Exponential Diversity Achieving Spatio–Temporal Power Allocation Scheme for Fading Channels

Vinod Sharma, *Senior Member, IEEE*, Karumbu Premkumar, and Raghava N. Swamy

Abstract—In this paper, we analyze optimal (in space and time) adaptive power transmission policies for fading channels when the channel-state information (CSI) at the transmitter (CSIT) and the receiver (CSIR) is available. The transmitter has a long-term (time) average power constraint. There can be multiple antennas at the transmitter and at the receiver. The channel experiences Rayleigh fading. We consider *beamforming* and *space–time coded* systems with perfect/imperfect CSIT and CSIR. The performance measure is the bit error rate (BER). We show that in both coded and uncoded systems, our power allocation policy provides *exponential diversity order* if perfect CSIT is available. We also show that, if the quality of CSIT degrades then the exponential diversity is retained in the low SNR region but we get only polynomial diversity in the high SNR region. Another interesting conclusion is that in case of imperfect CSIT and CSIR, knowledge of CSIT at the receiver is very important. Finally, for the optimal power control policy of the uncoded system we find the error-exponents which provide the rate versus diversity-order tradeoff for this policy. This tradeoff is of an entirely different nature than the well-known Zheng–Tse tradeoff.

Index Terms—Beamforming, channel-state information receiver (CSIR), channel-state information transmitter (CSIT), convolutional codes, error exponents, fading channels, power allocation, rate versus diversity order tradeoff, space–time codes.

I. INTRODUCTION

WIRELESS channels are essential to provide ubiquitous connectivity to the users. However, due to multipath fading, low bandwidth, and broadcast nature, providing Quality of Service (QoS) to users in such channels has been a challenge. Currently, a tremendous effort is being invested into increasing the capacity and reducing the bit error rate (BER) of wireless channels. A significant gain can be achieved by using multiple antennas at the transmitter and at the receiver (multiple-input–multiple-output (MIMO)) (see [6], [9], [14], [16], and [23]).

In an additive white Gaussian noise (AWGN) channel the BER decreases (for various error correcting codes and modulation schemes) exponentially with the signal-to-noise ratio (SNR). However, the average BER in a Rayleigh-fading channel decreases only *inversely* with SNR, at least at high SNR's. This indicates a severe degradation in performance caused by Rayleigh fading. This degradation can be partially mitigated by a fully interleaved convolution code of *minimum distance*, $d_{\min}$ (where the BER can decrease as $\text{SNR}^{-d_{\min}}$). However, this is done at the cost of reduced information rate, increased receiver complexity and increased decoding delay. Alternatively, the degradation in BER performance due to fading can be partially mitigated by using multiple antennas at the transmitter and the receiver. For example, if n_t antennas are used at the transmitter and n_r at the receiver, one can achieve BER decay at a rate $\text{SNR}^{-n_t n_r}$ (called a *diversity order* of $n_t n_r$) ([14], [16]). This diversity order can be achieved with Space-Time codes, even if the transmitter does not have any knowledge of the channel ([14], [16], [22]). If the transmitter also has CSI, even if not exact, it can be exploited to obtain further reduction in BER ([11], [12], [14], [16], [28]). But, the diversity order remains $n_t n_r$. It is generally believed to be the maximum diversity order one can achieve for a Rayleigh-fading channel (although if a convolutional code with minimum distance $d_{\min}$ and interleaving is used, one can get the diversity order of $d_{\min} n_t n_r$).

Power control in multiuser environment has been studied in [4].

In all the works cited above and in the references therein, the power allocation is among the different transmit antennas (space-only power allocation). This is apparently due to the fact that the power allocation in time does not yield much improvement in capacity for a Rayleigh-fading channel. However, it is shown in [2] and [19] that power allocation in time can provide significant reduction in BER (it can also substantially increase the outage capacity [26]). The results in [2] and [19] are for single-input–single-output (SISO) system. A close look at the results in [2] and [19] reveals that even for an SISO system with Rayleigh fading, the power allocation policies obtained in [2] and [19] provide *exponential diversity* order, i.e., the BER decreases exponentially with SNR, as in an AWGN channel. Parts of this paper appeared in ISIT 2004 and the 43rd Allerton Conf where we proved exponential diversity (see also [20]). Also recently in [13] a diversity order much higher than $n_t n_r$ is obtained for channels with CSIT.

This paper takes a close look at the exponential diversity order aspect of the policies provided in [2] and [19]. We also study the corresponding policies for the MIMO systems. For perfect CSIT case, our power allocation policy provides an exponential diversity order, i.e., the BER, $P_b \leq \alpha e^{-f(n_t, n_r)\text{SNR}}$, where α

V. Sharma and K. Premkumar are with the Department of Electrical Communication Engineering, the Indian Institute of Science (IISc), Bangalore 560012, India (e-mail: vinod@ece.iisc.ernet.in; kprem@ece.iisc.ernet.in).

R. N. Swamy was with the Department of Electrical Engineering, Indian Institute of Technology, Madras, India. He is now with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: hnraghava@gmail.com).

is a positive constant, and $f > 0$ is an increasing function of n_t and n_r. We have provided the lower bounds on the exponential diversity order we achieve.

We also study the more realistic scenario of partial CSI at the transmitter and/or receiver. We obtain a very interesting result that as the quality of CSIT degrades, the exponential diversity is observed in the low SNR regime but we get only polynomial diversity (as in previous studies) for high SNR case. However, our policies still provide significant improvements in BER as compared to previous studies ([11], [14], [16], [28]). Another interesting result we obtain is that in case of partial CSIT, it is important for the *receiver* to know CSIT.

In our earlier work [20], we studied the bounds on the *beamforming* approach in multiple-input–single-output (MISO) systems with perfect CSIT (and perfect CSIR). The present work is an extension of [20]. We provide the BER performance bounds for the perfect and imperfect CSIT and CSIR cases for SISO, MISO, and MIMO systems. Furthermore, the power allocation policy considered in [20] is an extension of [11], whereas here, we report that the power allocation policy of perfect CSIT case can be extended to the imperfect CSIT case (without any degradation in BER performance as compared to [20]).

For our optimal policy for an uncoded system we also obtain error exponents. This way we obtain rate versus diversity order tradeoff as studied in [27] (although our method of error exponents is different from that in [27]). The method of [27] has been used in [13] for channels with CSIT. Error exponents for MIMO channels (without CSIT) have been studied in [9], [25] while for SISO channels with CSIT in [1] and [15]. [1] and [15] do not provide diversity order for their power control policies.

We show a very interesting result that the expurgated bound ([7]) (or the ensemble average bound) decays exponentially with SNR in low and high rate region but not in the middle region. Furthermore, the exponential diversity is lost in the low rate region when CSIT is not perfect.

We obtain the results explicitly for BPSK modulation but most of the analysis can be extended to more general constellations in the same way.

The rest of the paper is organized as follows. In Section II, we explain the perfect CSIT system model and the optimization problem. We consider SISO, MISO, and MIMO systems and provide exponential diversity. In Section III, we study the system with imperfect CSIT. In Section IV, we study the case of imperfect CSIR. We study the BER performance of a space–time coded system with our power allocation policy in Section V. Section VI provides the error exponents and rate versus diversity tradeoff. Section VII concludes the paper.

II. SYSTEM WITH PERFECT CSIT

We consider a single user narrowband (*flat fading*) communication system employing n_t transmit antennas and n_r receive antennas. We describe the channel between ith receive antenna and jth transmit antenna by a complex Gaussian random variable h_{ij}. Thus, the $n_r \times n_t$ matrix, $\mathbf{H} = [h_{ij}]$ represents the channel. We assume independent Rayleigh fading on each of the diversity branches. Also, we assume the fading to vary independently from one symbol to another i.e., $\{\mathbf{H}_k : k \in \mathcal{Z}^+\}$

is an independent and identically distributed (i.i.d.) process (this represents fast fading or a slow fading channel with perfect interleaving). The additive noise, N, is temporally and spatially white with mean zero and follows a multivariate complex circular Gaussian distribution, $N \sim \mathcal{N}_C(0, \sigma^2 I_{n_r})$, i.e., $\mathrm{E}[NN^T] = 0$ and $\mathrm{E}[NN^\dagger] = \sigma^2 I_{n_r}$, where $\dagger$ denotes Hermitian. Coherent signalling is assumed.

In this section, we consider the perfect CSIT and CSIR case, where the channel $\mathbf{H}$ is known to the transmitter and receiver (generalization to imperfect CSIT and CSIR will be considered in Sections III and IV). In this case, the optimum (in the SNR or BER sense) power allocation in space is achieved by *beamforming* [14]. We call this as *Space-Only Power Allocation* (SOPA). It is to be noted that in SOPA, the total power transmitted in a symbol duration is constant. On the other hand, we use the CSIT, $\mathbf{H}$ to compute the power $P(\cdot)$ to be allocated in each symbol duration while satisfying the beamforming power allocation. We call this as *Space–Time Power Allocation* (STPA). Thus, in STPA, the total power transmitted in a symbol duration varies from one symbol to another, but the long term (time) average of transmit power is fixed.

We derive the optimal power allocation policy here. The output of the matched filter is sampled at symbol rate and the received complex signal vector at time k is given by

$$Y_k = \sqrt{P(\gamma)}\, \mathbf{H}_k\, W_k\, x_k + N_k, \quad k = 0, 1, 2, \ldots \quad (1)$$

where x_k is the transmitted symbol, W_k is the input beamforming weight vector, $P(\gamma)$ is the transmit power, and $\gamma = \|\mathbf{H}_k W_k\|^2 \mathrm{E}|x|^2/\sigma^2$ is the SNR at the output of the channel. We define $s = \mathrm{E}|x|^2/\sigma^2$, as the average SNR per branch. Because of the i.i.d. assumptions of $\mathbf{H}_k$ and N_k, the optimal power at time k does not need to depend upon previous decisions and hence we can drop the time index.

The maximum likelihood (ML) detection of x given Y corresponds to

$$\arg\min_x \left| x - \frac{(\mathbf{H}W)^\dagger Y}{\sqrt{P(\gamma)}\|\mathbf{H}W\|^2} \right|^2 = \arg\min_x |x - \hat{x}|^2 \quad (2)$$

where

$$\hat{x} \triangleq \frac{(\mathbf{H}W)^\dagger Y}{\sqrt{P(\gamma)}\|\mathbf{H}W\|^2} \sim \mathcal{N}_C\left(x, \frac{\sigma^2}{P(\gamma)\|\mathbf{H}W\|^2}\right).$$

Thus, the performance of the MIMO system (SISO and MISO systems are special cases) can be interpreted as the output of an AWGN channel with $\mathrm{SNR} = P(\gamma)\, s\, \|\mathbf{H}W\|^2$. The transmit weight vector W is chosen to maximize the output SNR subject to the average transmit power constraint. The optimum W is the normalized eigenvector of $\mathbf{H}^\dagger \mathbf{H}$ corresponding to its largest eigenvalue, λ ([14]) and $\|\mathbf{H}W\|^2 = \lambda$. For a given $\gamma (= s\lambda)$, the BER is given by

$$P_{b|\gamma} = c \cdot Q(\sqrt{g \cdot \gamma\, P(\gamma)}) \quad (3)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{y^2}{2}\right) dy$$

and c and g are constants. Typically, the value of g is related to the minimum distance in the constellation, and c is related to the number of constellation points that achieve this minimum distance. For $BPSK$ and $QPSK$ with gray coding $c = 1$ and $g = 2$ give the exact BER. For $M-ary\ QAM$ an accurate (for high SNR) and useful approximation of the BER is obtained ([21]) by setting $c = 4\frac{\sqrt{M}-1}{\sqrt{M}\log_2 M}$ and $g = \frac{3}{M-1}\log_2 M$. For $M-PSK$ and high SNR, an approximate BER can be obtained with $c = 2/\log_2 M$ and $g = 2\sin^2(\pi/M)\log_2 M$ ([21]).

The optimization problem of minimizing the BER given in (3) subject to the average transmit power constraint is

$$\min_{E_\gamma[P(\gamma)]\leq \bar{P}} \int_0^\infty cQ\left(\sqrt{g\gamma P(\gamma)}\right) f_\gamma(\gamma)d\gamma \qquad (4)$$

where f_γ is the pdf of γ.

In the following, we will take $\bar{P} = 1$. The solution of this problem will provide the optimal (in the BER sense) power allocation policy in space and time. Applying Lagrange's method to the above problem, we get the following family of unconstrained optimization problems parameterized by a multiplier $\mu > 0$

$$\min_P \int_0^\infty (cQ(\sqrt{g\gamma P(\gamma)}) + \mu[P(\gamma) - 1])f_\gamma(\gamma)\,d\gamma.$$

The above unconstrained minimization corresponds to minimizing $cQ(\sqrt{g\gamma P}) + \mu P$ for any given γ. The Q function is complicated, but can be replaced by an upper bound $\frac{1}{2}e^{-g\gamma P/2}$. Since $\frac{c}{2}e^{-g\gamma P/2} + \mu P$ is a convex function of P, the optimum P can be obtained by differentiating it w.r.t. P and equating to zero. This provides the optimum solution

$$P(\gamma) = \begin{cases} \frac{2/g}{\gamma}\ln\left(\frac{\gamma}{\gamma_0}\right), & \text{for } \gamma \geq \gamma_0 \\ 0, & \text{for } \gamma < \gamma_0 \end{cases} \qquad (5)$$

where γ_0 is found by solving

$$\int_{\gamma_0}^\infty P(\gamma)f_\gamma(\gamma)d\gamma = 1. \qquad (6)$$

It should be noted here that γ_0 depends on the average SNR per branch, s and the modulation scheme. Also, the transmitter should have the complete knowledge of the fading statistics to compute γ_0 and the instantaneous fade values to compute $P(\cdot)$. In the following we will call this policy STPA (space–time power allocation). As against this the policy that allocates power optimally among antennas but constant P from symbol to symbol, as done in previous studies ([14], [16]) will be called SOPA (space only power allocation).

The average BER is given by

$$P_b = \frac{1}{2}F_\gamma(\gamma_0) + \int_{\gamma_0}^\infty cQ\left(\sqrt{g\gamma P(\gamma)}\right) f_\gamma(\gamma)d\gamma \qquad (7)$$

where F_γ is the cumulative density function (cdf) of f_γ.

The SNR $\gamma = s\lambda$ (where $\lambda = ||\mathbf{H}W||^2$, the largest eigenvalue of $\mathbf{H}^\dagger\mathbf{H}$) and its distribution varies for the SISO, MISO, and MIMO cases. It is possible to compute (7) exactly in each of these cases (either by numerical computation or via computing the integral by a Monte-Carlo method). However, in the following sections, we will show that the BER decays

exponentially by providing explicit upper bounds. We study the cutoff γ_0/s and the bounds on the BER for SISO, MISO, and MIMO cases separately. We consider $BPSK$ modulation scheme throughout the paper (although, as mentioned above, by choosing c and g appropriately, we can consider other modulation schemes). This optimization problem was approximately solved for Convolutional and Turbo codes in [2].

A. SISO

We consider a system with single transmit antenna and single receive antenna (i.e., $n_t = n_r = 1$). The channel, $\mathbf{H}$ is a complex Gaussian random variable with mean 0 and variance 1. Here, the beamforming element $W = \mathbf{H}^*/|\mathbf{H}|$. Thus, the SNR of the channel is given by $\gamma = s|\mathbf{H}|^2 \sim \exp(1/s)$. The cutoff γ_0 is computed from

$$\int_{\gamma_0}^\infty \frac{1}{\gamma}\ln\left(\frac{\gamma}{\gamma_0}\right)\frac{1}{s}e^{-\gamma/s}\,d\gamma = 1.$$

Taking $x = \gamma/\gamma_0$, we obtain

$$\int_{x=1}^\infty \frac{\ln(x)e^{-(\gamma_0/s)x}}{x}dx = s.$$

From [8, p. 573, eq. 4.362 (2)][1]

$$\int_1^\infty \frac{e^{-\mu x}\ln(2x-1)}{x}\,dx = \frac{1}{2}[\text{Ei}(-\mu/2)]^2$$

where Re $\mu > 0$. Thus,

$$s < 0.5\left[-\text{E}_1\left(\frac{\gamma_0/s}{2}\right)\right]^2.$$

Using [3, Th. 2]

$$s \leq 0.5\left[\ln\left(1 - e^{-\frac{\gamma_0/s}{2}}\right)\right]^2.$$

Thus, using $-\ln(1-x) \leq kx$ (see Appendix I)

$$\frac{\gamma_0}{s} \leq 2ke^{-\sqrt{2s}}. \qquad (8)$$

Therefore, from (7), using the upper bound on Q-function

$$P_b \leq 0.5\left[1 - e^{-\gamma_0/s} + \int_{\gamma_0/s}^\infty e^{-\ln(\frac{s\beta}{\gamma_0})}\cdot e^{-\beta}d\beta\right]$$

$$\leq 0.5\left[1 - e^{-\gamma_0/s} - \frac{\gamma_0}{s}\ln\left(1 - e^{-\gamma_0/s}\right)\right]$$

$$\text{(Using [3, Th. 2])}$$

$$\leq 0.5\frac{\gamma_0}{s}[1 + k].$$

Therefore, from (8)

$$P_b \leq ke^{-\sqrt{2s}}[1 + k]. \qquad (9)$$

This shows exponential diversity.

Let us compare the performance of the above STPA scheme with SOPA (which for SISO is same as no power control). From [18, p. 818], for SOPA we obtain, $P_b = \frac{1}{2}(1 - \sqrt{\frac{s}{1+s}})$ which tends to 0 as $\frac{1}{2s}$ for $s \to \infty$.

---

[1]The exponential integral $\text{Ei}(-x) = -\int_x^\infty \frac{e^{-t}}{t}\,dt$ and the function $\text{E}_1(x) = \int_x^\infty \frac{e^{-t}}{t}\,dt$, where $x > 0$.
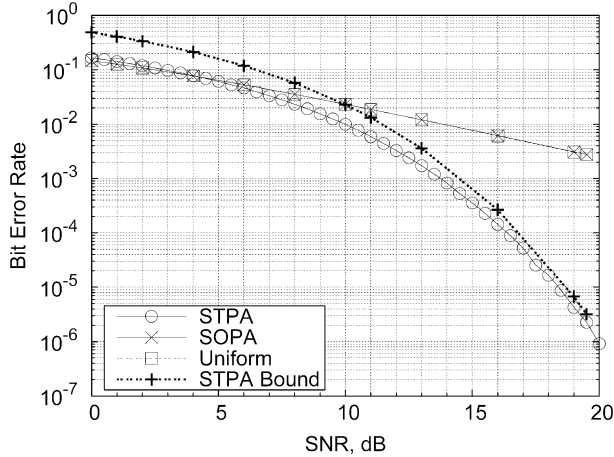
Fig. 1. BER versus average SNR for a SISO system with perfect CSIT.

In Fig. 1, we plot the upper bound given in (9) along with P_b for STPA and SOPA. One observes that at least for high SNR, the upper bound is quite tight. It is clear from Fig. 1 that the fall in BER with SNR is exponential in the case of STPA and linear in the case of SOPA. It is also evident from Fig. 1 that as SNR, s increases STPA provides significant performance gain as compared to SOPA. For example, to achieve a BER of 10^{-2}, STPA requires an SNR of 10 dB whereas SOPA requires an SNR of 16 dB, thus providing a power saving of 6 dB. For a more stringent BER requirement, the power saving by STPA is even more significant.

B. MISO

We consider a system with $n_t > 1$ transmit antennas and $n_r = 1$ receive antenna. The channel in this case is a $1 \times n_t$ vector, $\mathbf{H} \sim \mathcal{N}_C(0, I_{n_t})$. Here, the beamforming vector is $W = \mathbf{H}^\dagger/\|\mathbf{H}\|$, and the SNR of the channel is given by $\gamma = s\|\mathbf{H}W\|^2 = s\|\mathbf{H}\|^2 \sim \text{Erlang}_{n_t}(1/s)$. The cutoff γ_0 is given by

$$1 = \int_{\gamma_0}^{\infty} \frac{1}{\gamma} \ln\left(\frac{\gamma}{\gamma_0}\right) \frac{1}{s} \frac{e^{-\gamma/s}\left(\frac{\gamma}{s}\right)^{n_t-1}}{(n_t-1)!} \, d\gamma$$
$$= \frac{\left(\frac{\gamma_0}{s}\right)^{n_t}}{(n_t-1)!} \frac{1}{\gamma_0} \int_1^{\infty} e^{-(\gamma_0/s)x} x^{n_t-2} \ln(x) dx$$

where $x = \frac{\gamma}{\gamma_0}$.

We can easily show that the right side equals

$$\frac{1}{s(n_t-1)}\left[\int_{\gamma_0/s}^{\infty} \frac{e^{-t}}{t} dt + \sum_{k=0}^{n_t-3} \frac{e^{-\gamma_0/s}}{k+1} \sum_{i=0}^{k} \frac{(\gamma_0/s)^i}{i!}\right]$$
$$\leq \frac{1}{s(n_t-1)}\left[\int_{\gamma_0/s}^{\infty} \frac{e^{-t}}{t} dt + \sum_{k=0}^{n_t-3} \frac{e^{-\gamma_0/s}}{k+1} e^{\gamma_0/s}\right]$$
$$= \frac{1}{s(n_t-1)}\left[\int_{\gamma_0/s}^{\infty} \frac{e^{-t}}{t} dt + \delta\right]$$

where $\delta = \sum_{k=0}^{n_t-3} \frac{1}{k+1}$ (for $n_t = 2, \delta = 0$). Thus, using [3, Th. 2]

$$(n_t-1)s - \delta \leq -\ln(1 - e^{-\gamma_0/s})$$

and hence

$$e^{-(n_t-1)s+\delta} \geq 1 - e^{-\gamma_0/s}.$$

Using Appendix I, we get

$$\frac{\gamma_0}{s} \leq ke^{-(n_t-1)s+\delta}. \tag{10}$$

Therefore

$$P_b = \frac{1}{2}F_{\|\mathbf{H}\|^2}(\gamma_0/s)$$
$$+ \int_{\gamma_0/s}^{\infty} Q(\sqrt{2s\alpha P(s\alpha)}) f_{\|\mathbf{H}\|^2}(\alpha) d\alpha \tag{11}$$
$$\leq \frac{1}{2}F_{\|\mathbf{H}\|^2}(\gamma_0/s) + \int_{\gamma_0/s}^{\infty} \frac{1}{2}e^{-s\alpha P(s\alpha)} f_{\|\mathbf{H}\|^2}(\alpha) d\alpha$$
$$= 0.5\left[1 - \sum_{i=0}^{n_t-1} \frac{e^{-\gamma_0/s}(\gamma_0/s)^i}{i!}\right.$$
$$\left. + \int_{\alpha=\gamma_0/s}^{\infty} e^{-s\alpha \frac{1}{s\alpha}\ln\left(\frac{\alpha}{\gamma_0/s}\right)} \frac{e^{-\alpha}\alpha^{n_t-1}}{(n_t-1)!} d\alpha\right]$$
$$= 0.5\left[1 - \sum_{i=0}^{n_t-1} \frac{e^{-\gamma_0/s}(\gamma_0/s)^i}{i!} + \frac{\gamma_0}{s}\frac{e^{-\gamma_0/s}}{n_t-1} \sum_{i=0}^{n_t-2} \frac{(\gamma_0/s)^i}{i!}\right]$$
$$\leq 0.5\left[1 - e^{-\gamma_0/s} + \frac{\gamma_0/s}{n_t-1}\right] \tag{12}$$

and hence

$$P_b \leq 0.5\frac{n_t}{n_t-1}ke^{-(n_t-1)s+\delta}. \tag{13}$$

Thus, we see the exponential decay of P_s with SNR. As n_t increases, the rate of BER decay increases.

We compare the BER performance of the STPA scheme with the SOPA scheme. For SOPA scheme, the exact BER is (see [14, pp. 42–43])

$$P_b = \frac{1}{2^{n_t}}\left(1 - \sqrt{\frac{s}{1+s}}\right)^{n_t}$$
$$\times \sum_{k=0}^{n_t-1}\left[\frac{1}{2^k}\binom{n-1+k}{k}\cdot\left(1 + \sqrt{\frac{s}{1+s}}\right)^k\right]$$
$$= \frac{1}{2^{n_t}s^{n_t}} + O\left(\frac{1}{s^{2n_t}}\right)$$
$$\to \frac{1}{2^{n_t}s^{n_t}} \text{ as } s \to \infty.$$

We plot P_b from (11) in Fig. 2 for $n_t = 2$ and 3. For comparison, we have also included the curve for SOPA of [14]. One can see significant gains in BER for our policy as compared to the SOPA. The bound (12) is too loose and hence not provided.

It is worth mentioning here that the BER performance due to *Uniform Power Allocation* in MISO case is the same as that of SISO case evaluated by replacing s with $n_t s$, i.e.

$$P_b = \frac{1}{2}\left(1 - \sqrt{\frac{n_t s}{1+n_t s}}\right)$$
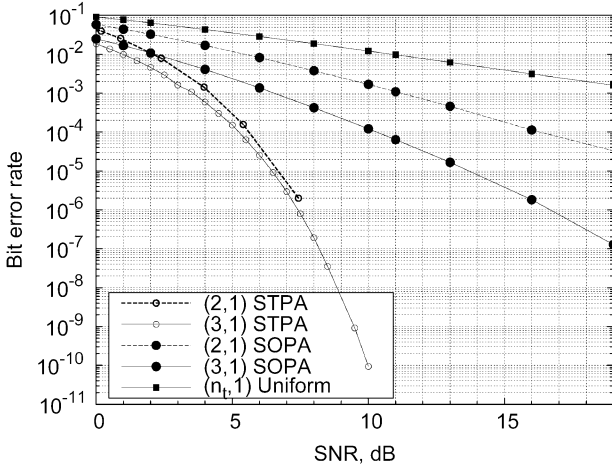$$\to \frac{1}{2n_t s} \text{ as } s \to \infty.$$

Fig. 2. BER versus average SNR for a MISO systems with perfect CSIT.

C. MIMO

For this system, the maximum SNR for an average transmit power constraint is achieved by beamforming in the direction of the eigenvector corresponding to the largest eigenvalue, λ of the Wishart distributed matrix, $\mathbf{H}^{\dagger}\mathbf{H}$. Thus, the SNR of the channel γ is $s\lambda = s\|\mathbf{H}W\|^2$. It is to be noted that the nonzero eigenvalues (and the distribution) of $\mathbf{H}^{\dagger}\mathbf{H}$ and $\mathbf{H}\mathbf{H}^{\dagger}$ are same. Hence, without loss of generality, we consider the case $n_t \geq n_r$ (i.e., the BER of the system with n_t transmit antennas and n_r receive antennas will be the same as that of the system with n_r transmit antennas and n_t receive antennas).

The density function of the largest eigenvalue of the Wishart distributed matrix is given by ([5])

$$f_\lambda(\lambda) = \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i^{m+1}\lambda^m e^{-i\lambda}}{m!} \qquad (14)$$

where $d_{i,m}$'s are constants which depend upon n_t and n_r. Therefore, the cutoff γ_0 is found by solving the equation

$$\sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i^{m+1}(\gamma_0/s)^m}{m!}$$
$$\times \int_1^\infty x^{m-1}e^{-i(\gamma_0/s)x}\ln(x)dx = s.$$

Solving this equation, we get (see Appendix II)

$$\frac{\gamma_0}{s} \lesssim k\exp(-(s-c)/K) \qquad (15)$$

where k was defined earlier and, c and K are defined in Appendix II (K and c defined in the Appendix are different for $n_t > n_r$ and $n_t = n_r$.) The average BER satisfies

$$P_b \leq \frac{1}{2}\int_0^{\gamma_0/s} f_\lambda(\lambda)d\lambda + \int_{\gamma_0/s}^\infty \frac{1}{2}e^{-s\lambda P(s\lambda)}f_\lambda(\lambda)d\lambda$$
$$\lesssim 0.5Kke^{-(s-c)/K}. \qquad (16)$$

TABLE I
$1/K$ FOR VARIOUS (n_t, n_r)

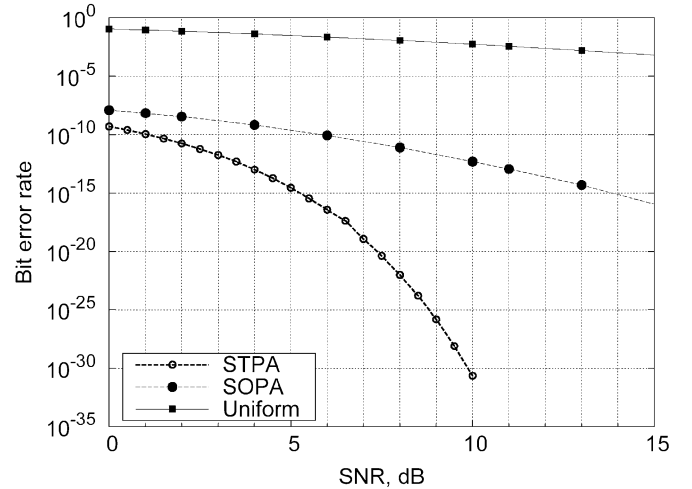| (n_t, n_r)	$1/K$	K
(2,2)	2.5887	0.3863
(3,2)	4.0000	0.2500
(4,2)	5.3333	0.1875
(6,2)	7.8769	0.1270
(8,2)	10.3261	0.0968
(4,3)	7.2605	0.1377
(5,3)	8.7465	0.1143
(8,4)	15.2426	0.0656

Fig. 3. BER versus average SNR for (4,2) MIMO systems with perfect CSIT.

The parameter $1/K$ for various values of (n_t, n_r) is numerically computed and shown in Table I.

The approximation given by (16) is very tight and is plotted in Fig. 3 for $(n_t, n_r) = (4, 2)$. It also shows the exponential decay of P_b with SNR and (using Table I) that the rate of decay increases with (n_t, n_r).

The exact BER expression for SOPA is given in [5], (19). The uniform power allocation case here is also equivalent to a SISO system evaluated by replacing s with $n_t n_r s$, i.e.

$$P_b = \frac{1}{2}\left(1 - \sqrt{\frac{n_t n_r s}{1 + n_t n_r s}}\right) \sim \frac{1}{2n_t n_r s} \text{ as } s \to \infty.$$

Fig. 3 shows the BER performance of a MIMO system with $n_t = 4$ and $n_r = 2$ for SOPA and uniform power allocation scheme also. It is clear from Fig. 3 that the fall in BER with SNR is exponential for STPA, polynomial for SOPA and linear for uniform power allocation. Also, the power saving due to STPA over the other schemes is substantial. For example, STPA achieves a BER of 10^{-10} at 1 dB whereas SOPA requires 6 dB to achieve the same BER. This gain margin increases with BER.

III. SYSTEM WITH IMPERFECT CSIT

The system model considered here is the same as in Section II, except that the channel-state information available at the transmitter is noisy. We assume that $\mathbf{H}_T$ is the transmitter's estimate

of the channel. $\mathbf{H}_T$ could be a delayed version or a finite precision representation or a noisy version of $\mathbf{H}$. We assume that $\mathbf{H}$ and $\mathbf{H}_T$ are jointly complex Gaussian with ρ_T the correlation between $\mathbf{H}$ and $\mathbf{H}_T$. These assumptions are commonly made in this literature (see the references provided).

Now the optimal transmit power will be the same as in (5) except that we use $\mathbf{H}_T$ instead of $\mathbf{H}$, i.e., compute $P(\gamma_T)$, where $\gamma_T = s\lambda_T$ and λ_T is the largest eigenvalue of $\mathbf{H}_T^\dagger \mathbf{H}_T$. Coherent signaling is assumed. The output of the matched filter is sampled at symbol duration and the received complex signal vector is given by

$$Y_k = \sqrt{P(\gamma_T)}\mathbf{H}_k W_{Tk}\, x_k + N_k, \quad k = 0, 1, 2, \ldots$$

where γ_T is the SNR estimate at the transmitter ($\gamma_T = s\lambda_T = s\|\mathbf{H}_T W_T\|^2$), $P(\gamma_T)$ is the instantaneous transmit power and W_{Tk} is the input weight vector (eigenvector of $\mathbf{H}_T^\dagger \mathbf{H}_T$ corresponding to the largest eigenvalue λ_T). Also, we define μ to be $\|\mathbf{H}W_T\|^2$. As before, we drop the time index k.

The imperfect CSIT is modeled as

$$\mathbf{H}_T = \rho_T \mathbf{H} + \sqrt{1 - \rho_T^2}\,\mathbf{E} \tag{17}$$

where $\mathbf{E}$ is the estimation error independent of $\mathbf{H}$ and has distribution $\sim \mathcal{N}(0, I_{n_t \cdot n_r})$. Equivalently, (as can be seen by the joint distribution of $(\mathbf{H}, \mathbf{H}_T)$)

$$\mathbf{H} = \rho_T \mathbf{H}_T + \sqrt{1 - \rho_T^2}\,\mathbf{E}' \tag{18}$$

where $\mathbf{E}'$ is the estimation error independent of $\mathbf{H}_T$ and has distribution $\sim \mathcal{N}(0, I_{n_t \cdot n_r})$ ($\mathbf{E}$ and $\mathbf{E}'$ are used interchangeably in the rest of the paper). (17) can be a reasonable model if $\mathbf{H}_T$ is a quantized version of $\mathbf{H}$ and then $\mathbf{E}$ is quantization noise. (18) represents the situation when $\mathbf{H}_T$ is a delayed version of $\mathbf{H}$ and $\mathbf{H_k}$ (the channel state during the kth symbol transmission) is an Autoregressive process.

We assume that the receiver has perfect CSI and also $\mathbf{H}_T$ the transmitter's knowledge of the channel. When $\mathbf{H}_T$ is a delayed or quantized version of $\mathbf{H}$, this assumption will be satisfied. The case when the receiver does not know $\mathbf{H}_T$ will be covered in the next section.

The average BER is given by

$$P_b = 0.5 \int_{\lambda_T = 0}^{\gamma_0/s} f_{\lambda_T}(\lambda_T)\, d\lambda_T + \int_{\mu = 0}^{\infty} \int_{\lambda_T = \gamma_0/s}^{\infty} Q$$
$$\times \left(\sqrt{2s\mu P(s\lambda_T)}\right) f_{\mu,\lambda_T}(\mu, \lambda_T)\, d\mu\, d\lambda_T. \tag{19}$$

The density function, $f_{\lambda_T}(\lambda_T)$ is the same as that of the perfect CSIT case. As with perfect CSIT (19) can be actually computed. But, we obtain bounds on it which provide a better insight on the behavior of P_b as the SNR changes. Since λ_T has the same distribution as λ corresponding to the perfect CSIT case, the bounds on the cutoff values, γ_0 computed for the perfect CSIT case hold good for the imperfect CSIT case. But, for the imperfect CSIT case, we also need to derive the conditional pdfs $f_{\mu \mid \lambda_T}(\mu \mid \lambda_T)$ for the SISO, MISO, and MIMO cases which we obtain in the following.

A. SISO

For $n_t = n_r = 1$, the (scalar) channel $\mathbf{H} \sim \mathcal{N}_C(0, 1)$. Since $\mathbf{H} = \rho_T \mathbf{H}_T + \sqrt{1 - \rho_T^2}\mathbf{E}$, $|\mathbf{H}|^2 = X^2 + Y^2$, where X and Y are the real and imaginary parts of $\mathbf{H}$. Clearly, X and Y are Gaussian distributed (given $\mathbf{H}_T$), with means, $m_x = \rho_T \text{Re}(\mathbf{H}_T)$ and $m_y = \rho_T \text{Im}(\mathbf{H}_T)$. Therefore, $m_x^2 + m_y^2 = \rho_T^2 |\mathbf{H}_T|^2$. The variance of X and Y is $\frac{1}{2}\sigma_e^2 = \frac{1}{2}(1 - \rho_T^2)$. Thus, the density function $f_{|\mathbf{H}|^2 \mid |\mathbf{H}_T|^2}(\alpha \mid \beta)$ follows a *noncentral chi-square* distribution

$$f_{|\mathbf{H}|^2 \mid |\mathbf{H}_T|^2}(\alpha \mid \beta) = \frac{1}{\sigma_e^2} e^{-(\beta \rho_T^2 + \alpha)/\sigma_e^2} \sum_{k=0}^{\infty} \frac{\left(\frac{-\alpha\beta\rho_T^2}{\sigma_e^4}\right)^k}{k!\Gamma(k+1)}$$

where Γ is the Gamma function. The average BER is

$$P_b \leq 0.5 F_{|\mathbf{H}_T|^2}(\gamma_0/s) + 0.5 \int_{\alpha=0}^{\infty} \int_{\beta=\gamma_0/s}^{\infty} e^{-s\alpha P(s\beta)} \tag{20}$$
$$\times\, f_{|\mathbf{H}|^2}(\beta) f_{|\mathbf{H}|^2 \mid |\mathbf{H}_T|^2}(\alpha \mid \beta)\, d\alpha\, d\beta$$
$$= 0.5[1 - e^{-\gamma_0/s}] + 0.5 \int_{\alpha=0}^{\infty} \int_{\beta=\gamma_0/s}^{\infty} e^{-\frac{\alpha}{\beta}\ln\left(\frac{s\beta}{\gamma_0}\right)}$$
$$\cdot\, e^{-\beta} \frac{1}{\sigma_e^2} e^{-(\beta\rho_T^2 + \alpha)/\sigma_e^2} \sum_{k=0}^{\infty} \frac{\left(\frac{-\alpha\beta\rho_T^2}{\sigma_e^4}\right)^k}{k!\Gamma(k+1)}\, d\alpha\, d\beta$$
$$= 0.5[1 - e^{-\gamma_0/s}] + 0.5\left(\frac{\gamma_0}{s}\right)^2 \int_{x=1}^{\infty} \frac{e^{-(1 + \frac{\rho_T^2}{\sigma_e^2})\frac{\gamma_0}{s}x}}{\frac{\gamma_0}{s} + \sigma_e^2 \frac{\ln(x)}{x}}$$
$$\times \exp\left(\frac{-\left(\frac{\gamma_0}{s}\right)^2 \frac{\rho_T^2}{\sigma_e^2} x}{\frac{\gamma_0}{s} + \sigma_e^2 \frac{\ln(x)}{x}}\right) dx$$
$$\leq 0.5\left[1 - e^{-\gamma_0/s} + \left(\frac{\gamma_0}{s}\right)^2 \int_{x=1}^{\infty} \frac{e^{-\alpha' x}}{\frac{\gamma_0}{s} + \sigma_e^2 \frac{\ln(x)}{x}}\, dx\right]$$
$$\leq 0.5\left[1 - e^{-\gamma_0/s} + \left(\frac{\gamma_0}{s}\right)\frac{e^{-\alpha'}}{\alpha'}\right]$$
$$\leq 0.5\frac{\gamma_0}{s}\left[1 + \frac{e^{-\alpha'}}{\alpha'}\right]$$
$$\leq 0.5 k(1+k) e^{-\sqrt{2s}}\left(1 + \frac{e^{-\alpha'}}{\alpha'}\right) \tag{21}$$

where

$$\alpha' = \left[1 + \frac{\rho_T^2}{\sigma_e^2}\right]\frac{\gamma_0}{s} + \frac{\left(\frac{\gamma_0}{s}\right)^2 \frac{\rho_T^2}{\sigma_e^2}}{\frac{\gamma_0}{s} + \sigma_e^2} \tag{22}$$

We plot in Fig. 4 the BER (calculated from (19)) of the power allocation policy defined in (5) for different values of ρ_T of a SISO system. Rayleigh fading is assumed and the average fading gain is taken to be one. It is evident from Fig. 4 that exponential diversity is retained for low $\text{SNR}(\leq 8\text{ dB})$ but is lost in the high SNR region (even for $\rho_T = 0.99$, which implies an error of the order of 15% in the estimation of $\mathbf{H}$). This means that the BER performance in low SNR region is inherently limited by the additive noise of the channel and in high SNR region it is limited by the channel estimation error. One interesting point to be deduced from here is that the finite bit
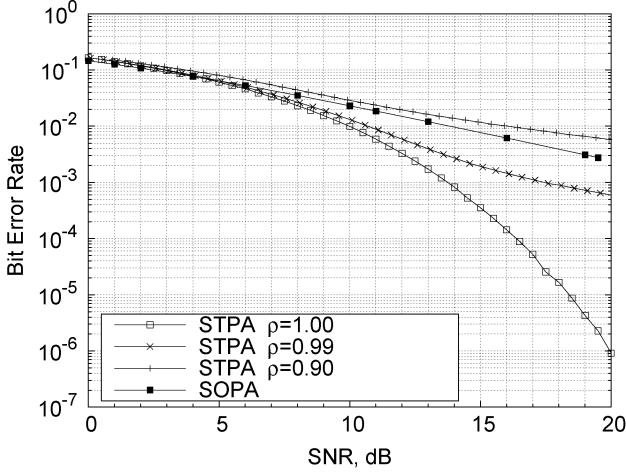
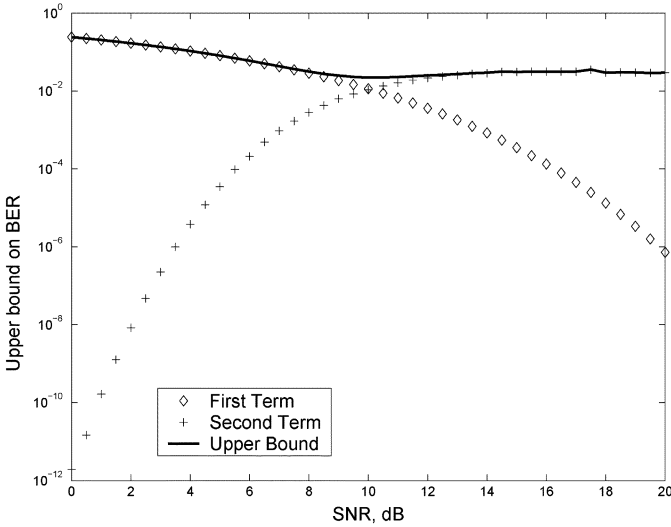Fig. 4.   BER versus average SNR for an SISO system with Imperfect CSIT.

Fig. 5.   Upper bound versus average SNR of SISO system.

precision of the feedback (channel estimate) can be coarser at low SNR but should be finer at high SNR.

The behavior of BER as a function of SNR for imperfect CSIT explained in the above paragraph is captured partly in the upper bound (21). Fig. 5 shows the behavior of the two terms in (21) as a function of SNR. It is evident from Fig. 5 that up to a certain SNR, the first term dominates and later on the second term. Thus, one can expect an exponential fall of BER in the low SNR region (which is caused by the channel noise, since this term is also present in the perfect CSIT case). As SNR increases, the effect of estimation error is more prominent and the BER fall in SNR becomes polynomial.

B. MISO

Let $\mathbf{H}_T = \rho_T \mathbf{H} + \sqrt{1 - \rho_T^2}\, \mathbf{E}$ be the CSI available at the transmitter. The transmitter does beamforming by choosing the transmit weight vector $W_T = \mathbf{H}_T^\dagger / \|\mathbf{H}_T\|$. The instantaneous transmit power is chosen as if $\mathbf{H}_T$ was the true channel.

Thus, the instantaneous transmit power is $P(s\|\mathbf{H}_T\|^2)$ (computed using (5)) and μ is given by $|\mathbf{H}W_T|^2 = |\mathbf{H}\mathbf{H}_T^\dagger / \|\mathbf{H}_T\|\|^2$. The SNR is $\gamma = \frac{1}{\sigma^2}|\frac{\mathbf{H}\mathbf{H}_T^\dagger}{\|\mathbf{H}_T\|}|^2 = s\mu$ and the BER for a given μ is

$$P_{b\,|\,\mu, \|\mathbf{H}_T\|^2} = Q\left(\sqrt{2s\mu P(s\|\mathbf{H}_T\|^2)}\right). \qquad (23)$$

From (10) we obtain the cutoff

$$\frac{\gamma_0}{s} \le k e^{-(n_t - 1)s + \delta}.$$

Let $\mu = X^2 + Y^2$, where X is the real part of $\mathbf{H}\mathbf{H}_T^\dagger / \|\mathbf{H}_T\|$ and Y is the imaginary part of $\mathbf{H}\mathbf{H}_T^\dagger / \|\mathbf{H}_T\|$, i.e., $X = \rho_T \|\mathbf{H}_T\| + \mathrm{Re}(\sqrt{1 - \rho_T^2}\, \mathbf{E}\mathbf{H}_T^\dagger / \|\mathbf{H}_T\|)$ and $Y = \mathrm{Im}(\sqrt{1 - \rho_T^2}\, \mathbf{E}\mathbf{H}_T^\dagger / \|\mathbf{H}_T\|)$. The random variables X and Y are real Gaussian variables (given $\mathbf{H}_T$) and let their means be m_X and m_Y respectively. Clearly, $m_X^2 + m_Y^2 = \|\mathbf{H}_T\|^2 \rho_T^2$ and the variance of X and Y is $\frac{1}{2}\sigma_e^2 = \frac{1}{2}(1 - \rho_T^2)$. The conditional pdf $f_{\mu\,|\,\|\mathbf{H}_T\|^2}(\lambda\,|\,\alpha)$ is *noncentral chi-square* distribution

$$f_{\mu\,|\,\|\hat{h}\|^2}(\lambda\,|\,\alpha) = \frac{1}{\sigma_e^2} e^{-(\alpha\rho_T^2 + \lambda)/\sigma_e^2} \sum_{k=0}^{\infty} \frac{\left(\frac{-\lambda\alpha\rho_T^2}{\sigma_e^4}\right)^k}{k!\Gamma(k+1)}.$$

The average BER is given by

$$P_b = 0.5 F_{\|\mathbf{H}_T\|^2}(\gamma_0/s) + \int_{\alpha=\gamma_0/s}^{\infty}\int_{\lambda=0}^{\infty} \qquad (24)$$
$$\times Q(\sqrt{2s\lambda P(s\alpha)})\, f_{\mu\,|\,\|\mathbf{H}_T\|^2}(\lambda\,|\,\alpha)d\lambda\, f_{\|\mathbf{H}_T\|^2}(\alpha)d\alpha$$
$$\le 0.5 F_{\|\mathbf{H}_T\|^2}(\gamma_0/s) + \int_{\alpha=\gamma_0/s}^{\infty}\int_{\lambda=0}^{\infty} \frac{1}{2}$$
$$\times e^{-s\lambda P(s\alpha)} f_{\mu\,|\,\|\mathbf{H}_T\|^2}(\lambda\,|\,\alpha)d\lambda\, f_{\|\mathbf{H}_T\|^2}(\alpha)d\alpha$$
$$= 0.5\left[1 - \sum_{i=0}^{n_t-1} \frac{e^{-\gamma_0/s}(\gamma_0/s)^i}{i!}\right]$$
$$+ 0.5 \int_{\alpha=\gamma_0/s}^{\infty} \frac{e^{-\alpha}\alpha^{n_t-1}}{(n_t-1)!}$$
$$\times \left[\int_{\lambda=0}^{\infty} e^{-\frac{\lambda}{\alpha}\ln\left(\frac{\alpha}{\gamma_0/s}\right)} \frac{1}{\sigma_e^2} e^{-(\alpha\rho_T^2 + \lambda)/\sigma_e^2}\right.$$
$$\left.\times \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{\lambda\alpha\rho_T^2}{\sigma_e^4}\right)^k}{k!\Gamma(k+1)} d\lambda\right] d\alpha$$
$$= 0.5\left[1 - \sum_{i=0}^{n_t-1} \frac{e^{-\gamma_0/s}(\gamma_0/s)^i}{i!}\right]$$
$$+ 0.5\frac{1}{\sigma_e^2}\int_{\alpha=\gamma_0/s}^{\infty} \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{\alpha\rho_T^2}{\sigma_e^4}\right)^k}{k!\Gamma(k+1)} e^{-\alpha\rho_T^2\sigma_e^2}$$
$$\times \left[\frac{k!}{\left(\frac{\ln\left(\frac{\alpha}{\gamma_0/s}\right)}{\alpha} + \frac{1}{\sigma_e^2}\right)^{k+1}}\right] \frac{e^{-\alpha}\alpha^{n_t-1}}{(n_t-1)!} d\alpha$$
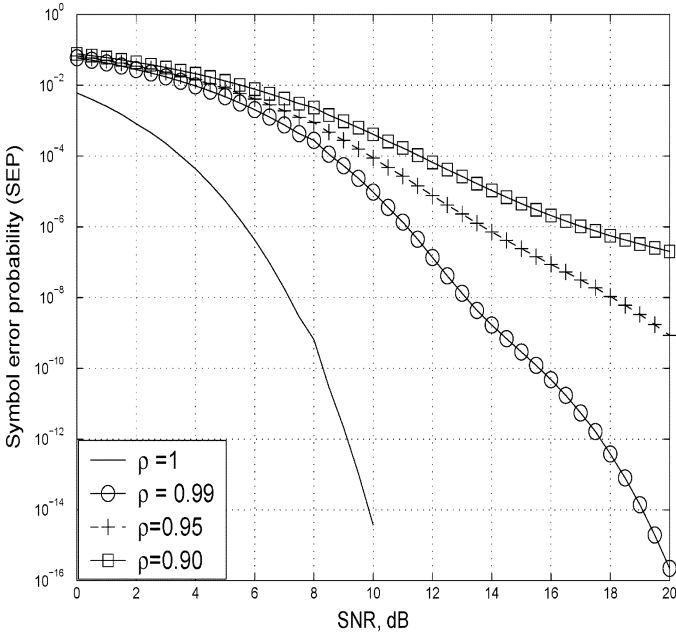
Fig. 6. BER versus average SNR for different values of ρ of MISO $(4, 1)$ system.

Fig. 7. BER versus average SNR for different values of ρ of MIMO $(4, 2)$ system.

$$= 0.5 \left[1 - \sum_{i=0}^{n_t-1} \frac{e^{-\gamma_0/s}(\gamma_0/s)^i}{i!} \right]$$

$$+ 0.5 \frac{(\frac{\gamma_0}{s})^{n_t+1}}{(n_t-1)!} \int_{x=1}^{\infty} \left(\frac{e^{-(1+\frac{\rho_T^2}{\sigma_e^2})\frac{\gamma_0}{s}x}}{\frac{\sigma_e^2 \ln(x)}{x} + \gamma_0/s} \right)$$

$$\times \exp \left(\frac{-\left(\frac{\gamma_0}{s}\right)^2 \frac{\rho_T^2}{\sigma_e^2}x}{\frac{\gamma_0}{s} + \sigma_e^2 \frac{\ln(x)}{x}} \right) x^{n_t-1} dx$$

$$\leq 0.5 \left[1 - \sum_{i=0}^{n_t-1} \frac{e^{-\gamma_0/s}(\gamma_0/s)^i}{i!} \right]$$

$$+ 0.5 \frac{(\frac{\gamma_0}{s})^{n_t+1}}{(n_t-1)!} \int_{x=1}^{\infty} e^{-\alpha' x} \frac{1}{\gamma_0/s} x^{n_t-1} dx$$

$$= 0.5 \left[1 - e^{-\gamma_0/s} \sum_{i=0}^{n_t-1} \frac{(\gamma_0/s)^i}{i!} \right]$$

$$+ 0.5 \left(\frac{\frac{\gamma_0}{s}}{\alpha'} \right)^{n_t} e^{-\alpha'} \sum_{j=0}^{n_t-1} \frac{\alpha'^j}{j!}$$

$$\leq 0.5 \left[1 - e^{-\gamma_0/s} \right] + 0.5 \left(\frac{\frac{\gamma_0}{s}}{\alpha'} \right)^{n_t} e^{-\alpha'} e^{\alpha'}$$

$$\leq 0.5 \left(\frac{\gamma_0}{s} \right) \left[1 + \frac{(\frac{\gamma_0}{s})^{n_t-1}}{(\Sigma \alpha' n_t)} \right]$$

$$\leq 0.5 k e^{-(n_t-1)s+\delta} \left[1 + k^{n_t-1} \frac{e^{-(n_t-1)^2 s + (n_t-1)\delta}}{\alpha'^{n_t}} \right] \quad (25)$$

where α' is given in (22).

The BER performance of $(4, 1)$ system is shown in Fig. 6. Here also, we observe the exponential fall of BER at low SNR. But, at high SNR the exponential decay is lost. This behavior can be explained from (25) in the same way as is done in Fig. 5.
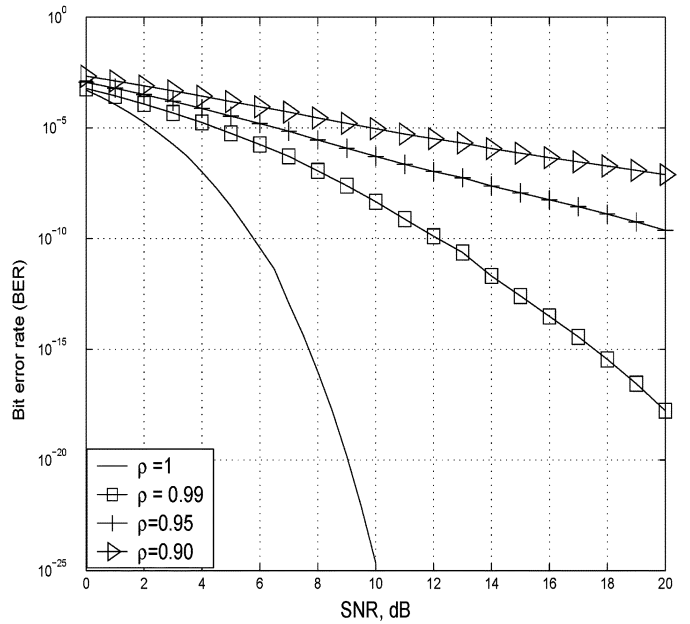
C. MIMO

As in the perfect CSIT case, the transmitter does beamforming by forming a transmit weight vector, W_T which is the eigenvector corresponding to the largest eigenvalue of $\mathbf{H}_T^{\dagger} \mathbf{H}_T$. Thus, the signal model is

$$Y = \sqrt{P(s\|\mathbf{H}_T W_T\|^2)} \mathbf{H} W_T x + N$$

and the power allocation policy is given by $P(s\|\mathbf{H}_T W_T\|^2)$ (see (5)). We define μ as $\|\mathbf{H} W_T\|^2$. The average BER is

$$P_b = \int_{\mu=0}^{\infty} \int_{\lambda'=0}^{\infty} Q\left(\sqrt{2s\mu P(s\lambda')} \right) f_{\mu,\lambda'}(\mu, \lambda') \, d\mu \, d\lambda'. \quad (26)$$

As the joint density function $f_{\mu,\lambda'}(\mu, \lambda')$ is not analytically tractable, we simulate the system for $n_t = 4$ and $n_r = 2$ for different values of ρ_T. The BER performance results are provided in Fig. 7.

The exponential fall in low SNR regime is observed in this case also. Similarly, one can observe that at high SNR, the imperfect CSIT causes the BER to fall polynomially with SNR. Comparing Figs. 6 and 7, we can see that as n_r increases from 1 to 2, there is a substantial increase in diversity gain. For $\rho_T = 1$, there is a gain in SNR of 3 dB for achieving a BER of 10^{-5} by the $(4, 2)$ system over the $(4, 1)$ system. The diversity gain diminishes as ρ_T decreases. For $\rho_T = 0.90$, we can see that the BER performance of $(4, 1)$ and $(4, 2)$ systems is almost the same.

IV. BEAMFORMING SYSTEM WITH IMPERFECT CSIT AND CSIR

The system model considered here is the same as in Section III except that the channel estimate at the receiver, $\mathbf{H}_R$ is not perfect. There are two possible scenarios that may

arise, viz. the receiver may or may not know the CSIT (H_T and W_T) which are necessary for estimation of the transmitted bit. The penalty for not knowing the CSIT is different in MIMO and MISO cases as we shall show. We consider MISO and MIMO systems with imperfect CSIT and imperfect CSIR in the following (SISO case is just a special case of MISO).

A. MISO

We consider a system with $n_t \geq 1$ transmit antennas and $n_r = 1$ receive antenna. The optimal beamforming vector is given by $W_T = \frac{\mathbf{H}_T^{\dagger}}{\|\mathbf{H}_T\|}$ and the SNR estimate at the transmitter, γ_T is $s\|\mathbf{H}_T\|^2$.

1) CSIT Not Known at the Receiver: We first consider the case when the receiver does not know CSIT. The decoding will be done as per (2) (with $\mathbf{H}_R$ and W_R used instead of $\mathbf{H}$ and W). It is to be noted that only the sign of the real part of $\frac{(\mathbf{H}_R W_R)^{\dagger} Y}{\sqrt{P(\gamma_R)}\|\mathbf{H}_R W_R\|^2}$ is of importance and not its magnitude. Also from (2) we can say that the imaginary part of the above quantity does not influence the decision, since the magnitude of the imaginary part of $\hat{x}$ does not depend on x. Furthermore, $\frac{(\mathbf{H}_R W_R)^{\dagger}}{\sqrt{P(\gamma_R)}\|\mathbf{H}_R W_R\|^2}$ will always be a positive number and so will have no role in the decoding process. Hence the channel state information at the receiver is not important for decoding nor does it influence the performance in single receiver antenna systems. When the CSIT is perfect, the analysis of Section II holds, but when the CSIT is imperfect, the analysis of Section III does not hold, because there it was assumed that the CSIT as well as perfect CSIR was available at the receiver. When the CSIT is imperfect and not available at the receiver, we analyze the performance here.

Based on the above discussion, we know that the sign of $Re(Y)$ is the decoding criterion. Thus the probability of error is

$$P_b = P[\mathrm{Re}(Y) < 0 | x = 1] P(x = 1)$$
$$+ P[\mathrm{Re}(Y) > 0 | x = -1] P(x = -1).$$

By symmetry

$$P_b = P[\mathrm{Re}(Y) < 0 | x = 1]$$
$$= P[\mathrm{Re}(\sqrt{P(\gamma_T)}\mathbf{H} W_T + N) < 0].$$

As per the model assumed for imperfect CSIT in Section III

$$\mathbf{H} = \rho_T \mathbf{H_T} + \sqrt{1 - {\rho_T}^2} \mathbf{E}$$

Hence

$$\mathbf{H} W_T = \rho_T \|\mathbf{H}_T\| + \sqrt{1 - {\rho_T}^2} \mathbf{E} W_T.$$

Since the components of $\mathbf{E}$ are independent complex Gaussian random variables, $\sqrt{1 - {\rho_T}^2} \mathbf{E} W_T = N'$ is circularly symmetric complex Gaussian given $\mathbf{H}_T$ (and hence W_T), (and has variance $= 1 - \rho_T^2$ (since $\|W_T\|^2 = 1$)). We assume that the channel noise N and N' are independent. Then

$$Y = \rho_T \|\mathbf{H}_T\| \sqrt{P(\gamma_T)} + N''$$

where N'' is complex Gaussian with variance $1/s + (1 - \rho_T^2) P(\gamma_T)$. Therefore the performance of this system
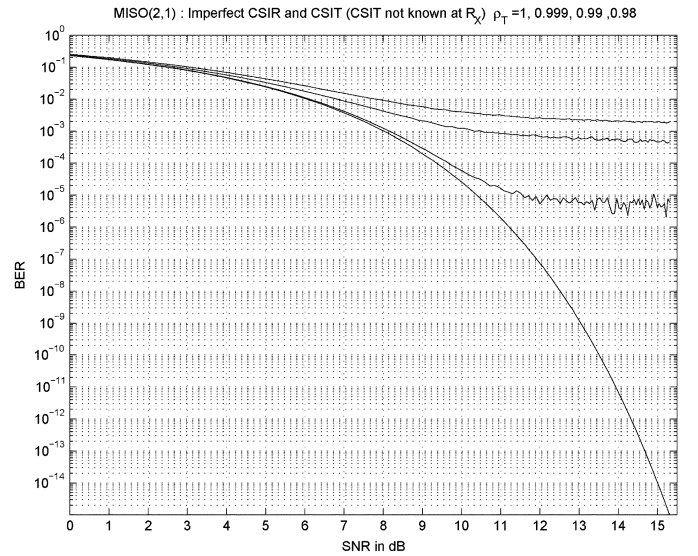

Fig. 8. Average BER versus SNR for different values of ρ_T of MISO $(2, 1)$ system, where the receiver does not know the CSIT.

is the same as that of an AWGN channel with SNR $= \frac{\rho_T^2 s \|\mathbf{H}_T\|^2 P(\gamma_T)}{1 + s P(\gamma_T)(1 - \rho_T^2)}$. Thus the average BER is

$$E_{\gamma_T} \left[Q \left(\sqrt{\frac{2\rho_T^2 \gamma_T P(\gamma_T)}{1 + s P(\gamma_T)(1 - \rho_T^2)}} \right) \right]. \quad (27)$$

The distribution of γ_T is given by the $\mathrm{Erlang}_{n_t}(1/s)$ distribution. We have numerically computed the above expression and plotted in Fig. 8. We observe that at smaller values of SNR, the exponential diversity is maintained but at large values of SNR, the BER is more or less constant. This can be predicted by having a closer look at (27). As $s \to 0$, (for $\ln \gamma_0 < 0$, which happens for $\mathrm{SNR} > 2$ dB for $n_t = 2, n_r = 1$) $s P(\gamma_T)(1 - \rho_T^2) \to 0$. Thus for small s, the term in the square root in (27) becomes approximately $2\rho_T^2 \gamma_T P(\gamma_T)$ and hence by the results of Section II, we get exponential diversity. As $s \to \infty$, $s P(\gamma_T)(1 - \rho_T^2) \to \infty$ and hence the term in the square root tends to $\frac{2\rho_T^2 \|\mathbf{H}_T\|^2}{(1 - \rho_T^2)}$. Thus for large s, the BER becomes approximately constant. For the intermediate values of channel SNR, we get diversity order changing from exponential to linear to sublinear to constant. The important parameter here is $s P(\gamma_T)(1 - \rho_T^2)$ the magnitude of which will decide the diversity order. An appropriate system can be designed such that at the working SNR levels, we can guarantee exponential diversity, but the price will be a higher value of ρ_T, that is a better feedback and estimation. Also if we need to operate in higher regions of SNR, we should have even higher values of ρ_T, while if we desire to operate at lower SNR levels, we can work with lower values of ρ_T.

2) CSIT Known at the Receiver: Now we consider the case when the CSIT ($\mathbf{H}_T$ and hence W_T) is known at the receiver. Suppose we decide to use the $\mathbf{H}_T$ for decoding in (2), we shall get exactly the same BER as in Section IV.A, because for BPSK, it is only the sign of the real part of $\frac{(\mathbf{H}_R W_R)^{\dagger} Y}{\sqrt{P(\gamma_R)}\|\mathbf{H}_R W_R\|^2}$ that matters and by using $\mathbf{H}_T$ and W_T instead of $\mathbf{H}_R$ and W_R, we do

not change the decision in anyway. But suppose we have a scenario where the receiver has a better channel estimate than the transmitter and also knows the CSIT. There arises a question as to which of $\mathbf{H}_R$ and $\mathbf{H}_T$ to use for decoding. We shall address this question here.

It is clear that we have to use W_T and not W_R for decoding in which case we get the same performance as in the previous section if we also use $\mathbf{H}_T$. Now we will use $\mathbf{H}_R$ and W_T and compare the BER with that of the previous section. The model assumed for $\mathbf{H}_R$ will be similar to the model for $\mathbf{H}_T$. Define

$$\theta = (\mathbf{H}_R W_T)^\dagger (HW_T \sqrt{P(\gamma_T)} + N).$$

The sign of the real part of θ will decide the decoding. Thus, by symmetry, the probability of error is,

$$P_b = P[\text{Re}(\theta) < 0 | x = 1].$$

Since the model for $\mathbf{H}_R$ is similar to that of $\mathbf{H}_T$, from (18)

$$\mathbf{H_R} = \rho_R \mathbf{H} + \sqrt{1 - \rho_R^2} \mathbf{E}$$

and given $x = 1$ was transmitted,

$$
\begin{aligned}
\theta = {} & \rho_R \|\mathbf{H}W_T\|^2 \sqrt{P(\gamma_T)} \\
& + \sqrt{1 - \rho_R^2} \sqrt{P(\gamma_T)} (\mathbf{E}W_T)^\dagger (\mathbf{H}W_T) \\
& + (\rho_R \mathbf{H}W_T)^\dagger N + \sqrt{(1 - \rho_R^2)} (\mathbf{E}W_T)^\dagger N.
\end{aligned}
$$

Given ($\mathbf{H}$ and $\mathbf{H}_T$), the last term in the above expression is super-Gaussian (e.g., tail of its cdf is lighter than a Gaussian distribution) and also has variance considerably less than the other Gaussian noise terms (since CSIR is better than CSIT, it is reasonable to assume that ρ_R is close to 1). Thus, it can be neglected. Hence, by arguments similar to Section IV-A, we can interpret it as an AWGN channel with noise term N' with variance

$$\rho_R^2 \frac{\|\mathbf{H}W_T\|^2}{s} + \left(1 - \rho_R^2\right) P(\gamma_T) \|\mathbf{H}W_T\|^2. \quad (28)$$

Thus, the average BER is given by

$$P_b = E_{\mathbf{H},\mathbf{H}_T} \left[Q \left(\sqrt{\frac{2sP(\gamma_T)\|\mathbf{H}W_T\|^2}{1 + sP(\gamma_T)\frac{1-\rho_R^2}{\rho_R^2}}} \right) \right] \quad (29)$$

and is plotted in Fig. 9 for $n_t = 2$ and $n_r = 1$. Again we see that the diversity order changes from exponential to sub-linear, for reasons similar to the ones mentioned in the previous section. If instead of $\mathbf{H}_R$ we use $\mathbf{H}_T$ for decoding then in (29) we will get ρ_T instead of ρ_R. Thus by examining the denominator of (29), we observe that when $\rho_R > \rho_T$ we can have a better performance if we decode using $\mathbf{H}_R$ as against using $\mathbf{H}_T$ and vice-versa.

We may conclude that for the single antenna receiver, CSIR gives an advantage only if we know the CSIT at the receiver and the CSIR is better than the CSIT. If the CSIR is exact, then knowing CSIT at the receiver provides polynomial diversity at
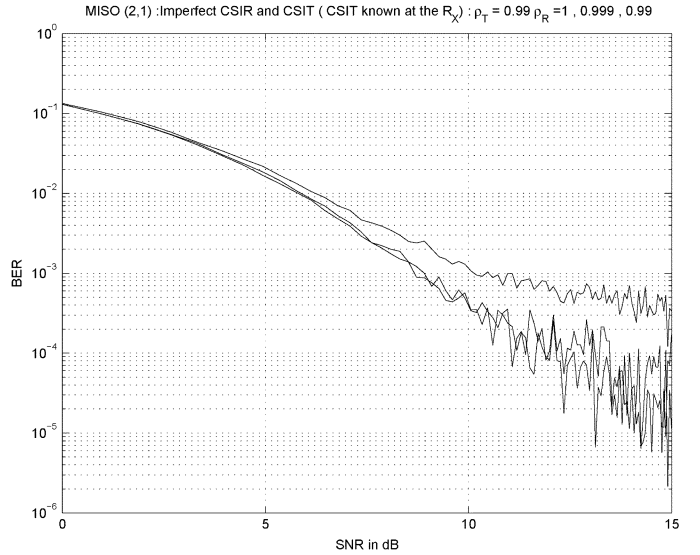

Fig. 9. Average BER versus SNR for different values of ρ_R with $\rho_T = 0.99$ for a MISO $(2, 1)$ system when the receiver knows CSIT.

high SNR as against constant (0 diversity) when CSIT is not known at the receiver.

B. MIMO

In this section, $n_t > 1$ and $n_r > 1$. The detection will be done by the rule (2), with $\mathbf{H}, W$ replaced by $\mathbf{H}_R$ and W_R

$$\arg \min_x \left| x - \frac{(\mathbf{H}_R W_R)^\dagger Y}{\sqrt{P(\gamma_R)}\|\mathbf{H}_R W_R\|^2} \right|^2 = \arg \min_x |x - \hat{x}|^2 \quad (30)$$

where W_R is the eigenvector of $\mathbf{H}_R^\dagger \mathbf{H}_R$ corresponding to its largest eigenvalue, $\lambda_R, \gamma_R = s\lambda_R$ is the SNR estimate at the receiver, and

$$\hat{x} \triangleq \frac{(\mathbf{H}_R W_R)^\dagger Y}{\sqrt{P(\gamma_R)}\|\mathbf{H}_R W_R\|^2} = r^\dagger Y \quad (31)$$

where r is the receive weight vector used for detection.

When CSIT is available at the receiver then the receiver may decide to use W_T and $\mathbf{H}_T$ in place of W_R and $\mathbf{H}_R$ in (30) and (31). In the following, we will also consider this possibility.

Since we consider the BPSK modulation scheme (x is real), it is enough to consider only the real part of (30) (the imaginary part offers no performance improvement). Thus the detection rule is

$$|1 - r^\dagger Y|^2 \underset{-1}{\overset{+1}{\lessgtr}} |1 + r^\dagger Y|^2.$$

We observe that by virtue of BPSK modulation, only the sign and not the magnitude of $Re(r^\dagger Y)$ is important. We analyze the BER performance for the cases when the CSIT is known and when the CSIT is not known at the receiver. The model assumed for CSIT and CSIR will be the same as in the previous sections.

1) CSIT Not Known at the Receiver: Since the receiver does not know CSIT, the decoding will have to be done using $\mathbf{H}_R$ and W_R. Unlike the MISO case, now the W_T is very crucial.
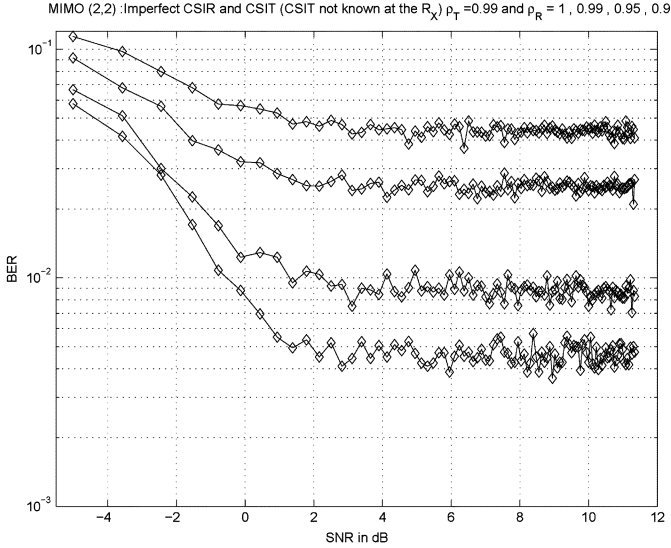
Fig. 10. Average BER versus SNR for a MIMO $(2,2)$ system with imperfect CSIT (CSIT not known at the R_X), imperfect CSIR.
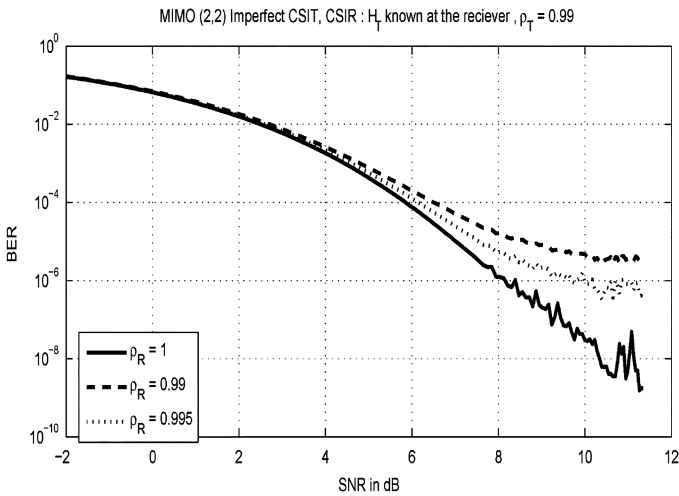

Fig. 11. BER versus average SNR for a MIMO $(2,2)$ system with imperfect CSIT (CSIT known at the R_X) and is used for decoding. CSIR quality is better than CSIT.

We simulate the system for $\rho_T = 0.99$ and $\rho_R = 0.9, 0.95, 0.99$ and 1. It is very clear from Fig. 10 (compare with Fig. 11) that even with perfect CSIR, if CSIT is not available at the receiver, a severe penalty has to be paid. The reason for this is that we do not know the beamformer exactly at the receiver. Most of the errors are due to this and not due to the channel noise. Thus when the transmit weight vector is not available at the receiver, we have to have a very good estimate of the channel at the transmitter ($\rho_T > 0.999$) as well as at the receiver ($\rho_R > 0.999$) to get a reasonable BER $\approx 10^{-4}$.

2) CSIT Known at the Receiver: Now the receiver knows the CSIT, and hence the transmit weight vector W_T. There arises a question as to which $\mathbf{H}$ ($\mathbf{H}_R$ or $\mathbf{H}_T$) should be used for decoding. We shall show that using the better estimate of $\mathbf{H}$ (higher ρ) whether it is the $\mathbf{H}_R$ or $\mathbf{H}_T$ will give a better average BER.

As said earlier, it is only the sign of $r^\dagger Y$ that will decide the bit and hence also the BER. Define $\theta = (\mathbf{H_R} W_T)^\dagger Y$. The probabilty of error (because of the symmetry), is given by

$$P_e = P[\mathrm{Re}(\theta) < 0 \,|\, x = 1]$$

if we use $\mathbf{H}_R$ and W_T for decoding. If $x = 1$ then

$$
\begin{aligned}
\theta &= (\mathbf{H}_R W_T)^\dagger (HW_T \sqrt{P(\gamma_T)} + N) \\
&= [(\rho_R \mathbf{H} + \sqrt{1 - \rho_R^2}\mathbf{E})W_T]^\dagger (HW_T \sqrt{P(\gamma_T)} + N) \\
&= \rho_R \|HW_T\|^2 \sqrt{P(\gamma_T)} + \rho_R (HW_T)^\dagger N \\
&\quad + \sqrt{1 - \rho_R^2}\sqrt{P(\gamma_T)}(\mathbf{E}W_T)^\dagger (HW_T) \\
&\quad + \sqrt{1 - \rho_R^2}(\mathbf{E}W_T)^\dagger N.
\end{aligned}
$$

Conditioned on $\mathbf{H}$ and $\mathbf{H}_T$, the final term is super-gaussian and can be neglected with respect to the other two Gaussian noise random variables since it has significantly smaller variance (especially when ρ_R is close to 1) at high SNR, the case of interest to us. As in Section III, the above System can be interpreted as an AWGN channel with output

$$\theta = \rho_R \|\mathbf{H}W_T\|^2 \sqrt{P(\gamma_T)} + N'$$

where N' is circularly symmetric complex Gaussian with variance

$$\sigma^2 = \rho_R^2 \frac{\|\mathbf{H}W_T\|^2}{s} + \left(1 - \rho_R^2\right) P(\gamma_T)\|\mathbf{H}W_T\|^2.$$

Therefore, the average BER is given by

$$E_{\mathbf{H}, \mathbf{H}_T}\left[Q\left(\sqrt{\frac{2s\|\mathbf{H}W_T\|^2 P(\gamma_T)}{1 + sP(\gamma_T)\frac{1 - \rho_R^2}{\rho_R^2}}} \right) \right].$$

If we use $\mathbf{H}_T$ instead of $\mathbf{H}_R$, we arrive at the same expression with ρ_T in place of ρ_R. Thus the choice of using $\mathbf{H}_R$ or $\mathbf{H}_T$ for decoding is to be decided by which of them is better. Similar to the arguments in Section IV-A, we can explain the diversity order going from exponential to constant. Fig. 11 has the simulation results. Comparing this with Fig. 10 shows the advantage of knowing CSIT at the receiver.

From the results of the above two sections, we observe that the knowledge of CSIT at the receiver is very important (especially for MIMO systems). The quality of CSIR becomes important only if it is better than the quality of CSIT available. However in the scenario where CSIT is not available at the receiver, one should use CSIR for MIMO systems although we have observed that for MISO, it is not useful at all. These results are for BPSK modulation. The conclusion may differ for other modulation schemes.

V. OSTBCS WITH IMPERFECT CSIT

In this section, we define the system model of a space-time coded system and analyze its BER performance with our power allocation policy defined in (5). We follow the approach given in

[11] and we borrow the results given in [11] for the sake of completeness. Let $\mathcal{C} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_K\}$ denote the set of codewords, where K is the number of codewords. We assume that the codewords are of length T, i.e., each codeword is described by a $n_t \times T$ matrix. Actually, the codeword $\mathbf{C}_k$ is generated by a Space-Time code, $\overline{\mathbf{C}}_k$ followed by a precoder matrix, $\mathbf{W}_T$, i.e., $\mathbf{C}_k = \mathbf{W}_T \overline{\mathbf{C}}_k$ ($\mathbf{W}_T$ is chosen such that it optimizes certain performance metric). The channel between the jth transmit antenna and the ith receive antenna, h_{ij} is modeled as a zero mean complex Gaussian random variable. Thus, the $n_r \times n_t$ matrix, $\mathbf{H}$ represents the MIMO channel. We assume a *block fading* channel where T successive realizations of the channel corresponding to one codeword are same and the channel varies independently from one codeword to another. Let $\mathbf{C} \in \mathcal{C}$ be the transmitted codeword. The received complex signal vectors corresponding to one codeword may then be arranged in an $n_r \times T$ matrix $\mathbf{Y}$, given by

$$\mathbf{Y} = \mathbf{H}\mathbf{W}_T\overline{\mathbf{C}} + \mathbf{N},$$

where $\mathbf{N} \in \mathbb{C}^{n_r \times T}$ represents the AWGN over T symbol durations at the n_r receive antennas. The ML decoding of $\mathbf{C}$ given $\mathbf{Y}$ corresponds to

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}\in\mathcal{C}} \|\mathbf{Y} - \mathbf{H}\mathbf{C}\|_F^2$$

where $\|\cdot\|_F$ represents the Frobenius norm.

Let $h = \text{vec}(\mathbf{H}^\dagger)$.[2] Let $\mathbf{H}_T$ be the channel estimate available at the transmitter. Also, let $\hat{h} = \text{vec}(\mathbf{H}_T) = [\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_{n_r n_t}]^T$. Suppose, $\mathbf{C}_k$ is the transmitted codeword. Then, $P(\mathbf{C}_k \to \mathbf{C}_l \mid \hat{h})$, the probability of wrongly decoding C_k as C_l given the side information $\hat{h}$ is called the pairwise error probability (PEP) between $\mathbf{C}_k$ and $\mathbf{C}_l$. The overall design goal is to find $\mathbf{W}_T$ which minimizes PEP with respect to the set of all codewords. This is equivalent to finding the optimum $\mathbf{W}_T$ which minimizes PEP for the worst-case pair $\mathbf{C}_k, \mathbf{C}_l$.

We start by conditioning on the true channel realization h and utilize a well-known upper bound on the Gaussian tail function to arrive at

$$P(\mathbf{C}_k \to \mathbf{C}_l \mid h, \hat{h}) \leq \frac{1}{2} e^{-d^2(\mathbf{C}_k, \mathbf{C}_l)/4\sigma^2} \tag{32}$$

where $d(\mathbf{C}_k, \mathbf{C}_l)$ is the Euclidean distance between the codewords $\mathbf{C}_k$ and $\mathbf{C}_l$

$$\begin{aligned}
d^2(\mathbf{C}_k, \mathbf{C}_l) &= \|\mathbf{H}(\mathbf{C}_k - \mathbf{C}_l)\|_F^2 \\
&= \text{tr}(\mathbf{H}(\mathbf{C}_k - \mathbf{C}_l)(\mathbf{C}_k - \mathbf{C}_l)^\dagger \mathbf{H}^\dagger) \\
&= \text{tr}(\mathbf{H}\, A(\mathbf{C}_k, \mathbf{C}_l)\, \mathbf{H}^\dagger) \\
&= \left(\text{vec}(A(\mathbf{C}_k, \mathbf{C}_l)\, \mathbf{H}^\dagger)\right)^\dagger \text{vec}(\mathbf{H}^\dagger) \\
&= h^\dagger (I_{n_r} \otimes A(\mathbf{C}_k, \mathbf{C}_l))\, h
\end{aligned}$$

where $\otimes$ represents the Kronecker product and $A(\mathbf{C}_k, \mathbf{C}_l) = (\mathbf{C}_k - \mathbf{C}_l)(\mathbf{C}_k - \mathbf{C}_l)^\dagger$.

[2]$\text{vec}(\cdot)$ is the vectorization operator which stacks the columns of its argument into a vector.

Since the true channel, h and the estimate $\hat{h}$ are jointly complex Gaussian, the pdf of h conditioned on $\hat{h}$ is given by

$$p_{h\mid\hat{h}}(h\mid\hat{h}) = \frac{e^{-(h-m_{h\mid\hat{h}})^\dagger R_{hh\mid\hat{h}}^{-1}(h-m_{h\mid\hat{h}})}}{\pi^{n_t n_r}\det(R_{hh\mid\hat{h}})}$$

where $m_{h\mid\hat{h}}$ and $R_{hh\mid\hat{h}}$ represent the conditional mean and covariance, respectively. Then from (32)

$$P(\mathbf{C}_k \to \mathbf{C}_l \mid \hat{h}) \leq \int \frac{1}{2} e^{-d^2(\mathbf{C}_k, \mathbf{C}_l)/4\sigma^2}\, p_{h\mid\hat{h}}(h\mid\hat{h})\, dh.$$

Let $\Psi(\mathbf{C}_k, \mathbf{C}_l) = (I_{n_r} \otimes A(\mathbf{C}_k, \mathbf{C}_l))/4\sigma^2 + R_{hh\mid\hat{h}}^{-1}$. The exponent of the above integral is

$$\begin{aligned}
&\frac{-h^\dagger(I_{n_r} \otimes A(\mathbf{C}_k, \mathbf{C}_l))\,h}{4\sigma^2} - (h - m_{h\mid\hat{h}})^\dagger R_{hh\mid\hat{h}}^{-1}(h - m_{h\mid\hat{h}}) \\
&= -h^\dagger\left(\Psi(\mathbf{C}_k, \mathbf{C}_l) - R_{hh\mid\hat{h}}^{-1}\right)h \\
&\quad - (h - m_{h\mid\hat{h}})^\dagger R_{hh\mid\hat{h}}^{-1}(h - m_{h\mid\hat{h}}) \\
&= -\left(h - \Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}\right)^\dagger \Psi(h - \Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}) \\
&\quad + m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}(\Psi^{-1} - R_{hh\mid\hat{h}})R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}.
\end{aligned}$$

Thus, the PEP between $\mathbf{C}_k$ and $\mathbf{C}_l$ is bounded by

$$\begin{aligned}
&P(\mathbf{C}_k \to \mathbf{C}_l \mid \hat{h}) \\
&\leq \frac{1}{2}\int \frac{e^{-(h-\Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}})^*\Psi(h-\Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}})}}{\pi^{n_t n_r}\det(R_{hh\mid\hat{h}})} \\
&\quad \times e^{m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}(\Psi^{-1}-R_{hh\mid\hat{h}})R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}\, dh \\
&= \frac{1}{2}\int \frac{e^{-(h-\Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}})^*\Psi(h-\Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}})}}{\pi^{n_t n_r}\det(R_{hh\mid\hat{h}})\det(\Psi^{-1})\det(\Psi)} \\
&\quad \times e^{m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}(\Psi^{-1}-R_{hh\mid\hat{h}})R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}\, dh \\
&= \frac{1}{2}\frac{e^{m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}(\Psi^{-1}-R_{hh\mid\hat{h}})R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}}{\det(R_{hh\mid\hat{h}})\det(\Psi)}.
\end{aligned}$$

From the above equation, we see that the objective function to minimize the PEP is given by

$$\begin{aligned}
V(\mathbf{C}_k, \mathbf{C}_l) &= \frac{e^{m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}(\Psi^{-1}-R_{hh\mid\hat{h}})R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}}{\det(R_{hh\mid\hat{h}})\det(\Psi)} \\
&= \left\{\frac{e^{m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}\Psi^{-1}R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}}{\det(\Psi)}\right\} \\
&\quad \times \frac{e^{-m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}}{\det(R_{hh\mid\hat{h}})} \\
&\triangleq \left\{e^{l(\mathbf{C}_k, \mathbf{C}_l)}\right\}\frac{e^{-m_{h\mid\hat{h}}^* R_{hh\mid\hat{h}}^{-1}m_{h\mid\hat{h}}}}{\det(\Psi)\det(R_{hh\mid\hat{h}})}.
\end{aligned}$$

Hence, we need to minimize $l(\mathbf{C}_k, \mathbf{C}_l)$ for the worst case codeword pairs $(\mathbf{C}_k, \mathbf{C}_l)$. This optimization problem is solved by Jongren [11]. Since, [11] solves the above optimization problem

with $\text{trace}(\mathbf{W}\mathbf{W}^\dagger) = 1$ as a constraint ($\mathbf{W}$ is the precoder matrix), the solution of this problem will provide the precoder matrix $\mathbf{W}$, which will allocate the power optimally in space. The second term in the above equation is a constant and does not depend on the power allocation policy. We combine our power allocation policy given by (5) with that of Jongren's policy to get an optimal power allocation in space and time for Orthogonal Space–Time Block Codes (OSTBCs).

We now consider the Alamouti code and find the optimal precoder matrix W_T for it. For the Alamouti code

$$\bar{\mathbf{C}} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}.$$

The optimal $\mathbf{W}_T$ (for SOPA) as obtained in [11], is

$$\mathbf{W}_{T,\,\text{SOPA}} = \frac{1}{\sqrt{|\hat{h}_1|^2 + |\hat{h}_2|^2}} \begin{bmatrix} -\hat{h}_2^* & \hat{h}_1 \\ \hat{h}_1^* & \hat{h}_2 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}$$

where λ_1 and λ_2 are obtained as follows. Let $\alpha = 1 - \rho_T^2, \eta = \frac{1}{\rho^2}$ $\kappa = \alpha(2 + \alpha\eta)$ and

$$\mu = \frac{\eta(3\kappa + |\rho|^2\|\hat{h}\|^2 + \sqrt{6\kappa|\rho|^2\|\hat{h}\|^2 + |\rho|^4\|\hat{h}\|^4 + \kappa^2})}{2(2 + \alpha\eta)^2}.$$

Also let $\lambda = \frac{1}{\mu} - \frac{1}{\alpha\eta}$. If $\lambda > 0$ then set $\lambda_1 = \lambda$ and $\lambda_2 = 1 - \lambda$. If $\lambda \le 0$ then set $\lambda_1 = 0$ and $\lambda_2 = 1$.

The optimum $\mathbf{W}_T$ for Space–Time power allocation is given by

$$\mathbf{W}_{T,\text{STPA}} = \frac{\sqrt{P\left(s(|\hat{h}_1|^2 + |\hat{h}_2|^2)\right)}}{\sqrt{|\hat{h}_1|^2 + |\hat{h}_2|^2}} \begin{bmatrix} -\hat{h}_2^* & \hat{h}_1 \\ \hat{h}_1^* & \hat{h}_2 \end{bmatrix}$$
$$\times \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}.$$

Fig. 12 shows the BER performance of a coded system with an outer rate-$1/2$ *convolutional code* and an inner *Alamouti code* for $n_t = 2$ and $n_r = 1$. We observe that for the perfect CSIT case, our power allocation policy provides *exponential order diversity* gain which is substantially more than the conventional *space-only* (SOPA) and *uniform* power allocation schemes. Also, when the quality of CSIT degrades ($\rho \ne 1$), the exponential diversity is lost at high SNR for our policy. Although, as for the uncoded system, we still achieve exponential diversity at low SNR (up to 9 dB in the systems we studied). In Fig. 12, we note that at $\rho = 1$ and $P_e = 10^{-3}$ the gain in STPA compared to SOPA is 2.5 dB and at $\rho = 0.99$ the gain is 3.2 dB. As, ρ can be as large as 0.99 for practical systems (the channel estimation in GSM is done by a 26-bit midamble sequence; a ML estimate of H from this sequence can provide ρ as large as 0.99 for an SNR of 10 dB) our power allocation policy provides significant improvement in BER performance over the existing policies.
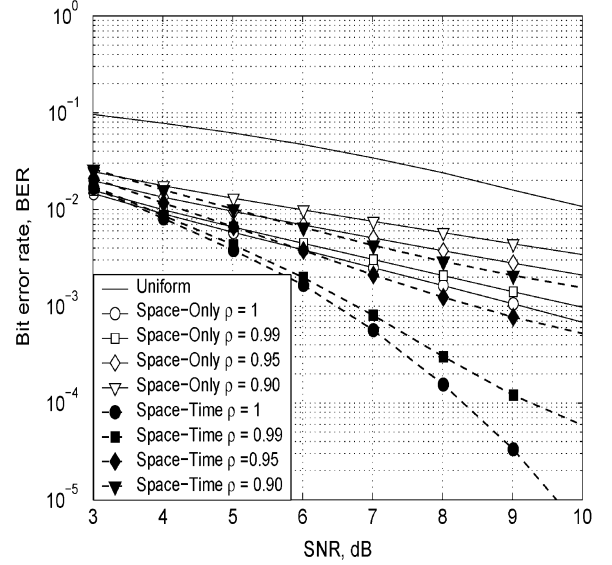

Fig. 12. BER versus average SNR for an Alamouti and convolutionally coded $(G(D) = \{1 + D^2, 1 + D + D^2\})$ $(2, 1)$ system with Imperfect CSIT.

VI. BLOCK CODED SYSTEM: ERROR EXPONENTS

In this section we provide error exponents for our power control policy (5). Even though it is not optimized for error exponents, one expects good performance. By studying the error exponents we can obtain results independently of the modulation and coding scheme. Also, this will give a Tradeoff between the rate and diversity-order. We will show that the diversity-order for each of the SISO, MISO and MIMO systems is exponential in most of the rate region, but the exact order does depend upon the rate as well as on n_t and n_r. Furthermore the exponential diversity is obtained in low rate region (region 1) and high rate (region 3) but not in the middle (region 2). This is rather counter intuitive. Also exponentiality is lost in region 1 (but not in region 3) if exact CSIT is not available.

In the following, we first study the systems with exact CSIT and CSIR and then with inexact CSIT. If the block error probability decays as $\exp(-\alpha N s^\beta)$, (where N is the block length) then we will call it exponential diversity of order (β, α).

A. Perfect CSIT and CSIR

In this section, we assume perfect CSIT and CSIR. For simplicity we first consider the SISO system, i.e., $n_t = n_r = 1$. For a discrete input, memoryless channel, by Gallager [7] (see also [24]) for rate R and block length N, the ensemble average bound on block error probability P_E is given by

$$P_E < e^{-NE(R)} \tag{33}$$

where

$$E(R) = \max_q \max_{0 \le \rho \le 1} [E_0(\rho, q) - \rho R],$$
$$E_0(\rho, q) = -\ln \int_y [\Sigma_x (q(x)p(y/x))^{1/(1+\rho)}]^{1+\rho} dy \tag{34}$$

q is a distribution on the input alphabet and $p(y/x)$ is the channel transition function. For small R, bound (33) is not tight. In that region, the expurgated average bound is tighter and is given as

$$P_E < e^{-NE_{\text{ex}}(R)} \qquad (35)$$

where

$$E_{\text{ex}}(R) = \max_q \sup_{\rho \geq 1}[E_x(\rho, q) - \rho R] \qquad (36)$$

and

$$E_x(\rho, q) = -\rho \ln \left[\Sigma_x \Sigma_{x'} q(x) q(x') \right.$$
$$\left. \times \left[\int_y \sqrt{p(y/x) p(y/x')} \right]^{1/\rho} \right].$$

For binary input channels to which we will limit ourselves in the rest of the paper, the maximizing input distribution q in (33) and (35) is $(\frac{1}{2}, \frac{1}{2})$.

The overall rate region from 0 to capacity can be divided into three regions. In region 1 (low rate region) defined by $0 \leq R \leq \partial E_x(\rho, q)/\partial \rho \,|_{\rho=1}$ for $q = (\frac{1}{2}, \frac{1}{2})$, the expurgated bound (35) provides the tighter bound. In region 2, defined by

$$\left. \frac{\partial E_x(\rho, q)}{\partial \rho} \right|_{\rho=1} \leq R \leq \left. \frac{\partial E_0(\rho, q)}{\partial \rho} \right|_{\rho=1}$$

with $q = (\frac{1}{2}, \frac{1}{2})$, the two bounds coincide and the high rate region (region 3), which is defined by $\frac{\partial E_0(\rho, q)}{\partial \rho}|_{\rho=1} \leq R \leq C$, the bound (33) is tighter.

For the wireless fading channel with fade γ and transmit power $P(\gamma)$, in the bound (33) $E_0(\rho, q)$ becomes

$$E_0(\rho, q)$$
$$= -\ln \left[\int_\gamma \int_y \Sigma_x (q(x)(p(y/x, \gamma))^{\frac{1}{1+\rho}} f(\gamma) d\gamma dy \right]^{1+\rho} \qquad (37)$$

and in the bound (35), we use

$$E_x(\rho, q) = -\rho \ln \left[\Sigma_x \Sigma_{x'} (q(x) q(x') \right.$$
$$\left. \times \left[\int_\gamma \int_y \sqrt{p(y/x, \gamma) p(y/x', \gamma)} f(\gamma) d\gamma dy \right]^{\frac{1}{\rho}} \right]. \qquad (38)$$

In the rest of the section, we explicitly study these bounds and error exponents as functions of rate R and SNR. This will provide the diversity versus multiplexing tradeoff. First we study the SISO system, then the MISO (SIMO will have identical results) system and finally the general MIMO system. The partial CSIT case will be studied in Section VI-B.

1) SISO System: In this section we study the bounds (33) and (35) when $n_t = n_r = 1$. We will consider the three rate regions mentioned above separately.

In region 1, we consider the expurgated bound. For binary input, symmetric output channels, this bound is given by the following set of equations (see, e.g., [24])

$$E_{\text{ex}}(R) = -\delta(R) \ln Z, \quad R = \ln 2 - \mathcal{H}(\delta(R)) \qquad (39)$$

where

$$\mathcal{H}(x) = -x \ln x - (1-x) \ln(1-x),$$
$$Z = \int_0^\infty d\gamma \int dy \sqrt{(p(y/x_1, \gamma) p(y/x_2, \gamma))} f(\gamma)$$
$$= \int_0^\infty \exp(-\gamma P(\gamma)) f(\gamma) d\gamma \qquad (40)$$

$P(\gamma)$ is given by (5) and γ is exponentially distributed with mean s. The RHS of (40) equals

$$\int_0^{\frac{\gamma_0}{s}} \frac{\exp(-\frac{\gamma}{s})}{s} d\gamma + \int_{\frac{\gamma_0}{s}}^\infty \exp\left(-\ln\left(\frac{\gamma}{\gamma_0}\right)\right) \exp\left(-\frac{\gamma}{s}\right) \frac{1}{s} d\gamma$$
$$= \left(1 - \exp\left(-\frac{\gamma_0}{s}\right)\right) + \frac{\gamma_0}{s} E_1\left(\frac{\gamma_0}{s}\right).$$

Then (from [3, Th. 2])

$$Z \leq \frac{\gamma_0}{s}\left[1 - \log\left(1 - e^{-\frac{\gamma_0}{s}}\right)\right] \leq \frac{\gamma_0}{s}[1 + k] \qquad (41)$$

where k is defined in Appendix I.

We have shown in (8)

$$\frac{\gamma_0}{s} \leq 2k e^{-\sqrt{2s}}.$$

Therefore, from (41)

$$Z \leq 2k(1+k) e^{-\sqrt{2s}}$$

and from (39)

$$P_E < [(2k(1+k))^N \exp(-\sqrt{2s}N)]^{\delta(R)}. \qquad (42)$$

One observes that the block error probability decreases as $\exp(-\sqrt{2s}N\delta(R))$. Thus for rate R we get exponential diversity of order $(\frac{1}{2}, \sqrt{2s}\delta(R))$ within region 1. The diversity versus rate tradeoff is captured by the function $\delta(R)$. For the binary channel the three regions versus SNR are plotted in Fig. 13 and $\delta(R)$ is plotted in Fig. 14. We observe that $\delta(R)$ is decreasing with R, the decrease being more rapid in the beginning. In Fig. 15 we plot the upper bound (42) along with that for regions 2 and 3.

It is interesting to compare the above diversity versus rate tradeoff with that of the channel without fading and Rayleigh-fading channel without power control. For the AWGN channel the expurgated bound in region 1 is $\exp(-N\delta(R)s)$ and for the Rayleigh channel $(1+s)^{-N\delta(R)}$, providing the diversity orders exponential $(1, \delta(R))$ and polynomial $\delta(R)$ respectively. One observes that the same function $\delta(R)$ is providing the diversity versus rate tradeoff in each case.
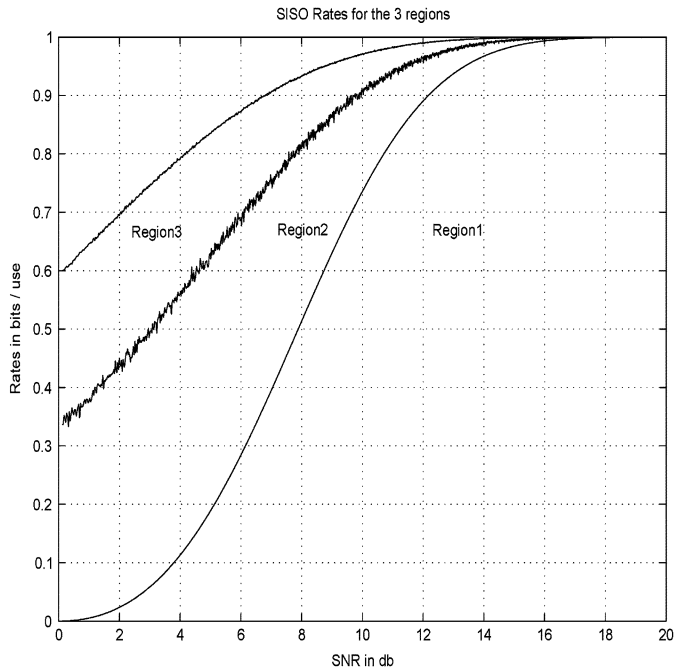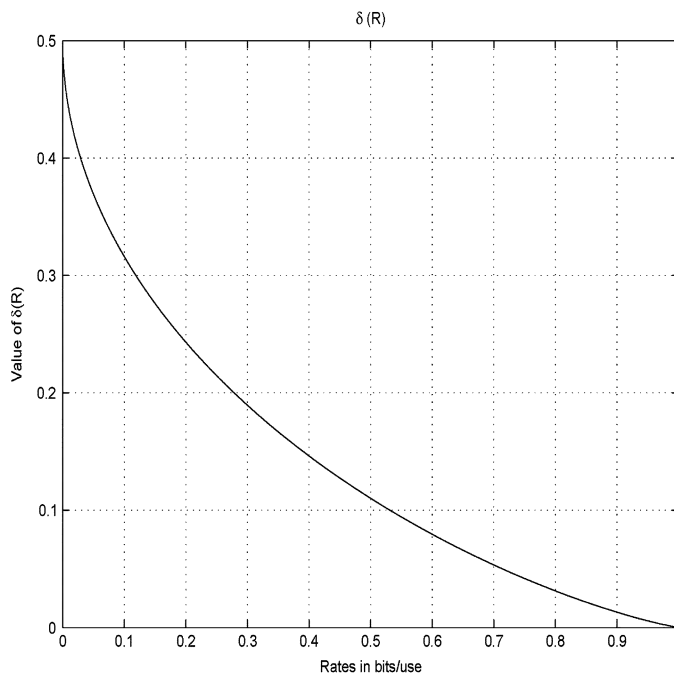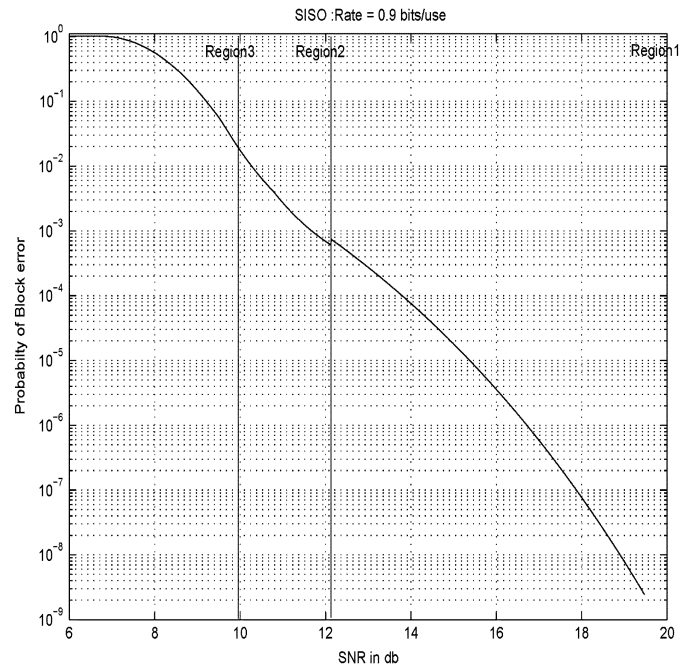
Fig. 13. Three rate regions for SISO channel with perfect CSIT and CSIR.

Fig. 15. Upper bound on P_E for SISO, perfect CSIT and CSIR, $R = 0.9$, $N = 128$.

Fig. 14. $\delta(R)$ for error exponents in Region 1 for perfect CSIT and CSIR.

Comparing (42) with $N = 1$ with the expression for BPSK in (9), one finds striking similarity. This similarity will continue to hold for MISO and MIMO systems also in region 1 even though the rate 1 of uncoded BPSK is actually outside the capacity region.

Next consider region 2. For regions 2 and 3 it is sufficient to consider the bound (33). In region 2, by definition, the maximizing ρ is 1. Therefore

$$E(R) = E_0(1, q) - R. \qquad (43)$$

From (37)

$$E_0(\rho, q) = -\ln(Z')$$

where

$$Z' = \sum_y \left[\sum_x q(x) p(y/x)^{\frac{1}{2}} \right]^2 = \frac{1 + Z}{2}. \qquad (44)$$

This is a useful result and will be used in future for extending the results of region 1 to region 2. Thus

$$P_E \leq 2^{-N} e^{NR} (1 + 2k(1 + k) e^{-\sqrt{2s}})^N. \qquad (45)$$

Let us look at the diversity order of this bound. To see the dependence of the term in bracket on SNR, observe that the first term is constant $= 1$ and $k \geq 1$. Thus the second term is $\geq 4 e^{-\sqrt{2s}}$. For low SNR (s close to zero) the second term dominates and we see exponential diversity. But at a very low SNR, (for a given R) the system may be in region 3. As s increases the first term starts becoming important and rate of decay will decrease. At very high SNR, the bound will become almost constant and diversity will be almost zero. In between, one will see diversity order decrease from exponential to polynomial to sub-linear before it becomes zero. In fact, as we see from Fig. 15, (in Fig. 15 we plot the upper bound (33) itself. This shows that this characteristic is not only in our bound (45) but also in (33) itself) it is only the polynomial and sublinear diversity that we see in region 2. However, as the above argument shows, the diversity one sees in region 2 will also depend upon R (depending upon R one may or may not see exponential and sub-linear diversity in region 2). This argument also shows that one may expect exponential diversity in region 1 and region 3. Furthermore, from
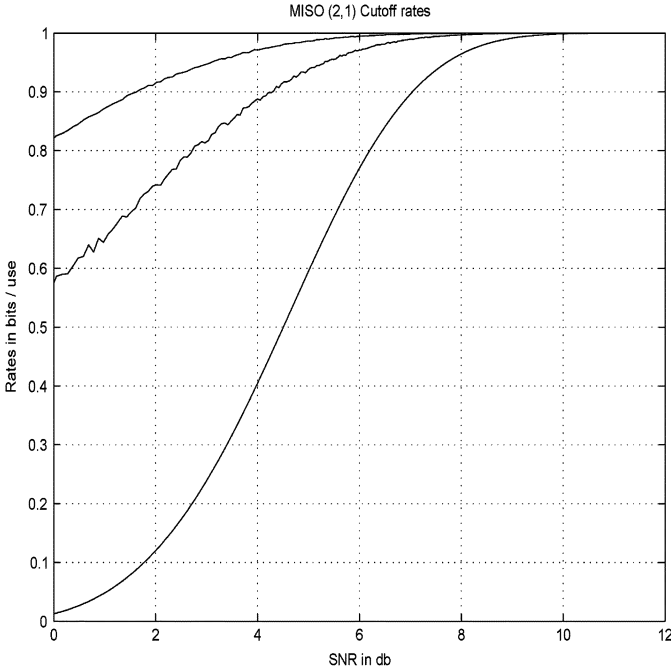
Fig. 16. Three rate regions for MISO $(n_t, n_r) = (2, 1)$ channel with perfect CSIT and CSIR.

(45) we see that the diversity order depends on R only to the extent when the system comes out of region 2.

Again we compare our results with that of channel without fading and with the Rayleigh-fading channel without power control. The upper bound on the P_E for the two cases are $2^{-N} e^{NR} (1 + e^{-s})^N$ and $2^{-N} e^{NR} (1 + \frac{1}{1+s})^N$ respectively. In both the cases we have two terms and their contribution to the overall bound can be explained as above.

Finally consider region 3. In this region $E(R) = E_0(p, q) - \rho R$, $R = \frac{\partial E_0}{\partial \rho}$ at the maximizing ρ and $E_0 = -\ln Z'$ where

$$Z' = \int_{-\infty}^{\infty} dy \int_0^{\infty} d\gamma f(\gamma)$$
$$\times \left[\frac{1}{2} \left(p(y/x_1)^{\frac{1}{1+\rho}} + \frac{1}{2} p(y/x_2)^{\frac{1}{1+\rho}} \right) \right]^{1+\rho}.$$

The corresponding upper bound on the block error probability is plotted in Fig. 15, and one sees an exponential decay here. Although we do not have explicit diversity versus rate tradeoff here (of course it can be computed from Fig. 15) the discussion for region 2 above suggested exponential diversity.

2) MISO System: In this section we consider the system with $n_t > 1, n_r = 1$. The SIMO system with $n_t = 1, n_r > 1$ has behavior similar to the corresponding MISO system. The optimal beam former $W = H^{\dagger}/\|H\|$ and SNR $\gamma = s\|HW\|^2$ is Erlang distributed with parameter n_t and mean s. Now the three regions are plotted in Fig. 16 (for $n_t = 2, n_r = 1$). One sees here that region 3 has become even less visible.

With the changes mentioned above the upper bounds on the block error probability in the three regions can be studied as in Section VI-A1.

Region 1
Using the relation (40) and knowing that the fading parameter γ is $\text{Erlang}_{n_t}(1/s)$ distributed

$$Z = \int_0^{\frac{\gamma_0}{s}} \frac{e^{(\frac{-\gamma}{s})} (\frac{\gamma}{s})^{n_t - 1}}{s(n_t - 1)!} d\gamma$$
$$+ \int_{\frac{\gamma_0}{s}}^{\infty} e^{-\ln(\frac{\gamma}{\gamma_0})} \frac{e^{(\frac{-\gamma}{s})} (\frac{\gamma}{s})^{n_t - 1}}{s(n_t - 1)!} d\gamma$$
$$= 1 - e^{-\frac{\gamma_0}{s}} \sum_{i=0}^{n_t - 1} \frac{(\frac{\gamma_0}{s})^i}{i!} + \left(\frac{\gamma_0}{s} \right)^{n_t} \int_1^{\infty} \frac{e^{(\frac{\gamma_0 x}{s})} x^{n_t - 2}}{(n_t - 1)!}$$
$$= 1 - e^{-\frac{\gamma_0}{s}} \sum_{i=0}^{n_t - 1} \frac{(\frac{\gamma_0}{s})^i}{i!} + \frac{\frac{\gamma_0}{s} e^{-\frac{\gamma_0}{s}}}{n_t - 1} \sum_{i=0}^{n_t - 2} \frac{(\frac{\gamma_0}{s})^i}{i!}$$
$$\leq 1 - e^{-\gamma_0/s} + \frac{\gamma_0}{s(n_t - 1)}$$
$$\leq \frac{\gamma_0}{s} \left(\frac{n_t}{n_t - 1} \right).$$

Applying (11) to the right-hand side, implies

$$Z \leq k e^{-(n_t - 1)s + \alpha} \left(\frac{n_t}{n_t - 1} \right)$$

which yields the upper bound

$$P_E \leq \begin{cases} e^{-N\delta(R)[s(n_t - 1) - \ln(\frac{n_t k}{n_t - 1}) - \alpha)]}, & \text{for } n_t \geq 3 \\ e^{-N\delta(R)[s - \ln 2k]}, & \text{for } n_t = 2 \end{cases} \quad (46)$$

where $\alpha = \sum_{k=0}^{n_t - 3} \frac{1}{k+1}$ (for $n_t = 2, \alpha = 0$). One observes a diversity order exponential $(1, \delta(R)(n_t - 1))$ and the diversity versus rate tradeoff in this region is captured by $\delta(R)$.

Region 2
Using the relation (44), we arrive at the bound for region 2

$$P_E \leq e^{-NE(R)}$$
$$\leq \exp \left[-N \left[\ln 2 - \ln \left(1 + \frac{n_t}{n_t - 1} k \right. \right. \right.$$
$$\left. \left. \left. \times e^{-[(n_t - 1)s - \alpha]} \right) - R \right] \right].$$

The upper bound for $n_t = 2$ is plotted in Fig. 17. The diversity order in the linear part here is about 1.5 the diversity order of $n_t = 1$.

Region 3
The plot of the upper bound on P_E for $n_t = 2$ is provided in Fig. 17. The region 3 is non vanishing only up to $s = 7$ dB (see Fig. 16). One sees that the upper bound is exponential till this point and then becomes linear (in region 2).

3) MIMO: The results for the three regions are summerized as follows:
Region 1
Using $f_\gamma(\gamma)$ from (14) in

$$Z = \int_0^{\gamma_0} f_\gamma(\gamma) d\gamma + \int_{\gamma_0}^{\infty} e^{-\ln(\gamma/\gamma_0)} f_\gamma(\gamma) d\gamma \quad (47)$$

one observes that this expression is very similar to (16). Thus the results of Appendix II can be used to show
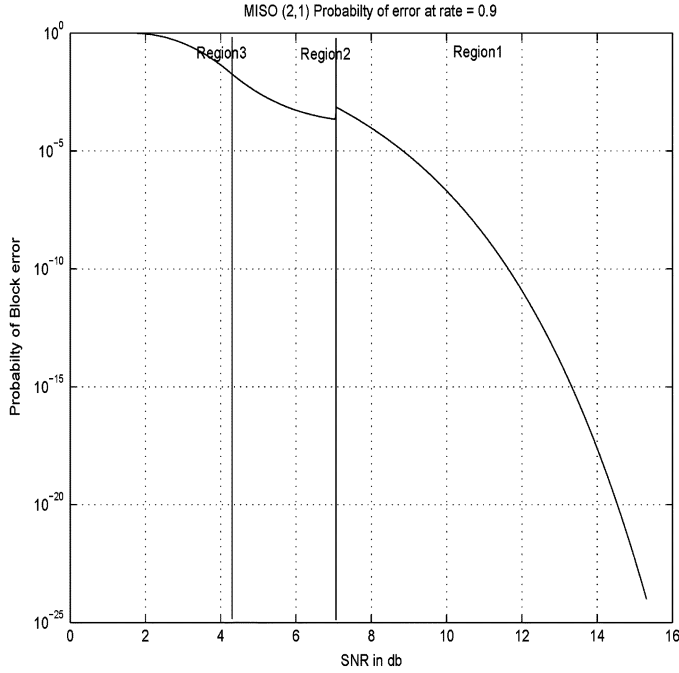
$$Z \stackrel{<}{\sim} K k e^{-(s-c)/K}.$$

Fig. 17. Upper bound on P_E for MISO $(2, 1)$, perfect CSIT and CSIR, $R = 0.9, N = 128$.

Fig. 18. Three rate regions for MIMO $(n_t, n_r) = (2, 2)$ channel with perfect CSIT and CSIR.

Therefore, from (35)

$$P_E \lesssim \left[Kk e^{-(s-c)/K} \right]^{N\delta(R)} \tag{48}$$

Hence the exponential diversity order is $\left(1, \frac{1}{K}\delta(R)\right)$. Here again $\delta(R)$ captures the diversity multiplexing tradeoff.

Region 2

Using relation (44) we obtain

$$P_E \le 2^{-N} e^{NR}[1 + Ke^{c/K}e^{-s/K}]^N.$$

The behavior of this bound can be explained as for the SISO system.

Region 3

In Region 3, we compute numerically the bounds as in the SISO case and find that we obtain exponential decay of block error rate. The plot of P_E for $n_t = 2, n_r = 2$ is provided in Fig. 19 and the rate regions are given in Fig. 18. The region 3 is nonvanishing up to $s = 4$ dB.

B. System With Imperfect CSIT

We follow the notation of the model in Section III. The CSIT is $\mathbf{H}_T$, which is available at the receiver and the receiver also knows $\mathbf{H}$.

1) SISO System: As in Section III-A, Z defined in (40) can be upper bounded as

$$Z \le ke^{-\sqrt{2s}} + ke^{-\sqrt{2s}}\frac{e^{-\alpha'}}{\alpha'} \tag{49}$$

where α' and σ_e^2 are defined in Section III-A. This upper bound is shown in Fig. 5.

We compute the error exponent for different values of SNR at rate 0.6 in Fig. 20. We find that even with $\rho_T = 0.99$, the
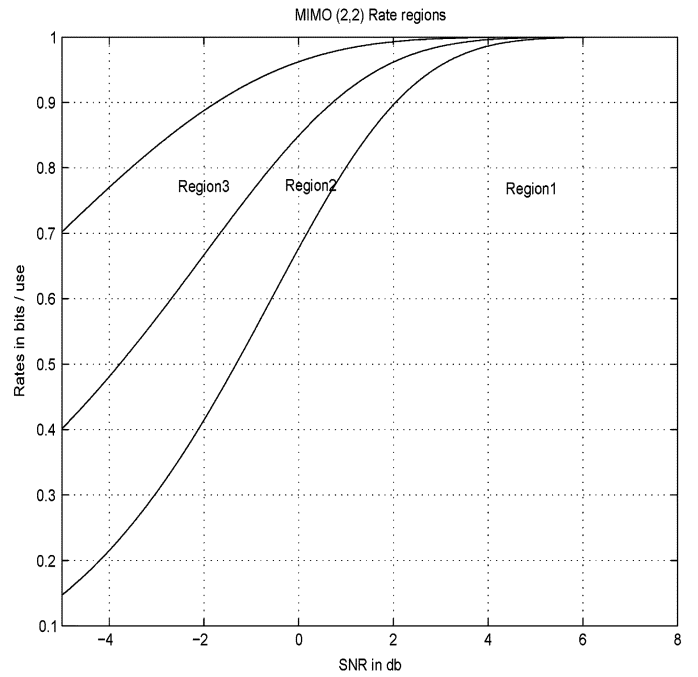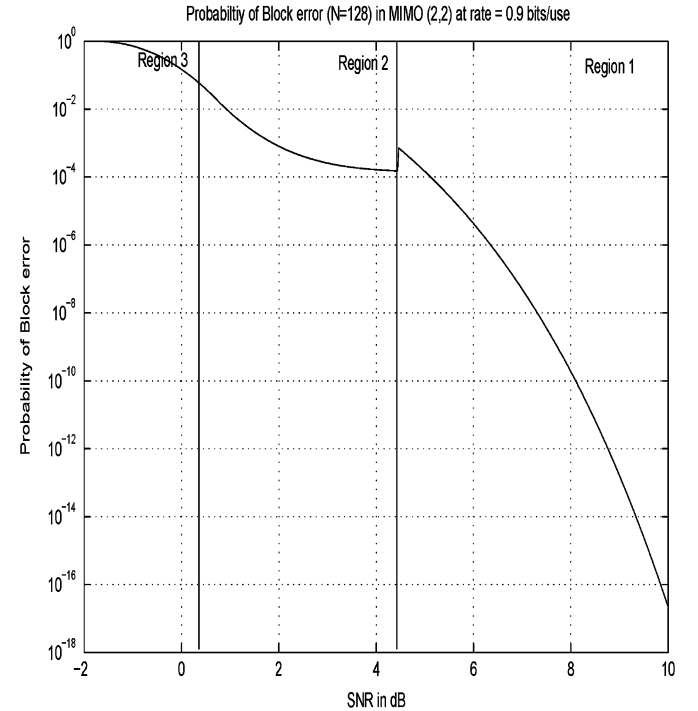

Fig. 19. Upper bound on P_E for MIMO (2,2), perfect CSIT and CSIR, $R = 0.9, N = 128$.

exponential nature of decay in the high SNR region, i.e., region 1 is lost. It is also observed that the quantitative character of the exponent in regions 2 and 3 is unchanged due to imperfect CSIT.

We find that though the bound (49) is loose for high SNR, it gives an intuitive understanding as to what actually happens in Fig. 20. There are two terms in the bound, the first one decays
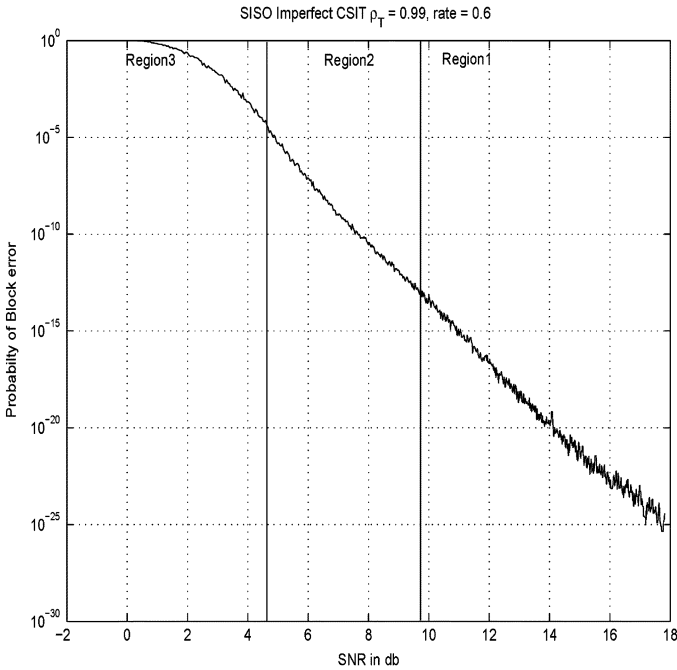
Fig. 20. Computed block error probability for SISO with Imperfect CSIT ($\rho_T = 0.99$).

Fig. 21. Computed block error probability for MISO $(2, 1)$ with Imperfect CSIT ($\rho_T = 0.99$).

Fig. 22. Computed block error probability for MIMO $(2, 2)$ with Imperfect CSIT ($\rho_T = 0.99$).

as the bound for Perfect CSIT, but the second (extra) term, due to the Imperfect CSIT, is an increasing function of SNR and becomes significant in the high SNR regime (see Fig. 5). This explains the loss of exponential diversity in the high SNR regime (region 1). This extra term can be controlled if we have a ρ_T very close to 1. So the better the CSIT, the greater is the region in which we have exponential diversity. This also explains why the quantitative character of the exponent does not change at low SNR (region 2 and region 3). We may conclude that we need to have a very good estimate of the channel ($\rho_T \sim 0.999$) to achieve exponential diversity in region 1.

2) MISO System: The power allocation in space and time is determined by the channel estimate $\mathbf{H}_T$. The distribution of $|H_T|^2$ is the same as in perfect CSIT case, i.e., $\mathrm{Erlang}_{n_t}(1/s)$. The conditional distribution of $|H|^2$ given $|H_T|^2$ is given by the *noncentral chisquare* as in the case of SISO. On computing the ensemble and expurgated bounds, we observe that with a ρ_T of 0.99 we do not obtain an exponential diversity in region1. However, as in SISO, the quantitative nature of the bounds for regions 2 and 3 are not affected. The actual calculations of the ensemble and expurgated bounds are plotted in Fig. 21 for $n_t = 2, n_r = 1$.

3) MIMO System: In a MIMO system also the results are similar to that of SISO and MISO. To obtain exponential diversity in region 1, we need a very good CSIT, else we may have to be satisfied with a polynomial decay. The computed block error probabilty is provided in Fig. 22 for $n_t = 2, n_r = 2$.

VII. CONCLUSION

We have analyzed the power allocation problem in space and time for Rayleigh-fading channels. Similar Analysis can also be done for other fading distributions. We have shown that the optimal p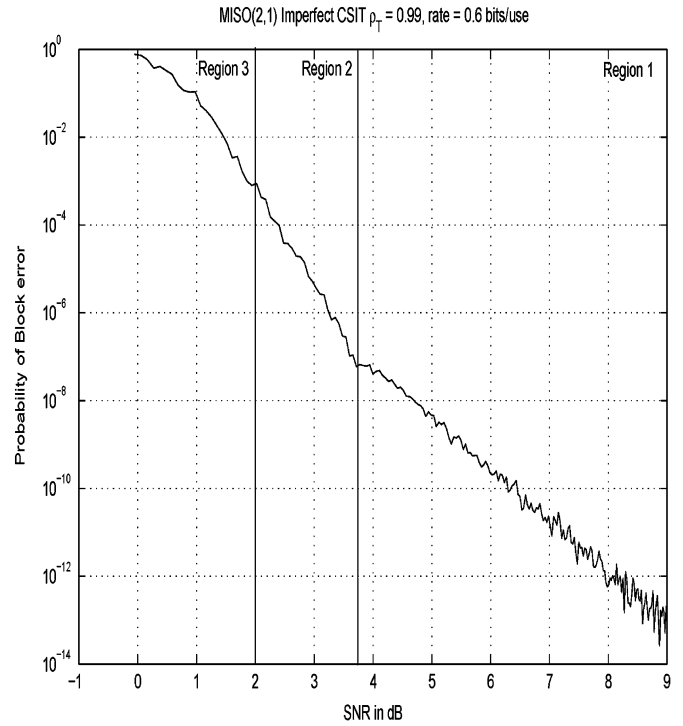ower allocation in space and time substantially reduces the BER. We observed that in the beamforming case, when perfect CSIT is available, our power allocation policy provides *exponential order diversity* gain for both coded and uncoded systems. The diversity gain that we achieve is substantially more than the conventional Space-Only and uniform power allocation schemes. Also, when the quality of CSIT degrades, the exponential diversity is lost at high SNR. The interesting observa-

tion here is we still achieve exponential diversity at low SNR. We also study the BER performance of S-T codes (Alamouti code) with our Space-Time power allocation scheme. We have observed that our power allocation policy improves the BER performance of S-T codes also. Infact, the performance of S-T codes for imperfect CSIT case is better than the power allocation policy given by [11] and the beamforming as observed in [16] and [20] for space-only power control. Another interesting conclusion is that in case of imperfect CSIT and CSIR, it is more important for the receiver to know the CSIT than the actual channel state.
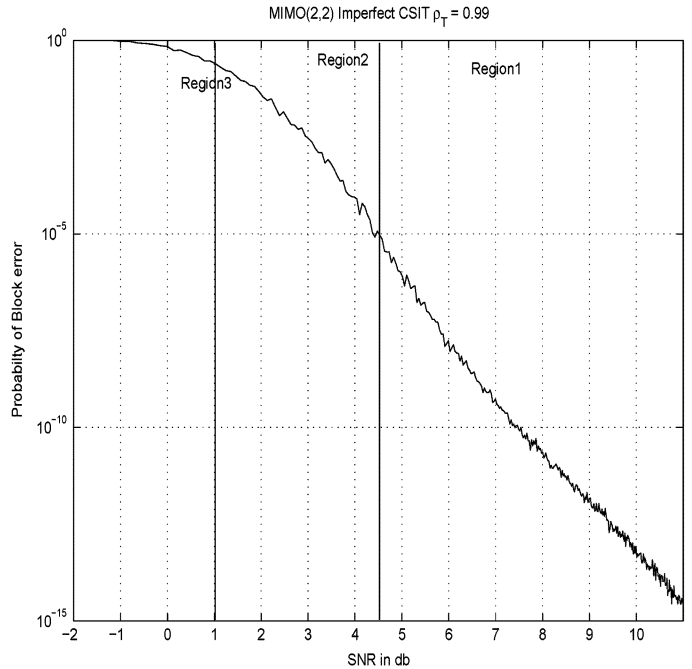
We have also obtained error exponents for our power control policy. We obtain exponential diversity of the expurgated/ensemble average bound in low and high rate regions but not in the intermediate region. The exponential diversity is lost in the low rate region if the CSIT is not perfect.

APPENDIX I
AN INEQUALITY

Let $f(x) = -\ln(1-x)$ defined in the interval $x \in [0,1)$. It is a convex increasing function of x. Thus, for all $x \in [0,x_0], (x_0 \ll 1), f(x) \leq kx$, where k depends on x_0 and is given by $k = f(x_0)/x_0$. We use this bound extensively in our derivations.

APPENDIX II
UPPER BOUND ON P_b FOR MIMO, PERFECT CSIT, CSIR

First we obtain the upper bound for $n_t > n_r$; then for $n_t = n_r$.

Case 1: $n_t > n_r$
From (7)

$$P_b = 0.5 \left(\int_0^{\gamma_0} f_\gamma(\gamma) d\gamma + \int_{\gamma_0}^{\infty} e^{-\ln(\gamma/\gamma_0)} f_\gamma(\gamma) d\gamma \right).$$

Using (14)

$$
\begin{aligned}
\int_0^{\gamma_0} f_\gamma(\gamma) d\gamma &= \int_0^{\gamma_0} \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \\
&\quad \times \frac{i^{m+1}(\gamma/s)^m e^{-i\gamma/s}}{m!} d\gamma \\
&= \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i^{m+1}}{m! s^m} \\
&\quad \times \int_0^{\gamma_0} \gamma^m e^{-i\gamma/s} d\gamma \\
&= \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} s \\
&\quad \times \left[1 - e^{-i\gamma_0/s} \left[\sum_{k=0}^{m} \frac{(i\gamma_0/s)^k}{k!} \right] \right]. \quad (50)
\end{aligned}
$$

Now consider the integral

$$
\begin{aligned}
\int_{\gamma_0}^{\infty} \frac{1}{\gamma} f_\gamma(\gamma) d\gamma &= \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \\
&\quad \times \int_{\gamma_0}^{\infty} \frac{1}{\gamma} \frac{i^{m+1}(\gamma/s)^m e^{-i\gamma/s}}{m!} d\gamma \\
&= \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i^{m+1}}{s^m m!} \\
&\quad \times \int_{\gamma_0}^{\infty} \gamma^{m-1} e^{-i\gamma/s} d\gamma \\
&= \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m} \\
&\quad \times \left[e^{-i\gamma_0/s} \left[\sum_{k=0}^{m} \frac{(i\gamma_0/s)^k}{k!} \right] \right]. \quad (51)
\end{aligned}
$$

On computation, we find that $\gamma_0/s \ll 1$ (for example, for SNR ≥ 3 dB, $\frac{\gamma_0}{s} \leq 0.01$). Then using (7), (50), and (51)

$$
\begin{aligned}
P_b &= \frac{1}{2s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \\
&\quad \times \left\{ s \left(1 - e^{-i\gamma_0/s} \left[\sum_{k=0}^{m} \frac{(i\gamma_0/s)^k}{k!} \right] \right) \right. \\
&\quad \left. + \frac{i\gamma_0}{m} \left[e^{-i\gamma_0/s} \left[\sum_{k=0}^{m} \frac{(i\gamma_0/s)^k}{k!} \right] \right] \right\} \\
&\approx \frac{1}{2s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \left\{ s(1 - e^{-i\gamma_0/s}[1]) \right. \\
&\quad \left. + \frac{i\gamma_0}{m} [e^{-i\gamma_0/s}[e^{i\gamma_0/s}]] \right\} \\
&\approx \frac{\gamma_0}{2s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} i + \frac{\gamma_0}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}.
\end{aligned}
$$

Since the first term is zero

$$P_b \approx \frac{\gamma_0}{2s} K \quad (52)$$

where

$$K = \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}. \quad (53)$$

Now let us calculate γ_0/s. By the average power constraint

$$
\begin{aligned}
1 &= E[P(\gamma)] \\
&= \int_0^{\gamma_0} \frac{1}{\gamma} \ln\left(\frac{\gamma}{\gamma_0}\right) \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \\
&\quad \times \frac{i^{m+1}(\gamma/s)^m e^{-i\gamma/s}}{m!} d\gamma
\end{aligned}
$$

which can be shown to be equal to

$$\frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{\alpha^m i}{m!} \int_1^\infty x^{m-1} e^{-\alpha x} \ln(x) dx$$

where $\alpha = i\gamma_0/s$. We can rewrite it as

$$1 = \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}$$
$$\times \left(\int_\alpha^\infty \frac{e^{-t}}{t} dt + e^{-\alpha} \sum_{k=0}^{m-2} \frac{1}{k+1} \sum_{i=0}^k \frac{\alpha^i}{i!} \right). \quad (54)$$

Using the fact that $\gamma_0/s \ll 1$

$$E_1\left(\frac{\gamma_0}{s}\right) = \int_{\gamma_0/s}^\infty \frac{e^{-t}}{t} dt$$
$$= \int_{i\gamma_0/s}^\infty \frac{e^{-t}}{t} + \int_{\gamma_0/s}^{i\gamma_0/s} \frac{e^{-t}}{t}$$
$$\approx \int_{i\gamma_0/s}^\infty \frac{e^{-t}}{t} + \int_{\gamma_0/s}^{i\gamma_0/s} \frac{1}{t}$$
$$= E_1\left(i\frac{\gamma_0}{s}\right) + \ln(i). \quad (55)$$

Thus from (54)

$$s \approx \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}$$
$$\times \left(E_1\left(\frac{\gamma_0}{s}\right) - \ln(i) + e^{-\alpha} \sum_{k=0}^{m-2} \frac{1}{k+1} \sum_{i=0}^k \frac{\alpha^i}{i!} \right)$$
$$= K E_1\left(\frac{\gamma_0}{s}\right) + c \quad (56)$$

and hence

$$\frac{(s-c)}{K} \approx E_1\left(\frac{\gamma_0}{s}\right) \quad (57)$$

where K was defined earlier and

$$c = \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}$$
$$\times \left(e^{-\alpha} \sum_{k=0}^{m-2} \frac{1}{k+1} \sum_{i=0}^k \frac{\alpha^i}{i!} - \ln(i) \right)$$
$$\approx \sum_{i=1}^{n_r} \sum_{m=n_t-n_r}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m} \left(\sum_{k=0}^{m-2} \frac{1}{k+1} - \ln(i) \right). \quad (58)$$

From [3, Th. 2], $E_1(\frac{\gamma_0}{s})$ is upper bounded by $-\ln(1 - e^{-\gamma_0/s})$. We find from numerical calculations that the approximations are tighter than the bound, and hence the bound shall override the approximations. Strictly speaking these are not bounds, but approximations to the bounds and we denote this relation by $\lesssim$. Thus

$$\frac{\gamma_0}{s} \lesssim k \exp\left(-\frac{s-c}{K}\right) \quad (59)$$

where k was defined in Appendix I. Hence from (53)

$$P_b \lesssim \frac{Kk}{2} \exp\left(-\frac{s-c}{K}\right). \quad (60)$$

The k for values of SNR ≥ 0 dB is very close to 1 and can be left out in most of the cases.

Case 2: $n_t = n_r$

The analysis is similar to Case 1 until (50).

Now consider the integral

$$\int_{\gamma_0}^\infty \frac{1}{\gamma} f_\gamma(\gamma) d\gamma = \frac{1}{s} \sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i^{m+1}}{s^m m!}$$
$$\times \int_{\gamma_0}^\infty \gamma^{m-1} e^{-i\gamma/s} d\gamma$$
$$+ \frac{\gamma_0}{s} \sum_{i=1}^{n_r} d_{i,0} i E_1(i\gamma_0/s).$$

By (55)

$$\frac{\gamma_0}{s} \sum_{i=1}^{n_r} d_{i,0} i E_1(i\gamma_0/s) \approx \frac{\gamma_0}{s} \sum_{i=1}^{n_r} d_{i,0} i (E_1(\gamma_0/s) - \ln(i))$$
$$= -\frac{\gamma_0}{s} \sum_{i=1}^{n_r} d_{i,0} i \ln(i)$$

since the first term is 0 (see [17], Appendix IV). Thus

$$\int_{\gamma_0}^\infty \frac{\gamma_0}{\gamma} f_\gamma(\gamma) d\gamma \approx \frac{\gamma_0}{s} \left\{ \sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m} \right.$$
$$\times \left[e^{-i\gamma_0/s} \left[\sum_{k=0}^m \frac{(i\gamma_0/s)^k}{k!} \right] \right]$$
$$\left. - \sum_{i=1}^{n_r} d_{i,0} i \ln(i) k \right\}$$
$$\approx K \frac{\gamma_0}{s} \quad (61)$$

where

$$K = \sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m} - \sum_{i=1}^{n_r} d_{i,0} i \ln(i). \quad (62)$$

Let us now evaluate γ_0/s. By the average power constraint, using results from Case 1

$$1 = E[P(\gamma)]$$
$$= \frac{1}{s} \left[\sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m} \right.$$
$$\times \left(\int_\alpha^\infty \frac{e^{-t}}{t} dt + e^{-\alpha} \sum_{k=0}^{m-2} \frac{1}{k+1} \sum_{i=0}^k \frac{\alpha^i}{i!} \right)$$
$$\left. + \sum_{i=1}^{n_r} i d_{i,0} \int_1^\infty \frac{e^{(\frac{-i\gamma_0 x}{s})} \ln(x)}{x} dx \right].$$

Now using [8], (55), and the fact that $\sum_{i=1}^{n_r} i d_{i,m}$ is zero (see [17], Appendix IV)

$$\sum_{i=1}^{n_r} d_{i,0} i \int_1^\infty \frac{e^{\left(\frac{-i\gamma_0 x}{s}\right)} \ln(x)}{x} dx$$

$$\leq \sum_{i=1}^{n_r} d_{i,0} \frac{i}{2} \left[E_1\left(\frac{\gamma_0}{s}\right) - \ln\left(\frac{i}{2}\right) \right]^2$$

$$= \sum_{i=1}^{n_r} d_{i,0} \frac{i}{2} \left\{ (\ln(i/2))^2 - 2 E_1\left(\frac{\gamma_0}{s}\right) \ln(i/2) \right\}$$

$$= \sum_{i=1}^{n_r} d_{i,0} \frac{i}{2} \left\{ (\ln(i/2))^2 - 2 E_1\left(\frac{\gamma_0}{s}\right) \ln(i) \right\}.$$

We find by numerical calculation that the approximations are tighter than bounds. Thus the bounds shall override the approximations. Therefore

$$s \lesssim \sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}$$

$$\times \left(\int_\alpha^\infty \frac{e^{-t}}{t} dt + e^{-\alpha} \sum_{k=0}^{m-2} \frac{1}{k+1} \sum_{i=0}^k \frac{\alpha^i}{i!} \right)$$

$$+ \sum_{i=1}^{n_r} d_{i,0} \frac{i}{2} \left\{ (\ln(i/2))^2 - 2 E_1\left(\frac{\gamma_0}{s}\right) \ln(i) \right\}.$$

On simplification, we get

$$E_1\left(\frac{\gamma_0}{s}\right) \gtrsim \frac{s-c}{K}$$

where K is given by (62) and

$$c = \sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m}$$

$$\times \left(e^{-\alpha} \sum_{k=0}^{m-2} \frac{1}{k+1} \sum_{i=0}^k \frac{\alpha^i}{i!} - \ln(i) \right)$$

$$+ \sum_{i=1}^{n_r} d_{i,0} \frac{i}{2} (\ln(i/2))^2$$

$$\approx \sum_{i=1}^{n_r} \sum_{m=1}^{(n_t+n_r)i-2i^2} d_{i,m} \frac{i}{m} \left(\sum_{k=0}^{m-2} \frac{1}{k+1} - \ln(i) \right)$$

$$+ \sum_{i=1}^{n_r} d_{i,0} \frac{i}{2} (\ln(i/2))^2.$$

Since the expressional form of the bound of $E_1\left(\frac{\gamma_0}{s}\right)$ is the same as in Case 1 the same expression will follow and so we obtain

$$P_b \lesssim \frac{Kk}{2} \exp\left(-\frac{s-c}{K} \right). \tag{63}$$

REFERENCES

[1] W. K. M. Ahmed and P. J. Mclane, "On the error exponents for memoryless flat fading channels with channel-state-information feedback," *IEEE Commun. Lett.*, vol. 3, pp. 49–51, 1999.

[2] S. S. Ahuja and V. Sharma, "Optimal power control for convolutional and turbo codes over fading channels," in *Proc. IEEE Conf. GLOBECOM'02*.

[3] H. Alzer, "On some inequalities for the incomplete gamma function," *Mathematics of Computation, AMS* vol. 66, pp. 771–778, 1997 [Online]. Available: http://www.ams.org/mcom/1997-66-218/

[4] H. Dai, L. Mailaender, and H. V. Poor, "CDMA downlink transmision with transmit antenna arrays and power control in multipath channels," *EURASIP J. Wireless Commun. Networks*, vol. 1, pp. 32–45, 2004.

[5] P. A. Dighe, R. K. Mallik, and S. S. Jamuar, "Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, pp. 694–703, Apr. 2003.

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[7] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[8] I. S. Gradshteyn and M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic, 2000.

[9] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP J. Appl. Signal Process.*, vol. 3, pp. 316–329, 2002.

[10] J. F. Hayes, "Adaptive feedback communication," *IEEE Trans. Commun. Technol.*, vol. Com-16, pp. 29–34, Feb. 1968.

[11] G. Jongren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 48, pp. 611–627, Mar. 2002.

[12] E. A. Jorswieck, A. Sezgin, H. Boche, and E. Costa, "Optimal transmit strategies in MIMO Ricean channels with MMSE receiver," in *IEEE Veh. Technol. Conf.*, 2004, vol. 5, pp. 3787–3791.

[13] A. Khoshnevis and A. Sabharwal, "On diversity and multiplexing gain of multi antenna systems with transmitter channel information," in *Proc. Allerton Conference*, 2004.

[14] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2003.

[15] K. N. Lau, "Optimal partial feedback design for SISO block fading channels," in *Proc IEEE WCNC*, 2003.

[16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2003.

[17] K. Premkumar, V. Sharma, and A. Rangarajan, Exponential diversity achieving spatio-temporal power allocation scheme for fading channels DRDO-IISc Program on Mathematical Engineering, IISc, Bangalore 560 012, India, Tech. Rep., TR-PME-2004-08 [Online]. Available: http://www.pal.ece.iisc.ernet.in/PAM

[18] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[19] A. Rangarajan, V. Sharma, and S. K. Singh, "Information-theoretic and communication-theoretic optimal power allocation for fading channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Yokohama, Japan, 2003, p. 246.

[20] A. Rangarajan and V. Sharma, "Achieving exponential diversity order in Rayleigh-fading channels," in *Proc. Nat. Conf. Commun. (NCC)*, Bangalore, Jan. 2004.

[21] M. K. Simon and M. S. Alouini, *Digital Communications Over Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley, 2004.

[22] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance analysis and code construction," *IEEE Trans. Inf. Theory*, vol. IT-44, pp. 744–765, 1998.

[23] I. E. Telatar, "Capacity of multi-antenna Gaussian channel," *European Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[24] A. J. Viterbi and J. K. Omura, *Principles of Digital Modulation and Coding*. New York: McGraw-Hill, 1979.

[25] X. Wu and R. Srikant, "MIMO channels in low SNR regime; communication rate, error exponents and signal peakiness," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2004.

[26] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel error," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2203–2214, 2006.

[27] L. Zheng and D. N. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–1096, 2003.

[28] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1673–1690, 2003.