-round to be sent, the long-term average throughput η is strictly less than R_1. The gap between the two quantities, however, diminishes as the SNR grows. This is due to the fact that at high SNRs, most of the messages are decoded error-free after the first round of transmission and the ARQ retransmission-rounds are used only for those *rare* events in which the message does not get through with only one round of transmission. Recognizing the operational significance of η, in the following we state the TRT for ARQ channels in terms of η, rather than R_1.

Theorem 6 *The optimal throughput-reliability tradeoff for the coherent block-fading MIMO ARQ channel with m transmit antennas, n receive antennas, L maximum number of transmission-rounds, under the long-term static channel and short-term power constraint assumptions is given by*[5]

$$\lim_{\substack{\rho\to\infty \\ (\eta,\rho)\in\mathcal{R}_{ls}(k)}} \frac{\log P_e(\eta,\rho) - c_{ls}(k)\eta}{\log\rho} = -g_{ls}(k), \tag{48}$$

where η denotes the long-term average data rate. In (48), $\mathcal{R}_{ls}(k)$, $c_{ls}(k)$ and $g_{ls}(k)$ are defined by

$$\mathcal{R}_{ls}(k) \triangleq \begin{cases} \{(\eta,\rho)|(k+1)L > \frac{\eta}{\log\rho} > kL\} & \text{for} \quad k\in\mathbb{Z}, \lfloor\frac{\min\{m,n\}}{L}\rfloor > k \geq 0 \\ \{(\eta,\rho)|\min\{m,n\} > \frac{\eta}{\log\rho} > \lfloor\frac{\min\{m,n\}}{L}\rfloor L\} & \text{for} \quad k = \lfloor\frac{\min\{m,n\}}{L}\rfloor \end{cases},$$

$$c_{ls}(k) \triangleq \frac{c(k)}{L} \quad \text{and} \quad g_{ls}(k) \triangleq g(k),$$

respectively. $c(k)$ and $g(k)$ are given by (7) and (8).

[5]The subscript "ls" stands for long-term static.

Proof: (Sketch) The proof follows the same lines as that of Theorem 5 in [7]. In particular, the converse is obtained by lower-bounding the error probability of the ARQ protocol with that of a ML decoder that operates on the whole codeword $\{\mathbf{X_p}\}_{p=1}^{L}$. The achievability, on the other hand, is established through the use of an ensemble of Gaussian code-books, along with a bounded-distance decoder. The main idea here is to differentiate between the undetected-errors (i.e., the ones for which the receiver sends back an ACK signal) and those errors that the decoder makes after requesting L rounds of transmission. It can then be shown that, through judicious choice of decoder threshold-distance, the latter error type becomes dominant, and hence, the lower and upper bounds become tight. It is then straightforward to argue the existence of codes in the ensemble that perform at least as well as the ensemble average.

4 Conclusions

We developed a new formulation for the asymptotic interplay between R, $\log\rho$ and $P_e(R,\rho)$ in block-fading MIMO channels. Our formulation, through relaxing the restriction imposed by the notion of multiplexing gain, sheds more light on the *worth* of a 3 dB SNR gain in block-fading MIMO channels at high SNRs. We also presented numerical results that validated the predictions of our formulation. For our predictions to be accurate at finite SNR's, we only require the operating point to be well within certain operating regions in the $R-\log\rho$ plane. Characterization of the performance in the transitional regions remains an open problem. Another problem of definite interest is establishing the relation between the TRT and DMT for a general channel and/or protocol.

5 Acknowledgment

The authors would like to thank Profs. G. Caire and M. O. Damen for inspiring discussions.

6 Appendix

6.1 Proof of Theorem 3

The proof follows that of Theorem 2 in [1]. In particular, we prove (42) in two steps. The first step is to show that

$$\liminf_{\substack{\rho\to\infty \\ (R,\rho)\in\mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log\rho} \geq -g(k). \tag{49}$$

This is done by taking steps similar to those outlined by Lemma 5 in [1]. More specifically, let us denote the ML error probability, conditioned on a certain channel realization H, by $P_{E|H}$. It then follows that

$$\begin{aligned} P_e(R,\rho) &= \mathrm{E}_H\{P_{E|H}(R,\rho)\}, \\ &= \int P_{E|H}(R,\rho)p(H)\mathrm{d}H. \end{aligned}$$

Thus

$$P_e(R,\rho) \geq \int_{\mathcal{C}} P_{E|H}(R,\rho)p(H)\mathrm{d}H, \tag{50}$$

where $\mathcal{C}$ denotes any subset of the set of all channel realizations. Let us define $\mathcal{C}_\epsilon$ as the set of channel realizations for which the conditional ML error probability cannot be made smaller than ϵ, i.e.

$$\mathcal{C}_\epsilon \triangleq \{H | P_{E|H}(R,\rho) \geq \epsilon\}. \tag{51}$$

From (50) and (51) one concludes that

$$P_e(R,\rho) \geq \epsilon P_{\mathcal{C}_\epsilon}(R,\rho).$$

This means that

$$\frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq \frac{\log P_{\mathcal{C}_\epsilon}(R,\rho) - c(k)R}{\log \rho} + \frac{\log \epsilon}{\log \rho},$$

where $c(k)$ is given by (7), or

$$\liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq \liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_{\mathcal{C}_\epsilon}(R,\rho) - c(k)R}{\log \rho}, \tag{52}$$

where $\mathcal{R}_\delta(k)$ is given by (14). Now, application of Fano's inequality reveals that

$$P_{E|H}(R,\rho) \geq 1 - \frac{\mathrm{I}(\mathbf{x};\mathbf{y}|\mathbf{H}=H)}{R} - \frac{1}{Rl},$$

where l denotes the codeword length [1]. This, together with (51), means that

$$\{H | 1 - \frac{\mathrm{I}(\mathbf{x};\mathbf{y}|\mathbf{H}=H)}{R} - \frac{1}{Rl} \geq \epsilon\} \subseteq \mathcal{C}_\epsilon,$$

which using (4) results in

$$P_{\mathcal{C}_\epsilon}(R,\rho) \geq P_o((1-\epsilon-\frac{1}{Rl})R,\rho) \quad \text{or}$$

$$\liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_{\mathcal{C}_\epsilon}(R,\rho) - c(k)R}{\log \rho} \geq \liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}_\delta(k)}} \frac{\log P_o((1-\epsilon)R,\rho) - c(k)R}{\log \rho}. \tag{53}$$

Now, from (52) and (53), together with the fact that both, δ and ϵ can be made arbitrarily small, we conclude that

$$\liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \geq \liminf_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} \frac{\log P_o(R,\rho) - c(k)R}{\log \rho}.$$

But, from Theorem 2 (refer to (5)), we know that the right-hand side equals $-g(k)$ (given by (8)). This proves (49) and thus completes the first step.

The second step in proving (42) is to show that

$$\limsup_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} \frac{\log P_e(R,\rho) - c(k)R}{\log \rho} \leq -g(k), \tag{54}$$

provided that the codeword length, l, satisfies $l \geq m + n - 1$. To prove this, we follow the steps of Theorem 2 in [1] and consider a Gaussian code-book with 2^{Rl} codewords of length l. It is straightforward to verify that the ML error probability, conditioned on a certain channel realization, is upper-bounded by

$$\begin{aligned} P_{E|H}(R,\rho) &\leq 2^{Rl} \det(I_n + \frac{\rho}{2m} HH^H)^{-l}, \\ &= 2^{Rl} \prod_{i=1}^{q} (1 + \frac{\rho}{2m}\mu_i)^{-l}, \end{aligned} \tag{55}$$

where $q \triangleq \min\{m,n\}$ and $\mu_q \geq \cdots \geq \mu_1 \geq 0$ represent the ordered eigenvalues of HH^H. The change of variables

$$\gamma_i \triangleq \frac{\log(1 + \frac{\rho}{2m}\mu_i)}{R},$$

changes (55) into

$$P_{E|\gamma}(R,\rho) \leq 2^{(1-\sum_{i=1}^{q}\gamma_i)Rl}. \tag{56}$$

Next we define $\mathcal{D}$ as

$$\mathcal{D} \triangleq \{\gamma | \gamma_q \geq \cdots \geq \gamma_i \geq 0, 1 - \sum_{i=1}^{q} \gamma_i \geq 0\}.$$

Referring to (56) reveals that $\mathcal{D}$ consists of those channel realizations for which the upper-bound on the ML error probability cannot be made arbitrarily small, even through the use of infinitely long codewords. For these channel realizations, we upper-bound $P_{E|\gamma}(R,\rho)$ by 1, i.e.

$$\begin{aligned} P_e(R,\rho) &= P_{E,\mathcal{D}^c}(R,\rho) + P_{E,\mathcal{D}}(R,\rho), \\ P_e(R,\rho) &\leq P_{E,\mathcal{D}^c}(R,\rho) + P_{\mathcal{D}}(R,\rho), \end{aligned} \tag{57}$$

where $\mathcal{D}^c$ denotes the complement of $\mathcal{D}$. Let us first focus on $P_{\mathcal{D}}(R,\rho)$. Realizing that $\mathcal{D}$ is precisely the same set as $\mathcal{A}$ (refer to (13)) and that, up to a scaling factor, $p(\gamma)$ is identical to $p(\alpha)$, it follows immediately that

$$\limsup_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} \frac{\log P_{\mathcal{D}}(R,\rho) - c(k)R}{\log \rho} \leq -g(k). \tag{58}$$

Now, turning our attention back to $P_{E,\mathcal{D}^c}(R,\rho)$, we realize that

$$\lim_{\substack{\rho \to \infty \\ (R,\rho) \in \mathcal{R}(k)}} P_{E,\mathcal{D}_1^c}(R,\rho) 2^{-c(k)R} = 0, \tag{59}$$

which means that

$$\limsup_{\substack{\rho\to\infty \\ (R,\rho)\in\mathcal{R}(k)}} \frac{\log P_{E,\mathcal{D}^c}(R,\rho)-c(k)R}{\log\rho} = \limsup_{\substack{\rho\to\infty \\ (R,\rho)\in\mathcal{R}(k)}} \frac{\log P_{E,\mathcal{D}_2^c}(R,\rho)-c(k)R}{\log\rho}, \tag{60}$$

where

$$\mathcal{D}_1^c \triangleq \{\gamma\notin\mathcal{D}|\gamma_q > \frac{\log\rho}{R}\} \text{ and } \mathcal{D}_2^c \triangleq \{\gamma\notin\mathcal{D}|\gamma_q \le \frac{\log\rho}{R}\}.$$

Notice that (59) holds for exactly the same reason as (22) does (refer to the comment after (22)). Now, using (56), we have

$$\begin{aligned} P_{E,\mathcal{D}_2^c}(R,\rho)2^{-c(k)R} &= 2^{-c(k)R}\int_{\mathcal{D}_2^c} P_{E|\gamma}(R,\rho)p(\gamma)\mathrm{d}\gamma, \\ &\le KR^q\rho^{-mn}2^{-c(k)R}\int_{\mathcal{D}_2^c} 2^{[f(\gamma)+l(1-\sum_{i=1}^q\gamma_i)]R}\mathrm{d}\gamma, \end{aligned}$$

where

$$f(\gamma) \triangleq 2^{\sum_i(|m-n|+2i-1)\gamma_i R}.$$

Thus

$$P_{E,\mathcal{D}_2^c}(R,\rho)2^{-c(k)R} \le KR^q\rho^{-mn}2^{(f_2-c(k))R}\mathrm{Vol}\{\mathcal{D}_2^c\}, \tag{61}$$

where

$$f_2 \triangleq \sup_{\mathcal{D}_2^c} f(\gamma) + l(1-\sum_{i=1}^q\gamma_i).$$

Realizing that for $l \ge m+n-1$, the supremum occurs at $\gamma=\gamma^*$, such that $1-\sum\gamma_i^*=0$, f_2 can be easily derived from (24) and (25) by simply plugging in $\epsilon=0$. Therefore, (61) gives

$$\limsup_{\substack{\rho\to\infty \\ (R,\rho)\in\mathcal{R}(k)}} \frac{\log P_{E,\mathcal{D}_2^c}(R,\rho)-c(k)R}{\log\rho} \le -g(k). \tag{62}$$

Now, from (62), (60) and (58), we conclude (54), which together with (49), proves (42). Since we proved (42) using an *ensemble* of Gaussian codes, it follows that, for any code-length $l \ge m+n-1$, there exists at least a code, for which (42) holds. This completes the proof.

6.2 Proof of Theorem 4

Realizing that the V-BLAST protocol essentially transforms the $m\times m$ MIMO channel into a multiple-access channel with m single-antenna users and a destination with m receive antennas, we prove (43) by following the same lines as that of Theorem 2 in [4].

In particular, let E_S denote the event that a certain decoder makes errors in decoding the codewords transmitted by a subset S of the antennas. It then follows that

$$\sum_{S \neq \varnothing} \Pr\{E_S\} \geq P_e(R, \rho). \tag{63}$$

It is also clear that

$$P_e(R, \rho) \geq \Pr\{E_{S^*}\}, \tag{64}$$

where S^* denotes any non-empty subset of $\{1, \cdots, m\}$. Let us define S^* as the non-empty subset of $\{1, \cdots, m\}$, such that for all other non-empty subsets S, we have

$$\liminf_{\substack{\rho \to \infty \\ (\frac{|S|}{m}R, \rho) \in \mathcal{R}_S(k) \\ (\frac{|S^*|}{m}R, \rho) \in \mathcal{R}_{S^*}(k^*)}} \Big(g_S(k) - g_{S^*}(k^*)\Big) - \Big(\frac{|S|}{m}c_S(k) - \frac{|S^*|}{m}c_{S^*}(k^*)\Big)\frac{R}{\log \rho} \geq 0. \tag{65}$$

In (65), $|S|$ denotes the cardinality of set S. Also, $\mathcal{R}_S(k)$, $c_S(k)$ and $g_S(k)$ denote $\mathcal{R}(k)$, $c(k)$ and $g(k)$, as defined by (6), (7) and (8), for a MIMO channel with $|S|$ transmit and m receive antennas. The proof of (43) then follows in three steps. First, we prove that for any decoder

$$\liminf_{\substack{\rho \to \infty \\ (\frac{|S^*|}{m}R, \rho) \in \mathcal{R}_{S^*}(k^*)}} \frac{P_e(R, \rho) - \frac{|S^*|}{m}c_{S^*}(k)R}{\log \rho} \geq -g_{S^*}(k^*). \tag{66}$$

This follows immediately from (64) and the fact that $\Pr\{E_{S^*}\}$ upper-bounds the ML error probability of a MIMO channel with $|S^*|$ transmit antennas, m receive antennas and rate $\frac{|S^*|}{m}R$. The second step in proving (43) is to show that there exists a code, along with a decoder, for which

$$\limsup_{\substack{\rho \to \infty \\ (\frac{|S^*|}{m}R, \rho) \in \mathcal{R}_{S^*}(k^*)}} \frac{P_e(R, \rho) - \frac{|S^*|}{m}c_{S^*}(k^*)R}{\log \rho} \leq -g_{S^*}(k^*). \tag{67}$$

We prove this by showing that the error probability of the joint ML decoder, averaged over the ensemble of Gaussian codes, satisfies (67). The existence of the desired code then follows from the fact that there exist codes in the ensemble that perform at least as well as the average. For this purpose, assume that each of the antennas uses a Gaussian code-book of code-word length $l = 2m + 1$ and size $2^{\frac{R}{m}l}$ codewords. It then follows from Theorem 3, (refer to (42)), that

$$\lim_{\substack{\rho \to \infty \\ (\frac{|S|}{m}R, \rho) \in \mathcal{R}_S(k)}} \frac{\log \Pr\{E_S\} - \frac{|S|}{m}c_S(k)R}{\log \rho} = -g_S(k), \quad \forall S \neq \varnothing, \tag{68}$$

which means

$$\lim_{\substack{\rho\to\infty \\ (\frac{|S|}{m}R,\rho)\in\mathcal{R}_S(k) \\ (\frac{|S^*|}{m}R,\rho)\in\mathcal{R}_{S^*}(k^*)}} \left[\frac{\log \Pr\{E_S\}/\Pr\{E_{S^*}\}}{\log\rho} + \Big(g_S(k) - g_{S^*}(k^*)\Big) - \Big(\frac{|S|}{m}c_S(k) - \frac{|S^*|}{m}c_{S^*}(k^*)\Big)\frac{R}{\log\rho}\right] = 0. \quad (69)$$

Now, (69), together with (65), results in

$$\limsup_{\substack{\rho\to\infty \\ (\frac{|S|}{m}R,\rho)\in\mathcal{R}_S(k) \\ (\frac{|S^*|}{m}R,\rho)\in\mathcal{R}_{S^*}(k^*)}} \frac{\log \Pr\{E_S\}/\Pr\{E_{S^*}\}}{\log\rho} \le 0, \quad \forall S \neq \varnothing. \quad (70)$$

Returning to (63), we have

$$\frac{\log(1+\sum_{S\neq S^*,\varnothing}\frac{\Pr\{E_S\}}{\Pr\{E_{S^*}\}})}{\log\rho} + \frac{\log\Pr\{E_{S^*}\} - \frac{|S^*|}{m}c_{S^*}(k^*)R}{\log\rho} \ge \frac{\log P_e(R,\rho) - \frac{|S^*|}{m}c_{S^*}(k^*)R}{\log\rho}.$$

Taking the lim sup of both sides, together with (70) and (68) results in (67). Notice that (66) and (67) mean that

$$\lim_{\substack{\rho\to\infty \\ (\frac{|S^*|}{m}R,\rho)\in\mathcal{R}_{S^*}(k^*)}} \frac{P_e(R,\rho) - \frac{|S^*|}{m}c_{S^*}(k)R}{\log\rho} = -g_{S^*}(k^*). \quad (71)$$

The third and last step in proving (43) is to show that $|S^*| = 1$. One can prove this directly using the definitions of $\mathcal{R}(k)$, $c(k)$ and $g(k)$ (equations (6), (7) and (8)). However, we choose to do this using observations (33) and (34). In particular, notice that based on these observations, (65) reduces to finding the subset S^*, for which

$$d_S(\frac{|S|}{m}r) - d_{S^*}(\frac{|S^*|}{m}r) \ge 0, \quad \forall S \neq \varnothing,$$

where $d_S(r)$ represents the diversity-multiplexing tradeoff for a MIMO system with $|S|$ transmit and m receive antennas. From the proof of Theorem 3 in [4], we know that $|S^*| = 1$. Now, this together with (71) results in (43) and thus completes the proof.

References

[1] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna Channels," *IEEE Trans. Info. Theory*, 49:1073-1096, May 2003.

[2] H. Yao, G. W. Wornell, "Structured Space-time Block Codes with Optimal Diversity-Multiplexing Tradeoff and Minimum Delay," *IEEE Globecom 03*, 4:1941-1945, Dec. 2003.

[3] G. Foschini, G. Golden, R. Valenzuela and P. Wolniansky, "Simplified Processing for High Spectral Efficiency Wireless Communication Employing Multi-Element Arrayas," *IEEE Jour. Select. Areas on Comm.*, 17:1841-1852, Nov. 1999.

[4] D. N. C. Tse, P. Viswanath and L. Zheng, "Diversity-multiplexing Tradeoff in Multiple-Access Channels," *IEEE Trans. Info. Theory*, 50:1859-1874, Sept. 2004.

[5] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. Info. Theory*, 45:1456-1467, July 1999.

[6] S. Alamouti, "A Simple Transmitter Diversity Scheme for Wireless Communications," *IEEE Jour. Select. Areas on Comm.*, 16:1451-1458, Oct. 1998.

[7] H. El Gamal, G. Caire and M. O. Damen, "The MIMO ARQ Channel: Diversity-Multiplexing-Delay Tradeoff," *Submitted to the IEEE Trans. Info. Theorey.*

[8] L. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information-Theoretic Considerations in Cellular Mobile Radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359-378, May. 1994.

[9] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Europ. Trans. Telecomm.*, vol. 10, pp. 585-595, Nov./Dec. 1999.

[10] K. Azarian, "Cooperative Channels: Protocols and Analysis," *Ph.D. Dissertation*, The Ohio State University, Columbus, OH, Aug. 2006.

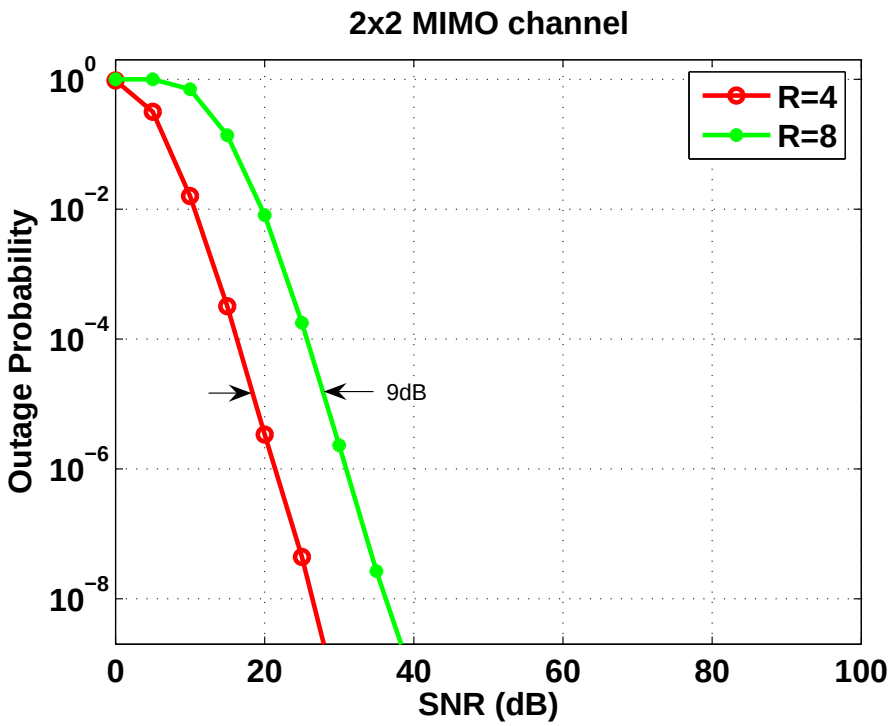

Figure 1: Outage curves corresponding to $R = 4, 8$ bpcu, for a 2×2 MIMO channel.

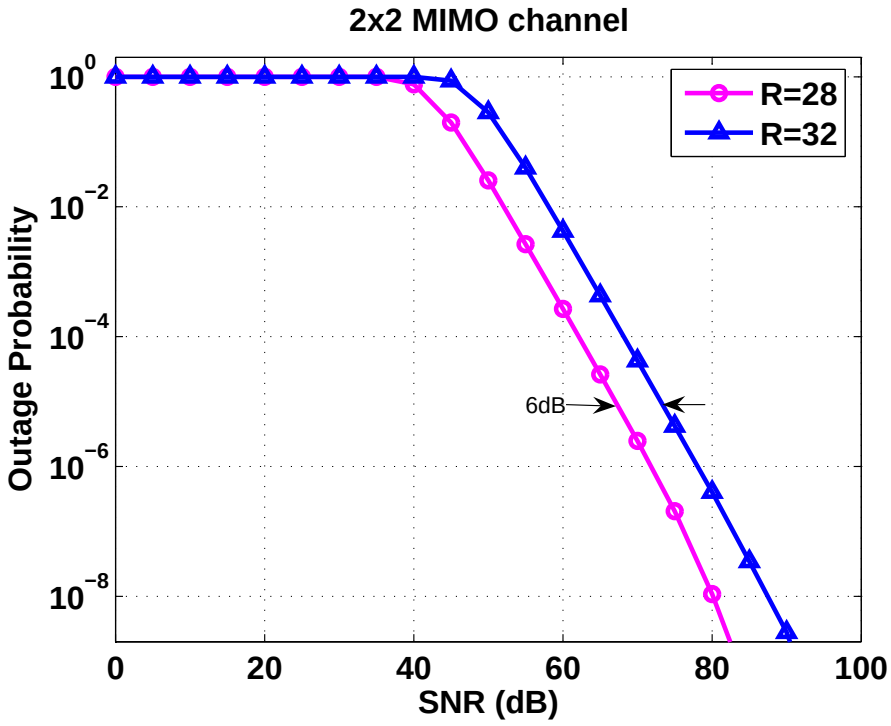

Figure 2: Outage curves corresponding to $R = 28, 32$ bpcu, for a 2×2 MIMO channel.

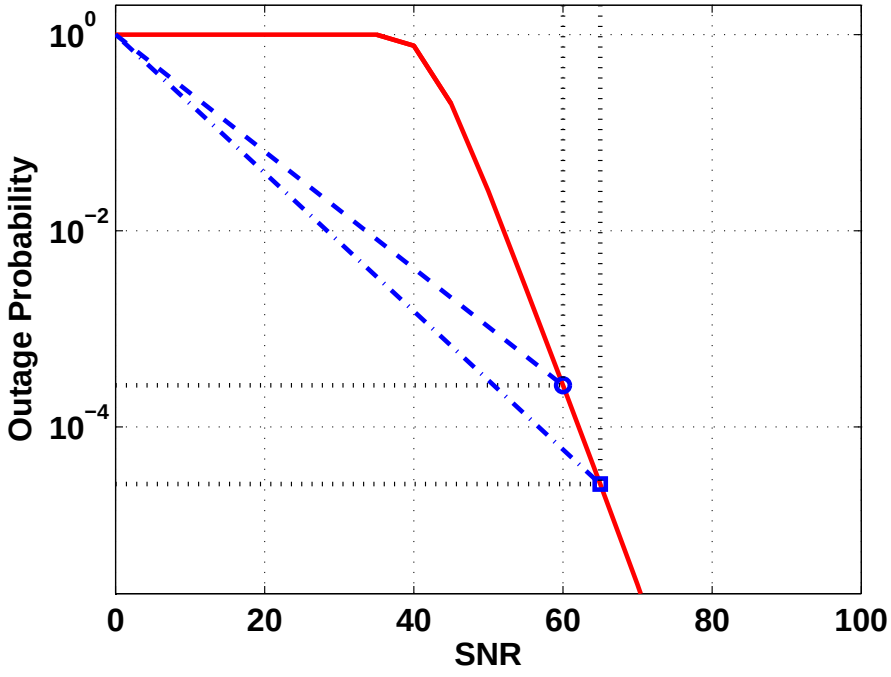

Figure 3: A *heuristic* method for computing the slope of the outage probability curve (for a *fixed* transmission rate), from the DMT (refer to (33)). The slope of the dashed and dot-dashed lines are given by $-d(\frac{R}{\log \rho})$ and $-d(\frac{R}{\log \rho + \Delta \log \rho})$, respectively.

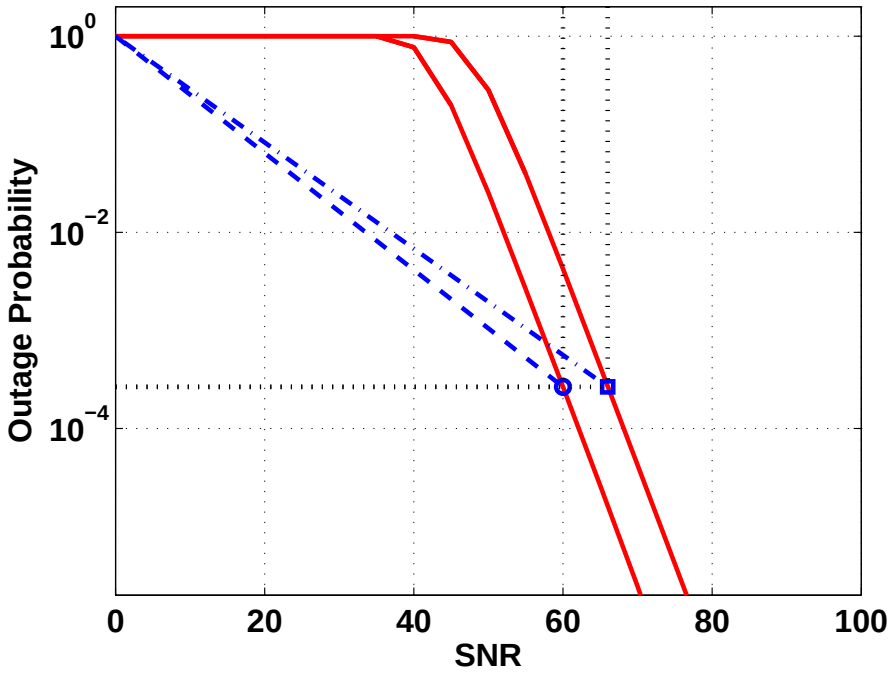

Figure 4: A *heuristic* method for computing the horizontal spacing between the outage probability plots (corresponding to two *fixed* transmission rates), from the DMT (refer to (34)). The slope of the dashed and dot-dashed lines are given by $-d(\frac{R}{\log \rho})$ and $-d(\frac{R+\Delta R}{\log \rho + \Delta \log \rho})$, respectively.

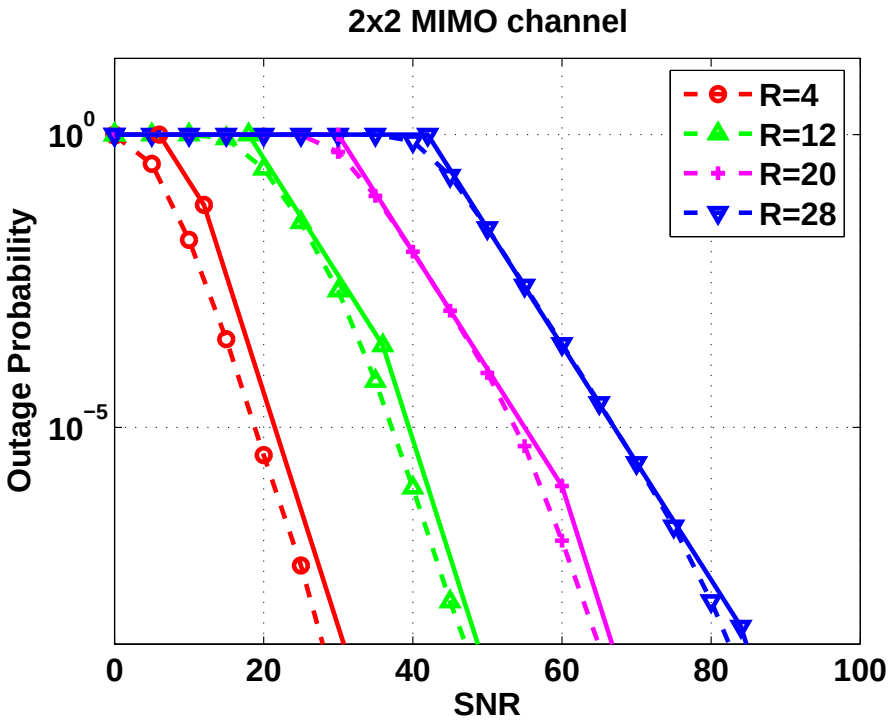

Figure 5: Outage probability curves for the 2×2 MIMO channel (dashed), along with the piecewise linear approximation (solid) suggested by the TRT (refer to (41)).

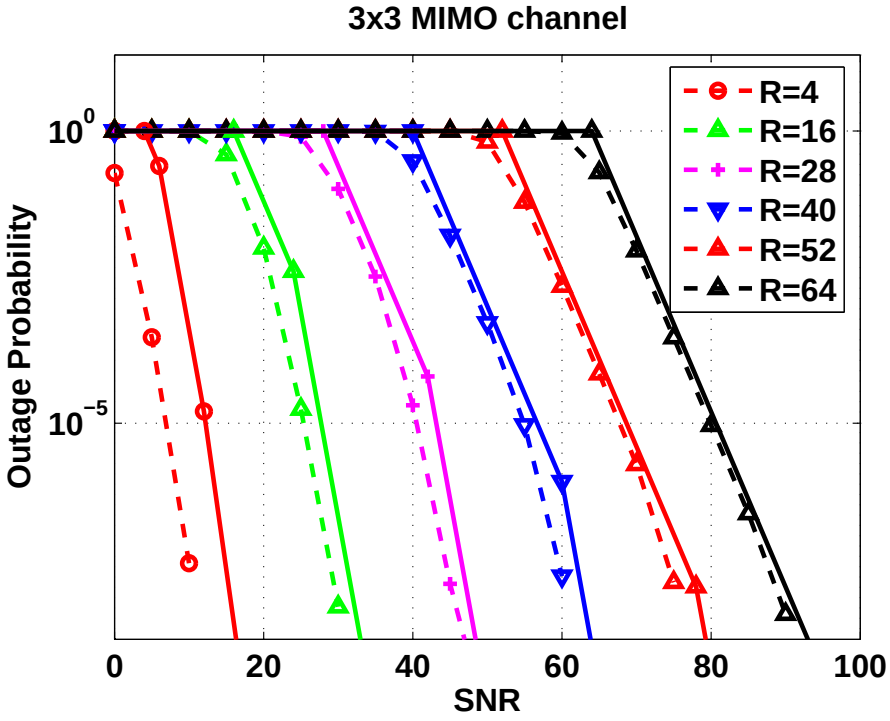

Figure 6: Outage probability curves for the 3×3 MIMO channel (dashed), along with the piecewise linear approximation (solid) suggested by the TRT (refer to (41)).

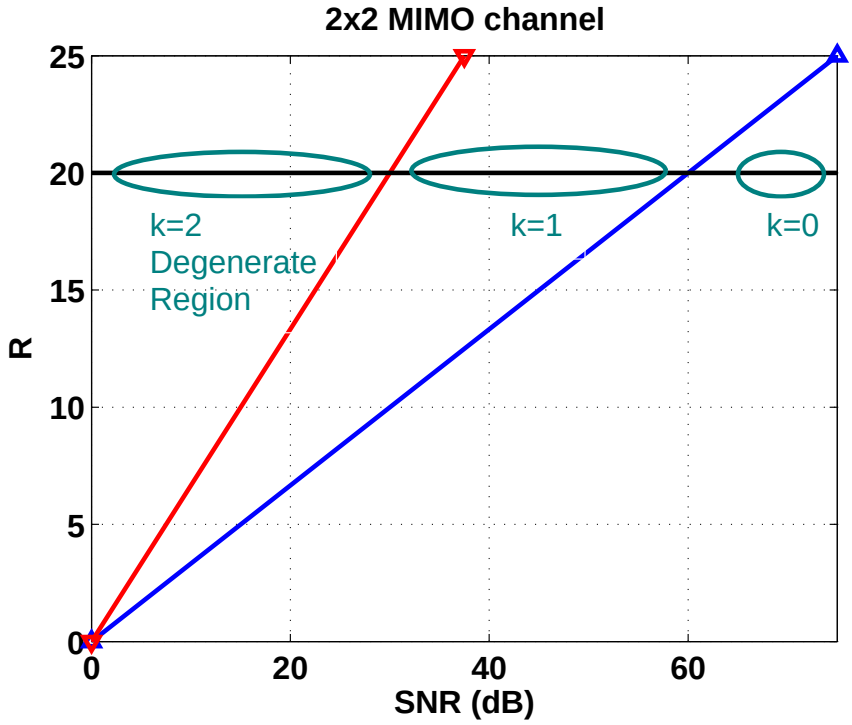

Figure 7: The constant rate trajectory with $R = 20$ bpcu passes through different operating regions in a 2×2 MIMO system

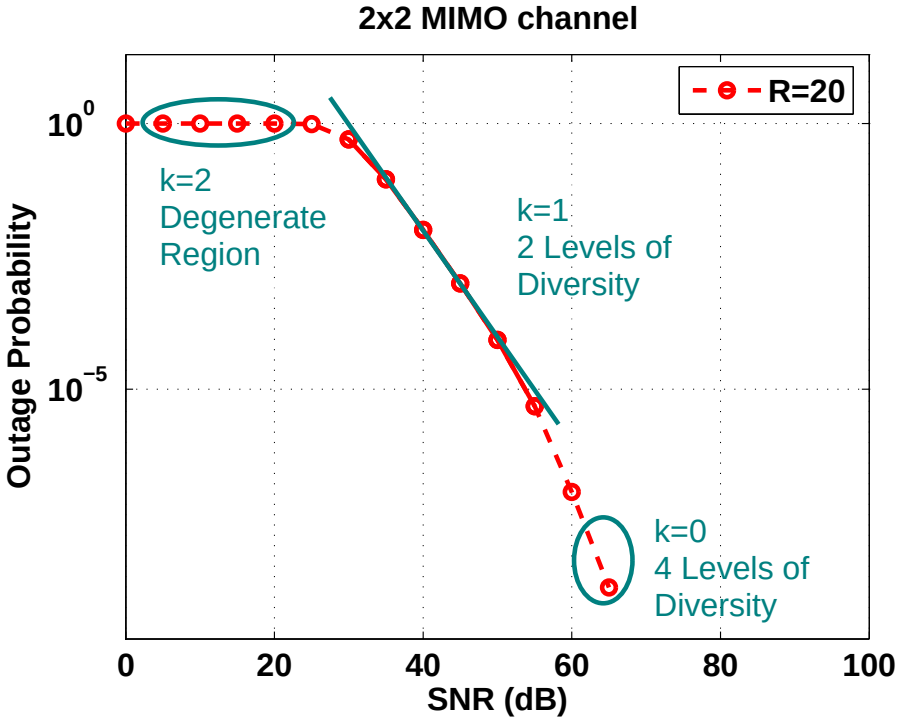

Figure 8: Outage curves corresponding to $R = 20$ bpcu for a 2×2 MIMO channel. The solid segment corresponds to the $\mathcal{R}(1)$ operating region.

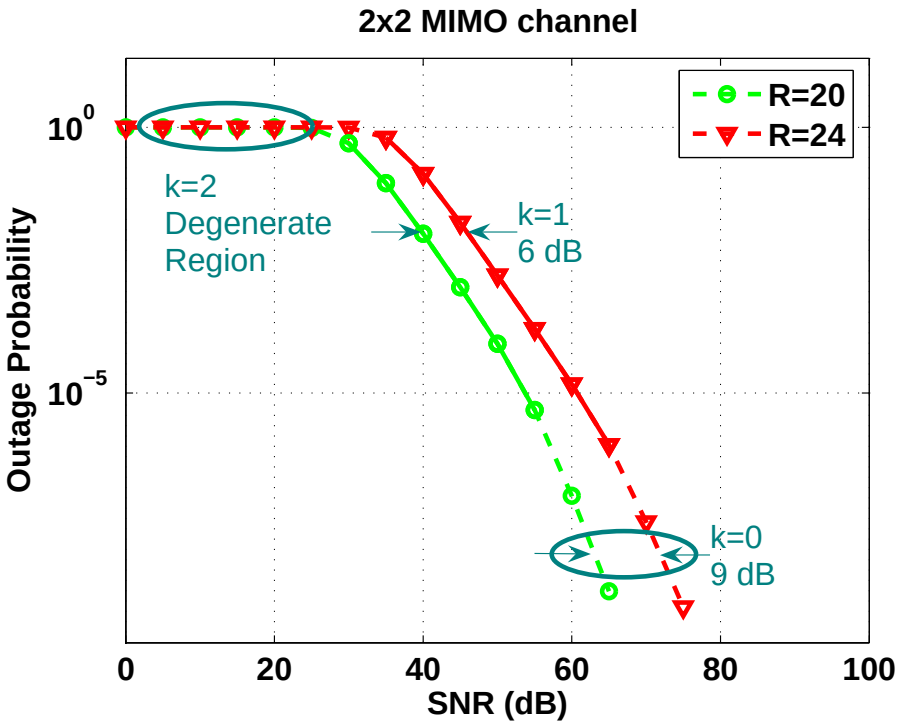

Figure 9: Outage curves corresponding to $R = 20, 24$ bpcu for a 2×2 MIMO channel. The solid segments correspond to the $\mathcal{R}(1)$ operating region.

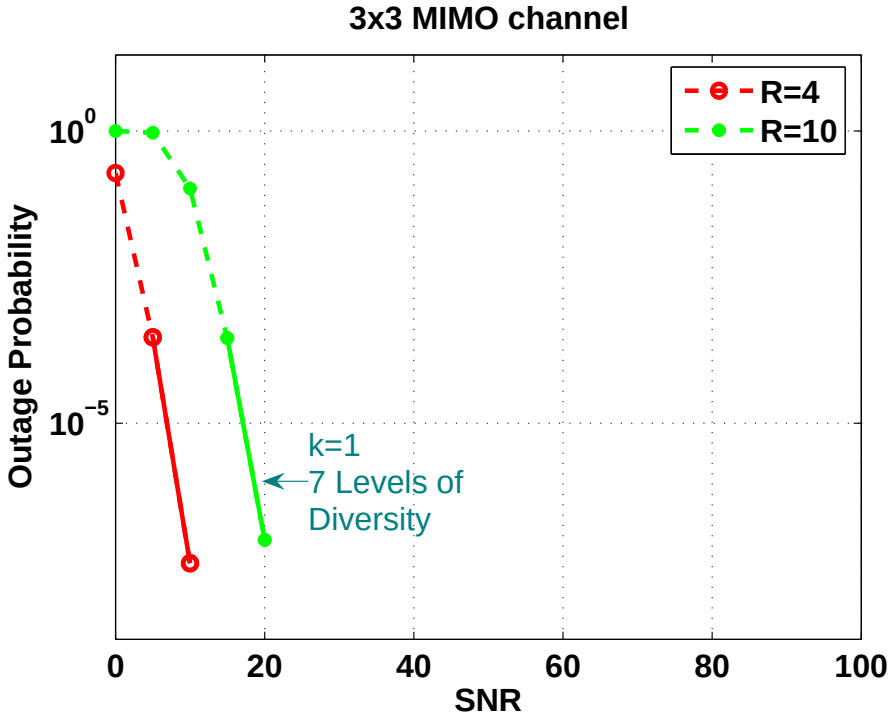

Figure 10: Outage curves corresponding to $R = 4, 10$ bpcu for a 3×3 MIMO channel. The solid segments correspond to the $\mathcal{R}(1)$ operating region.

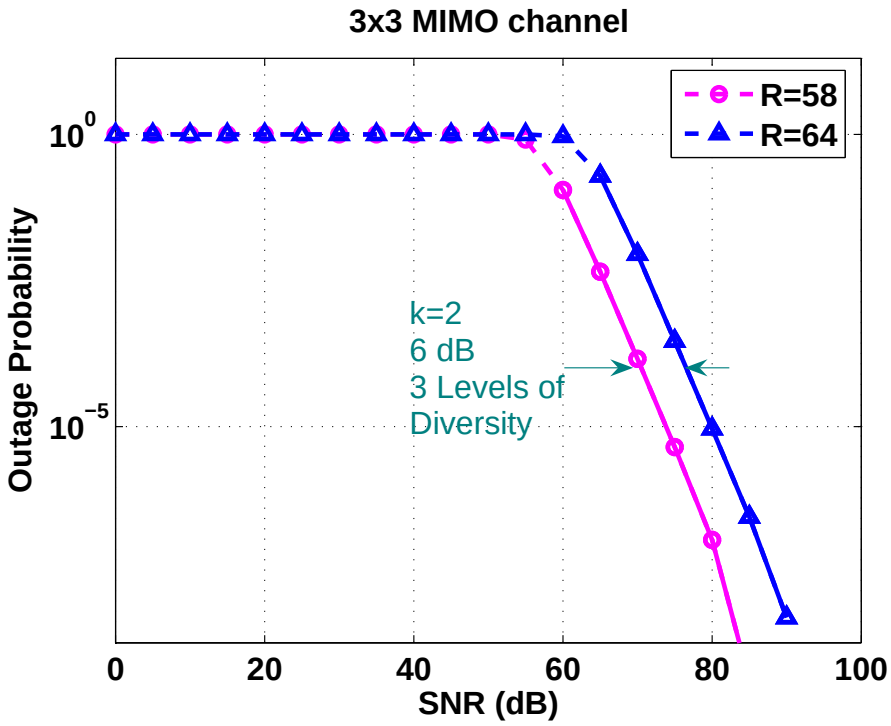

Figure 11: Outage curves corresponding to $R = 58, 64$ bpcu for a 3×3 MIMO channel. The solid segments correspond to the $\mathcal{R}(2)$ operating region.

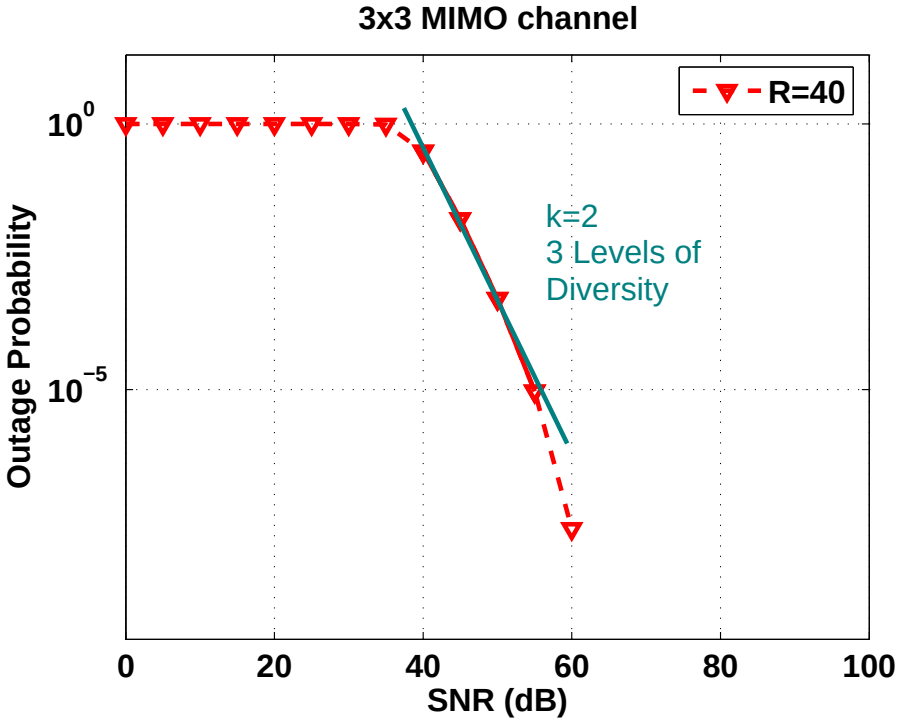

Figure 12: Outage curves corresponding to $R = 40$ bpcu for a 3×3 MIMO channel. The solid segment corresponds to the $\mathcal{R}(2)$ operating region.

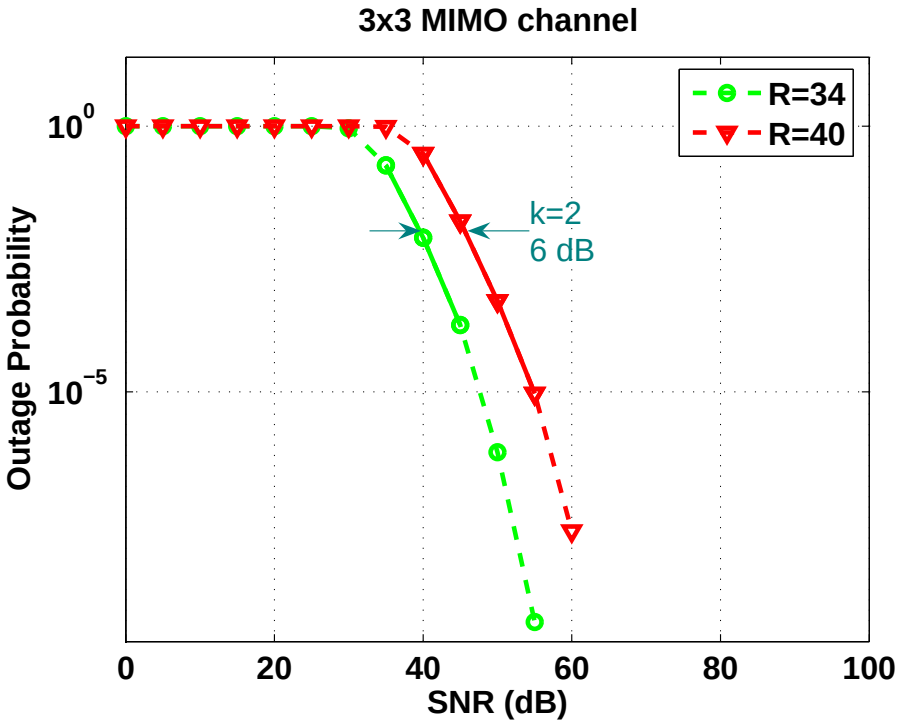

Figure 13: Outage curves corresponding to $R = 34, 40$ bpcu for a 3×3 MIMO channel. The solid segments correspond to the $\mathcal{R}(2)$ operating region.

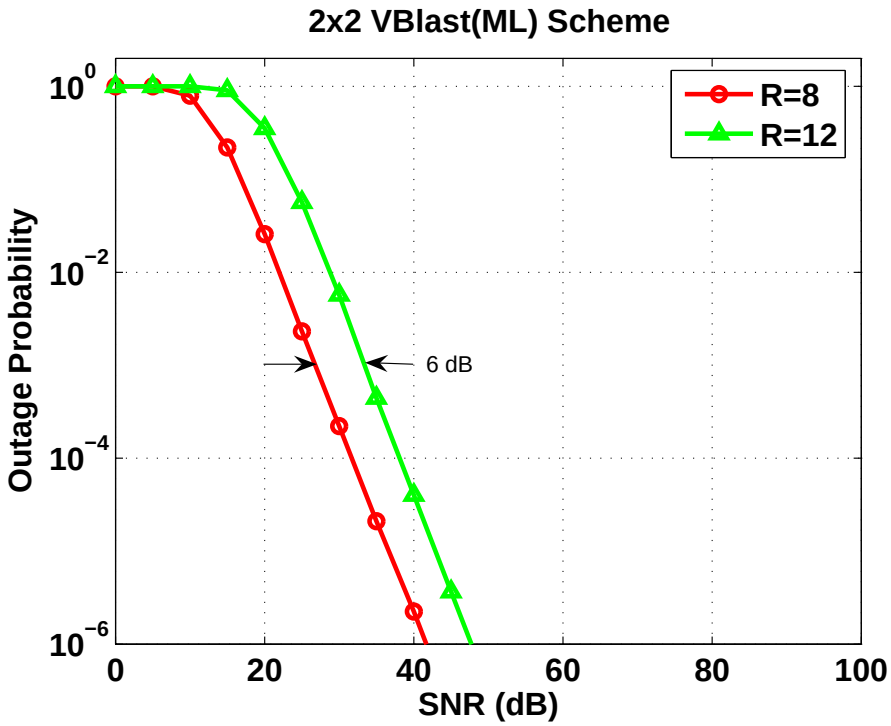

Figure 14: Outage curves corresponding to $R = 8, 12$ bpcu for a 2×2 V-BLAST (ML decoding) scheme.

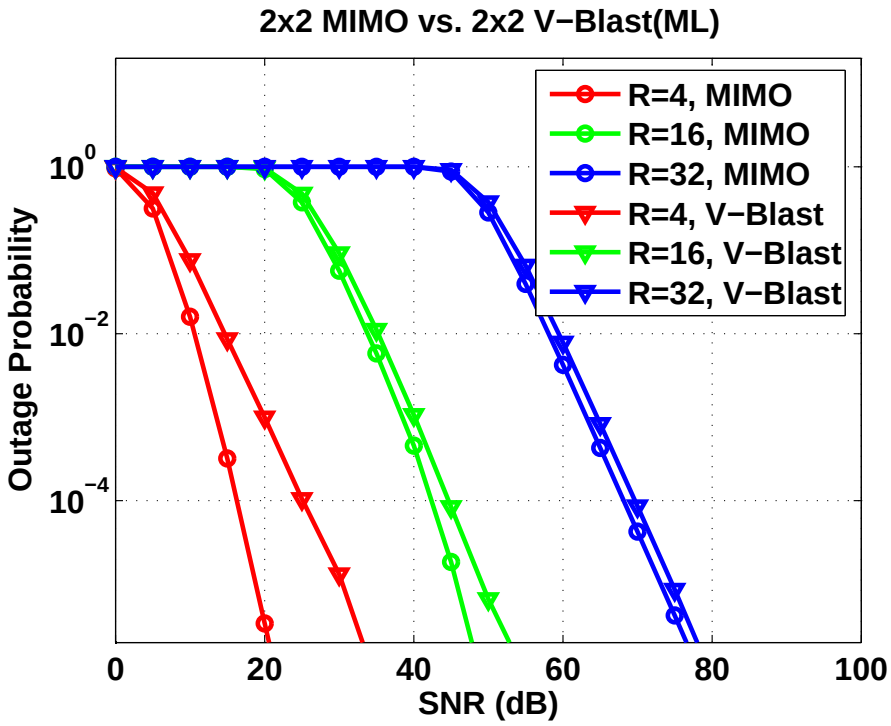

Figure 15: Comparison of outage curves corresponding to $R = 4, 16, 32$ bpcu for the 2×2 MIMO channel and the V-BLAST (ML decoding) scheme.

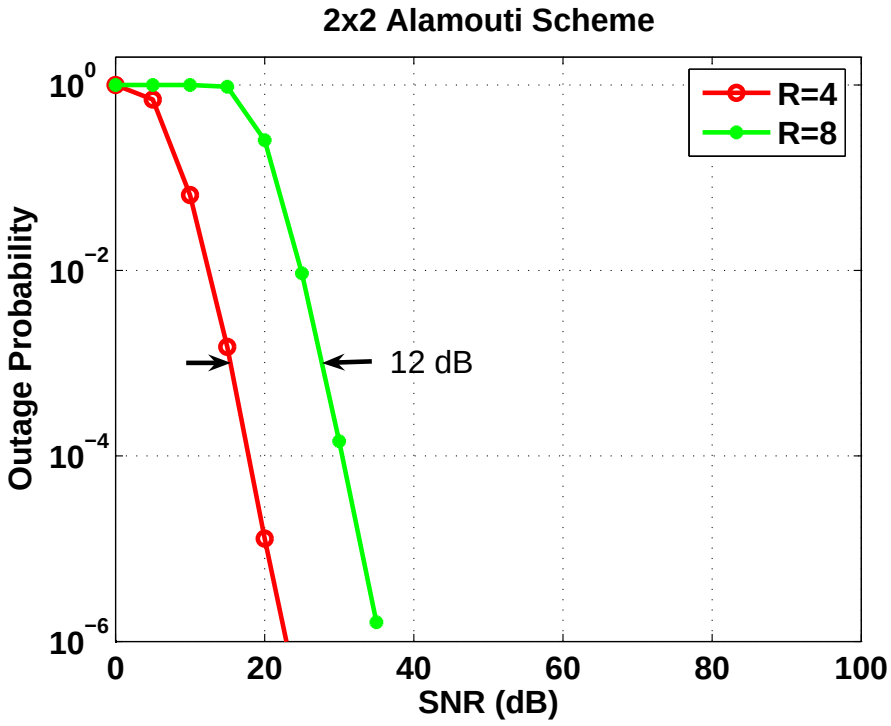

Figure 16: Outage curves corresponding to $R = 4, 8$ bpcu for the 2×2 Alamouti scheme.

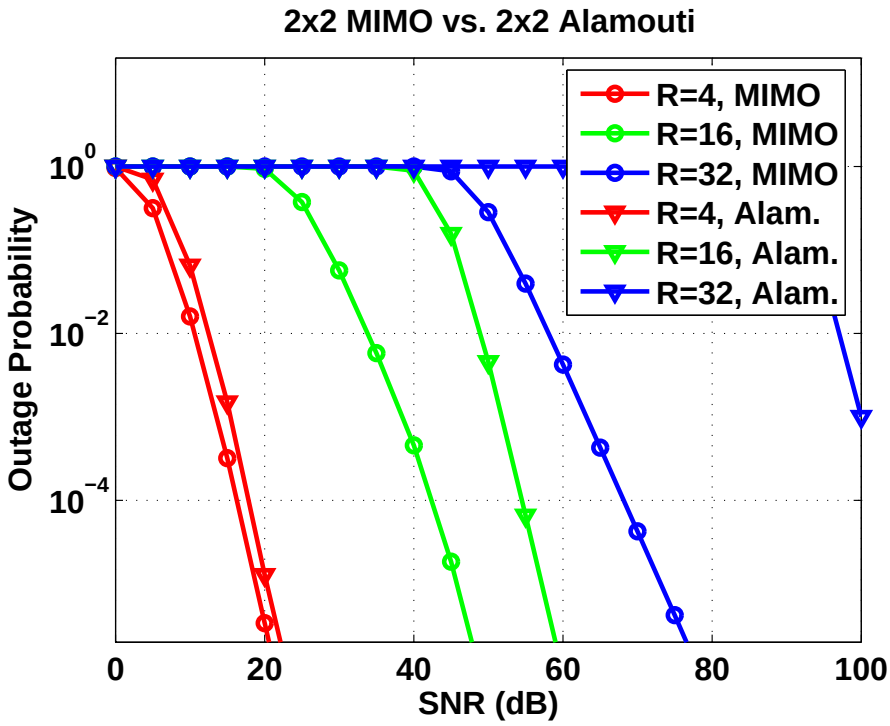

Figure 17: Comparison of outage curves corresponding to $R = 4, 16, 32$ bpcu for the 2×2 MIMO channel and the Alamouti scheme.