Spatial Multiplexing Architectures with Jointly Designed Rate-Tailoring and Ordered BLAST Decoding
Part II: A Practical Method for Rate and Power Allocation

Yi Jiang Mahesh K. Varanasi

Abstract

The study of the class of new spatial multiplexing architectures (SMAs) is continued. As introduced in Part I of this paper, the SMAs consist of joint design of rate and power allocation to the spatially-multiplexed substreams at the transmitter and ordered nulling/canceling detection/decoding at the receiver. This was studied in Part I under the diversity-multiplexing tradeoff framework. Here the more detailed and practical problem of allocating rates and powers across the transmit antennas is investigated to minimize the overall system (uncoded or outage) error probability. Since the layer gains are unavailable to the transmitter, the rates and powers must be allocated based on the *statistics* of the layer gains. However, the channel dependent ordering rules make the precise distributions of the layer gains complicated or intractable. To solve this problem, a simple, yet effective, four-parameter hyperbola model is proposed to closely approximate the error probability of each layer. With this approximation, the computation of rate and power allocation according to the criterion of minimizing the maximal error probability of all the substreams is given. Simulation results validate the superior performance of the proposed SMAs, especially that of the Greedy ordering Rate Tailored SMA (GRT-SMA). Although the rate and power allocation method is obtained under the assumption of iid Rayleigh fading channel, the proposed SMAs also work well in other types of fading channels (such as in correlated and Rician channels).

Index Terms: Error analysis, fading channels, MIMO, rate allocation, power allocation, space-time architectures, spatial multiplexing, finite-rate feedback.

I. Introduction

This paper continues the study of the class of new spatial multiplexing architectures (SMA) with jointly designed rate and power allocation at the transmitter and ordered nulling/canceling or BLAST (Bell Labs layered Space-Time) detection/decoding at the receiver. Part-I of this paper [1] introduced the basic concept of the SMAs, and proved the significant improvement in the diversity-multiplexing (D-M) gain tradeoff [2] that accrues from such a joint design. The Greedy ordering Rate Tailored SMA

This work is supported in part by the National Science Foundation Grant CCF-0423842, CCF-0434410. Part of this paper was presented at Conference on Information Sciences and Systems 2006.

The authors are with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: yjiang@dsp.colorado.edu; varanasi@dsp.colorado.edu).

(GRT-SMA) was also introduced and proved to be optimal in terms of D-M tradeoff among the class of SMAs. The general problem is that, given a detection/decoding ordering rule (which specifies a member of the class of SMAs), the rate and power allocation must be computed to minimize the overall system error probability. This error probability could be uncoded bit error rate (BER) or the outage probability. In deriving the D-M gain tradeoff in Part I [1], the optimal allocation of multiplexing gains to the K layers were specified. However, the multiplexing gain metric is too coarse for practical system design and it is also limited to just the high SNR regime. Moreover, the D-M tradeoff analysis is insensitive to power disparities and hence it cannot be used as a design tool for spatial power allocation.

As described in Part I, the receiver only feeds back the detection/decoding ordering. The exact values of the layer gains are unknown to the transmitter. Therefore the rate and power allocation must be calculated based on the statistical information of the layer gains. However, the ordering rules render the distribution functions of the layer gains complicated (as in norm ordering) or intractable (as in greedy ordering) [1, Section IV]. The key contribution of this paper is to overcome this difficulty.

In particular, a simple, yet effective, four-parameter hyperbola model is proposed to closely approximate the error probability (either uncoded error or outage probability) of each layer. Next, the problem of optimal allocation of rates and powers according to the minimax uncoded BER optimality criterion is re-stated in terms of the hyperbola model and the resulting non-convex optimization problem is shown to have a unique solution. Coded SMAs are also considered where rates are allocated per layer to equalize the per layer outage probabilities to a target level, again using the hyperbola model, and then the power allocations are optimized to yield the minimum total power to support a target rate R. It is worth emphasizing that the power and rate allocation only needs to be calculated *offline* once and for all. This generates a lookup table for online use.

Furthermore, the outage capacity behavior at high SNR of SMAs with joint ordered decoding and rate/power allocations is analyzed and shown to that an outage capacity that behaves as $N \log \rho +$ const. (where $N = \min\{M_t, M_r\}$) which shows that the behavior is the same as that of the optimal unconstrained architecture analyzed in [3]. Moreover, an explicit formula for the constant term is obtained in terms of the distribution functions of channel gains which yields an optimality criterion for the choice of ordering rule that maximizes outage capacity at high SNR (which is however intractable) and gives insight about why the GRT-SMA gives superior performance.

Simulation results are presented to validate the superior performance of the proposed SMAs, especially the GRT-SMA. The remarkable performance of GRT-SMA illustrates the significance of transmitter and

receiver *collaboration* through a few bits of feedback. Simulations results also show that the rate and power allocations obtained under the assumption of i.i.d. Rayleigh fading channel are robust in that they also work very well in spatially correlated Rayleigh and Rician fading channels.

More broadly speaking, when a strict error probability analysis is intractable in fading channel communications that employ complex signal processing methods, one may be able to resort to the closed-form four-parameter model presented in this paper to optimize system performance. The value of the model may therefore extend well beyond the context of SMAs considered in this work.

The remainder of this paper is organized as follows. Section II introduces the hyperbola model of error probability, including uncoded BER and outage probability, over fading channels. In Sections III and IV, the problem of optimal allocation of rates and powers according to the minimax optimality criteria are analytically re-stated in terms of the hyperbola model and solutions to the resulting optimization problems are provided. In Section IV-B, the behavior of the outage capacity of the proposed class of SMAs is given in terms of the distribution functions of the channel gains from which an insight about the superior performance of the GRT-SMA is drawn. Extensive numerical examples are given (including coded SMAs with outer trellis coded modulation) in Section V to demonstrate the superior performance of the SMAs, especially the GRT-SMA. Finally, Section VI concludes this paper.

II. Hyperbola Model of Error Probability and Outage Probability

As stated earlier, the statistics of layer gains and hence exact formulas for the error (outage) probability are either too complicated or unavailable. Here, we present a simple yet highly accurate four-parameter hyperbola model to approximate the error (outage) probability as a function of input SNR.

With input SNR ρ and target rate R (in bps/Hz), the channel is said to be in *outage* if the channel cannot support the target rate, i.e., the channel gain

$$h \in \left\{|h|^2 : \log_2(1+|h|^2\rho) < R\right\} = \left\{|h|^2 : |h|^2 < \frac{2^R-1}{\rho}\right\}.$$

For a fading channel with random gain h, the outage probability decreases like $\rho^{-\mathsf{D}}$ as ρ increases, where D is the diversity gain associated with that channel. Correspondingly, with uncoded quadrature amplitude modulation (QAM), the average BER diminishes like $\rho^{-\mathsf{D}}$ over a fading channel.

Figures 1−3 show the uncoded BER and outage probability of the layers obtained using zero forcing V-BLAST (ZF-VB) detector. For all the three figures, the MIMO channel is an iid Rayleigh fading channel with $M_t = 4$ transmit antennas and $M_r = 4$ receive antennas. In Figure 1, the dots represent the uncoded BER performance the four layers obtained via ZF-VB with *norm ordering rule*, given that

16-QAM is used at each layer and there is no error propagation. Note that the uncoded BERs can be calculated as the expectation $P_{b,i} = \mathbb{E}_{r_{ii}^2}\left[P_b\left(r_{ii}^2\rho_i, M\right)\right]$, where r_{ii} is the ith layer gain, and $P_b\left(\gamma, M\right)$ is the BER of an M-QAM input with instantaneous SNR γ (see [4] for an explicit formula for $P_b\left(\gamma, M\right)$). As the distributions of the layer gains for the norm-ordering rule are known (cf. [1, Theorem IV.1]), the BER can be calculated through numerical integration. The dots in Figure 2 are the uncoded BERs of the four layers obtained via ZF-VB with *greedy ordering rule*. The inputs are BPSK symbols. In this case, the distributions of r_{ii}^2's are unknown except for $i = 1$. The BERs of the ith layer ($i \neq 1$) are estimated by averaging the simulated BERs over 10^8 Monte Carlo trials.

With input SNR ρ and target rate R, the outage probability of the ith layer is

$$P_{\text{outage},i}(\rho_i, R_i) = \mathbb{P}\left(|r_{ii}|^2 < \frac{2^{R_i} - 1}{\rho_i}\right). \tag{1}$$

In [5], the authors define the *normalized* SNR $\rho_{\text{norm}} \triangleq \frac{\rho}{2^R - 1}$ to measure the gap between the input SNR ρ and the minimal SNR required for the target rate R over an AWGN channel with unit channel gain. Using this terminology, we rewrite (1) as

$$P_{\text{outage},i}(\rho_{\text{norm}}) = \mathbb{P}\left(|r_{ii}|^2 < \rho_{\text{norm}}^{-1}\right) = F_{|r_{ii}|^2}(\rho_{\text{norm}}^{-1}), \tag{2}$$

where $F_{|r_{ii}|^2}(x)$ is the cumulative distribution function (cdf) of $|r_{ii}|^2$. The dots in Figure 3 show the outage probabilities of the four layers obtained via ZF-VB detector with *greedy ordering rule*. The plot corresponding to the *norm ordering rule* is similar which we omit here.

Since the closed-form probability density functions (pdf) of the layer gains (r_{ii}^2's) obtained via Norm QR decomposition are known, $P_{\text{outage},i}(\rho_{\text{norm}})$ can be calculated by numerical integration. But for the ZF-VB with greedy detection ordering, the pdfs of r_{ii}^2's are unknown except for $i = 1$. Hence in Figure 3, the outage probabilities of the ith layer ($i > 1$) are obtained via averaging the simulated BERs over 10^7 Monte Carlo trials. It can be seen from Figures 1−3 that the logarithms of the uncoded BER and channel (layer) outage probability decrease in a linear manner as $\log\rho \to \infty$. The steepness of decreasing slope is measured by diversity gain. According to [1, Theorems IV.1, IV.3], the four layers of the ZF-VB detector based on Norm QR have diversity gains $\mathsf{D}_1 = 16, \mathsf{D}_2 = 3, \mathsf{D}_3 = 2, \mathsf{D}_4 = 1$, while the ZF-VB detector based on Greedy QR yields four layers with diversity gain $\mathsf{D}_1 = 16, \mathsf{D}_2 = 9, \mathsf{D}_3 = 4, \mathsf{D}_4 = 1$. The three figures validate this diversity gain analysis. On the other hand, as $\rho \to 0$, or equivalently $\log\rho \to -\infty$, the error probabilities level to a constant.

Based on the above observations, we propose a heuristic model to quantify both uncoded BERs and channel outage probability as functions of input SNR. In the case of quantifying uncoded BERs, let us

define

$$x \triangleq 10\log_{10}\rho \quad \text{and} \quad y \triangleq \log_{10} P_e(\rho, M). \tag{3}$$

Similarly, to quantify channel outage probability, we define

$$x \triangleq 10\log_{10}\rho_{\text{norm}} \quad \text{and} \quad y \triangleq \log_{10} P_{\text{outage}}(\rho_{\text{norm}}), \tag{4}$$

Then x and y defined in (3) are the X and Y-coordinates of Figures 1 and 2. And those defined in (4) are the coordinates of Figure 3. We propose the four-parameter hyperbola model:

$$y(x) = -\frac{d}{20}\left[(x-c) + \sqrt{(x-c)^2 + a}\right] - b. \tag{5}$$

The solid lines in Figures 1−3 are obtained by fitting the dots using the model (5). The fitting parameters (a, b, c, d) are estimated by minimizing the sum of the squared fitting error. Because for ZF-VB with greedy ordering, the BERs of all the layers except the first one are estimated based on 10^8 Monte Carlo trials, the estimates are not reliable when BER is very small (say, $< 10^{-10}$), as shown in Figure 2. In these cases, we discard such outliers in the high SNR regime when applying the curve fitting. Otherwise, the hyperbolas fit the actual BERs and outage probabilities very closely in Figures 1−3.

We give some insights into this model to understand why the hyperbola model fits the experimental data so well. Indeed, the hyperbola of (5) has two asymptotes:

$$y = -b \ \text{ as } \ x \to -\infty \quad \text{and} \quad y = -\frac{d}{10}(x-c) - b \ \text{ as } \ x \to +\infty. \tag{6}$$

Obviously, by changing c and b, one can shift the curve in the horizontal and vertical directions, respectively. Hence, the parameter c is related to the *coding gain*. It is easy to verify that $\lim_{x\to\infty} \frac{dy(x)}{dx} = -\frac{d}{10}$. According to the diversity gain definition [6]

$$\mathsf{D} = -\lim_{x\to\infty} \frac{\log \mathbb{P}(\text{error})}{\log \rho} = -\lim_{x\to\infty} \frac{dy(x)}{dx/10} = d. \tag{7}$$

Therefore the parameter d stands for the *diversity gain*. The parameter a controls the curvature of the hyperbola between the two asymptotes. In other words, a affects the BER/outage probability in the low to moderate SNR regime. Since $\lim_{x\to-\infty} \log_{10} P_{\text{outage}}(\rho_{\text{norm}}) = \log_{10} 1 = 0$ and $\lim_{x\to-\infty} \log_{10} P_e(\rho, M) = \log_{10} 0.5 = -0.301$, one may conjecture according to (6) that the optimal $b = 0$ for the outage probability case and $b = 0.301$ for the uncoded BER case. However, it is still beneficial to keep b as an undetermined parameter since it helps to reduce further the fitting error. After all, the hyperbola model is just a heuristic approximation rather than a rigorous theoretical expression.

For any fading channel, once the four parameters (a, b, c, d) are known, we can analytically approximate the BERs and outage probabilities of the channel. Although this model is heuristic, the extensive numerical experiments show that this four-parameter model is surprisingly accurate for a wide variety of channel fading statistic, including Rayleigh, Rician, and Nakagami-m fading channels. It facilitates a unified solution to the rate and power allocation, as we shall see in the next section.

III. Optimal Rate/Power Allocation

In this section, we formulate the problem of optimal rate and power allocation across the $N(\triangleq \min\{M_r, M_t\})$ transmit antennas given rate and power constraint. Applying the hyperbola model, we develop a joint rate and power allocation algorithm to minimize the system error probability, including uncoded BER and outage probability.

A. Optimization Problem Formulation

We first consider the BER criterion. If the error propagation effect is ignored, the average uncoded BER of the ith layer, with input SNR ρ_i and M_i-QAM, can be expressed as $\mathbb{E}_{r_{ii}^2}\left[P_b\left(r_{ii}^2\rho_i, M_i\right)\right]$, where $P_b\left(r_{ii}^2\rho_i, M_i\right)$ is defined in (**??**). We propose to determine the optimal rate and power allocation by solving the following optimization problem:

$$\begin{array}{ll} \min_{\rho_i, M_i} & \max_{i:M_i>0}\left\{\mathbb{E}_{r_{ii}^2}\left[P_b\left(r_{ii}^2\rho_i, M_i\right)\right]\right\} \\ \text{subject to} & \sum_{i=1}^{N}\rho_i = \rho \\ & \sum_{i=1}^{N}\log_2 M_i = R, \end{array} \tag{8}$$

where, in view of practical system design, we constrain constellations to be square QAMs with $M_i \in \{0, 4, 16, 64, 256\}$. The reasons of adopting the minimax criterion are two-fold. First, the exact expression of the overall uncoded BER is intractable due to error propagation effects in the V-BLAST detector. Second, as we use independent scalar coding for each layer, the whole frame of data transmission fails if one layer is in error. This argument motivates the optimization of the weakest layer in use.

Let us now consider outage probability criterion. Let R_i be the rate allocated to the ith layer. The ith layer is in outage if $\log_2(1 + r_{ii}^2\rho_i) < R_i$, of which the probability is

$$P_{\mathcal{O},i}\left(\rho_i, R_i\right) \triangleq \mathbb{P}(\log_2(1 + r_{ii}^2\rho_i) < R_i). \tag{9}$$

Similar to (8), given the constraint of overall rate R and overall input SNR ρ, the optimal allocation of

rate/power is determined by solving

$$\begin{array}{ll} \min_{\rho_i, R_i} & \max_{i:R_i>0} \left\{P_{\mathcal{O},i}\left(\rho_i, R_i\right)\right\} \\ \text{subject to} & \sum_{i=1}^{N} \rho_i = \rho \ , \ \rho_i \geq 0, \\ & \sum_{i=1}^{N} R_i = R \ , \ R_i \in \mathcal{S} \cup \{0\}, \end{array} \tag{10}$$

where $\mathcal{S}$ is a finite set of allowable rates. Using the definition of normalized SNR $\rho_{\text{norm},i} = \frac{\rho_i}{2^{R_i}-1}$, we can rewrite (10) as

$$\begin{array}{ll} \min_{\rho_i, R_i} & \max_{i:R_i>0} \left\{\mathbb{P}\left(r_{ii}^2 < \rho_{\text{norm},i}^{-1}\right)\right\} \\ \text{subject to} & \sum_{i=1}^{N} (2^{R_i}-1)\rho_{\text{norm},i} = \rho \ , \ \rho_{\text{norm},i} \geq 0 \\ & \sum_{i=1}^{N} R_i = R \ , \ R_i \in \mathcal{S} \cup \{0\} \end{array} \tag{11}$$

Because the analytical expressions of $\mathbb{E}_{r_{ii}^2}\left[P_b\left(r_{ii}^2\rho_i, M_i\right)\right]$ and $\mathbb{P}\left(r_{ii}^2 < \rho_{\text{norm},i}^{-1}\right)$ are usually intractable, we apply the hyperbola model to approximate the cost function, which greatly simplifies the numerical solution to the optimization problems.

B. Numerical Optimization

Let the set of feasible rate tuples be $\mathcal{M} \triangleq \left\{\{M_i\}_{i=1}^N : \sum_{i=1}^N \log_2 M_i = R, M_i \in \{0, 4, 16, 64, 256\}\right\}$. Fix a feasible constellation tuple $\{M_i\}_{i=1}^N$, use the hyperbola model and the definitions in (3), and reformulate the optimal rate and power allocation problem (8) into two steps. First, we solve the following problem

$$\begin{array}{ll} \min_{x_i} \max_{i:M_i>0} & \left\{10^{-\frac{d_i}{20}\left[x_i - c_i + \sqrt{(x_i-c_i)^2+a_i}\right]-b_i}\right\} \\ \text{subject to} & \sum_{i:M_i>0} 10^{\frac{x_i}{10}} = \rho. \end{array} \tag{12}$$

Note that the hyperbola parameters, a_i, b_i, c_i, d_i, depend on the size of QAM. Hence M_i is relevant in the cost function. In the second step, we let $\{M_i\}_{i=1}^N$ go over the feasible set $\mathcal{M}$ and for each constellation tuple we solve (12). We then record the constellation tuple $\{M_i\}_{i=1}^N$ and the associated $\{x_i\}_{i=1}^N$ which yields the smallest cost function.

Now consider (12). Denoting $\mathcal{I} = \{i : M_i > 0\}$ and replacing x_i by $10\log_{10} w_i$, we reformulate (12) as

$$\begin{array}{ll} \min_{x_i, y} & y \\ \text{subject to} & 10^{-\frac{d_i}{20}\left[10\log_{10} w_i - c_i + \sqrt{(10\log_{10} w_i - c_i)^2+a_i}\right]-b_i} - y \leq 0, \quad i \in \mathcal{I} \\ & \sum_{i\in\mathcal{I}} w_i = \rho. \end{array} \tag{13}$$

Recall that a function is convex if and only if its Hessian matrix is positive semi-definite (psd) [7]. It can be verified that the constraint function

$$f_i(w_i, y) \triangleq 10^{-\frac{d_i}{20}\left[10\log_{10} w_i - c_i + \sqrt{(10\log_{10} w_i - c_i)^2+a_i}\right]-b_i} - y,$$

is not necessarily convex as $\frac{\partial^2 f_i(w_i,y)}{\partial w_i^2} < 0$ for some $a_i > 0$ and $x > 0$. Therefore (13) is not a convex optimization problem. However, we can still obtain a simple solution to (13). We observe that at an optimal solution to (13), all the inequality constraints $f_i(w_i, y) \le 0$ should be active, i.e., $f_i(w_i, y) = 0$ for $\forall\, i \in \mathcal{I}$. The argument is as follows. We can always decrease y such that there is at least one active inequality constraint. Suppose at an optimal solution there exists an inactive constraint $f_i(w_i, y) < 0$ for some $i \in \mathcal{I}$. Because $f_i(w_i, y)$ is a continuous decreasing function of w_i, there exists $\delta > 0$ such that $f_i(w_i - \delta, y) < 0$ still holds. Denote $j_1, \cdots, j_L$ as the indices of L active constraints ($L < K \triangleq |\mathcal{I}|$). Without violating the power constraint, we increase w_j to $w_j + \delta/L$ for $j \in \{j_1, \cdots, j_L\}$. But then $f_j(w_j + \delta/L, y) < 0$ for any $j \in \{j_1, \cdots, j_L\}$, i.e., all the K inequality constraints are inactive. We can therefore further reduce y for a better solution, which contradicts the optimality assumption. Now we have proven that at the optimal solution to (13), all the K inequality constraints are active, i.e.,

$$10^{-\frac{d_i}{20}\left[10\log_{10} w_i - c_i + \sqrt{(10\log_{10} w_i - c_i)^2 + a_i}\right] - b_i} - y = 0, \quad i \in \mathcal{I}.$$

Denote

$$\lambda \triangleq \log_{10} y = -\frac{d_i}{20}\left[10\log_{10} w_i - c_i + \sqrt{(10\log_{10} w_i - c_i)^2 + a_i}\right] - b_i, \quad i \in \mathcal{I}. \tag{14}$$

The equations lead to

$$w_i = 10^{\frac{1}{10}\left[c_i + \frac{a_i d_i}{40(\lambda + b_i)} - \frac{10(\lambda + b_i)}{d_i}\right]}. \tag{15}$$

Inserting (15) into the power constraint, we obtain

$$\sum_{i \in \mathcal{I}} 10^{\frac{1}{10}\left[c_i + \frac{a_i d_i}{40(\lambda + b_i)} - \frac{10(\lambda + b_i)}{d_i}\right]} = \rho. \tag{16}$$

For $a_i > 0$ and $d_i > 0$, the left hand side of (16) is a monotonous increasing function of λ given $\lambda + b_i < 0$. This constraint is innocuous as $\lambda + b_i = -\frac{d_i}{20}\left[x_i - c_i + \sqrt{(x_i - c_i)^2 + a_i}\right] < 0$. According to the hyperbola model, $b_i = \lim_{x \to -\infty} \log_{10} P_{\text{outage}}(\rho_{\text{norm}}) = 0$ in the case of outage probability, or $b_i = \lim_{x \to -\infty} \log_{10} P_e(\rho, M) = -0.301$ in the case of uncoded BER. Hence in theory $b_i = b$ for $\forall\, i$. (In practice, b_i's are slightly different due to the fitting error.) As

$$\lim_{\lambda \uparrow -b} \sum_{i \in \mathcal{I}} 10^{\frac{1}{10}\left[c_i + \frac{a_i d_i}{40(\lambda + b)} - \frac{10(\lambda + b)}{d_i}\right]} = 0 \quad \text{and} \quad \lim_{\lambda \to -\infty} \sum_{i \in \mathcal{I}} 10^{\frac{1}{10}\left[c_i + \frac{a_i d_i}{40(\lambda + b)} - \frac{10(\lambda + b)}{d_i}\right]} = \infty, \tag{17}$$

we conclude that there is a unique solution to (16). We can solve λ from (16) using Newton's iterative method or bisection method. Consequently, w_i is obtained according to (15).

It is even simpler to solve the dual problem: to minimize the input power required to meet a prescribed BER. Given a target BER P_b, we obtain the power w_i by inserting $\lambda = \log_{10} P_b$ into (15), which depends

on the hyperbola parameters (a_i, b_i, c_i, d_i) and in turn depends on M_i. The optimal constellation tuple is determined by finding the one yielding the minimal $\rho = \sum_{i=1}^{K} 10^{\frac{x_i}{10}}$. The actual probability of error is not exactly P_b due to the error-propagation effect. Some margin of input power is required to compensate for error propagation.

With the hyperbola model, the procedure of solving the optimization problem (8) can be equally applied to solve (11). We omit it here to avoid repetition. Furthermore, the optimization problems (8) and (11) are applicable to the SMA with any detection ordering. Given a detector ordering rule, which yields N layers with layer gains r_{ii}^2, $1 \le i \le N$, one simply uses the associated hyperbola model parameters (a_i, b_i, c_i, d_i) in the above described procedure to solve (8) or (11).

It is worthwhile emphasizing that the above algorithm only needs to be implemented *offline* once and for all, which generates a lookup table including input SNR constraint (or target error probability) and the rates and powers allocated to the K substreams. Indeed, the online computational complexity is quite small. Assuming that the transmitter knows the average input SNR ρ (which varies much slower than the channel gains), it allocates power and rates to K substreams according to the pre-generated lookup table. At the receiver side, for each channel realization the receiver determines the detection ordering according to some ordering rule, e.g., the greedy ordering rule introduced in [1, Section IV]. Then the receiver feeds the ordering information back to the transmitter. If K transmit antennas are to be used, then the ordering information can be encoded by $\log_2 (M_t!/(M_t - K)!)$ bits. Based on the ordering information, the transmitter maps the K substreams to K transmit antennas. Compared to the conventional V-BLAST algorithm, the only added complexity of the proposed schemes is to maintain a lookup table and a small amount of feedback.

Table I presents the lookup table for an iid Rayleigh channel with $M_r = M_r = 4$. The overall rate constraint $R = 16$ and the allocated rates and powers are calculated with BER criterion. The optimal rate allocation is 6/6/4/0 over the whole range of input SNR $14 \sim 30$ dB. That is, only $K = 3$ transmit antennas are in use with 64-QAM over the first and second layers and 16-QAM over the third layer. Note that the numbers therein are in absolute value, not in dB. In this case, the receiver needs to feed $\log_2 M_t! = 4.59$ bits per channel realization back to the transmitter.

IV. Analysis of Outage Probability and Outage Capacity

Outage probability and ϵ-outage capacity are the performance measures of fundamental importance as they represent the performance limit of a system using capacity-achieving coding. In this section,

TABLE I

INPUT SNRS $\rho_i, 1 \le i \le 4$, FOR GREEDY ORDERING (THE OPTIMAL RATE ALLOCATION: 6/6/4/0)

Layer	SNR = 14 dB	16 dB	= 18 dB	20 dB	22 dB	24 dB	26 dB	28 dB	30 dB
1	5.6594	8.9944	14.3925	22.9163	35.7918	54.0829	78.3608	108.6637	144.7930
2	11.6497	18.6142	29.3691	45.7216	70.1770	106.0488	157.4286	229.0046	326.0209
3	7.8098	12.2024	19.3342	31.3621	52.5233	91.0571	162.3178	293.2890	529.1864
4	0	0	0	0	0	0	0	0	0

we apply the hyperbola model to analyze the outage probability and outage capacity of the proposed class of SMAs. We assume the input constellation to be of infinitesimal granularity. Our goal here is to obtain the fundamental performance limit and hence gain more insights into the proposed schemes.

A. Outage Probability

An SMA system (with independent coding per layer) is in outage if and only if there is at least one layer in outage (see (9)). Therefore, the outage probability of the overall system is

$$P_{\mathcal{O},\mathrm{SMA}} = \mathbb{P}\left(\bigcup_{i:R_i>0}\left\{r_{ii}^2 < \frac{2^{R_i}-1}{\rho_i}\right\}\right). \tag{18}$$

In contrast, the fundamental outage probability of the MIMO channel with uniform power allocation is

$$P_{\mathcal{O},\mathrm{opt}} = \mathbb{P}\left(\log_2\left|\mathbf{I}+\frac{\rho}{N}\mathbf{H}\mathbf{H}^*\right| < \sum_{i=1}^{N} R_i\right), \tag{19}$$

i.e., the probability of mutual information of the channel being less than the target rate $R = \sum_{i=1}^{N} R_i$.

In Section III, we have studied allocating rate and power such that all the layers in use have the same error/outage probability. Following this criterion, we pose the constraint that

$$\mathbb{P}\left(\left\{r_{ii}^2 < \frac{2^{R_i}-1}{\rho_i}\right\}\right) = 10^{\lambda}, \quad i \in \mathcal{I} \tag{20}$$

for some $\lambda < 0$, where $\mathcal{I} = \{i : R_i > 0\}$. It is easy to see that $10^{\lambda} \le P_{\mathcal{O},\mathrm{RTVB}} \le K10^{\lambda}$, where $K = |\mathcal{I}|$. In practical system design one should consider this and have some margin for the design parameter λ.

It follows from (20) that

$$\frac{2^{R_i}-1}{\rho_i} = F_{r_{ii}^2}^{-1}(10^{\lambda}), \quad \text{and} \quad R_i = \log_2(1+\rho_i F_{r_{ii}^2}^{-1}(10^{\lambda})), \tag{21}$$

where $F^{-1}_{r^2_{ii}}(x)$ is the inverse function of the cdf of r^2_{ii}, i.e., $\int_{-\infty}^{F^{-1}_{r^2_{ii}}(x)} f_{r^2_{ii}}(t)dt = x$. The function $F^{-1}_{r^2_{ii}}$ can also be represented by the hyperbola model. With $y_i(x) \triangleq \log_{10} P_{\mathcal{O},i} = \lambda$ we obtain from the hyperbola model (see (5)) that

$$x_i = c_i + \frac{a_i d_i^2 - 400(b_i+\lambda)^2}{40 d_i (b_i+\lambda)},$$

where a_i, b_i, c_i, d_i are the hyperbola parameters corresponding to the ith layer. Defining the normalized SNR $\rho_{\text{norm},i} = \frac{\rho_i}{2^{R_i}-1}$, we obtain from (21) and the definition in (4) that

$$F^{-1}_{r^2_{ii}}(10^{\lambda}) = \rho^{-1}_{\text{norm},i} = 10^{-\frac{x_i}{10}} = 10^{-\left[\frac{c_i}{10} + \frac{a_i d_i^2 - 400(b_i+\lambda)^2}{400 d_i (b_i+\lambda)}\right]}. \tag{22}$$

Now we are ready to study the following problem: given the target outage probability 10^{λ}, what is the minimum power allocation required to achieve the target rate R ? The answer is obtained by solving

$$\begin{array}{ll} \min_{\rho_i} & \sum_{i=1}^N \rho_i \\ \text{subject to} & \sum_{i=1}^N \log_2(1+\rho_i F^{-1}_{r^2_{ii}}(10^{\lambda})) = R\ ,\ \rho_i \geq 0, \end{array} \tag{23}$$

The solution is the well-known "water filling" power allocation [8]

$$\rho_i(\mu) = \left(\mu - \frac{1}{F^{-1}_{r^2_{ii}}(10^{\lambda})}\right)^+, \tag{24}$$

where $(x)^+ = \max\{0, x\}$ and the Lagrange multiplier μ is chosen such that $\sum_{i=1}^N \log_2(1+\rho_i(\mu) F^{-1}_{r^2_{ii}}(10^{\lambda})) = R$. For any outage probability $P_{\mathcal{O},i} = 10^{\lambda}$, the water filling algorithm can be used to find the optimal power allocation and overall input power. Hence the $P_{\mathcal{O},i}$-vs-ρ curves can be easily obtained.

We see that an SMA converts the original MIMO channel into parallel scalar channels with *virtual* channel gains $F^{-1}_{r^2_{ii}}(10^{\lambda}), i = 1, \cdots, N$, determined by the target outage probability. For small target outage probability, the virtual channel gain is small (recall that $F^{-1}_{r^2_{ii}}(x)$ is an increasing function), which requires more input power to achieve the target rate. Moreover, if the ith layer gain is statistically larger than the jth layer, [1] then $F^{-1}_{r^2_{ii}} \geq F^{-1}_{r^2_{jj}}$. According to (24), $\rho_i \geq \rho_j$. Hence the water filling algorithm tends to put more power to the statistically stronger layers, which is intuitively pleasing.

B. Outage Capacity

We now study the outage capacity of the proposed schemes at high SNR. Given the target outage probability ϵ, the subchannel gains $F^{-1}_{r^2_{ii}}(\epsilon)$ are fixed. As the input SNR $\rho \to \infty$, all the N subchannels

[1] We say X is statistically larger than Y if $P(X < a) \leq P(Y < a)$ for $\forall a \in \mathbb{R}$.

will be in use. In this case $\sum_{i=1}^{N}\left(\mu-\frac{1}{F^{-1}_{r_{ii}^2}(\epsilon)}\right)=\rho$,, i.e., $\mu=\frac{\rho}{N}+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{F^{-1}_{r_{ii}^2}(\epsilon)}$, and the water filling power allocation is

$$\rho_i=\mu-\frac{1}{F^{-1}_{r_{ii}^2}(\epsilon)}=\frac{\rho}{N}+\frac{1}{N}\sum_{i=1}^{N}F^{-1}_{r_{ii}^2}(\epsilon)-\frac{1}{F^{-1}_{r_{ii}^2}(\epsilon)},\ i=1,\cdots,N. \tag{25}$$

Therefore ϵ-outage capacity is

$$\begin{aligned} C(\epsilon) &= \sum_{i=1}^{N}\log_2(1+\rho_i F^{-1}_{r_{ii}^2}(\epsilon)) \\ &= N\log_2\left(\frac{\rho}{N}+\frac{1}{N}\sum_{i=1}^{N}F^{-1}_{r_{ii}^2}(\epsilon)\right)+\sum_{i=1}^{N}\log_2 F^{-1}_{r_{ii}^2}(\epsilon) \qquad (26)\\ &= N\log_2\left(\frac{\rho}{N}\right)+\sum_{i=1}^{N}\log_2 F^{-1}_{r_{ii}^2}(\epsilon)+O\left(\frac{1}{\rho}\right),\ \text{as } \rho\to\infty. \qquad (27)\end{aligned}$$

The above expression shows that at high SNR, the ϵ-outage capacities of the SMAs are different from each other by a constant depending on $\sum_{i=1}^{N}\log_2 F^{-1}_{r_{ii}^2}(\epsilon)$. Hence to improve the ϵ-capacity, one should design a decoding ordering rule which maximizes $\sum_{i=1}^{N}\log_2 F^{-1}_{r_{ii}^2}(\epsilon)$. The problem of obtaining the ordering rule that maximizes ϵ-outage capacity is still open. Note that if r_{ii}^2 is statistically larger than $\tilde{r}_{ii}^2$, then $F^{-1}_{r_{ii}^2}(\epsilon)\geq F^{-1}_{\tilde{r}_{ii}^2}(\epsilon)$. Recall from the recursive procedures described in [1, Section IV] that the Greedy QR "greedily" attempts to make the squared layer gains r_{ii}^2 as large as possible. Hence, the Greedy QR can be regarded as an algorithm which maximizes (27) greedily. This observation explains to some extent the superior performance of GRT-SMA.

Again, we note that the actual system outage probability is greater than ϵ but less than $N\epsilon$. Indeed, based on this observation it is easy to show that

$$\begin{aligned} &N\log_2\left(\frac{\rho}{N}\right)+\sum_{i=1}^{N}\log_2 F^{-1}_{r_{ii}^2}(\epsilon/N)+O\left(\frac{1}{\rho}\right)\leq R(\epsilon)\\ &\leq N\log_2\left(\frac{\rho}{N}\right)+\sum_{i=1}^{N}\log_2 F^{-1}_{r_{ii}^2}(\epsilon)+O\left(\frac{1}{\rho}\right),\ \text{as } \rho\to\infty. \qquad (28)\end{aligned}$$

As a benchmark, if $M_t=N=\min\{M_t,M_r\}$, an asymptotically tight lower bound on the outage capacity of the MIMO channel with an unconstrained architecture is derived in [3], which is

$$C_{\text{opt}}(\epsilon)=N\log_2\left(\frac{\rho}{N}\right)+\log_2 C_\infty(\epsilon)+o(1),\ \text{as } \rho\to\infty, \tag{29}$$

where $C_\infty(\epsilon)=\sup\{y:\mathbb{P}(|\mathbf{H}^*\mathbf{H}|<y)\leq\epsilon\}$. The determinant of the Wishart matrix $\mathbf{H}^*\mathbf{H}\in\mathbb{C}^{N\times N}$ can be written as the product of N independent Chi-square random variables. Therefore $C_\infty(\epsilon)$ can be

calculated via numerical integration [3]. Hence at high SNR regime, the ϵ-outage capacity of a SMA are different from $C_{\text{opt}}(\epsilon)$ by some constant which is determined by the cdfs of the layer gains.

V. Numerical Examples

In this section, we present several numerical examples to demonstrate the effectiveness of GRT-SMA.

We include the uniform channel decomposition (UCD) scheme [9] as a benchmark. The UCD scheme also exploits the collaboration between transmitter and receiver through a feedback channel. One implementation of the UCD scheme can be regarded as a precoded MMSE-V-BLAST which has the following nice feature; combined with the precoder, the MMSE-V-BLAST equalizer converts, in a capacity lossless manner, a MIMO channel into multiple *identical* layers with equal output SNRs. The UCD scheme can achieve the optimal diversity-multiplexing (D-M) tradeoff. The drawback of UCD is that it requires feeding back a unitary precoder matrix, which in principle requires infinitely many feedback bits.

In the first example, we consider an iid Rayleigh flat fading channel with $M_t = 4$ and $M_r = 4$. We compare uncoded BER performances of the GRT-SMA and NRT-SMA against the MMSE-V-BLAST with ordered detection, the fixed order RT-VB scheme [10], and UCD. For UCD and the conventional V-BLAST schemes, uniform power and rate (16-QAM) are allocated to each antenna. The GRT-SMA employs the rate and power allocation given in Table I. The simulation results are presented in Figure 4. We see that the diversity gain of each scheme agrees with the D-M gain tradeoff Analysis of [1, Section III]. There is about 1 to 2 dB gap between the GRT-SMA and UCD. Note that in UCD, the receiver needs to feed a unitary precoder matrix back to the transmitter, while for the SMAs, the feedback is 4.59 bits per channel realization. Hence compared to UCD, GRT-SMA trades a very mild performance loss as a price for a significant reduction of feedback overhead.

We then apply the SMAs to non-Rayleigh fading channels. Two typical simulations results are presented here. One is based on spatially correlated channel matrix generated as follows

$$\mathbf{H} = \mathbf{R}_r^{1/2} \tilde{\mathbf{H}} \mathbf{R}_t^{1/2}, \tag{30}$$

where $\tilde{\mathbf{H}}$ is an iid Rayleigh fading channel matrix with $\mathbf{R}_r$ having unit diagonal elements and equal off diagonal elements equal to 0.3 and $\mathbf{R}_t$ having unit diagonal elements and equal off diagonal elements equal to 0.7. The other example is based on the Rician distribution generated by the model

$$\mathbf{H} = \sqrt{\frac{\kappa}{1+\kappa}} \mathbf{\Theta} + \sqrt{\frac{1}{1+\kappa}} \tilde{\mathbf{H}}. \tag{31}$$

Here $\mathbf{\Theta}$ is a random matrix with iid elements whose amplitudes are one and phases are uniformly distributed between 0 and 2π. In the simulation, we set the K-factor $\kappa = 5$. For both examples, the GRT-SMA still adopts the rate and power allocation given in Table I, which is clearly suboptimal since Table I is obtained under the assumption of iid Rayleigh fading. The simulation results are given in Figure 5. Figure 5(a) shows that the gap between UCD and the GRT-SMA with Greedy Ordering is slightly larger compared to Figure 4. However, the relative advantage of the GRT-SMA over the open-loop schemes remains significant. In Figure 5(b) the gap between GRT-SMA and UCD is still less than 2 dB. Hence we have seen that the GRT-SMA is robust against the inaccurate assumption of the channel fading statistics.

Figure 6 compares the outage probabilities of the NRT-SMA, the GRT-SMA, the fixed order RT-VB schemes, which are given in (18), with the channel outage probability given in (19). The rate/power allocation algorithm introduced in Section IV is used for the SMAs as well as the fixed order RT-VB schemes. The outage probabilities are averaged over 10^7 independent channel realizations. We note that the GRT-SMA suffers about 2 dB loss compared to the fundamental limit. However, it has comparable performance to the V-BLAST scheme with a maximum likelihood (ML) receiver (denoted VB+ML), in particular, GRT-SMA has significantly higher diversity gain than VB+ML. Figure 7 illustrates the gap between the overall system outage probability (18) with the nominal outage probability of each layer 10^{λ} (20). This figure gives us an idea about how much margin one should allow to design for a nominal outage probability.

Figure 8 compares the ϵ-outage capacities of the SMAs, the fixed order RT-VB scheme, and the channel outage capacity. The results are obtained via Monte Carlo trials of the iid Rayleigh fading channel. As predicted at the end of Section IV, the SMAs have the same rate of increase of the ϵ-capacity as $\rho \to \infty$, while there are constant gaps between those of the SMAs and the optimal. The GRT-SMA has higher ϵ-capacity than NRT-SMA, which in turn is better than the fixed order RT-VB. We also note that compared to the optimal, the rate loss of the SMAs is smaller in the low SNR regime.

Finally, we combine the GRT-SMA scheme with trellis coded modulation (TCM) [11]. The standard 64-state Ungerboeck code is applied to all the layers in use. The rates and powers are allocated according to Table I, i.e., we use the 64-QAM on the first two layers and 16-QAM on the third layer. We apply three TCM codes with code rate 5/6, 5/6, and 3/4, respectively. The total information rate is therefore $5 + 5 + 3 = 13$ bps/Hz. We take each packet to contain 10400 information bits. The packet error rate (PER) curve is shown in Figure 9. Also included as benchmark is the fundamental outage probability

and the outage probability of the system applying GRT-SMA. It can be seen that the PER is only about 3 dB from the outage probability of GRT-SMA and about 5 dB away from the fundamental outage probability at PER= 0.1. This gap can be further reduced with better codes.

VI. Conclusion and Discussion

In this two-part paper, a framework is developed for jointly designing channel-dependent ordered decoding at the receiver and rate and power allocation at the transmitter. The joint design is facilitated by feeding a few ($\leq \log_2(M_t!)$) bits from receiver back to the transmitter with regard to the decoding order. This framework encompasses a class of new SMAs which admit independent scalar coding for the multiplexed substreams. The D-M gain tradeoffs of the class of SMAs is analyzed. Two SMAs based on two special decoding orderings are proposed. One is called the Norm ordering Rate Tailored SMA (NRT-SMA), and the other the Greedy ordering Rate Tailored SMA (GRT-SMA). The latter is shown to have the optimal diversity-multiplexing (D-M) gain tradeoff among the class, which in turn is quite close to the fundamental D-M gain tradeoff of the MIMO channel. Compared to the classic V-BLAST, the only added complexity of the proposed SMAs is that (i) for each channel realization the receiver feeds no more than $\log_2(M_t!)$ bits on the decoding order back to the transmitter, and (ii) the transmitter applies rate and power allocation according to a lookup table. Although the added complexity is modest, the proposed SMAs, especially the GRT-SMA has dramatic performance improvement over classical V-BLAST, which is proven by both theoretical analyses and numerical experiments.

Because the SMAs apply independent coding for the multiplexed substreams, they are also applicable to the multi-access channel (MAC) communications. The superior D-M tradeoff performance of the GRT-VB suggests that there are MAC schemes which have better D-M tradeoff than that given in [12][13]. This is not surprising though. The D-M tradeoff in [12][13] applies to the scenario where multi-users compete with each other without any coordination. For the proposed class of SMAs, however, the substreams are transmitted in a somewhat coordinated manner, facilitated by the low rate feedback. The remarkable performance of the GRT-SMA clearly demonstrates the benefit of CSI feedback.

In pursuing an simple solution to the optimal rates and power allocation problem, we proposed a four-parameter hyperbola model which can be used to closely approximate the error (outage) probability. This result may be of independent interest and find applications in other problems involving the optimization of wireless communication systems.

References

[1] Y. Jiang and M. Varanasi, "A class of spatial multiplexing architecture combining rate-tailored transmission and ordered BLAST detection Part I: On ordered detection," *IEEE Transactions on Wireless Communications*, Submitted. 2006.

[2] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[3] N. Prasad and M. K. Varanasi, "Outage capacities of space-time architecture," *Information Theory Workshop, San Antonio, Texas*, October 24-29 2004.

[4] K. Cho and D. Yoon, "On the general BER expression of one and two-dimensional amplitude modulations," *IEEE Transactions on Communications*, vol. 50, pp. 1074–1080, July 2002.

[5] J. G. D. Forney and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *IEEE Transactions on Information Theory*, October 1998.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge Press, 2005.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. 2002.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc, 1991.

[9] Y. Jiang, J. Li, and W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Transactions on Signal Processing*, vol. 53, pp. 4283 – 4294, November 2005.

[10] N. Prasad and M. K. Varanasi, "Analysis of decision feedback detection for MIMO rayleigh fading channels and the optimization of rate and power allocations," *IEEE Transactions on Information Theory*, vol. 50, pp. 1009–1025, June 2004.

[11] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets part i: introduction," *IEEE Communications Magazine*, vol. 25-2, pp. 5–11, 1987.

[12] N. Prasad and M. K. Varanasi, "Outage analysis and optimization for multiaccess/V-BLAST architecture over MIMO Rayleigh fading channels," *41st Annual Allerton Conf. on Comm. Control, and Comput., Monticello, IL*, October 2003.

[13] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple access channel," *IEEE Transactions on Information Theory*, vol. 50, pp. 1859–1874, September 2004.

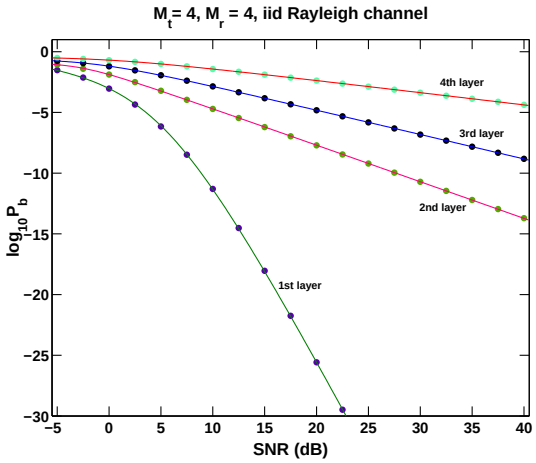


Fig. 1. BER-vs-SNR curves of the layers with 16-QAM input. The layers are obtained via Norm QR decomposition. The dots "•" are plotted according to numerical integration of $\mathbb{E}_{r_{ii}^2}\left[P_b\left(r_{ii}^2\rho_i, M_i\right)\right]$. The solid lines are the fitting hyperbolas. $M_t = 4$ and $M_r = 4$.

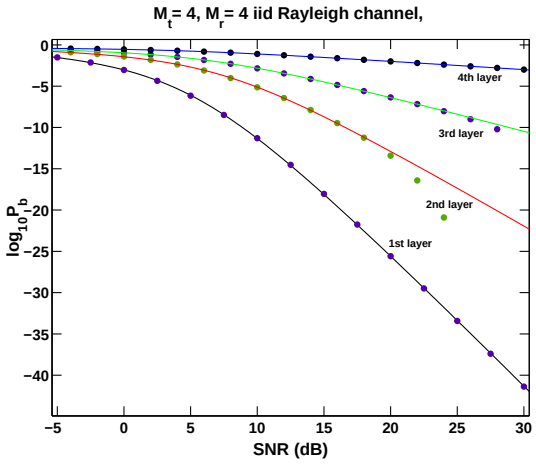


Fig. 2. BER-vs-SNR curves of the layers with BPSK input. The layers are obtained via Greedy QR decomposition. The dots "•" are plotted via numerical integration (the first layer) or 10^8 Monte Carlo trials (the other layers). The solid lines are the fitting hyperbolas. $M_t = M_r = 4$.

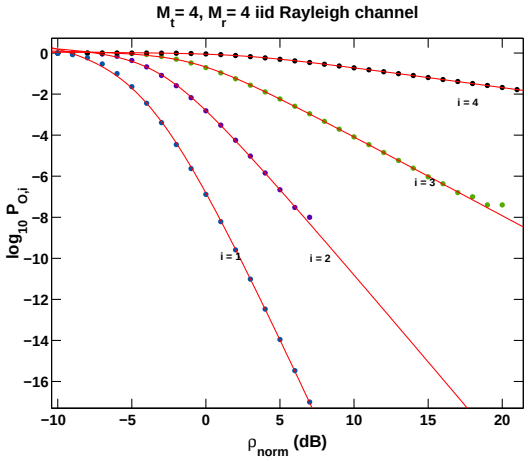


Fig. 3. Outage probabilities of the four layers obtained through Greedy QR. The dots "•" are plotted according to the estimation of $\mathbb{P}\left(r_{ii}^2 < \rho_{\mathrm{norm,i}}^{-1}\right)$ via averaging 10^7 Monte Carlo trials. The solid lines are the fitting hyperbolas.

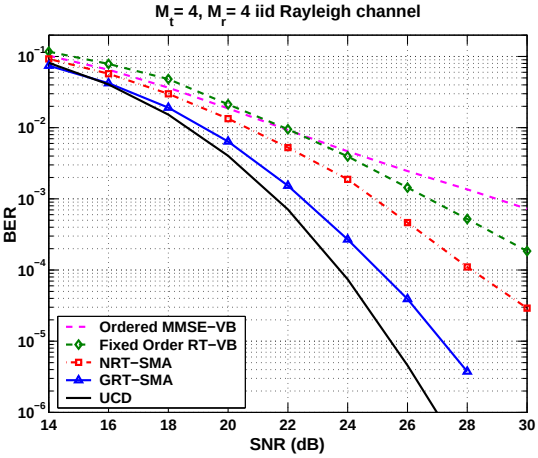
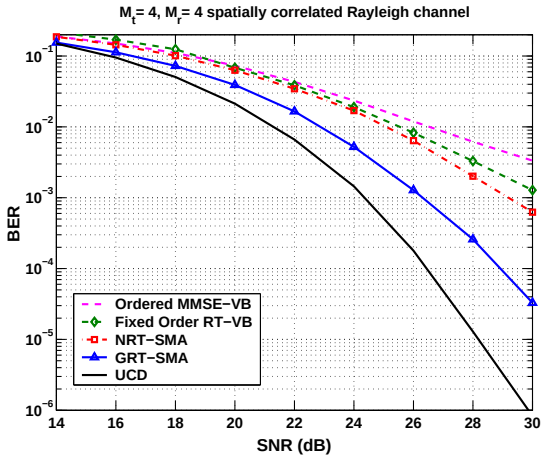
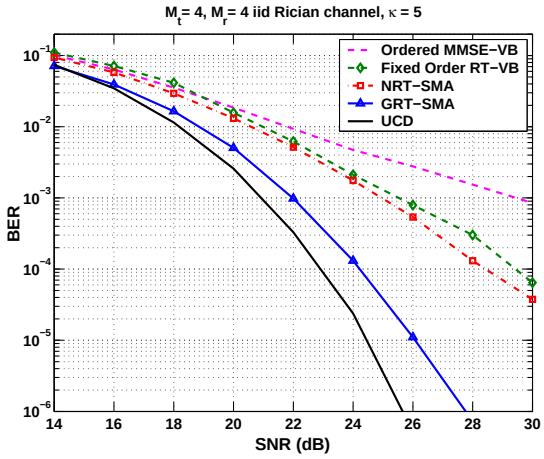


Fig. 4. BER performances of the ordered MMSE-SMA, fixed ordering RT-SMA the NRT-SMA, GRT-SMA, and the UCD scheme Results are based on 5000 Monte Carlo trials of an iid Rayleigh fading channel with $M_t = 4$ and $M_r = 4$.



(a)



(b)

Fig. 5. $M_t = 4$ and $M_r = 4$. (a) spatially correlated fading (b) Rician fading channel with K-factor $\kappa = 5$.

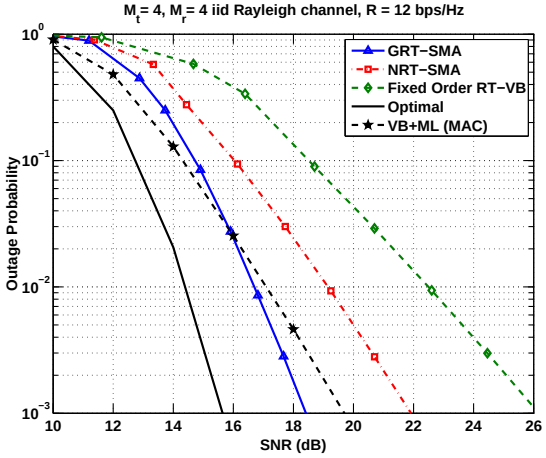


Fig. 6. Comparison of outage probability performances of the three RT-SMA schemes with the optimal channel outage probability of single user MIMO and MAC MIMO. $M_t = 4$ and $M_r = 4$ iid Rayleigh channel.

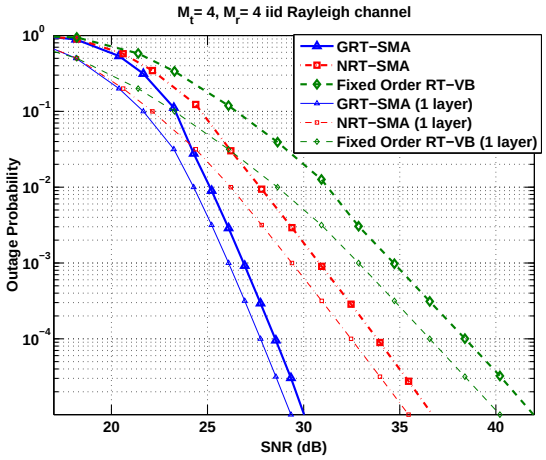


Fig. 7. Comparison of overall system outage probability with the outage probability of a single layer. $M_t = 4$ and $M_r = 4$ iid Rayleigh channel. $R = 20$.

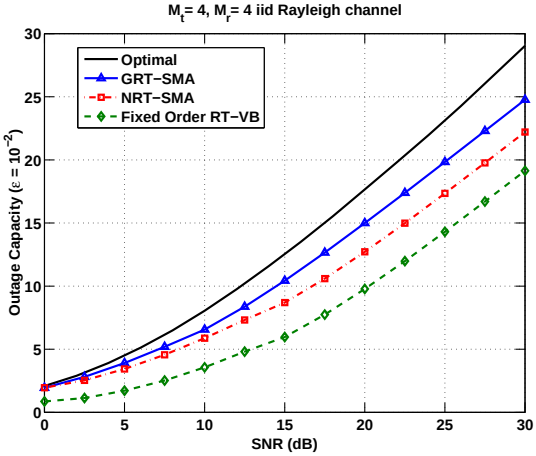


Fig. 8. Comparison of ϵ-outage capacity ($\epsilon = 10^{-2}$) of the three RT-SMA schemes with the channel outage capacity. $M_t = 4$ and $M_r = 4$ iid Rayleigh channel.

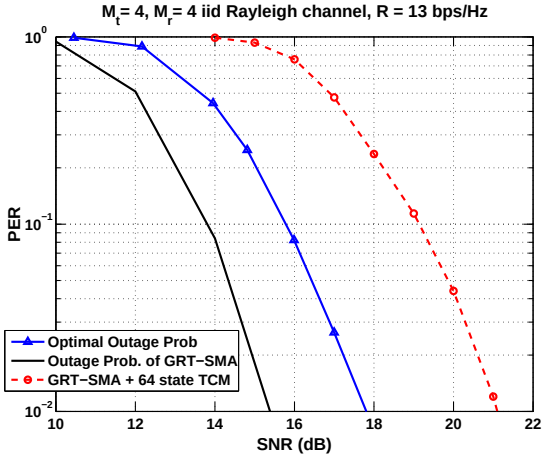


Fig. 9. Packet error rate of GRT-SMA using 64-state TCM. $M_t = 4$ and $M_r = 4$ iid Rayleigh channel. Packet length is 10400 bits.