Joint Source-Channel Codes for MIMO Block Fading Channels

Deniz Gündüz, Elza Erkip

Abstract

We consider transmission of a continuous amplitude source over an L-block Rayleigh fading $M_t \times M_r$ MIMO channel when the channel state information is only available at the receiver. Since the channel is not ergodic, Shannon's source-channel separation theorem becomes obsolete and the optimal performance requires a joint source -channel approach. Our goal is to minimize the expected end-to-end distortion, particularly in the high SNR regime. The figure of merit is the distortion exponent, defined as the exponential decay rate of the expected distortion with increasing SNR. We provide an upper bound and lower bounds for the distortion exponent with respect to the bandwidth ratio among the channel and source bandwidths. For the lower bounds, we analyze three different strategies based on layered source coding concatenated with progressive, superposition or hybrid digital/analog transmission. In each case, by adjusting the system parameters we optimize the distortion exponent as a function of the bandwidth ratio. We prove that the distortion exponent upper bound can be achieved when the channel has only one degree of freedom, that is $L = 1$, and $\min\{M_t, M_r\} = 1$. When we have more degrees of freedom, our achievable distortion exponents meet the upper bound for only certain ranges of the bandwidth ratio. We demonstrate that our results, which were derived for a complex Gaussian source, can be extended to more general source distributions as well.

Index Terms

Broadcast codes, distortion exponent, diversity-multiplexing gain tradeoff, hybrid digital/analog coding, joint source-channel coding, multiple input- multiple output (MIMO), successive refinement.

I. INTRODUCTION

Recent advances in mobile computing and hardware technology enable transmission of rich multimedia contents over wireless networks. Examples include digital TV, voice and video transmission over cellular and wireless LAN networks, and sensor networks. With the high demand for such services, it becomes crucial to identify the system limitations, define the appropriate performance metrics, and to design

wireless systems that are capable of achieving the best performance by overcoming the challenges posed by the system requirements and the wireless environment. In general, multimedia wireless communication requires transmitting analog sources over fading channels while satisfying the end-to-end average distortion and delay requirements of the application within the power limitations of the mobile terminal.

Multiple antennas at the transceivers have been proposed as a viable tool that can remarkably improve the performance of multimedia transmission over wireless channels. The additional degrees of freedom provided by multiple input multiple output (MIMO) system can be utilized in the form of spatial multiplexing gain and/or spatial diversity gain, that is, either to transmit more information or to increase the reliability of the transmission. The tradeoff between these two gains is explicitly characterized as the diversity-multiplexing gain tradeoff (DMT) in [1]. How to translate this tradeoff into an improved overall system performance depends on the application requirements and limitations.

In this paper, we consider transmission of a continuous amplitude source over a MIMO block Rayleigh fading channel. We are interested in minimizing the end-to-end average distortion of the source. We assume that the instantaneous channel state information is only available at the receiver (CSIR). We consider the case where K source samples are to be transmitted over L fading blocks spanning N channel uses. We define the *bandwidth ratio* of the system as

$$b = \frac{N}{K} \text{ channel uses per source sample,} \tag{1}$$

and analyze the system performance with respect to b. We assume that K is large enough to achieve the rate-distortion performance of the underlying source, and N is large enough to design codes that can achieve all rates below the instantaneous capacity of the block fading channel.

We are particularly interested in the high SNR behavior of the expected distortion (ED) which is characterized by the *distortion exponent* [2]:

$$\Delta = - \lim_{SNR \to \infty} \frac{\log ED}{\log SNR}. \tag{2}$$

Shannon's fundamental source-channel separation theorem does not apply to our scenario as the channel is no more ergodic. Thus, the optimal strategy requires a joint source-channel coding approach. The minimum expected end-to-end distortion depends on the source characteristics, the distortion metric, the power constraint of the transmitter, the joint compression, channel coding and transmission techniques used.

Since we are interested in the average distortion of the system, this requires a strategy that performs 'well' over a range of channel conditions. Our approach is to first compress the source into multiple layers, where each layer successively refines the previous layers, and then transmit these layers at varying rates, hence providing unequal error protection so that the reconstructed signal quality can be adjusted to the

instantaneous fading state without the availability of the channel state information at the transmitter (CSIT). We consider transmitting the source layers either progressively in time, *layered source coding with progressive transmission* (LS), or simultaneously by superposition, *broadcast strategy with layered source* (BS). We also discuss a hybrid digital-analog extension of LS called *hybrid-LS* (HLS).

The characterization of distortion exponent for fading channels has recently been investigated in several papers. Distortion exponent is first defined in [2], and simple transmission schemes over two parallel fading channels are compared in terms of distortion exponent. Our prior work includes maximizing distortion exponent using layered source transmission for cooperative relay [3], [4], [5], for SISO [6], for MIMO [7], and for parallel channels [8]. Holliday and Goldsmith [9] analyze high SNR behavior of the expected distortion for single layer transmission over MIMO without explicitly giving the achieved distortion exponent. Hybrid digital-analog transmission, first proposed in [10] for the Gaussian broadcast channel, is considered in terms of distortion exponent for MIMO channel in [11].

Others have focused on minimizing the end-to-end source distortion for general SNR values [15], [16]. Recently, the LS and BS strategies introduced here have been analyzed for finite SNR and finite number of source layers in [17]-[22].

This paper derives explicit expressions for the achievable distortion exponent of LS, HLS and BS strategies, and compares the achievable exponents with an upper bound derived by assuming perfect channel state information at the transmitter. Our results reveal the following:

- LS strategy, which can easily be implemented by concatenating a layered source coder with a MIMO channel encoder that time-shares among different code rates, improves the distortion exponent compared to the single-layer approach of [9] even with limited number of source layers. However, the distortion exponent of LS still falls short of the upper bound.

- While the hybrid digital-analog scheme meets the distortion exponent upper bound for small bandwidth ratios as shown in [11], the improvement of hybrid extension of LS (HLS) over pure progressive layered digital transmission (LS) becomes insignificant as the bandwidth ratio or the number of digital layers increases.

- Transmitting layers simultaneously as suggested by BS provides the optimal distortion exponent for all bandwidth ratios when the system has one degree of freedom, i.e., for single block MISO/SIMO systems, and for high bandwidth ratios for the general MIMO system. Hence, for the mentioned cases the problem of characterizing the distortion exponent is solved.

- There is a close relationship between the DMT of the underlying MIMO channel and the achievable distortion exponent of the proposed schemes. For LS and HLS, we are able to give an explicit characterization of the achievable distortion exponent once the DMT of the system is provided. For

BS, we enforce successive decoding at the receiver and the achievable distortion exponent closely relates to the '*successive decoding diversity-multiplexing tradeoff*' which will be rigorously defined in Section V.

- The correspondence between source transmission to a single user with unknown noise variance and multicasting to users with different noise levels [10] suggests that, our analysis would also apply to the multicasting case where each receiver has the same number of antennas and observes an independent block fading Rayleigh channel possibly with a different mean. Here the goal is to minimize the expected distortion of each user. Alternatively, each user may have a static channel, but the channel gains over the users may be randomly distributed with independent Rayleigh distribution, where the objective is to minimize the source distortion averaged over the users.

- While minimizing the end-to-end distortion for finite SNR is still an open problem, in the high SNR regime we are able to provide a complete solution in certain scenarios. Using this high SNR analysis, it is also possible to generalize the results to non-Gaussian source distributions. Furthermore, LS and BS strategies motivate source and channel coding strategies that are shown to perform very well for finite SNRs as well [17]-[22].

We use the following notation throughout the paper. $E[\cdot]$ is the expectation, $f(x) \doteq g(x)$ is the exponential equality defined as $\lim_{x \to \infty} \frac{\log f(x)}{\log g(x)} = 1$, while $\dot{\geq}$ and $\dot{\leq}$ are defined similarly. Vectors and matrices are denoted with bold characters, where matrices are in capital letters. $[\cdot]^T$ and $[\cdot]^\dagger$ are the transpose and the conjugate transpose operations, respectively. $\mathbf{tr}(\mathbf{A})$ is the the trace of matrix $\mathbf{A}$. For two Hermitian matrices $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive-semidefinite. $(x)^+$ is x if $x \geq 0$, and 0 otherwise. We denote the set $\{[x_1, \ldots, x_n] : x_i \in \mathbb{R}_+, \forall i\}$ by $\mathbb{R}^{n+}$.

II. SYSTEM MODEL

We consider a discrete time continuous amplitude (analog) source $\{s_k\}_{k=1}^{\infty}$, $s_k \in \mathbb{R}$ available at the transmitter. For the analysis, we focus on a memoryless, i.i.d., complex Gaussian source with independent real and imaginary components each with variance 1/2. We use the distortion-rate function of the complex Gaussian source $D(R) = 2^{-R}$ where R is the source coding rate in bits per source sample, and consider compression strategies that meet the distortion-rate bound. Although in Sections III-VI we use properties of this complex Gaussian source (such as its distortion-rate function and successive refinability), in Section VIII we prove that our results can be extended to any complex source with finite second moment and finite differential entropy, with squared-error distortion metric. As stated in Section I, we assume that K source samples are transmitted in N channel uses which corresponds to a bandwidth ratio of $b = N/K$. In all our derivations we allow for an arbitrary bandwidth ratio $b > 0$.

We assume a MIMO block fading channel with M_t transmit and M_r receive antennas. The channel model is

$$\mathbf{y}[i] = \sqrt{\frac{SNR}{M_t}}\mathbf{H}[i]\mathbf{x}[i] + \mathbf{z}[i], \qquad i = 1, \ldots, N \tag{3}$$

where $\sqrt{\frac{SNR}{M_t}}\mathbf{x}[i]$ is the transmitted signal at time i, $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{C}^{M_r \times N}$ is the complex Gaussian noise with i.i.d entries $\mathcal{CN}(0, 1)$, and $\mathbf{H}[i] \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix at time i, which has i.i.d. entries with $\mathcal{CN}(0, 1)$. We have an L-block fading channel, that is, the channel observes L different i.i.d. fading realizations $\mathbf{H}_1, \ldots, \mathbf{H}_L$ each lasting for N/L channel uses. Thus we have

$$\mathbf{H}\left[k\frac{N}{L} + 1\right] = \mathbf{H}\left[k\frac{N}{L} + 2\right] = \ldots = \mathbf{H}\left[(k+1)\frac{N}{L}\right] = \mathbf{H}_{k+1}, \tag{4}$$

for $k = 0, \ldots, L-1$ assuming N/L is integer. The realization of the channel matrix $\mathbf{H}_i$ is assumed to be known by the receiver and unknown by the transmitter, while the transmitter knows the statistics of $\mathbf{H}_i$. The codeword, $\mathbf{X} = \left[\mathbf{x}[1], \ldots, \mathbf{x}_L\right] \in \mathbb{C}^{M_t \times N}$ is normalized so that it satisfies $tr(E[\mathbf{X}^\dagger \mathbf{X}]) \leq M_t N$. We assume Gaussian codebooks which can achieve the instantaneous capacity of the MIMO channel. We define $M_* = \min\{M_t, M_r\}$ and $M^* = \max\{M_t, M_r\}$.

The source is transmitted through the channel using one of the joint source-channel coding schemes discussed in this paper. In general, the source encoder matches the K-length source vector $\mathbf{s}^K = [s_1, \ldots, s_K]$ to the channel input $\mathbf{X}$. The decoder maps the received signal $\mathbf{Y} = \left[\mathbf{y}[1], \ldots, \mathbf{y}[N]\right] \in \mathbb{C}^{M_r \times N}$ to an estimate $\hat{\mathbf{s}} \in \mathbb{C}^K$ of the source. Average distortion $ED(SNR)$ is defined as the average mean squared error between $\mathbf{s}$ and $\hat{\mathbf{s}}$ at average channel signal-to-noise ratio SNR, where the expectation is taken with respect to all source samples, channel realizations and the channel noise. The exact expression of $ED(SNR)$ for the strategies introduced will be provided in the respective sections.

As mentioned in Section I, we are interested in the high SNR behavior of the expected distortion. We optimize the system performance to maximize the *distortion exponent* defined in Eqn. (2). A distortion exponent of Δ means that the expected distortion decays as $SNR^{-\Delta}$ with increasing SNR when SNR is high.

In order to obtain the end-to-end distortion for our proposed strategies, we will need to characterize the error rate of the MIMO channel. Since we are interested in the high SNR regime, we use the outage probability, which has the same exponential behavior as the channel error probability [1]. For a family of codes with rate $R = r \log SNR$, r is defined as the multiplexing gain of the family, and

$$d(r) = -\lim_{SNR \to \infty} \frac{\log P_{out}(SNR)}{\log SNR} \tag{5}$$

as the diversity advantage, where $P_{out}(SNR)$ is the outage probability of the code. The diversity gain $d^*(r)$ is defined as the supremum of the diversity advantage over all possible code families with

multiplexing gain r. In [1], it is shown that there is a fundamental tradeoff between multiplexing and diversity gains, also known as the diversity-multiplexing gain tradeoff (DMT), and this tradeoff is explicitly characterized with the following theorem.

Theorem 2.1: (Corollary 8, [1]) For an $M_t \times M_r$ MIMO L-block fading channel, the optimal tradeoff curve $d^*(r)$ is given by the piecewise-linear function connecting the points $(k, d^*(k))$, $k = 0, 1, \ldots, M_*$, where

$$d^*(k) = L(M_t - k)(M_r - k). \tag{6}$$

III. DISTORTION EXPONENT UPPER BOUND

Before we study the performance of various source-channel coding strategies, we calculate an upper bound for the distortion exponent of the MIMO L-block fading channel, assuming that the transmitter has access to perfect channel state information at the beginning of each block. Then the source-channel separation theorem applies for each block and transmission at the highest rate is possible with zero outage probability.

Theorem 3.1: For transmission of a memoryless i.i.d. complex Gaussian source over an L-block $M_t \times M_r$ MIMO channel, the distortion exponent is upper bounded by

$$\Delta^{UB} = L \sum_{i=1}^{M_*} \min\left\{ \frac{b}{L}, 2i - 1 + |M_t - M_r| \right\}. \tag{7}$$

Proof: Proof of the theorem can be found in Appendix I. ∎

Note that, increasing the number of antennas at either the transmitter or the receiver by one does not provide an increase in the distortion exponent upper bound for $b < L(1 + |M_t - M_r|)$, since the performance in this region is bounded by the bandwidth ratio. Adding one antenna to both sides increases the upper bound for all bandwidth ratios, while the increase is more pronounced for higher bandwidth ratios. The distortion exponent is bounded by the highest diversity gain LM_tM_r.

In the case of $M \times 1$ MISO system, and alternatively $1 \times M$ SIMO system, the upper bound can be simplified to

$$\Delta^{UB}_{MISO/SIMO} = \min\{b, LM\}. \tag{8}$$

We next discuss how a simple transmission strategy consisting of single layer digital transmission performs with respect to the upper bound. In single layer digital transmission, the source is first compressed at a specific rate bR, the compressed bits are channel coded at rate R, and then transmitted over the channel. This is the approach taken in [2] for two-parallel channels, in [3] for cooperative relay channels, and in [9] for the MIMO channel to transmit an analog source over a fading channel. Even though
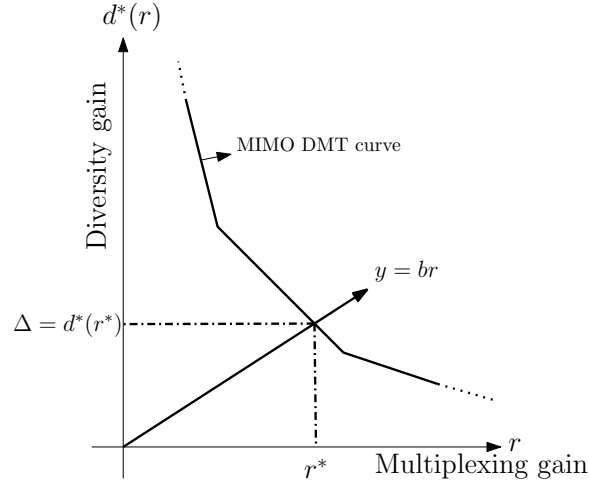
Fig. 1. A geometric interpretation illustrating the optimal multiplexing gain for a single layer source-channel coding system. The intersection point of the DMT curve and the line $y=br$ gives the optimal multiplexing gain- distortion exponent pair.

compression and channel coding are done separately, the rate is a common parameter that can be chosen to minimize the end-to-end distortion. Note that the transmitter choses this rate R without any channel state information.

The expected distortion of single layer transmission can be written as

$$ED(R, SNR) = (1 - P_{out}(R, SNR))D(bR) + P_{out}(R, SNR), \quad (9)$$

where $P_{out}(R, SNR)$ is the outage probability at rate R for given SNR, and $D(R)$ is the distortion-rate function of the source. Here we assume that, in case of an outage, the decoder simply outputs the mean of the source leading to the highest possible distortion of 1 due to the unit variance assumption. At fixed SNR, there is a tradeoff between reliable transmission over the channel (through the outage probability), and increased fidelity in source reconstruction (through the distortion-rate function). This suggests that there is an optimal transmission rate that achieves the optimal average distortion. For any given SNR this optimal R can be found using the exact expressions for $P_{out}(R, SNR)$ and $D(R)$.

In order to study the distortion exponent, we concentrate on the high SNR approximation of Eqn. (9). To achieve a vanishing expected distortion in the high SNR regime we need to increase R with SNR. Scaling R faster than $O(\log SNR)$ would result in an outage probability of 1, since the instantaneous channel capacity of the MIMO system scales as $M_* \log SNR$. Thus we assign $R = r \log SNR$, where $0 \le r \le M_*$. Then the high SNR approximation of Eqn. (9) is

$$
\begin{aligned}
ED &\doteq D(bR) + P_{out}(R), \\
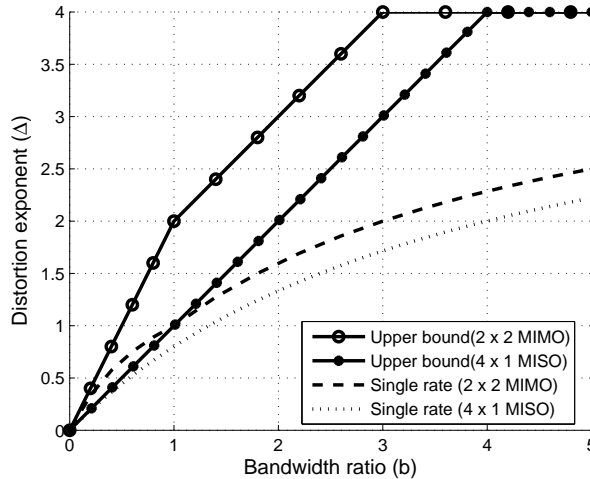&\doteq SNR^{-br} + SNR^{-d^*(r)}.
\end{aligned}
\quad (10)
$$

Fig. 2. Upper bound and single layer achievable distortion exponents for 2×2 and 4×1 MIMO systems.

Of the two terms, the one with the highest SNR exponent would be dominant in the high SNR regime. Maximum distortion exponent is achieved when both terms have the same SNR exponent. Then the optimal multiplexing gain r^* satisfies

$$\Delta \triangleq br^* = d^*(r^*), \tag{11}$$

where Δ is the corresponding distortion exponent. Eqn. (11) suggests an optimal operating point on the DMT curve to maximize the distortion exponent of the single layer scheme.

Figure 1 shows a geometric illustration of the optimal multiplexing gain and the corresponding distortion exponent. A similar approach was taken in [9] for single layer transmission with the restriction of integer multiplexing gains, and later extended to all multiplexing gains in [11]. However, as we argue next, even when all multiplexing gains are considered, this single layer approach is far from exploiting all the resources provided by the system.

In Figure 2, we illustrate the distortion exponent upper bound and the distortion exponent of the single layer scheme for 4×1 MISO and 2×2 MIMO systems. We observe a significant gap between the upper bounds and the single layer distortion exponents in both cases for all bandwidth ratios. This gap gets larger with increasing degrees of freedom and increasing bandwidth ratio.

The major drawback of the single layer digital scheme is that it suffers from the threshold effect, i.e., error probability is bounded away from zero or an outage occurs when the channel quality is worse than a certain threshold, which is determined by the attempted rate. Furthermore, single layer digital transmission cannot utilize the increase in the channel quality beyond this threshold. Lack of CSIT

makes only a statistical optimization of the compression/transmission rate possible. To make the system less sensitive to the variations in the channel quality, we will concentrate on layered source coding where the channel codewords corresponding to different layers are assigned different rates. Using the successive refinability of the source, we transmit more important compressed bits with higher reliability. The additional refinement bits are received when the channel quality is high. This provides adaptation to the channel quality without the transmitter actually knowing the instantaneous fading levels. We argue that, due to the exponential decay of the distortion-rate function in general, layering increases the overall system performance from the distortion exponent perspective. Our analysis in the following sections proves this claim.

IV. Layered Source Coding with Progressive Transmission and Hybrid Digital-Analog Extension

The first source-channel coding scheme we consider is based on compression of the source in layers, where each layer is a refinement of the previous ones, and transmission of these layers successively in time using channel codes of different rates. We call this scheme *layered source coding with progressive transmission* (LS). This classical idea, mostly referred as progressive coding, has been used to various extents in the image and video standards such as JPEG2000 and MPEG-4. After analyzing the distortion exponent of LS in Section IV-A, in Section IV-B we consider a hybrid digital-analog extension called *hybrid LS* (HLS) where the error signal is transmitted without coding. In this section we analyze single block fading, i.e., $L = 1$, for clarity of presentation. Generalization to the multiple block case ($L > 1$) will be a straightforward extension of the techniques presented here and will be briefly discussed in Section VI.

A. Layered Source Coding with Progressive Transmission (LS)

We assume that the source encoder has n layers with each layer transmitted over the channel at rate R_i bits per channel use in $t_i N$ channel uses for $i = 1, 2, \ldots, n$, with $\sum_{i=1}^{n} t_i = 1$. This is illustrated in Fig. 3(a). We assume that $t_i N$ is large enough to approach the instantaneous channel capacity. For each layer this corresponds to a source coding rate of $b t_i R_i$ bits per sample, where b is the bandwidth ratio defined in (1). The ith layer is composed of the successive refinement bits for the previous $i - 1$ layers. The transmission power is kept constant for each layer, so the optimization variables are the rate vector $\mathbf{R} = [R_1, \ldots, R_n]$ and the channel allocation vector $\mathbf{t} = [t_1, \ldots, t_n]$.

Let P_{out}^i denote the outage probability of layer i, i.e., $P_{out}^i = Pr\{C(\mathbf{H}) < R_i\}$. Using successive refinability of the complex Gaussian source [23], the distortion achieved when the first i layers are

(a) LS strategy with n layers.

(b) HLS strategy with n layers for $b > 1/M_*$.

(c) BS strategy with n layers. Layers are transmitted simultaneously with total power allocated among them.
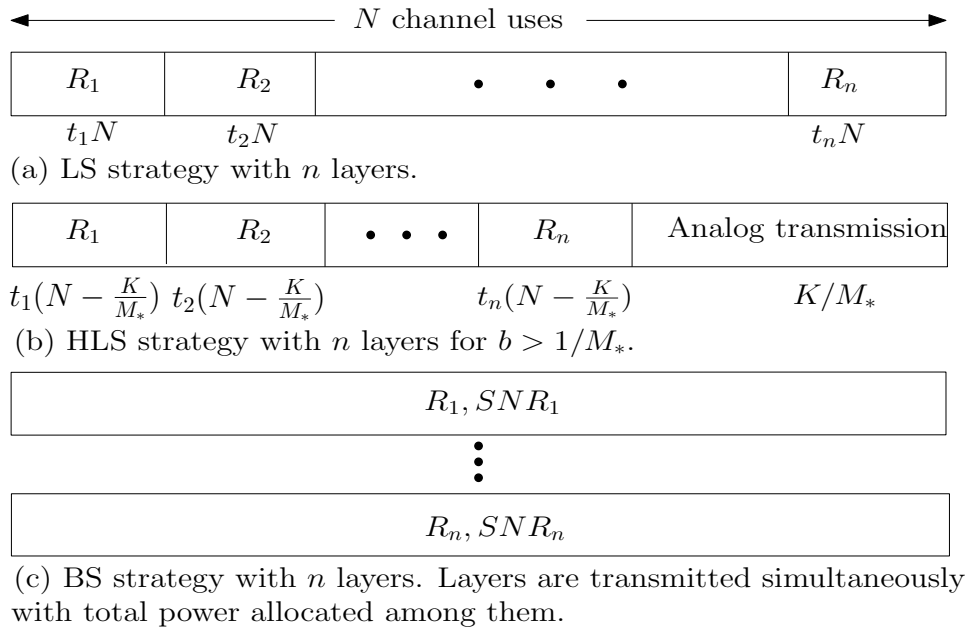
Fig. 3.   Channel and power allocation for different transmission strategies explored in the paper.

successfully decoded is

$$
\begin{aligned}
D_i^{LS} &= D\left(b\sum_{k=1}^{i} t_k R_k\right), \\
&= 2^{-b\sum_{k=1}^{i} t_k R_k},
\end{aligned}
\tag{12}
$$

with $D_0^{LS} = 1$. Note that due to successive refinement source coding, a layer is useless unless all the preceding layers are received successfully. This imposes a non-decreasing rate allocation among the layers, i.e., $R_i \leq R_j$ for any $j > i$. Then the expected distortion (ED) for such a rate allocation can be written as

$$
ED(\mathbf{R}, \mathbf{t}, SNR) = \sum_{i=0}^{n} D_i^{LS} \cdot \left(P_{out}^{i+1} - P_{out}^{i}\right),
\tag{13}
$$

where we define $P_{out}^0 = 0$ and $P_{out}^{n+1} = 1$.

The minimization problem to be solved is

$$
\begin{aligned}
\min_{\mathbf{R}, \mathbf{t}} \quad & ED(\mathbf{R}, \mathbf{t}, SNR) \\
\text{s.t.} \quad & \sum_{i=1}^{n} t_i = 1, \\
& t_i \geq 0, \text{ for } i = 1, \ldots, n \\
& 0 \leq R_1 \leq R_2 \leq \cdots \leq R_n.
\end{aligned}
\tag{14}
$$

This is a non-linear optimization problem which can be untractable for a given SNR. An algorithm solving the above optimization problem for finite SNR is proposed in [17]. However when we focus on the high SNR regime and compute the distortion exponent Δ, we will be able to obtain explicit expressions.

In order to have a vanishing expected distortion in Eqn. (13) with increasing SNR, we need to increase the transmission rates of all the layers with SNR as argued in the single layer case. We let the multiplexing gain vector be $\mathbf{r} = [r_1, \ldots, r_n]^T$, hence $\mathbf{R} = \mathbf{r} \log SNR$. The ordering of rates is translated into multiplexing gains as $0 \leq r_1 \leq \cdots \leq r_n$. Using the DMT of the MIMO system under consideration and the distortion-rate function of the complex Gaussian source, we get

$$
\begin{aligned}
ED(\mathbf{R}, SNR) &\doteq \sum_{k=0}^{n} \left[SNR^{-d^*(r_{k+1})} - SNR^{-d^*(r_k)} \right] SNR^{-b\sum_{i=1}^{k} t_i r_i} \\
&\doteq \sum_{k=0}^{n} SNR^{-d^*(r_{k+1})} SNR^{-b\sum_{i=1}^{k} t_i r_i} \\
&\doteq SNR^{\max_{0 \leq k \leq n} \left\{ -d^*(r_{k+1}) - b\sum_{i=1}^{k} t_i r_i \right\}},
\end{aligned}
\tag{15}
$$

where $d^*(r_{n+1}) = 0$, and the last exponential equality arises because the summation will be dominated by the slowest decay in high SNR regime. Then the optimal LS distortion exponent can be written as

$$
\Delta_n^{LS} = \max_{\mathbf{r}, \mathbf{t}} \min_{0 \leq k \leq n} \left\{ d^*(r_{k+1}) + b\sum_{i=1}^{k} t_i r_i \right\}
\tag{16}
$$

$$
\text{s.t.} \quad \sum_{i=1}^{n} t_i = 1,
$$

$$
t_i \geq 0, \text{ for } i = 1, \ldots, n
$$

$$
0 \leq r_1 \leq r_2 \leq \cdots \leq r_n \leq M_*.
$$

Assuming a given channel allocation among n layers, i.e., $\mathbf{t}$ is given, the Karush-Kuhn-Tucker (KKT) conditions for the optimization problem of (16) lead to:

$$
bt_n r_n = d^*(r_n),
\tag{17}
$$

$$
d^*(r_n) + bt_{n-1} r_{n-1} = d^*(r_{n-1}),
\tag{18}
$$

$$
\cdots
$$

$$
d^*(r_2) + bt_1 r_1 = d^*(r_1),
\tag{19}
$$

where the corresponding distortion exponent is $\Delta_n^{LS} = d^*(r_1)$.

The equations in (17)-(19) can be graphically illustrated on the DMT curve as shown in Fig. 4. This illustration suggests that, for given channel allocation, finding the distortion exponent in n-layer LS can be formulated geometrically: We have n straight lines each with slope bt_i for $i = 1, \ldots, n$, and each line
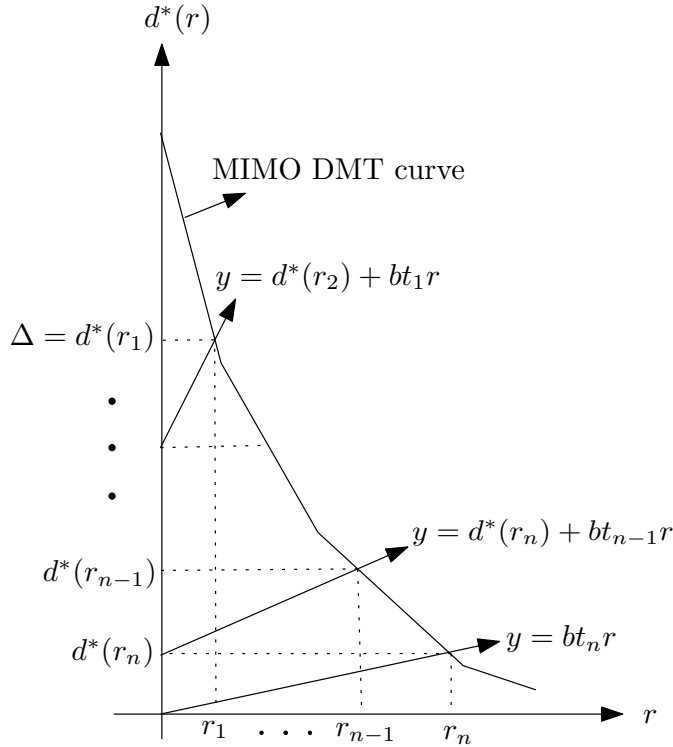
Fig. 4. Rate allocation for the source layers of LS illustrated on DMT curve of the MIMO channel.

intersects the y-axis at a point with the same ordinate as the intersection of the previous line with the DMT curve. Although the total slope is always equal to b, the more layers we have, the higher we can climb on the tradeoff curve and obtain a larger Δ.

The distortion exponent of LS in the limit of infinite layers provides a benchmark for the performance of LS in general. The following lemma will be used to characterize the optimal LS distortion exponent in the limit of infinite layers.

Lemma 4.1: In the limit of infinite layers, i.e., as $n \to \infty$, the optimal distortion exponent for LS can be achieved by allocating the channel equally among the layers.

Proof: Proof of the lemma can be found in Appendix II. ∎

The next theorem provides an explicit characterization of the asymptotic optimal LS distortion exponent Δ^{LS} (in the case of infinite layers) for an $M_t \times M_r$ MIMO system.

Theorem 4.2: Let the sequence $\{c_i\}$ be defined as $c_0 = 0$, $c_i = c_{i-1} + (|M_r - M_t| + 2i - 1) \ln\left(\frac{M_* - i + 1}{M_* - i}\right)$ for $i = 1, \ldots, M_* - 1$ and $c_{M_*} = \infty$. The optimal distortion exponent of infinite layer LS is given by:

$$\Delta^{LS} = \sum_{i=1}^{p-1} \ (|M_r - M_t| + 2i - 1)$$
$$+ (M_* - p + 1)(|M_r - M_t| + 2p - 1)(1 - e^{-\frac{b - c_{p-1}}{|M_r - M_t| + 2p - 1}}),$$

for $c_{p-1} \leq b < c_p$, $p = 1, \ldots, M_*$.

Proof: Proof of the theorem can be found in Appendix III. ∎

Corollary 4.3: For a MISO/SIMO system, we have

$$\Delta_{MISO/SIMO}^{LS} = M^*(1 - e^{-b/M^*}). \tag{20}$$

Illustration of Δ^{LS} for some specific examples as well as comparison with the upper bound and other strategies is left to Section VII. However, we note here that, although LS improves significantly compared to the single layer scheme, it still falls short of the upper bound. Nevertheless, the advantage of LS is the simple nature of its transceivers. We only need layered source coding and rate adaptation among layers while power adaptation is not required.

Another important observation is that, the geometrical model provided in Fig. 4 and Theorem 4.2 easily extends to any other system utilizing LS once the DMT is given. This is done in [4], [5] for a cooperative system, and will be carried out to extend the results to multiple block fading ($L > 1$) and parallel channels in Section VI.

B. Hybrid Digital-Analog Transmission with Layered Source (HLS)

In [11], the hybrid digital-analog technique proposed in [10] is analyzed in terms of the distortion exponent, and is shown to be optimal for bandwidth ratios $b \leq 1/M_*$. For higher bandwidth ratios, while the proposed hybrid strategy improves the distortion exponent compared to single layer digital transmission, its performance falls short of the upper bound. Here, we show that, combining the analog transmission with LS further improves the distortion exponent for $b > 1/M_*$. We call this technique *hybrid digital-analog transmission with layered source* (HLS). We will show that, introduction of the analog transmission will improve the distortion exponent compared to LS with the same number of digital layers, however; the improvement becomes insignificant as the number of layers increases.

For $b \geq 1/M_*$, we divide the N channel uses into two portions. In the first portion which is composed of $N - K/M_*$ channel uses, n source layers are channel coded and transmitted progressively in time in the same manner as LS. The remaining K/M_* channel uses are reserved to transmit the error signal in an analog fashion described below. Channel allocation for HLS is illustrated in Fig. 3(b).

Let $\bar{\mathbf{s}} \in \mathbb{C}^K$ be the reconstruction of the source $\mathbf{s}$ upon successful reception of all the digital layers. We denote the reconstruction error as $\mathbf{e} \in \mathbb{C}^K$ where $\mathbf{e} = \mathbf{s} - \bar{\mathbf{s}}$. This error is mapped to the transmit antennas where each component of the error vector is transmitted without coding in an analog fashion, just by scaling within the power constraint. Since $rank(\mathbf{H}) \leq M_*$, degrees of freedom of the channel is at most M_* at each channel use. Hence, at each channel use we utilize M_* of the M_t transmit antennas and in K/M_* channel uses we transmit all K components of the error vector $\mathbf{e}$. HLS encoder is shown in Fig. 5.
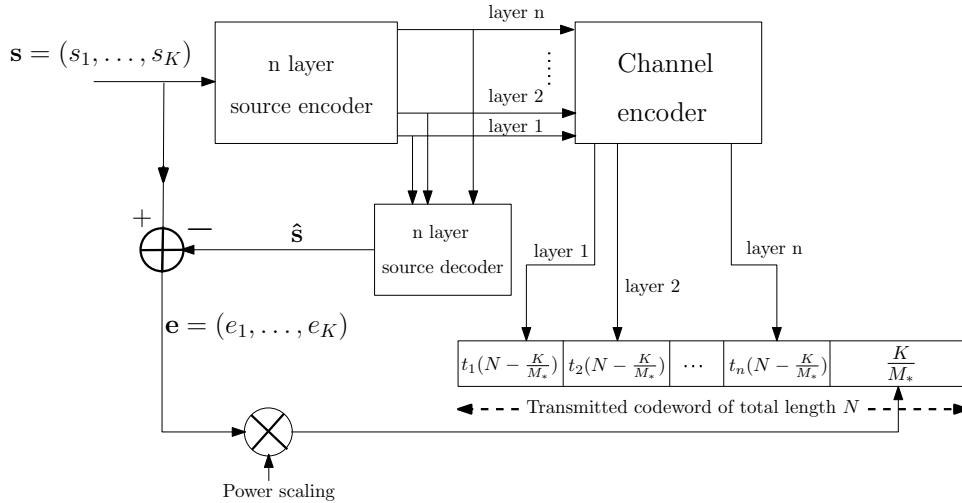
Fig. 5. Encoder for the n layer HLS for $b > 1/M_*$.

Receiver first tries to decode all the digitally transmitted layers as in LS, and in case of successful reception of all the layers, it forms the estimate $\bar{\mathbf{s}} + \tilde{\mathbf{e}}$, where $\tilde{\mathbf{e}}$ is the linear MMSE estimate of $\mathbf{e}$. This analog portion is ignored unless all digitally transmitted layers are successfully decoded at the destination.

The expected distortion for n-layer HLS can be written as

$$ED(\mathbf{R}, SNR) = \sum_{i=0}^{n-1} D_i^{HLS} \cdot \left(P_{out}^{i+1} - P_{out}^i \right) + \int_{\mathcal{A}^c} D_a^{HLS}(\bar{\mathbf{H}}) p(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \tag{21}$$

where $P_{out}^0 = 0$, $P_{out}^{n+1} = 1$, $D_0^{HLS} = 0$,

$$D_i^{HLS} = D\left(\left(b - \frac{1}{M_*} \right) \sum_{k=1}^i t_k R_k \right) \text{ for } i = 1, \ldots, n \tag{22}$$

and

$$D_a^{HLS}(\bar{\mathbf{H}}) = \frac{D_n^{HLS}}{M_*} \sum_{i=1}^{M_*} \frac{1}{1 + \frac{SNR}{M_*} \bar{\lambda}_i}, \tag{23}$$

where $\mathcal{A}$ denotes the set of channel states at which the n'th layer is in outage, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_{M_*}]$ are the eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$, $\bar{\mathbf{H}}$ is the $M_r \times M_*$ constrained channel matrix, and $\bar{\boldsymbol{\lambda}} = [\bar{\lambda}_1, \ldots, \bar{\lambda}_{M_*}]$ are the eigenvalues of $\bar{\mathbf{H}}^\dagger \bar{\mathbf{H}}$.

Note that the expected distortion of HLS in Eqn. (21) contains two terms. The first term which consists of the finite sum can be obtained similar to LS by using appropriate source coding rates. The following lemma will be used to characterize the high SNR behavior of the second term.

Lemma 4.4: Suppose that the transmission rate of the n-th layer for HLS is $R = r \log SNR$, where $r \le M_*$ is the multiplexing gain and that $\mathcal{A}$ denotes the outage event for this layer. If the average

signal-to-noise ratio for the analog part is SNR, we have

$$\int_{\mathcal{A}^c} \frac{1}{M_*} \sum_{i=1}^{M_*} \frac{1}{1 + \frac{SNR}{M_*}\bar{\lambda}_i} p(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \quad \dot{\leq} \quad SNR^{-1}. \tag{24}$$

Proof: Proof of the lemma can be found in Appendix IV. ∎

Then the second term of the expected distortion in (21) can be shown to be exponentially less than or equal to

$$SNR^{-\left[1 + (b - \frac{1}{M_*})\sum_{k=1}^{n} t_k\right]}. \tag{25}$$

Note that at high SNR, both LS and HLS have very similar ED expressions. Assuming the worst SNR exponent for the second term, the high SNR approximation of Eqn. (21) can be written as in Eqn. (15), except for the following: i) the bandwidth ratio b in (15) is replaced by $b - \frac{1}{M_*}$, and ii) the n-th term in (15) is replaced by (25). Hence, for a given time allocation vector $\mathbf{t}$, we obtain the following set of equations for the optimal multiplexing gain allocation:

$$1 + \left(b - \frac{1}{M_*}\right) t_n r_n = d^*(r_n), \tag{26}$$

$$d^*(r_n) + \left(b - \frac{1}{M_*}\right) t_{n-1} r_{n-1} = d^*(r_{n-1}), \tag{27}$$

$$\dots$$

$$d^*(r_2) + \left(b - \frac{1}{M_*}\right) t_1 r_1 = d^*(r_1), \tag{28}$$

where the corresponding distortion exponent is again $\Delta_n^{HLS} = d^*(r_1)$. Similar to LS, this formulation enables us to obtain an explicit formulation of the distortion exponent of infinite layer HLS using the DMT curve. For brevity we omit the general MIMO HLS distortion exponent and only give the expression for 2×2 MIMO and general MISO/SIMO systems for comparison.

Corollary 4.5: For 2×2 MIMO, HLS distortion exponent with infinite layers for $b \geq 1/2$ is given by

$$\Delta^{HLS} = 1 + 3[1 - e^{-\frac{1}{3}(b - \frac{1}{2})}].$$

Corollary 4.6: For a MISO/SIMO system utilizing HLS, we have (for $b \geq 1$)

$$\Delta_{MISO/SIMO}^{HLS} = M^* - (M^* - 1)e^{-(b-1)/M^*}. \tag{29}$$

Proof: For MISO/SIMO, we have $M_* = 1$. For n layer HLS with equal time allocation, using Lemma 3.1 in Appendix III we obtain the distortion exponent

$$\hat{\Delta}_{MISO/SIMO,n}^{HLS} = M^* - (M^* - 1)\left(\frac{1}{1 + \frac{b-1}{nM^*}}\right)^n. \tag{30}$$

Since equal channel allocation in the limit of infinite layers is optimal, taking the limit as $n \to \infty$, we obtain (29). ∎

Comparing $\Delta_{MISO/SIMO}^{HLS}$ with $\Delta_{MISO/SIMO}^{LS}$ in Corollary 4.3, we observe that

$$\Delta_{MISO/SIMO}^{HLS} - \Delta_{MISO/SIMO}^{LS} = e^{-b/M^*}[M^* - (M^* - 1)e^{1/M^*}]. \tag{31}$$

For a given MISO/SIMO system with a fixed number of M^* antennas, the improvement of HLS over LS exponentially decays to zero as the bandwidth ratio increases. Since we have $b \geq 1/M_*$, the biggest improvement of HLS compared to LS is achieved when $b = 1/M_* = 1$, and is equal to $e^{-1/M^*}[M^* - (M^* - 1)e^{1/M^*}]$. This is a decreasing function of M^*, and achieves its highest value at $M^* = 1$, i.e., SISO system, at $b = 1$, and is equal to $1/e$. Illustration of Δ^{HLS} for some specific examples as well as a comparison with the upper bound and other strategies is left to Section VII.

V. Broadcast Strategy with Layered Source (BS)

In this section we consider superimposing multiple source layers rather than sending them successively in time. We observe that this leads to higher distortion exponent than LS and HLS, and is in fact optimal in certain cases. This strategy will be called '*broadcast strategy with layered source*' (BS).

BS combines broadcasting ideas of [24] -[27] with layered source coding. Similar to LS, source information is sent in layers, where each layer consists of the successive refinement information for the previous layers. As in Section IV we enumerate the layers from 1 to n such that the ith layer is the successive refinement layer for the preceding $i-1$ layers. The codes corresponding to different layers are superimposed, assigned different power levels and sent simultaneously throughout the whole transmission block. Compared to LS, interference among different layers is traded off for increased multiplexing gain for each layer. We consider successive decoding at the receiver, where the layers are decoded in order from 1 to n and the decoded codewords are subtracted from the received signal.

Similar to Section IV we limit our analysis to single block fading ($L = 1$) scenario and leave the discussion of the multiple block case to Section VI. We first state the general optimization problem for arbitrary SNR and then study the high SNR behavior. Let $\mathbf{R} = [R_1, R_2, \ldots, R_n]^T$ be the vector of channel coding rates, which corresponds to a source coding rate vector of $b\mathbf{R}$ as each code is spread over the whole N channel uses. Let $\mathbf{SNR} = [SNR_1, \ldots, SNR_n]^T$ denote the power allocation vector for these layers with $\sum_{i=1}^n SNR_i = SNR$. Fig. 3(c) illustrates the channel and power allocation for BS. For $i = 1, \ldots, n$ we define

$$\overline{SNR}_i = \sum_{j=i}^n SNR_j. \tag{32}$$

The received signal over N channel uses can be written as

$$\mathbf{Y} = \mathbf{H} \sum_{i=1}^n \sqrt{\frac{SNR_i}{M_t}} \mathbf{X}_i + \mathbf{Z}, \tag{33}$$

where $\mathbf{Z} \in \mathbb{C}^{M_r \times N}$ is the additive complex Gaussian noise. We assume each $\mathbf{X}_i \in \mathbb{C}^{M_t \times N}$ is generated from i.i.d. Gaussian codebooks satisfying $\mathbf{tr}(E[\mathbf{X}_i \mathbf{X}_i^\dagger]) \leq M_t N$. Here $\sqrt{\frac{SNR_i}{M_t}} \mathbf{X}_i$ carries information for the i-th source coding layer. For $k = 1, \ldots, n$, we define

$$\bar{\mathbf{X}}_k = \sum_{j=k}^n \sqrt{\frac{SNR_j}{M_t}} \mathbf{X}_j, \tag{34}$$

and

$$\mathbf{Y}_k = \mathbf{H}\bar{\mathbf{X}}_k + \mathbf{Z}. \tag{35}$$

Note that $\mathbf{Y}_k$ is the remaining signal at the receiver after decoding and subtracting the first $k-1$ layers. Denoting $\mathcal{I}(\mathbf{Y}_k; \mathbf{X}_k)$ as the mutual information between $\mathbf{Y}_k$ and $\mathbf{X}_k$, we can define the following outage events,

$$\mathcal{A}_k = \left\{ \mathbf{H} : \mathcal{I}(\mathbf{Y}_k; \mathbf{X}_k) < R_k \right\}, \tag{36}$$

$$\mathcal{B}_k = \bigcup_{i=1}^k \mathcal{A}_i, \tag{37}$$

and the corresponding outage probabilities

$$P_{out}^k = Pr\left\{ \mathbf{H} : \mathbf{H} \in \mathcal{A}_k \right\}, \tag{38}$$

$$\bar{P}_{out}^k = Pr\left\{ \mathbf{H} : \mathbf{H} \in \mathcal{B}_k \right\}. \tag{39}$$

We note that P_{out}^k denotes the probability of outage for layer k given that the decoder already has access to the previous $k-1$ layers. On the other hand, $\bar{P}_{out}^k$ is the overall outage probability of layer k in case of successive decoding, where we assume that if layer k cannot be decoded, then the receiver will not attempt to decode the subsequent layers $i > k$. Then the expected distortion for n-layer BS using successive decoding can be written as

$$ED(\mathbf{R}, \mathbf{SNR}) = \sum_{i=0}^n D_i^{BS}(\bar{P}_{out}^{i+1} - \bar{P}_{out}^i), \tag{40}$$

where

$$D_i^{BS} = D\left(b \sum_{k=1}^i R_k \right),$$

$\bar{P}_{out}^0 = 0$, $\bar{P}_{out}^{n+1} = 1$, and $D_0^{BS} = 1$. Various algorithms solving this optimization problem are proposed in [15],[18]-[22].

The following definition will be useful in characterizing the distortion exponent of BS.

Definition 5.1: The '*successive decoding diversity gain*' for layer k of the BS strategy is defined as the high SNR exponent of the outage probability of that layer using successive decoding at the receiver.

The successive decoding diversity gain can be written as

$$d_{sd}(r_k) \triangleq - \lim_{SNR \to \infty} \frac{\log \bar{P}_{out}^k}{\log SNR}$$

Note that the successive decoding diversity gain for layer k depends on the power and multiplexing gain allocation for layers $1, \ldots, k-1$ as well as layer k itself. However, we drop the dependence on the previous layers for simplicity.

For any communication system with DMT characterized by $d^*(r)$, the successive decoding diversity gain for layer k satisfies $d_{sd}(r_k) \leq d^*(r_1 + \ldots + r_k)$. Concurrent work by Diggavi and Tse [28], coins the term '*successive refinability of the DMT curve*' when this inequality is satisfied with equality, i.e. multiple layers of information simultaneously operate on the DMT curve of the system. Our work, carried out independently, illustrates that combining successive refinability of the source and the successive refinability of the DMT curve leads to an optimal distortion exponent in certain cases.

From (40), we can write the high SNR approximation for ED as below.

$$\begin{aligned} ED & \doteq \sum_{i=0}^{n} \bar{P}_{out}^{i+1} D_i^{BS}, \\ & \doteq \sum_{i=0}^{n} SNR^{-d_{sd}(r_{i+1})} SNR^{-b \sum_{j=1}^{i} r_j}. \end{aligned} \tag{41}$$

Then the distortion exponent is given by

$$\Delta_n^{BS} = \min_{0 \leq i \leq n} \left\{ d_{sd}(r_{i+1}) + b \sum_{j=1}^{i} r_j \right\}. \tag{42}$$

Note that, while the DMT curve for a given system is enough to find the corresponding distortion exponent for LS and HLS, in the case of BS, we need the successive decoding DMT curve.

Next, we propose a power allocation among layers for a given multiplexing gain vector. For a general $M_t \times M_r$ MIMO system, we consider multiplexing gain vectors $\mathbf{r} = [r_1, \ldots, r_n]$ such that $r_1 + \cdots + r_n \leq 1$. This constraint ensures that we obtain an increasing and nonzero sequence of outage probabilities $\{P_{out}^k\}_{k=1}^n$. We impose the following power allocation among the layers:

$$\overline{SNR}_k = SNR^{1-(r_1 + \cdots + r_{k-1} + \epsilon_{k-1})}, \tag{43}$$

for $k = 2, \ldots, n$ and $0 < \epsilon_1 < \cdots < \epsilon_{n-1}$.

Our next theorem computes the successive decoding diversity gain obtained with the above power allocation. We will see that the proposed power allocation scheme results in successive refinement of the DMT curve for MISO/SIMO systems. By optimizing the multiplexing gain $\mathbf{r}$ we will show that the optimal distortion exponent for MISO/SIMO meets the distortion exponent upper bound.

Theorem 5.1: For $M_t \times M_r$ MIMO, the successive decoding diversity gain for the power allocation in (43) is given by

$$d_{sd}(r_k) = M^* M_*(1 - r_1 - \cdots - r_{k-1}) - (M^* + M_* - 1)r_k. \tag{44}$$

Proof: Proof of the theorem can be found in Appendix V. ∎

Corollary 5.2: The power allocation in (43) results in the successive refinement of the DMT curve for MISO/SIMO systems.

Proof: For MISO/SIMO we have $M_* = 1$. By Theorem 5.1 we have $d_{sd}(r_k) = M^*(1 - r_1 - \cdots - r_k) = d^*(r_1 + \cdots + r_k)$. Thus all simultaneously transmitted n layers operate on the DMT curve. ∎

Using the successive decoding DMT curve of Theorem 5.1, the next theorem computes an achievable distortion exponent for BS by optimizing the multiplexing gain allocation among layers.

Theorem 5.3: For $M_t \times M_r$ MIMO, n-layer BS with power allocation in (43) achieves a distortion exponent of

$$\Delta_n^{BS} = b\frac{(M_t - k)(M_r - k)(1 - \eta_k^n)}{(M_t - k)(M_r - k) - b\eta_k^n}, \tag{45}$$

for $b \in \left[(M_t - k - 1)(M_r - k - 1), (M_t - k)(M_r - k)\right)$, $k = 0, \ldots, M_* - 1$, where

$$\eta_k = 1 + \frac{b - (M_t - k - 1)(M_r - k - 1)}{M_t + M_r - 2k - 1} > 0, \tag{46}$$

and

$$\Delta_n^{BS} = \frac{n(M_t M_r)^2}{nM_t M_r + M_t + M_r - 1}, \tag{47}$$

for $b \geq M_t M_r$.

In the limit of infinite layers, BS distortion exponent becomes

$$\Delta^{BS} = \begin{cases} b & \text{if } b < M_t M_r, \\ M_t M_r & \text{if } b \geq M_t M_r. \end{cases} \tag{48}$$

Hence, BS is distortion exponent optimal for $b \geq M_t M_r$.

Proof: Proof of the theorem can be found in Appendix VI. ∎

Corollary 5.4: For a MISO/SIMO system, the n-layer BS distortion exponent achieved by the power allocation in (43) is

$$\Delta_{MISO/SIMO,n}^{BS} = M^*\left(1 - \frac{1 - b/M^*}{1 - (b/M^*)^{n+1}}\right). \tag{49}$$

In the limit of infinite layers, we obtain

$$\Delta_{MISO/SIMO}^{BS} = \begin{cases} b & \text{if } b < M^*, \\ M^* & \text{if } b \geq M^*. \end{cases} \tag{50}$$
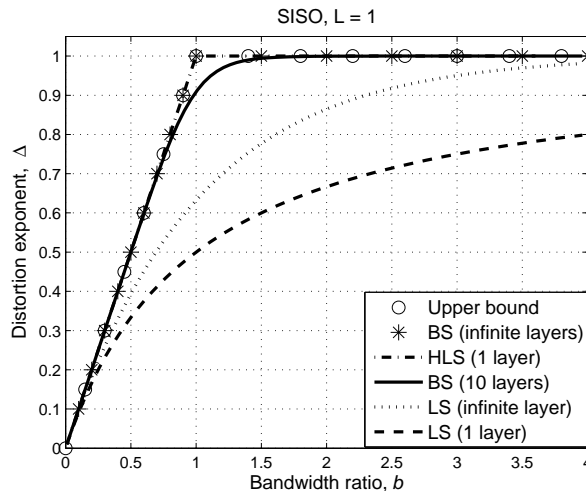
Fig. 6.   Distortion exponent vs. bandwidth ratio for SISO channel, $L = 1$.

BS with infinite source layers meets the distortion exponent upper bound of MISO/SIMO given in (8) for all bandwidth ratios, hence is optimal. Thus, (50) fully characterizes the optimal distortion exponent for MISO/SIMO systems.

Recently, [12], [13] reported improved BS distortion exponents for general MIMO by a more advanced power allocation strategy. Also, while a successively refinable DMT would increase the distortion exponent, we do not know whether it is essential to achieve the distortion exponent upper bound given in Theorem 3.1. However, successive refinement of general MIMO DMT has not been established [28], [29].

As in Section IV, the discussion of the results is left to Section VII.

VI.   MULTIPLE BLOCK FADING CHANNEL

In this section, we extend the results for $L = 1$ to multiple block fading MIMO, i.e., $L > 1$. As we observed throughout the previous sections, the distortion exponent of the system is strongly related to the maximum diversity gain available over the channel. In the multiple block fading scenario, channel experiences L independent fading realizations during N channel uses, so we have L times more diversity as reflected in the DMT of Theorem 2.1. The distortion exponent upper bound in Theorem 3.1 promises a similar improvement in the distortion exponent for multiple block fading. However, we note that increasing L improves the upper bound only if the bandwidth ratio is greater than $|M_t - M_r| + 1$, since the upper bound is limited by the bandwidth ratio, not the diversity at low bandwidth ratios.

Following the discussion in Section IV, extension of LS to multiple block fading is straightforward.
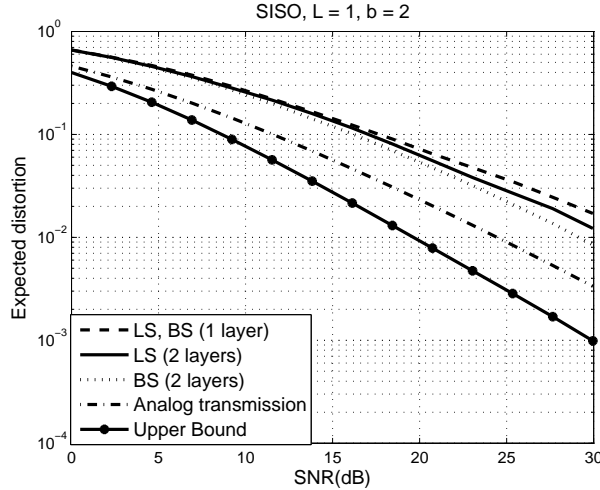
Fig. 7. Expected distortion vs. SNR plots for $b = 2$. The topmost curve LS, BS (1 layer) corresponds to single layer transmission.

As before, each layer of the successive refinement source code should operate on a different point of the DMT. This requires us to transmit codewords that span all channel realizations, thus we divide each fading block among n layers and transmit the codeword of each layer over its L portions.

Without going into the details of an explicit derivation of the distortion exponent for the multiple block fading case, as an example, we find the LS distortion exponent for 2×2 MIMO channel with $L = 2$ as

$$\Delta^{LS} = \begin{cases} 4(1 - \exp(-b/2)) & \text{if} \quad 0 < b \leq 2\ln 2, \\ 2 + 6\left[1 - \exp\left(-(\frac{b}{6} - \frac{\ln 2}{3})\right)\right] & \text{if} \quad b > 2\ln 2. \end{cases} \tag{51}$$

For HLS, the threshold bandwidth ratio $1/M_*$ is the same as single block fading as it depends on the channel rank per channel use, not the number of different channel realizations. For $b \geq 1/M_*$, extension of HLS to multiple blocks can be done similar to LS. As an example, for 2-block Rayleigh fading 2×2 MIMO channel, the optimal distortion exponent of HLS for $b \geq 1/2$ is given by

$$\Delta^{HLS} = \begin{cases} 1 + 3\left[1 - \exp\left(-\frac{1}{2}(b - \frac{1}{2})\right)\right] & \text{if} \quad 1/2 \leq b \leq \frac{1}{2} + 2\ln\frac{3}{2}, \\ 2 + 6\left[1 - \exp\left(-\frac{1}{6}(b - \frac{1}{2} - 2\ln\frac{3}{2})\right)\right] & \text{if} \quad b > \frac{1}{2} + 2\ln\frac{3}{2}. \end{cases} \tag{52}$$

For BS over multiple block fading MIMO, we use a generalization of the power allocation introduced in Section V in (43). For L-block fading channel and for $k = 2, \ldots, n$ let

$$\overline{SNR}_k = SNR^{1 - L(r_1 + \cdots + r_{k-1} - \epsilon_{k-1})}, \tag{53}$$

with $0 < \epsilon_1 < \cdots < \epsilon_{n-1}$ and imposing $\sum_{i=1}^{n} r_i \leq 1/L$. Using this power allocation scheme, we obtain the following distortion exponent for BS over L-block MIMO channel.

Theorem 6.1: For L-block $M_t \times M_r$ MIMO, BS with power allocation in (53) achieves the following distortion exponent in the limit of infinite layers.

$$\Delta^{BS} = \begin{cases} b/L & \text{if} \quad b < L^2 M_t M_r, \\ L M_t M_r & \text{if} \quad b \geq L^2 M_t M_r. \end{cases} \tag{54}$$

This distortion exponent meets the upper bound for $b \geq L^2 M_t M_r$.

Proof: Proof of the theorem can be found in Appendix VII. ∎

The above generalizations to multiple block fading can be adapted to parallel channels through a scaling of the bandwidth ratio by L [8]. Note that, in the block fading model, each fading block lasts for N/L channel uses. However, for L parallel channels with independent fading, each block lasts for N channel uses instead. Using the power allocation in (43) we can get achievable BS distortion exponent for parallel channels. We refer the reader to [8] for details and comparison. Detailed discussion and comparison of the L-block LS, HLS and BS distortion exponents are left to Section VII.

Distortion exponent for parallel channels has also been studied in [2] and [14] both of which consider source and channel coding for two parallel fading channels. The analysis of [2] is limited to single layer source coding and multiple description source coding. Both schemes perform worse than the upper bound in Theorem 3.1 and the achievable strategies presented in this paper. Particularly, the best distortion exponent achieved in [2] is by single layer source coding and parallel channel coding which is equivalent to LS with one layer. In [14], although 2-layer successive refinement and hybrid digital-analog transmission are considered, parallel channel coding is not used, thus the achievable performance is limited. The hybrid scheme proposed in [14] is a repetition based scheme and cannot improve the distortion exponent beyond single layer LS.

VII. DISCUSSION OF THE RESULTS

This section contains a discussion and comparison of all the schemes proposed in this paper and the upper bound. We first consider the special case of single-input single-output (SISO) system. For a SISO single block Rayleigh fading channel, the upper bound for optimal distortion exponent in Theorem 3.1 can be written as

$$\Delta = \begin{cases} b & \text{if} \quad b < 1, \\ 1 & \text{if} \quad b \geq 1. \end{cases} \tag{55}$$

This optimal distortion exponent is achieved by BS in the limit of infinite source coding layers (Corollary 5.2) and by HLS [11]. In HLS, pure analog transmission is enough to reach the upper bound when $b \geq 1$ [6], while the hybrid scheme of [11] achieves the optimal distortion exponent for $b < 1$.
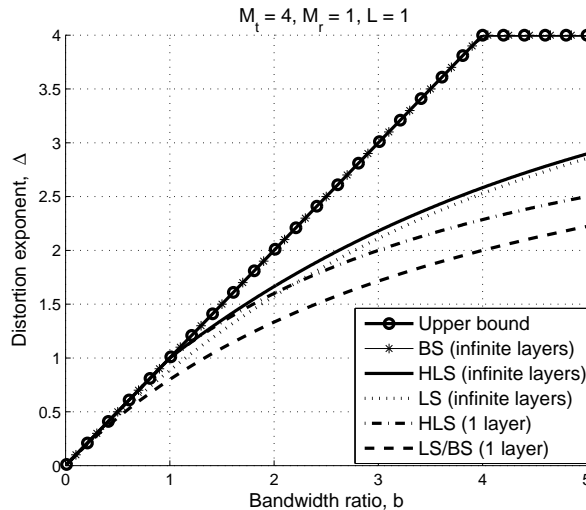
Fig. 8. Distortion exponent vs. bandwidth ratio for 4×1 MIMO, $L = 1$.

The distortion exponent vs. bandwidth ratio of the various schemes for the SISO channel, $L = 1$ are plotted in Fig. 6. The figure suggests that while BS is optimal in the limit of infinite source layers, even with 10 layers, the performance is very close to optimal for almost all bandwidth ratios.

For a SISO channel, when the performance measure is the expected channel rate, most of the improvement provided by the broadcast strategy can be obtained with two layers [30]. However, our results show that when the performance measure is the expected end-to-end distortion, increasing the number of superimposed layers in BS further improves the performance especially for bandwidth ratios close to 1.

In order to illustrate how the suggested source-channel coding techniques perform for arbitrary SNR values for the SISO channel, in Fig. 7 we plot the expected distortion vs. SNR for single layer transmission (LS, BS with 1 layer), LS and BS with 2 layers, analog transmission and the upper bound for $b = 2$. The results are obtained from an exhaustive search over all possible rate, channel and power allocations. The figure illustrates that the theoretical distortion exponent values that were found as a result of the high SNR analysis hold, in general, even for moderate SNR values.

In Fig. 8, we plot the distortion exponent versus bandwidth ratio of 4×1 MIMO single block fading channel for different source-channel strategies discussed in Section IV-V as well as the upper bound. As stated in Corollary 5.2, the distortion exponent of BS coincides with the upper bound for all bandwidth ratios. We observe that HLS is optimal up to a bandwidth ratio of 1. This is attractive for practical applications since only a single coded layer is used, while BS requires many more layers to be superimposed. However, the performance of HLS degrades significantly beyond $b = 1$, making BS more
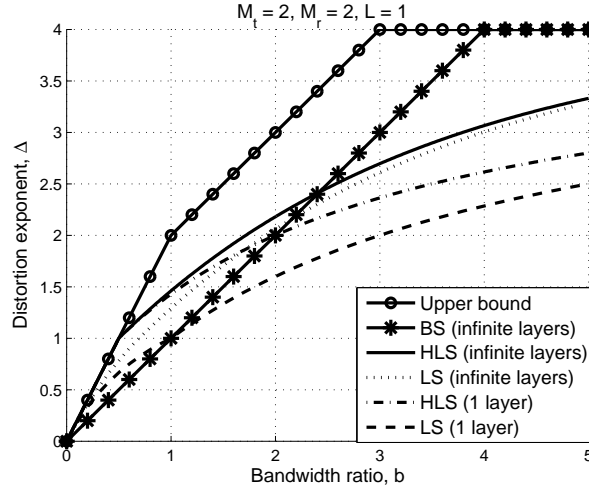
Fig. 9. Distortion exponent vs. bandwidth ratio for 2×2 MIMO, $L = 1$.
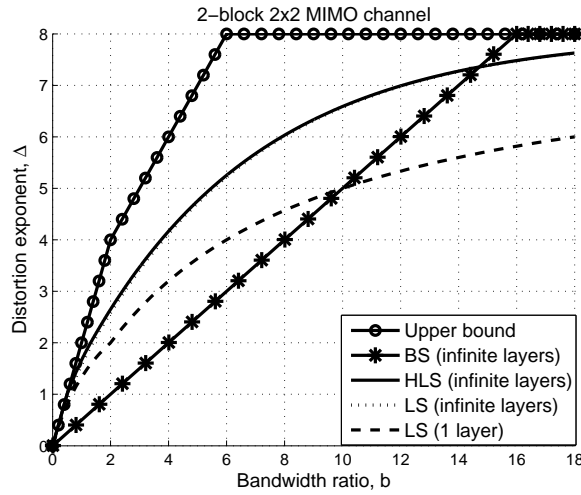

Fig. 10. Distortion exponent vs. bandwidth ratio for 2x2 MIMO, $L = 2$.

advantageous in this region. Pure analog transmission of the source samples would still be limited to a distortion exponent of 1 as in SISO, since linear encoding/decoding can not utilize the diversity gain of the system. More advanced nonlinear analog schemes which would take advantage of the diversity gain and achieve an improved distortion exponent may be worth exploring. While LS does not require any superposition or power allocation among layers, and only uses a digital encoder/decoder pair which can transmit at variable rates, the performance is far below BS. Nevertheless, the improvement of infinite layer LS compared to a single layer strategy is significant.

We plot the distortion exponent versus bandwidth ratio for 2×2 MIMO with $L = 1$ in Fig. 9. We

observe that BS is optimal for $b \geq 4$ and provides the best distortion exponent for $b > 2.4$. For bandwidth ratios $1/2 < b < 4$, none of the strategies discussed in this paper achieves the upper bound.

Note that, both for 4×1 MISO and 2×2 MIMO, when $b \geq 1/M^*$ the gain due to the analog portion, i.e., gain of HLS compared to LS, is more significant for one layer and decreases as the number of layers goes to infinity. Furthermore, at any fixed number of layers, this gain decays to zero with increasing bandwidth ratio as well. We conclude that for general MIMO systems, when the bandwidth ratio is high, layered digital transmission with large number of layers results in the largest improvement in the distortion exponent.

In Figure 10 we plot the distortion exponent for a 2-block 2×2 MIMO channel. We observe that the improvement of HLS over LS, both operating with infinite number of layers, is even less significant than the single block case. However, HLS can still achieve the optimal distortion exponent for $b < 1/2$. Although BS is optimal for $b \geq L^2 M_t M_r = 16$, both this threshold of $L^2 M_t M_r$ and the gap between the upper bound and BS performance below this threshold increases as L increases.

VIII. Generalization to Other Sources

Throughout this paper, we have used a complex Gaussian source for clarity of the presentation. This assumption enabled us to use the known distortion-rate function and to utilize the successive refinable nature of the complex Gaussian source. In this section we argue that our results hold for any memoryless source with finite differential entropy and finite second moment under squared-error distortion.

Although it is hard to explicitly obtain the rate-distortion function of general stationary sources, lower and upper bounds exist. Under the mean-square error distortion criteria, the rate-distortion function $R(D)$ of any stationary continuous amplitude source X is bounded as [31]

$$R_L(D) \leq R(D) \leq R_G(D), \tag{56}$$

where $R_L(D)$ is the Shannon lower bound and $R_G(D)$ is the rate-distortion function of the complex Gaussian source with the same real and imaginary variances.

Further in [32] it is shown that the Shannon lower bound is tight in the low distortion $(D \to 0)$, or, the high rate $(R \to \infty)$ limit when the source has a finite moment and finite differential entropy. We have

$$\lim_{R \to \infty} D(R) - \frac{e^{2h(X)}}{2\pi e} 2^{-R} = 0. \tag{57}$$

The high rate approximation of the distortion-rate function can be written as $D(R) = 2^{-R+O(1)}$, where $O(1)$ term depends on the source distribution but otherwise independent of the compression rate. Since in our distortion exponent analysis we consider scaling of the transmission rate, hence the source coding rate,

logarithmically with increasing SNR, the high resolution approximations are valid for our investigation. Furthermore the $O(1)$ terms in the above distortion-rate functions do not change our results since we are only interested in the SNR exponent of the distortion-rate function.

Although most sources are not successively refinable, it was proven in [33] that all sources are nearly successively refinable. Consider n layer source coding with rate of R_i bits/sample for layer $i = 1, \ldots, n$. Define D_i as the distortion achieved with the knowledge of first i layers and $W_i = R_i - R(D_i)$ as the rate loss at step i, where $R(D)$ is the distortion-rate function of the given source. Throughout the paper we used the fact that this rate loss is 0 for the Gaussian source [23]. Now we state the following result from [33] to argue that our high SNR results hold for sources that are nearly successively refinable as well.

Lemma 8.1: (Corollary 1, [33]) For any $0 < D_n < ... < D_2 < D_1$, $(n \geq 2)$ and squared error distortion, there exists an achievable M-tuple $(R_1, \ldots, R_n)$ with $W_k \leq 1/2$, $k \in \{1, ..., n\}$.

This means that to achieve the distortion levels we used in our analysis corresponding to each successive refinement layer, we need to compress the source layer at a rate that is at most 1 bits/sample[1] greater than the rates required for a successively refinable source. This translates into the distortion rate function as an additional $O(1)$ term in the exponent, which, as argued above, does not change the distortion exponent results obtained in this paper. These arguments together suggest that relaxing the Gaussian source assumption alters neither the calculations nor the results of our paper.

IX. Conclusion

We considered the fundamental problem of joint source-channel coding over block fading MIMO channels in the high SNR regime with no CSIT and perfect CSIR. Although the general problem of characterizing the achievable average distortion for finite SNR is still open, we showed that we can completely specify the high SNR behavior of the expected distortion in various settings. Defining the distortion exponent as the decay rate of average distortion with SNR, we provided a distortion exponent upper bound and three different lower bounds.

Our results reveal that, layered source coding with unequal error protection is critical for adapting to the variable channel state without the availability of CSIT. For the proposed transmission schemes, depending on the bandwidth ratio, either progressive or simultaneous transmission of the layers perform better. However, for single block MISO/SIMO channels, BS outperforms all other strategies and meets the upper bound, that is, BS is distortion exponent optimal for MISO/SIMO.

---

[1] We have $W_k \leq 1$ due to complex source assumption.

APPENDIX I

PROOF OF THEOREM 3.1

Here we find the distortion exponent upper bound under short-term power constraint assuming availability of the channel state information at the transmitter (CSIT)[2]. Let $C(\mathbf{H})$ denote the capacity of the channel with short-term power constraint when CSIT is present. Note that $C(\mathbf{H})$ depends on the channel realizations $\mathbf{H}_1, \ldots, \mathbf{H}_L$. The capacity achieving input distribution at channel realization $\mathbf{H}_j$ is Gaussian with covariance matrix $\mathbf{Q}_j$. We have

$$
\begin{aligned}
C(\mathbf{H}) &= \frac{1}{L} \sum_{j=1}^{L} \sup_{\mathbf{Q}_j \succeq 0, \sum_{j=1}^{L} \mathbf{tr}(\mathbf{Q}_j) \leq LM_t} \log \det \left(\mathbf{I} + \frac{SNR}{M_t} \mathbf{H}_j \mathbf{Q}_j \mathbf{H}_j^\dagger \right), \\
&\leq \frac{1}{L} \sum_{j=1}^{L} \sup_{\mathbf{Q}_j \succeq 0, \mathbf{tr}(\mathbf{Q}_j) \leq LM_t} \log \det \left(\mathbf{I} + \frac{SNR}{M_t} \mathbf{H}_j \mathbf{Q}_j \mathbf{H}_j^\dagger \right), \\
&\leq \frac{1}{L} \sum_{j=1}^{L} \log \det(\mathbf{I} + L \cdot SNR \mathbf{H}_j \mathbf{H}_j^\dagger),
\end{aligned}
\tag{58}
$$

where the first inequality follows as we expand the search space, and the second inequality follows from the fact that $LM_t \mathbf{I} - \mathbf{Q}_j \succeq 0$ when $\mathbf{tr}(\mathbf{Q}_j) \leq LM_t$ and $\log \det(\cdot)$ is an increasing function on the cone of positive-definite Hermitian matrices. Then the end-to-end distortion can be lower bounded as

$$
D(\mathbf{H}) = 2^{-bC(\mathbf{H})} \geq \prod_{j=1}^{L} [\det(\mathbf{I} + L \cdot SNR \mathbf{H}_j \mathbf{H}_j^\dagger)]^{-b/L}.
\tag{59}
$$

We consider expected distortion, where the expectation is taken over all channel realizations and analyze its high SNR exponent to find the corresponding distortion exponent. We will follow the technique used in [1]. Assume without loss of generality that $M_t \geq M_r$. Then from Eqn. (59) we have

$$
\begin{aligned}
D(\mathbf{H}) &\geq \prod_{j=1}^{L} [\det(\mathbf{I} + L \cdot SNR \mathbf{H}_j \mathbf{H}_j^\dagger)]^{-b/L}, \tag{60} \\
&\geq \prod_{j=1}^{L} \prod_{i=1}^{M_r} (1 + L \cdot SNR \lambda_{ji})^{-b/L}, \tag{61}
\end{aligned}
$$

where $\lambda_{j1} \leq \lambda_{j1} \leq \cdots \leq \lambda_{jM_r}$ are the ordered eigenvalues of $\mathbf{H}_j \mathbf{H}_j^\dagger$ for block $j = 1, \ldots, L$. Let $\lambda_{ji} = SNR^{-\alpha_{ji}}$. Then we have $(1 + L \cdot SNR \lambda_{ji}) \doteq SNR^{(1-\alpha_{ji})^+}$.

---

[2]We note that a similar upper bound for $L = 1$ is also given in [11]. We derive it for the L-block channel here for completeness.

The joint pdf of $\boldsymbol{\alpha_j} = [\alpha_{j1}, \ldots, \alpha_{jM_t}]$ for $j = 1, \ldots, L$ is

$$
\begin{aligned}
p(\boldsymbol{\alpha}_j) \;=\; & K_{M_t, M_r}^{-1} (\log SNR)^{M_r} \prod_{i=1}^{M_r} SNR^{-(M_t - M_r + 1)\alpha_{ji}} \\
& \cdot \left[\prod_{i<k} (SNR^{\alpha_{ji}} - SNR^{\alpha_{jk}})^2 \right] \exp\left(-\sum_{i=1}^{M_r} SNR^{\alpha_{ji}} \right),
\end{aligned}
\tag{62}
$$

where K_{M_t, M_r} is a normalizing constant. We can write the expected end-to-end distortion as

$$
E[D(\mathbf{H})] \;\doteq\; \int D(\mathbf{H}) p(\boldsymbol{\alpha}_1) \ldots p(\boldsymbol{\alpha}_L) d\boldsymbol{\alpha}_1 \ldots d\boldsymbol{\alpha}_L,
\tag{63}
$$

$$
\;\dot{\geq}\; \left[\int \prod_{i=1}^{M_r} (1 + SNR\lambda_{1i})^{-b/L} p(\boldsymbol{\alpha}_1) d\boldsymbol{\alpha}_1 \right]^L,
\tag{64}
$$

where we used the fact that $\boldsymbol{\alpha}_j$'s are i.i.d. for $j = 1, \ldots, L$ and that the constant multiplicative term in front of SNR does not affect the exponential behavior. Since we are interested in the distortion exponent, we only need to consider the exponents of SNR terms. Following the same arguments as in [1] we can make the following simplifications.

$$
\begin{aligned}
\int \prod_{i=1}^{M_r} (1 + SNR\lambda_{1i})^{-b/L} p(\boldsymbol{\alpha}_1) d\boldsymbol{\alpha}_1 \;\doteq\; & \int_{\mathcal{R}^{n+}} \prod_{i=1}^{M_r} (1 + SNR^{1-\alpha_{1i}})^{-b/L} \\
& \cdot \prod_{i=1}^{M_r} SNR^{-(M_t - M_r + 1)\alpha_{1i}} \cdot \prod_{i<k} (SNR^{-\alpha_{1i}} - SNR^{-\alpha_{1k}})^2 d\boldsymbol{\alpha}_1. \\
\;\doteq\; & \int_{\mathcal{R}^{n+}} \prod_{i=1}^{M_r} SNR^{-\frac{b}{L}(1-\alpha_{1i})^+} \prod_{i=1}^{M_r} SNR^{-(2i-1+M_t-M_r)\alpha_{1i}} d\boldsymbol{\alpha}_1. \\
\;\doteq\; & \int_{\mathcal{R}^{n+}} \prod_{i=1}^{M_r} SNR^{-(2i-1+M_t-M_r)\alpha_{1i} - \frac{b}{L}(1-\alpha_i)^+} d\boldsymbol{\alpha}, \\
\;\doteq\; & SNR^{-\Delta_1}
\end{aligned}
$$

Again following the arguments of the proof of Theorem 4 in [1], we have

$$
\Delta_1 \;=\; \inf_{\boldsymbol{\alpha} \in \mathcal{R}^{n+}} \sum_{i=1}^{M_r} (2i - 1 + M_t - M_r)\alpha_{1i} + \frac{b}{L}(1 - \alpha_{1i})^+.
\tag{65}
$$

The minimizing $\boldsymbol{\alpha}_1$ can be found as

$$
\alpha_{1i} = \begin{cases} 0 & \text{if } \; \frac{b}{L} < 2i - 1 + M_t - M_r \\[2mm] 1 & \text{if } \; \frac{b}{L} \geq 2i - 1 + M_t - M_r. \end{cases}
\tag{66}
$$

Letting $E[D(\mathbf{H})] \dot{\geq} SNR^{-\Delta^{UB}}$, we have $\Delta^{UB} = L\Delta_1$, and

$$
\Delta^{UB} = L \sum_{i=1}^{M_r} \min\left\{ \frac{b}{L}, 2i - 1 + M_t - M_r \right\}.
\tag{67}
$$

Similar arguments can be made for the $M_t < M_r$ case, completing the proof.

APPENDIX II

PROOF OF LEMMA 4.1

Let $\mathbf{t}^*$ be the optimal channel allocation vector and $\mathbf{r}^*$ be the optimal multiplexing gain vector for n layers. For any $\varepsilon > 0$ we can find $\tilde{\mathbf{t}}$ with $\tilde{t}_i \in \mathbb{Q}$ and $\sum_{i=1}^n \tilde{t}_i = 1$ where $|t_i^* - \tilde{t}_i| < \varepsilon$. Let $\tilde{t}_i = \gamma_i / \rho_i$ where $\gamma_i \in \mathbb{Z}, \rho_i \in \mathbb{Z}$ and $\theta = LCM(\rho_1, \ldots, \rho_n)$ is the least common multiple of $\rho_1, \ldots, \rho_n$. Now consider the channel allocation $\hat{\mathbf{t}} = [1/\theta, \ldots, 1/\theta]^T$, which divides the channel into θ equal portions and the multiplexing gain vector

$$\hat{\mathbf{r}} \;=\; [\underbrace{r_1^*, \ldots, r_1^*}_{\theta \tilde{t}_1 \text{ times}}, \underbrace{r_2^*, \ldots, r_2^*}_{\theta \tilde{t}_2 \text{ times}}, \ldots, \underbrace{r_n^*, \ldots, r_n^*}_{\theta \tilde{t}_n \text{ times}}]^T \tag{68}$$

Due to the continuity of the outage probability and the distortion-rate function, this allocation which consists of θn layers achieves a distortion exponent arbitrarily close to the n-layer optimal one as $\varepsilon \to 0$.

Note that $\{\Delta_n^{LS}\}_{n=1}^\infty$ is a non-decreasing sequence since with n layers it is always possible to assign $t_n = 0$ and achieve the optimal performance of $n - 1$ layers. On the other hand, using Theorem 3.1, it is easy to see that $\{\Delta_n^{LS}\}$ is upper bounded by $d^*(0)$, hence its limit exists. We denote this limit by Δ^{LS}. If we define $\hat{\Delta}_n^{LS}$ as the distortion exponent of n-layer LS with equal channel allocation, we have $\hat{\Delta}_n^{LS} \leq \Delta_n^{LS}$. On the other hand, using the above arguments, for any n there exists $m \geq n$ such that $\hat{\Delta}_m^{LS} \geq \Delta_n^{LS}$. Thus we conclude that

$$\lim_{n \to \infty} \hat{\Delta}_n^{LS} = \lim_{n \to \infty} \Delta_n^{LS} = \Delta^{LS}$$

Consequently, in the limit of infinite layers, it is sufficient to consider only the channel allocations that divide the channel equally among the layers.

APPENDIX III

PROOF OF THEOREM 4.2

We will use geometric arguments to prove the theorem. Using Lemma 4.1, we assume equal channel allocation, that is, $\mathbf{t} = [\frac{1}{n}, \ldots, \frac{1}{n}]$. We start with the following lemma.

Lemma 3.1: Let l be a line with the equation $y = -\alpha(x - M)$ for some $\alpha > 0$ and $M > 0$ and let l_i for $i = 1, \ldots, n$ be the set of lines defined recursively from n to 1 as $y = (b/n)x + d_{i+1}$, where $b > 0$, $d_{n+1} = 0$ and d_i is the y−component of the intersection of l_i with l. Then we have

$$d_1 = M\alpha \left[1 - \left(\frac{\alpha}{\alpha + b/n} \right)^n \right]. \tag{69}$$

with

$$\lim_{n \to \infty} d_1 = M\alpha \left(1 - e^{-b/\alpha} \right). \tag{70}$$
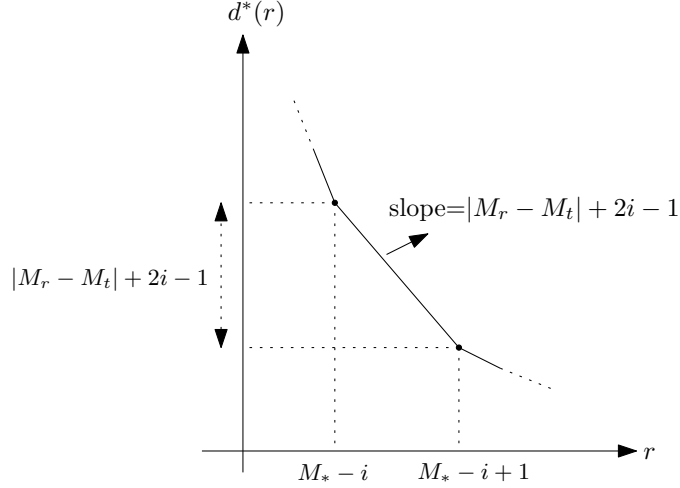
Fig. 11. DMT curve of an $M_t \times M_r$ MIMO system is composed of $M_* = \min\{M_t, M_r\}$ line segments, of which the i'th one is shown in the figure.

Proof: If we solve for the intersection points sequentially we easily find

$$d_k - d_{k+1} = M \frac{b}{n} \left(\frac{\alpha}{\alpha + b/n} \right)^{n-k+1}, \tag{71}$$

for $k = 1, \ldots, n$, where $d_{n+1} = 0$. Summing up these terms, we get

$$d_k = M\alpha \left[1 - \left(\frac{\alpha}{\alpha + b/n} \right)^{n-k+1} \right]. \tag{72}$$

∎

In the case of a DMT curve composed of a single line segment, i.e., $M_* = 1$, using Lemma 3.1 we can find the distortion exponent in the limit of infinite layers by letting $M = 1$ and $\alpha = M^*$. However, for a general $M_t \times M_r$ MIMO system the tradeoff curve is composed of M_* line segments where the ith segment has slope $|M_r - M_t| + 2i - 1$, and abscissae of the end points $M_* - i$ and $M_* - i + 1$ as in Fig. 11. In this case, we should consider climbing on each line segment separately, one after another in the manner described in Lemma 3.1 and illustrated in Fig. 4. Then, each break point of the DMT curve corresponds to a threshold on b, such that it is possible to climb beyond a break point only if b is larger than the corresponding threshold.

Now let $M = M_* - i + 1$, $\alpha = |M_r - M_t| + 2i - 1$ in Lemma 3.1 and in the limit of $n \to \infty$, let $k_i n$ be the number of lines with slopes b/n such that we have $d_n = |M_r - M_t| + 2i - 1$. Using the limiting form of Eqn. (72) we can find that

$$k_i = \frac{|M_r - M_t| + 2i - 1}{b} \ln \left(\frac{M_* - i + 1}{M_* - i} \right). \tag{73}$$

This gives us the proportion of the lines that climb up the pth segment of the DMT curve. In the general MIMO case, to be able to go up exactly to the pth line segment, we need to have $\sum_{j=1}^{p-1} k_j < 1 \leq \sum_{j=1}^{p} k_j$. This is equivalent to the requirement $c_{p-1} < b \leq c_p$ in the theorem.

To climb up each line segment, we need $k_i n$ lines (layers) for $i = 1, \ldots, p-1$, and for the last segment we have $(1 - \sum_{j=1}^{p-1} k_j)n$ lines, which gives us an extra ascent of

$$(M_* - p + 1)(|M_r - M_t| + 2p - 1)(1 - e^{-\frac{bk_p}{|M_r - M_t| + 2p - 1}})$$

on the tradeoff curve. Hence the optimal distortion exponent, i.e., the total ascent on the DMT curve, depends on the bandwidth ratio and is given by Theorem 4.2.

APPENDIX IV
PROOF OF LEMMA 4.4

As in the proof of Theorem 3.1 in Appendix I, we let $\lambda_i = SNR^{-\alpha_i}$ and $\bar{\lambda}_i = SNR^{-\beta_i}$ for $i = 1, \ldots, M_*$. The probability densities of $\boldsymbol{\lambda}$ and $\bar{\boldsymbol{\lambda}}$ and their exponents $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are given in Appendix I. Note that since $\bar{\mathbf{H}}$ is a submatrix of $\mathbf{H}$, $\boldsymbol{\lambda}$ and $\bar{\boldsymbol{\lambda}}$ as well as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are correlated. Let $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$ be the joint probability density of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. If $M_t \leq M_r$, $\mathbf{H}$ and $\bar{\mathbf{H}}$ coincide and $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}}$, $\boldsymbol{\alpha} = \boldsymbol{\beta}$.

We can write

$$\int_{\alpha \in \mathcal{A}^c} \frac{1}{M_*} \sum_{i=1}^{M_*} \frac{1}{1 + \frac{SNR}{M_*}\bar{\lambda}_i} p(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \doteq \int_{\alpha \in \mathcal{A}^c} \sum_{i=1}^{M_*} SNR^{-(1-\beta_i)^+} p(\boldsymbol{\alpha}) d\boldsymbol{\alpha}, \tag{74}$$

$$\doteq \int_{\alpha \in \mathcal{A}^c} SNR^{-(1-\beta_{max})^+} p(\boldsymbol{\alpha}) d\boldsymbol{\alpha}, \tag{75}$$

$$\doteq \int_{\alpha \in \mathcal{A}^c} SNR^{-(1-\beta_{max})^+} \int_{\beta} p(\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\beta} d\boldsymbol{\alpha}, \tag{76}$$

$$\doteq \int_{\beta} SNR^{-(1-\beta_{max})^+} \int_{\alpha \in \mathcal{A}^c} p(\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\alpha} d\boldsymbol{\beta}, \tag{77}$$

$$\leq \int_{\beta} SNR^{-(1-\beta_{max})^+} p(\boldsymbol{\beta}) d\boldsymbol{\beta}, \tag{78}$$

$$\doteq SNR^{-\mu}, \tag{79}$$

where

$$\mu = \inf_{\beta \in \mathbb{R}^{M_*+}} (1 - \beta_{max})^+ + \sum_{i=1}^{M_*} (2i - 1)\beta_i. \tag{80}$$

The minimizing $\tilde{\boldsymbol{\beta}}$ satisfies $\tilde{\beta}_1 \in [0, 1]$ and $\tilde{\beta}_2 = \cdots = \tilde{\beta}_{M_*} = 0$, and we have $\mu = 1$.

APPENDIX V

PROOF OF THEOREM 5.1

The mutual information between $\mathbf{Y}_k$ and $\mathbf{X}_k$ defined in (35) can be written as

$$
\begin{aligned}
\mathcal{I}(\mathbf{Y}_k; \mathbf{X}_k) &= \mathcal{I}(\mathbf{Y}_k; \mathbf{X}_k, \bar{\mathbf{X}}_{k+1}) - \mathcal{I}(\mathbf{Y}_k; \bar{\mathbf{X}}_{k+1} | \mathbf{X}_k), \quad &(81) \\
&= \log \det \left(\mathbf{I} + \frac{\overline{SNR}_k}{M_t} \mathbf{H}\mathbf{H}^\dagger \right) - \log \det \left(\mathbf{I} + \frac{\overline{SNR}_{k+1}}{M_t} \mathbf{H}\mathbf{H}^\dagger \right), \\
&= \log \frac{\det \left(\mathbf{I} + \frac{\overline{SNR}_k}{M_t} \mathbf{H}\mathbf{H}^\dagger \right)}{\det \left(\mathbf{I} + \frac{\overline{SNR}_{k+1}}{M_t} \mathbf{H}\mathbf{H}^\dagger \right)}. \quad &(82)
\end{aligned}
$$

For layers $k = 1, \ldots, n-1$, and the multiplexing gain vector $\mathbf{r}$ we have

$$
\begin{aligned}
P_{out}^k &= Pr \left\{ \mathbf{H} : \log \frac{\det \left(\mathbf{I} + \frac{\overline{SNR}_k}{M_t} \mathbf{H}\mathbf{H}^\dagger \right)}{\det \left(\mathbf{I} + \frac{\overline{SNR}_{k+1}}{M_t} \mathbf{H}\mathbf{H}^\dagger \right)} < r_k \log SNR \right\} \\
&= Pr \left\{ \mathbf{H} : \frac{\prod_{i=1}^{M_*}(1 + \frac{\overline{SNR}_k}{M_t}\lambda_i)}{\prod_{i=1}^{M_*}(1 + \frac{\overline{SNR}_{k+1}}{M_t}\lambda_i)} < SNR^{r_k} \right\}, \quad &(83)
\end{aligned}
$$

and

$$
P_{out}^n = \left\{ \mathbf{H} : \prod_{i=1}^{M_*} \left(1 + \frac{\overline{SNR}_n}{M_t}\lambda_i \right) < SNR^{r_n} \right\}. \quad (84)
$$

where $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{M_*}$ are the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$ ($\mathbf{H}^\dagger\mathbf{H}$) for $M_t \ge M_r$ ($M_t < M_r$). Let $\lambda_i = SNR^{-\alpha_i}$. Then for the power allocation in (43), conditions in Eqn. (83) and Eqn. (84) are, respectively, equivalent to

$$
\sum_{i=1}^{M_*}(1 - r_1 - \cdots - r_{k-1} - \epsilon_{k-1} - \alpha_i)^+ - \sum_{i=1}^{M_*}(1 - r_1 - \cdots - r_k - \epsilon_k - \alpha_i)^+ < r_k,
$$

and

$$
\sum_{i=1}^{M_*}(1 - r_1 - \cdots - r_{n-1} - \epsilon_{n-1} - \alpha_i)^+ < r_n.
$$

Using Laplace's method and following the similar arguments as in the proof of Theorem 4 in [1] we show that, for $k = 1, \ldots, n$,

$$
P_{out}^k \doteq SNR^{-d_k}, \quad (85)
$$

where

$$
d_k = \inf_{\boldsymbol{\alpha} \in \tilde{A}_k} \sum_{i=1}^{M_*} (|M_t - M_r| + 2i - 1)\, \alpha_i. \quad (86)
$$

For $k = 1, \ldots, n-1$

$$\tilde{A}_k = \left\{ \begin{array}{l} \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{M_*}] \in \mathbb{R}^{M_*+} : \alpha_1 \geq \cdots \geq \alpha_{M_*} \geq 0, \\ \sum_{i=1}^{M_*} (1 - r_1 - \cdots - r_{k-1} - \epsilon_{k-1} - \alpha_i)^+ - \sum_{i=1}^{M_*} (1 - r_1 - \cdots - r_k - \epsilon_k - \alpha_i)^+ < r_k \end{array} \right\}.$$

while

$$\tilde{A}_n = \{ \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{M_*}] \in \mathbb{R}^{M_*+} : \alpha_1 \geq \cdots \geq \alpha_{M_*} \geq 0, \\ \sum_{i=1}^{M_*} (1 - r_1 - \cdots - r_{n-1} - \epsilon_{n-1} - \alpha_i)^+ < r_n \}.$$

The minimizing $\tilde{\boldsymbol{\alpha}}$ for each layer can be explicitly found as

$$\tilde{\alpha}_i = 1 - r_1 - \cdots - r_{k-1} - \epsilon_{k-1}, \text{ for } i = 1, \ldots, M_* - 1,$$

and

$$\tilde{\alpha}_{M_*} = 1 - r_1 - \cdots - r_{k-1} - r_k - \epsilon_{k-1}.$$

Letting $\epsilon_k \to 0$ for $k = 1, \ldots, n-1$, we have

$$d_k = M^* M_* (1 - r_1 - \cdots - r_{k-1}) - (M^* + M_* - 1) r_k. \tag{87}$$

Note that the constraint $\sum_{i=1}^{n} r_i \leq 1$ makes the sequence $\{1 - r_1 - \cdots - r_k\}_{k=1}^{n}$ decreasing and greater than zero. Thus P_{out}^k constitutes an increasing sequence. Therefore, using (37) and (39), in the high SNR regime we have $\bar{P}_{out}^k \doteq P_{out}^k$, and $d_{sd}(r_k) = d_k$.

APPENDIX VI

PROOF OF THEOREM 5.3

Using the formulation of Δ_n^{BS} in (41) and successive decoding diversity gains of the proposed power allocation in Theorem 5.1, we find the multiplexing gain allocation that results in equal SNR exponents for all the terms in (41).

We first consider the case $b \geq (M_t - 1)(M_r - 1)$. Let

$$\eta_0 = \frac{b - (M_t - 1)(M_r - 1)}{M_t + M_r - 1} \geq 0. \tag{88}$$

For $0 \leq \eta_0 < 1$, we set

$$r_1 = \frac{M_t M_r (1 - \eta_0)}{M_t M_r - b \eta_0^n}, \tag{89}$$

$$r_i = \eta_0^{i-1} r_1, \text{ for } i = 2, \ldots, n. \tag{90}$$

If $\eta_0 \geq 1$, we set

$$r_1 = \cdots = r_n = \frac{M_t M_r}{n M_t M_r + M_t + M_r - 1}. \tag{91}$$

Next, we show that the above multiplexing gain assignment satisfies the constraint $\sum_{i=1}^{n} r_i \leq 1$. For $\eta_0 < 1$, $b \leq M_t M_r$ and

$$
\begin{aligned}
r_1 + \cdots + r_n &= r_1 \left(1 + \eta_0 + \cdots + \eta_0^{n-1}\right), \\
&= \frac{M_t M_r (1 - \eta_0^n)}{M_t M_r - b \eta_0^n}, \\
&\leq 1.
\end{aligned}
\tag{92}
$$

On the other hand, when $\eta_0 \geq 1$, we have $\sum_{i=1}^{n} r_i = \frac{n M_t M_r}{n M_t M_r + M_t + M_r - 1} < 1$.

Then the corresponding distortion exponent can be found as

$$
\Delta_n^{BS} = \begin{cases} b \frac{M_t M_r (1 - \eta_0^n)}{M_t M_r - b \eta_0^n} & \text{if} \quad (M_t - 1)(M_r - 1) \leq b < M_t M_r, \\ \frac{n (M_t M_r)^2}{n M_t M_r + M_t + M_r - 1} & \text{if} \quad b \geq M_t M_r. \end{cases}
\tag{93}
$$

For $(M_t - k - 1)(M_r - k - 1) \leq b < (M_t - k)(M_r - k)$, $k = 1, \ldots, M_* - 1$, we can consider the $(M_t - k) \times (M_r - k)$ antenna system and following the same steps as above, we obtain a distortion exponent of

$$
b \frac{(M_t - k)(M_r - k)(1 - \eta_k^n)}{(M_t - k)(M_r - k) - b \eta_k^n},
$$

where η_k is defined in (46).

In the limit of infinite layers, it is possible to prove that this distortion exponent converges to the following.

$$
\Delta^{BS} = \lim_{n \to \infty} \Delta_n^{BS} = \begin{cases} b & \text{if} \quad 0 \leq b < M_t M_r, \\ M_t M_r & \text{if} \quad b \geq M_t M_r. \end{cases}
\tag{94}
$$

APPENDIX VII

PROOF OF THEOREM 6.1

We transmit codewords of each layer across all fading blocks, which means that P_{out}^k in Eqn. (83) becomes

$$
P_{out}^k = Pr \left\{ (\mathbf{H}_1, \ldots, \mathbf{H}_L) : \frac{1}{L} \sum_{i=1}^{L} \log \frac{\det \left(\mathbf{I} + \frac{\overline{SNR_k}}{M_t} \mathbf{H}_i \mathbf{H}_i^\dagger \right)}{\det \left(\mathbf{I} + \frac{\overline{SNR_{k+1}}}{M_t} \mathbf{H}_i \mathbf{H}_i^\dagger \right)} < r_k \log SNR \right\},
$$

for $k = 1, \ldots, n - 1$, and P_{out}^n can be defined similarly.

For the power allocation in (53), above outage event is equivalent to

$$
\sum_{j=1}^{L} \sum_{i=1}^{M_*} (1 - L r_1 - \cdots - L r_{k-1} - L \epsilon_{k-1} - \alpha_{j,i})^+ - (1 - L r_1 - \cdots - L r_k - L \epsilon_k - \alpha_{j,i})^+ < L r_k.
$$

Following the same steps as in the proof of Theorem 5.1, we have

$$
P_{out}^k \doteq SNR^{-d_k},
\tag{95}
$$

where

$$d_k = \inf_{\boldsymbol{\alpha} \in \tilde{A}_k} \sum_{j=1}^{L} \sum_{i=1}^{M_*} \left(|M_t - M_r| + 2i - 1 \right) \alpha_{j,i}. \tag{96}$$

For $k = 1, \ldots, n-1$

$$\tilde{A}_k = \Bigg\{ \ \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{LM_*}] \in \mathbb{R}^{LM_*+} : \alpha_{j,1} \geq \cdots \geq \alpha_{j,M_*} \geq 0 \text{ for } j = 1, \ldots, L$$

$$\sum_{j=1}^{L} \sum_{i=1}^{M_*} (1 - Lr_1 - \cdots - Lr_{k-1} - L\epsilon_{k-1} - \alpha_{j,i})^+$$

$$-(1 - Lr_1 - \cdots - Lr_k - L\epsilon_k - \alpha_{j,i})^+ < Lr_k \Bigg\}.$$

The minimizing $\tilde{\boldsymbol{\alpha}}$ for each layer can be explicitly found as

$$\tilde{\alpha}_{j,i} = 1 - Lr_1 - \cdots - Lr_{k-1} - L\epsilon_{k-1}, \text{ for } j = 1, \ldots, L, \text{ and } i = 1, \ldots, M_* - 1,$$

and

$$\tilde{\alpha}_{j,M_*} = 1 - Lr_1 - \cdots - Lr_{k-1} - r_k - L\epsilon_{k-1}, \text{ for } j = 1, \ldots, L.$$

Then, letting $\epsilon_{n-1} \to 0$, the diversity gain d_k is obtained as

$$d_k = L \Bigg[M^* M_* (1 - Lr_1 - \cdots - Lr_{k-1}) - (M^* + M_* - 1) r_k \Bigg]. \tag{97}$$

We skip the rest of the proof as it closely resembles the proof of Theorem 5.3 in Appendix VI.

References

[1] L. Zheng, D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," IEEE Transactions on Information Theory, May 2003.

[2] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, "Source-Channel Diversity for Parallel Channels," IEEE Transactions on Information Theory, vol. 51, no. 10, pp. 3518-3539, Oct. 2005.

[3] D. Gündüz, E. Erkip, "Joint source-channel cooperation: Diversity versus spectral efficiency," Proceedings of 2004 ISIT, Chicago, June 2004.

[4] D. Gündüz, E. Erkip, "Source and channel coding for cooperative relaying," Proceedings of SPAWC, New York City, June 2005.

[5] D. Gündüz, E. Erkip, "Source and channel coding for cooperative relaying,", IEEE Transactions on Information Theory, Special Issue on Models, Theory and Codes for Relaying and Cooperation in Communication Networks, Oct. 2007.

[6] D. Gündüz, E. Erkip, "Source and channel coding for quasi-static fading channels," Proceedings of 39th Asilomar Conf., Pacific Grove, California, November 2005.

[7] D. Gündüz, E. Erkip, "Distortion exponent of MIMO fading channels," Proceedings of IEEE Information Theory Workshop, Punta del Este, Uruguay, March 2006.

[8] D. Gündüz, E. Erkip, "Distortion exponent of parallel channels," Proceedings of ISIT 2006, Seattle, WA, July 2006.

[9] T. Holliday and A. Goldsmith, "Joint source and channel coding for MIMO systems," Proc. of Allerton Conference on Communications, Control, and Computing, Oct. 2004, Allerton, IL.

[10] U. Mittal, N. Phamdo, "Hybrid digital-analog joint source-channel codes for broadcasting and robust communications," IEEE Transactions on Information Theory, May 2002.

[11] G. Caire, K. Narayanan , "On the SNR exponent of hybrid digital-analog space time coding," Proceedings of Allerton Conference, Allerton, IL, September 2005.

[12] K. Bhattad, K. R. Narayanan, G. Caire, "On the distortion exponent of some layered transmission schemes," Asilomar Conference, Nov. 2006.

[13] K. Bhattad, K. R. Narayanan, G. Caire, "On the distortion SNR exponent of some layered transmission schemes," submitted to IEEE Transactions on Information Theory.

[14] B. P. Dunn and J. N. Laneman, "Characterizing source-channel diversity approaches beyond the distortion exponent," in Proc. Allerton Conf. Communications, Control, and Computing, Monticello, IL, Sept. 2005.

[15] K. Zachariadis, M. L. Honig, A. K. Katsaggelos, "Source fidelity over fading channels: Erasure codes versus scalable codes", submitted to IEEE Transactions on Communications, July 2006.

[16] S. Sesia, G. Caire, G. Vivier, "Lossy transmission over slow-fading AWGN channels: a comparison of progressive, superposition and hybrid approaches," in Proc. of IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia, Sept. 2005. C. Tian,

[17] F. Etemadi, H. Jafarkhani, "Optimal layered transmission over quasi-static fading channels," Proc. of IEEE International Symposium on Information Theory (ISIT), Seattle, WA, July 2006.

[18] F. Etemadi, H. Jafarkhani, "Optimal rate and power allocation for layered transmission with superposition coding," IEEE Data Compression Conference (DCC '07), Salt Lake City, UT, March 2007.

[19] C. T. K. Ng, D. Gündüz, A. J. Goldsmith and E. Erkip, "Minimum expected distortion in Gaussian joint source-channel layered broadcast coding," in the Proc. of IEEE International Symposium on Information Theory (ISIT 2007), Nice, France, June 2007.

[20] C. T. K. Ng, D. Gündüz, A. J. Goldsmith and E. Erkip, "Recursive power allocation in Gaussian joint source-channel layered broadcast coding," in the Proceedings of IEEE International Conference on Communications (ICC 2007), Glasgow, Scotland, June 2007.

[21] C. T. K. Ng, D. Gündüz, A. J. Goldsmith, and E. Erkip, "Optimal power distribution and minimum expected distortion in Gaussian layered broadcast coding with successive refinement," Submitted to IEEE Transactions on Information Theory, May 2007.

[22] C. Tian, A. Steiner, S. Shamai, S. Diggavi, "Expected distortion for Gaussian source with a broadcast transmission strategy over a fading channel," in Proc. of IEEE Inform. Theory Workshop, Bergen, Norway, Sept. 2007.

[23] W. Equitz, T. Cover, "Successive refinement of information," IEEE Transactions on Information Theory, 37(2): 269-275, Mar. 1991.

[24] I. Kozintsev, K. Ramchandran, "Multiresolution joint source-channel coding using embedded constellations for power-constrained time-varying channels," In Proceedings of 1996 IEEE International Conference on Acoustics, Speech, and Signal Proc. (ICASSP), Atlanta, GA, May 1996.

[25] I. Kozintsev, K. Ramchandran, "Robust image transmission over energy-constrained time-varying channels using multiresolution joint source-channel coding," IEEE Transactions on Signal Processing, Volume: 46 Issue: 4, April 1998.

[26] S. Shamai, "A broadcast strategy for the Gaussian slowly fading channel," Proceedings of ISIT, p. 150, June 1997.

[27] S. Shamai, A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," IEEE Transactions on Information Theory, October 2003.

[28] S. Diggavi, D. Tse, "Fundamental limits of diversity-embedded codes over fading channels," Proceedings ISIT, Adelaide, Australia, September 2005.

[29] S. Diggavi, D. Tse, "On opportunistic codes and broadcast codes with degraded message set," in Proc. 2006 IEEE Information Theory Workshop (ITW 2006), Punta del Este, Uruguay, March 13-17, 2006.

[30] Y. Liu, K. N. Lau, O. Y. Takeshita, M. P. Fitz, "Optimal rate allocation for superposition coding in quasi-static fading channels," In Proceedings of ISIT, Lausanne, Switzerland, June 2002.

[31] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.

[32] T. Linder, R. Zamir, "On the asymptotic tightness of the Shannon lower bound", IEEE Transactions on Information Theory, pp. 2026-2031, Nov. 1994.

[33] L. Lastras, T. Berger, "All sources are nearly successively refinable," IEEE Transactions on Information Theory, 47(3): 918-926, March 2000.

Deniz Gündüz [S'02] received the B.S. degree in Electrical and Electronics Engineering from Middle East Technical University, Ankara, Turkey, in 2002 and the M.S. degree in Electrical and Computer Engineering from Polytechnic University, Brooklyn, NY, in 2004.

He is currently working toward his Ph.D. at Polytechnic University. In 2004, he was a summer researcher in the laboratory of information theory (LTHI) at EPFL in Lausanne, Switzerland. His research interests lie in the areas of communications theory and information theory with special emphasis on joint source-channel coding, cooperative communications and cross-layer design.

Mr. Gündüz coauthored the paper that received the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory.

Elza Erkip [S'93, M'96, SM'05] received the Ph.D. and M.S. degrees in Electrical Engineering from Stanford University, and the B.S. degree in Electrical and Electronics Engineering from Middle East Technical University, Turkey. She joined Polytechnic University in Spring 2000, where she is currently an Associate Professor of Electrical and Computer Engineering. She is a Visiting Professor at Princeton University for 2007-2008 academic year.

Dr. Erkip received the 2004 Communications Society Stephen O. Rice Paper Prize in the Field of Communications Theory and the NSF CAREER award in 2001. She is an Associate Editor of IEEE Transactions on Communications, a Publications Editor of IEEE Transactions on Information Theory and a Guest Editor of IEEE Signal Processing Magazine. She is the Technical Area Chair for the "MIMO Communications and Signal Processing" track of 41st Annual Asilomar Conference on Signals, Systems, and Computers, and the Technical Program Co-Chair of 2006 Communication Theory Workshop. Her research interests are in wireless communications, information theory and communication theory.